Alfredo Cesario
Frederick Marcus *Editors*

Cancer Systems Biology, Bioinformatics and Medicine

Research and Clinical Applications

Springer

Cancer Systems Biology, Bioinformatics and Medicine

Alfredo Cesario • Frederick B. Marcus
Editors

Cancer Systems Biology, Bioinformatics and Medicine

Research and Clinical Applications

Springer

Editors
Dr. Alfredo Cesario
IRCCS San Raffaele Pisana
Scientific Direction
Via di Valcannuta 247
00166 Rome, Italy and
Department of Surgery
Catholic University
Largo Agostino Gemelli, 1
00166 Rome, Italy
alfredo.cesario@sanraffaele.it
alfcesario@rm.unicatt.it

Dr. Frederick B. Marcus
Advanced Therapies and Systems Medicine
Health Research Directorate
European Commission
Brussels, Belgium
frederick.marcus@ec.europa.eu
frederick.b.marcus@gmail.com

ISBN 978-94-007-1566-0 e-ISBN 978-94-007-1567-7
DOI 10.1007/978-94-007-1567-7
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2011933914

We dedicate this book to all those suffering from cancer, and to their families, carers, nurses and physicians.

Preface

Nature of This Book

This book is a comprehensive and integrated textbook about systems approaches to cancer biology, bioinformatics and medicine, aimed at students, researchers and clinicians, that documents and builds upon many of the most recent advances in cancer science, technology, diagnosis, treatment and analysis methods. These approaches are applied to central processes in cancer origins and progression, and to translational applications including cancer diagnosis, drug development and treatment. The methods involve at least some level of computational analysis involving multiple interactions which are quantified, classified, analysed and modelled for optimized description. New genomic and proteomic measurements and cancerous phenotypes are causally linked (or capable of being so) at the molecular, cellular and tissue levels; research is directly and functionally connected with outcomes, using clinical observations as immediate feedback and input. Computational models of primary and secondary networks at the molecular and physiological levels can serve as test beds for developing and testing new biomarkers, drugs and treatment strategies, and enhance the efficiency, coherency and economy of the pre-clinical test phase and clinical trials.

These approaches transform fragmented investigations into integrated pipelines of research, discovery, monitoring and treatment, based on a quantitative understanding of various cancer mechanisms and their behaviour. Systems approaches thereby energize translational research, where fundamental knowledge of key processes is applied to clinical practice and personalized medicine.

Authors, Level of Presentation and Readership

Many of the 48 authors from 11 countries are world-renowned scientists, clinicians and organizational leaders in their fields, some of which they founded. Their chapters contain comprehensive presentations, pragmatic tools, guidance and inspiration for transforming cancer research and therapy, together with thousands of experimental, computational and clinical references.

The book is designed as a textbook for graduate students in their first and second year of PhD or MD studies and sufficiently advanced undergraduates in life sciences, biomedical and medical studies. It is also a comprehensive reference and survey for professional researchers of cancer systems approaches. Readers without a mathematical background can still benefit from most of the contents, which will allow them to understand the field, and to work closely with computational experts. In fact, independently of the systems aspects, a state-of-the-art description is provided of cancer as a multi-faceted disease, along with relevant research, diagnosis and treatment, so computational biologists will gain the biological and clinical background necessary for effectively working in cancer research. Doctors, clinicians, and university and industrial researchers will find essential references for their investigations and new directions in therapies.

Organisation and Contents

The book has a logically-developed, integrated and cross-referenced presentation, and contains (by chapters):

Part I—Introduction and Background

1. An overview of the topics covered in the book with a general introduction to the basis of systems biology, bioinformatics, and medicine approaches.
2. The morphology and pathology of the full range of cancers, including their grade and stage, plus the most prevalent cancers, their specific sub-classifications, implications for treatment, and the need to move to a systems context.

Part II—Laboratory, Clinical, Data and Educational Resources

3. Major advances in genomic and proteomic measurement technologies and their application to cancer research in the context of numerical modelling.
4. Cellular and tissue material essential for cancer research, model organisms and their modifications, and means of making optimum choices for elucidating biological function in cancer.
5. A wide range of genomic sequence and expression databases which form the basis of systems research in cancerous processes, and some major collaborative efforts in data generation.
6. Interdisciplinary skills for researchers and doctors, along with the educational processes and infrastructures that support basic investigations and clinical applications.

Part III—Bioinformatics and Systems Biology Analysis

7. A comprehensive description of the mathematical basis for systems approaches and their areas of application.
8. A guide to the wide range of computational tools and databases available and how to access and use them optimally.

9. Cancer progression in a systems context via modelling methods, demonstrating the power of these techniques.
10. Cellular pathways of cell-death processes and their relation to various cancer progression modes and interactions with therapies, also illustrating the benefits of collaborative and interdisciplinary research.
11. The utilization of gene expression data and biological motifs for diagnosis and analysis of tumour growth.
12. Evolutionary paradigms used to examine novel systems approaches to understanding tumour growth and reaction in the presence to treatment and bodily defences.

Part IV—Diagnosis and Treatment Applications

13. Biomarker development for diagnosis, prognosis and drug development, and their identification and correct interpretation to aid therapy.
14. Systems approaches for improving productivity and understanding at several stages in the drug development process, with examples of their use by the pharmaceutical industry.
15. Cancer chronotherapy and choices of combination therapies, dosage and timing, improving treatment outcomes during clinical trials, and computational frameworks for optimization of these choices.
16. An overview and synthesis of systems approaches to biomarkers, drug development and therapies, together with new directions concerning their application to clinical trials.
17. The role of robustness in understanding cancer as a systems disease, resistance to therapy, and new treatment paradigms.

Part V—Perspectives and Conclusions

18. Near and medium term prospects for advances in systems applications, and areas for progress, especially the application of synthetic biology technologies.
19. Conclusions from each chapter, integrated themes found in the book, and avenues for clinical practice.

Acknowledgements

We gratefully acknowledge the support of Springer in publishing this book, initially with Cristina Alves dos Santos, PhD, Publishing Editor, Cancer Research and Sara Huisman, Publishing Assistant, and later with Melania Ruiz, Publishing Editor and Ilse Hensen, Publishing Assistant, and publishing support from Crest Premedia Solutions in India. We also appreciate initial commentary from Dr. Bernard Mulligan. We are especially grateful to FBM's wife, Rosemary Marcus, D.Phil. (Oxford) in Modern Languages, for her excellent editing and reviewing contributions.

Contents

Chapter Authors

Atilla Altinok PhD Unité de Chronobiologie théorique, Université Libre de Bruxelles, Brussels, Belgium
e-mail: aaltinok@ulb.ac.be

Charles Auffray PhD Functional Genomics and Systems Biology for Health, CNRS Institute of Biological Sciences -7, rue Guy Moquet, BP8, 94801, Villejuif, France
e-mail: charles.auffray@vjf.cnrs.fr

Emmanuel Barillot PhD Bioinformatics and Computational Systems Biology of Cancer, Institut Curie - INSERM U900/Mines ParisTech, 26 rue d'Ulm, 75248 Paris Cedex 05, France
e-mail: emmanuel.barillot@curie.fr

Alvis Brazma PhD Microarray Informatics Group, European Bioinformatics Institute, Hinxton, Cambridge, UK
e-mail: brazma@ebi.ac.uk

Francesca Buffa PhD Department of Medical Oncology, Weatherall Institute of Molecular Medicine, University of Oxford, OX3 9DS, Oxford, UK
e-mail: francesca.buffa@imm.ox.ac.uk; aharris.lab@imm.ox.ac.uk

Julio Celis PhD Danish Institute for Cancer Biology, Copenhagen, Denmark
e-mail: jec@cancer.dk

Alfredo Cesario MD, PhD Deputy Scientific Director, IRCCS San Raffaele Pisana, Via di Valcannuta 247, 00166 Rome, Italy and Assistant Professor of Thoracic Surgery, Catholic University, Largo Agostino Gemelli 1, 00166, Rome, Italy
e-mail: alfredo.cesario@sanraffaele.it; alfcesario@rm.unicatt.it

Christophe Chassagnole PhD Physiomics Plc, Magdalen Centre, Oxford Science Park, Oxford, UK
e-mail: cchassagnole@physiomics-plc.com

Dirk Drasdo PhD Institut National de Recherche en Informatique et en Automatique (INRIA), Le Chesnay Cedex, France
e-mail: dirk.drasdo@inria.fr

David Fell PhD Department of Biological and Medical Sciences, Oxford Brookes University, Oxford, UK

Physiomics Plc, Magdalen Centre, Oxford Science Park, Oxford, UK
e-mail: dfell@brookes.ac.uk

Eric Fernandez PhD Physiomics Plc, Magdalen Centre, Oxford Science Park, Oxford, UK
e-mail: efernandez@physiomics-plc.com

David J. Galas PhD Professor & Senior Vice President, Institute for Systems Biology, 401 Terry Avenue N., Seattle, WA 98109, USA
e-mail: dgalas@systemsbiology.org

Robert A. Gatenby MD Departments of Radiology and Integrative Mathematical Oncology, Moffitt Cancer Center, Tampa, USA
e-mail: robert.gatenby@moffitt.org

Mathias Gehrmann PhD Siemens Medical Solutions Diagnostics GmbH, Leverkusen, Germany
e-mail: mathias.gehrmann@bayer.com

Albert Goldbeter PhD Unité de Chronobiologie théorique, Université Libre de Bruxelles, Brussels, Belgium
e-mail: agoldbet@ulb.ac.be

Adrian L. Harris FRCP Clinical Oncology, University of Oxford, Oxford, UK
Medical Oncology Unit, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK
e-mail: aharris.clin@cancer.org

Jan G. Hengstler MD Department of Toxicology, IfADo-Leibniz Research Centre for Working Environment and Human Factors, Technical University of Dortmund, Ardeystraβe 67, 44139 Dortmund, Germany
e-mail: hengstler@ifado.de

Stefan Höhme PhD Interdisciplinary Centre for Bioinformatics (IZBI), University of Leipzig, Leipzig, Germany
e-mail: hoehme@izbi.uni-leipzig.de

Leroy Hood MD PhD Institute of Systems Biology, 401 Terry Avenue North, Seattle, WA 98109, USA
e-mail: lhood@systemsbiology.org

Sergio Iadevaia PhD Division Chief of Genetics, and Associate Departments of Medicine and Bioengineering, University of California at San Diego, San Diego, USA
e-mail: siadevai@mdanderson.org

Trey Ideker PhD Division Chief of Genetics, and Associate Departments of Medicine and Bioengineering, University of California at San Diego, San Diego, USA
e-mail: siadevai@mdanderson.org

Hiroaki Kitano PhD Okinawa Institute of Science and Technology, Okinawa, Japan
e-mail: kitano@sbi.jp

Edda Klipp PhD Institute for Biology, Theoretical Biophysics, Humboldt-Universität zu Berlin, Invalidenstr. 42, 10115, Berlin, Germany
e-mail: edda.klipp@rz.hu-berlin.de

Falko Krause Humboldt-Universität zu Berlin, Berlin, Germany
e-mail: falko.krause@staff.hu-berlin.de

Bodo M. H. Lange PD, PhD Department of Vertebrate Genomics, Max-Planck Institute for molecular Genetics, Ihnestrasse 73, 14195, Berlin, Deutschland
e-mail: Lange_b@molgen.mpg.de

Hans Lehrach PhD Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin, Germany
e-mail: lehrach@molgen.mpg.de

Francis Lévi PhD INSERM, U776, Rythmes Biologiques et Cancers, 94807, Villejuif, France
e-mail: francis.levi@inserm.fr

Antonio Llombart-Bosch MD, PhD Department of Pathology, School of Medicine, University of Valencia, Valencia, Spain
e-mail: antonio.llombart@uv.es

Frederick B. Marcus DPhil Advanced Therapies and Systems Medicine, Health Research Directorate, European Commission, Brussels, Belgium
e-mail: frederick.marcus@ec.europa.eu; frederick.b.marcus@gmail.com

Gordon B. Mills MD, PhD Department of Systems Biology, University of Texas M.D. Anderson Cancer Center, Houston, Texas, USA
e-mail: gmills@mdanderson.org

David Orrell PhD Physiomics Plc, Magdalen Centre, Oxford Science Park, Oxford, UK
e-mail: dorrell@physiomics-plc.com

Prahlad Ram PhD Department of Systems Biology, University of Texas M.D. Anderson Cancer Center, Houston, Texas, USA
e-mail: pram@mdanderson.org

Ulrik Ringborg MD, PhD Cancer Center Karolinska, Department of Oncology-Pathology, Karolinska Institute, Stockholm, Sweden
e-mail: ulrik.ringborg@karolinska.se

Sergio Rutella MD, PhD Assistant Professor, Department of Haematology, Catholic University, Largo Agostino Gemelli 1, 00166 Rome, Italy and Director, Laboratory of Chronic Diseases and Systems Approaches, IRCCS San Raffaele Pisana, Via di Valcannuta 247, 00166 Rome, Italy
e-mail: srutella@rm.unicatt.it

Johan Rung PhD EMBL—European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1SD, Cambridge, UK
e-mail: johan@ebi.ac.uk

Marcus Schmidt PhD Department of Obstetrics and Gynecology, Medical School, University of Mainz, Mainz, Germany
e-mail: marcus.schmidt@frauen.klinik.uni-mainz.de

Dr. Michal R. Schweiger Max Planck Institute for Molecular Genetics, Department of Vertebrate Genomics, Berlin, Germany
e-mail: mschweig@molgen.mpg.de

Christopher Snell Physiomics Plc, Magdalen Centre, Oxford Science Park, Oxford, OX4 4GA, UK
e-mail: chris.m.snell@gmail.com

Joanna D. Stewart PhD Systems Toxicology, IfADo-Leibniz Research Centre for Working Environment and Human Factors, Technical University of Dortmund, Dortmund, Germany
e-mail: stewart@ifado.de

Adel B. Tabchy PhD Hematology and Oncology, University of Texas M.D. Anderson Cancer Center, Houston, Texas, USA
e-mail: atabchy@hotmail.com

Sandra Tomaszek MD Department of Thoracic Surgery, University Hospital Zurich, Raemistr. 101, 8091 Zurich, Switzerland
e-mail: sandra.tomaszek@usz.ch

Anna Tramontano PhD Department of Physics, Sapienza University of Rome, P.le Aldo Moro, 5, 00185 Rome, Italy
e-mail: Anna.Tramontano@uniroma1.it

Alfonso Valencia PhD Structural Biology and Biocomputing Programme of the Spanish National Cancer Research Centre (CNIO) and Spanish National Bioinformatics Institute (INB), Madrid, Spain
e-mail: valencia@cnio.es

Julio Vera PhD Department of Systems Biology and Bioinformatics, University of Rostock, 18051 Rostock, Germany
e-mail: julio.vera@uni-rostock.de

Dennis A. Wigle MD, PhD Division of Thoracic Surgery, Mayo Clinic Cancer Center, Rochester, Minnesota, USA
e-mail: wigle.dennis@mayo.edu

Olaf Wolkenhauer PhD Systems Biology and Informatics, University of Rostock, Rostock, Germany
e-mail: olaf.wolkenhauer@uni-rostock.de

Toru Yao PhD Genomic Sciences Center, RIKEN, Yokohama (Major), Japan
e-mail: yao@riken.jp

Boris Zhivotovsky PhD Institute of Environmental Medicine, Division of Toxicology, Karolinska Institutet, 17177, Stockholm, Sweden
e-mail: Boris.Zhivotovsky@ki.se

Part I
Introduction and Background

Chapter 1
Introduction to Systems Approaches to Cancer

Frederick B. Marcus and Alfredo Cesario

Abstract Despite major advances in research into and treatment of cancer, there remain exceptional difficulties arising from the disease's complexity and variability. These could be eased by integrated systems approaches to biology, bioinformatics and medicine. Such approaches have already proved successful in optimizing cancer-related research and clinical applications, building upon existing practice. In particular, they enhance the quantitative basis of knowledge and decisions; organize information in more accessible and applicable ways; extend understanding of complicated biological interactions; and provide numerical test beds for planning and implementing biomarker and drug development and treatment strategies.

1.1 Cancer and Systems Approaches

1.1.1 Nature and Causes of Cancer

Cancer is still a devastating disease after decades of research. Jemal (2009) estimates that one in four deaths in the United States is due to cancer. The most common (about half of all) cancers are prostate (men), breast (women), lung-bronchial, and colo-rectal. See also Cancer Facts (2010). Cancer research has a very high priority due to the impressive death toll, the enormous human suffering and the very high economic costs it entails. Progress has been made in the treatment of rare childhood cancers, in the early detection of some solid adulthood neoplasms (aided by breast, cervical and colonic cancer screening methodologies), in the identification of environmental factors with mutagenic potential (e.g. pollution and tobacco smoking), and in the improvement of long-term survival rates in some cancers. Disappointingly, little overall statistical change has been observed since 1975 in the global population-normalized death rates from all cancers (Jemal 2009).

If we consider the success stories of innovative cancer treatments (e.g. Trastuzumab-Herceptin© or Imatinib-Glivec©), these are effective in only a fraction of

F. B. Marcus (✉)
Advanced Therapies and Systems Medicine, Health Research Directorate,
Directorate General for Research, European Commission, Brussels, Belgium
e-mail: frederick.marcus@ec.europa.eu; frederick.b.marcus@gmail.com

A. Cesario, F. B. Marcus (eds.), *Cancer Systems Biology, Bioinformatics and Medicine,*
DOI 10.1007/978-94-007-1567-7_1,

patients (Slamon et al. 1987). Individual molecular characteristics of the disease, different from historical classification criteria (histology and staging), have a powerful stratification potential when the efficacy of treatment is concerned. Unfortunately, this evidence has sometimes been detected *ex post* (after the fact), where large scale clinical trials have delivered unexpected negative results, because of patient-to-patient differences in the molecular characteristics of neoplastic cell populations (INTACT1 and INTACT2).

Because cancer is a complicated, multistage disease (Cassidy et al. 2002), a more systematic approach is needed for understanding and improving cancer treatment (Cesario et al. 2004; Weinberg 2007). Its various mechanisms of onset and progression involve the biology and genetics of cells, tissues and organs, where a multi-role evolution occurs within a multidimensional framework, involving (as examples) tumour viruses, cellular oncogenes, growth factors and their receptors, cytoplasmic signalling circuitry, cell cycle controls, tumour suppressor genes, cell death mechanisms such as apoptosis, cell immortalization, tumorigenesis (also multistep), senescence, genomic integrity, developmental mechanisms, angiogenesis, lymphangiogenesis, metastasis and tumour immunology. The two main causes for cancer initiation are genetic predisposition and environmental influence. The processes of infection and inflammation are strongly interconnected in gene to environment dynamics. However, on a more analytical and molecular level, the ontogeny of cancer is not obvious. Both clinical as well as basic research suggests that cancer is the result of the accumulation and interaction of many factors that promote tumour initiation and subsequently conduce to growth and metastasis.

1.1.2 The Progression of Cancer

The key principles governing cancer progression were described by Hanahan and Weinberg (2000) and extended by Hahn and Weinberg (2002). Hanahan and Weinberg propose ‘that the vast catalogue of cancer cell genotypes is a manifestation of six essential alterations (“hallmarks”) in cell physiology that collectively dictate malignant growth:

1. Self-sufficiency in growth signals;
2. Insensitivity to growth-inhibitory (antigrowth) signals;
3. Evasion of programmed cell death (apoptosis);
4. Limitless replicate potential;
5. Sustained angiogenesis, and
6. Tissue invasion and metastasis.’

They also highlight genome instability as an enabling characteristic, including the variability of pathways leading to cancer, and the multiplicity of cell types within tumours. A seventh hallmark of cancer, cancer-related inflammation, has been proposed by Colotta et al. (2009).

1.1.3 Cancer: Clinical Background and Key Challenges

Systems approaches in biology, bioinformatics and medical applications should significantly enhance our understanding and its translation to applications. Conversely, information coming from clinical successes and failures (reverse translation) is needed to optimize model development. Despite the bewildering variety of cancer types, classifications, stages, diagnosis and treatments (see Chap. 2), many common areas exist which are amenable to modelling. However, field experience teaches that at the clinical level, strong divergences are in place which account for the vast heterogeneity in outcomes; hence detailed modelling will always require continuous iterations.

Investigative pathology remains the principal approach for classifying and grouping cancer tumours, leading to a coherent clinical approach. Whereas such classifications formerly involved microscope observation in histology and cytology, there is a modern tendency to consider cancer research purely at the sub-cellular level. This disregards information essential for classification and investigation, especially in terms of cells selected for examination of their intra- and intercellular reactions, most particularly during tumour growth and subsequent metastasis. A systems approach must therefore take full account of pathology observations.

It is nowadays commonly accepted that gene expression profiling (as obtained with microarray technology) may lead to the refining of classification criteria (Liu 2003; ICGC 2010), The greatest advances in terms of overall rate of cure have been recorded in some subtypes of lymphomas and leukaemias, and in breast cancer, where information derived from gene expression analysis has identified new diagnostic categories not obtainable from standard pathology. These subcategories often have distinct prognostic profiles. The same investigations have provided useful information for understanding of the cellular lineage of cancer types, highlighted the importance of biochemical pathways in determining the expression phenotype, and identified potential new diagnostic markers and therapeutic targets.

Taking as an example the genetic level of hematologic neoplasms, the delineation of various chromosomal translocations allows a clinically relevant classification of leukaemias and lymphomas (Chan 2001) and (Sen et al. 2002). In this context, genetic analysis further helps to obtain more precise (and clinically relevant) classification categories (Singer 1999). All of these classification criteria have led to useful stratification in the clinical decision-making process for systemic therapy (Abeloff et al. 2008, Chap. 30), and it is now common practice to prescribe agents that specifically target various cancer cell-specific pathways. On a molecular level, chemotherapy approaches are becoming very specific and tailored for each (sub) type of cancer. A systems approach is needed to further support this systematic prescription of treatments. Cell energetic metabolism, for example, is deemed crucial for the survival of a cancer cell, and targeting mitochondria, in the context of a comprehensive approach, has proved to be appropriate in the experimental model (Gogvadze et al. 2009).

Molecular targeted drugs aimed at signalling pathways or key parts of the cell cycle ought in principle to be of general validity, but in practice have proved useful

only in particular cancers. Therefore, the process of validation and prioritization of the highlighted and selected molecular targets, for any therapeutic development and implementation, depends on several criteria (chemical, biological and clinical) that in turn, need to be evaluated, integrated and harmonized towards a final classifier value. For example, the optimization of the properties of the kinase inhibitors identified by high-throughput screening (HTS) (Aherne et al. 2002) has led to perhaps the best-known example of a mechanism-based, genome-targeted cancer drug, Imatinib-Gleevec©. This compound inhibits the constitutive kinase activity of the Bcr-Abl oncoprotein responsible for driving malignancy in chronic myeloid leukaemia (Druker 2002). Modulating this mechanism by Imatinib has proved to be useful in several other forms of cancer. This experience, along with others, fosters and supports the adoption of a thorough analysis of medical mechanisms of action, using HTS interpreted by numerical modelling as a means of identifying those cancers (and neoplasms) for which a particular drug is most likely to benefit patients.

The clinician can then have a wide range of inputs to the decision matrix regarding therapeutic options for patients by using a systems approach to provide a coherent basis for an informed integration process. It is increasingly recognized (Cellis 2008) that treatment optimization relies upon the development of discovery-driven translational research processes that use knowledge-based, multidisciplinary approaches to derive new diagnostics and targeted treatments. The outcome of every strategy is then matched against the relevant benchmarks for that type of cancer and that particular patient. The cross fertilization leads to an upgraded clinical strategy.

1.1.4 Systems Biology Approaches to Cancer

Systems approaches involve mathematical and computational modelling in order to enhance understanding of a biological system through the interactions of its components. This understanding can then be expressed in qualitative and quantitative terms (Marcus 2008). The concept of a ‘biological system’ is as old as modern biology, e.g. the citric acid cycle in metabolism (Krebs 1953). However, the identification of components is only a first step. Once the pathway interactions are quantitatively known, they can be integrated into a dynamic simulation model, thereby constituting a systems approach. Similar techniques have long been used in physiology and in pharmacokinetic and pharmacodynamic modelling of the effects of medicines. Thus, a research process in any biological area can be streamlined, made more efficient and economical, by harnessing models derived from other related areas.

Key recent advances that make a modern systems approach necessary and ultimately possible involve ongoing development of high-throughput and/or quantitative technologies in many domains of biology, as well as of emerging computational technologies that allow new types of quantitative data to be generated and handled

with high power, precision and resolution. Furthermore, recent developments in complex systems theory provide a solid mathematical framework embodied in increasingly user-friendly and efficient tools needed to understand some of the dynamical phenomena observed in the living world. Analysis of just a small fraction of the available data has led to the realization that understanding of biology, health, disease and medicine requires modelling the processes involved. Even with our current and limited state of knowledge, impressive contributions have been made thereby to both fundamental understanding and direct applications to health. Carefully tailored models can already make useful predictions about disease processes and how medicines can be optimally applied, e.g. in cancer chronotherapy, constituting the beginnings of personalized medicine.

Researchers using systems approaches aim to identify areas that can combine experimental and clinical data with theoretical and computerized (*in silico*) models of cancer-related (e.g. signalling) pathways. An example is the modelling of a human metabolic network, with feedback incorporated from experimental data, to test responses to drugs and chemical treatments. Such models also have to address the multiscale and multilevel nature of a biological disease system involving molecular and cellular pathways.

1.1.5 Key Books and Reviews of Systems Approaches

Khalil and Hill (2005) show how systems biology for cancer can be implemented via mechanistic dynamical simulations and inferential data mining. Nagl's (2006) book, applying bioinformatics to cancer research, provides a description of various computational approaches, integrated informatics platforms, mathematical models, and computer simulation of tumours, structural bioinformatics, *in vivo* animal modelling, tissue resources and data. Hornberg et al. (2006) discuss how cancer must be regarded as a systems biology disease, since there is a great deal of interactive signalling between pathways, and cancer must be modelled at the network level. Sanga et al. (2006) give a review of mathematical modelling of cancer progression and response to chemotherapy. They argue that simulations of various aspects of cancer often need to function at multiple and interacting physiological levels. Materi and Wishart (2007) describe and assess the practical and theoretical underpinnings of commonly-used modelling approaches, including ordinary and partial differential equations, Petri nets, cellular automata, agent-based models and hybrid systems. Rosenfeld and Kapetanovic (2008) examine the role of systems approaches in cancer prevention. Price et al. (2008) focus on gene expression profile signatures which provide insights into the networks involved in the application of cancer diagnosis, patient stratification, and treatment management. They discuss emerging experimental technologies and computational modelling approaches to cancer stem cell challenges. Ptitsyn et al. (2008) investigate the identification of biomarkers via a systems approach relying on groups of interacting genes, leading to understanding of the cellular processes in metastatic progression. Seigneuric

et al. (2009) consider some of the challenges of data mining and integration of large amounts of expression measurements by cancer systems biology methods, as a chapter in Daskalaki (2009), where a wide range of systems biology applications to many fields of medicine is discussed. Faratian et al. (2009) examine how systems approaches reveal new strategies for personalizing cancer medicine, with examples in the area of drug resistance. Jain (2009) discusses personalized medicine approaches to cancer treatment in Chap. 10 of his book. Klipp et al. (2009) unify and demonstrate the extent of the availability of systems tools. Auffray et al. (2009) consider how the paradigm of systems medicine is determining the future of medical genomics and healthcare.

Most recently, Knox (2010) demonstrates new modelling techniques and nonlinear mathematics used to investigate gene-environment and epigenetic interactions and thereby to improve treatment efficacy. Kreeger and Lauffenburger (2010) discuss how significant advances can be obtained by applying computational modelling approaches to elucidate the pathways most critically involved in tumour formation and progression, impact of particular mutations on pathway operation, consequences of altered cell behaviour in tissue environments, and effects of molecular therapeutics. Wang (2010) gives descriptions of several specialized areas in cancer systems biology, presents an important complementary source to this book, and covers: theories of systems biology; overviews of basic cancer biology, genomics, cell signalling, and tumorigenesis, molecular mechanisms of cancer metastasis, molecular relationships between solid tumours, their microenvironments, and tumour blood vessels; computational tools and public data resources. The research of the Center for a Virtual Tumour (CViT) (2010), along with other worldwide projects, is summarized in a book by Deisboeck and Stamatakos (2010). They emphasize the need for multiscale approaches to modelling cancer in all its complexity. Mani (2010) presents a critical synthesis of an emerging class of methods that use systems biology, or networks of gene interactions inside the cell, to help characterize cancer progression. The yearly major International Conferences on Systems Biology, most recently ICSB (2010), have many papers dedicated to cancer research.

Extensive techniques for clinical oncology applications impose a strong requirement of proof that new approaches and new techniques can actually result in better understanding and translation to applications. The massive scale of existing cancer research is exemplified by Weinberg (2007) for biology and Abeloff et al. (2008) for clinical practice. Nevertheless, many types of research data, when considered individually, are not sufficiently integrated or interpreted to permit a proper understanding of the real situation in cells and in cancer progression. The conclusions of a recent workshop (Aebersold et al. 2009) support the idea that 'Systems Biology approaches can indeed advance cancer research, having already proved successful in a very wide variety of cancer-related areas, and are likely to prove superior to many current research strategies.' Important advances have already been demonstrated in areas such as models for gene regulation networks (Hache et al. 2009), cell growth (Godoy et al. 2009), cancer treatment (Fulda 2009) and effects of drug treatments (Hoffmann et al. 2008).

1.1.6 Importance of Legal and Ethical Considerations

Systems approaches offer many new possibilities; hence special attention must be paid to legal, supervisory and ethical aspects. This book does not discuss these aspects, since they are major areas themselves. Cancer prognosis, drug development and treatment are highly subject to legal requirements. New procedures need to be ethical, to conform to relevant laws, and to be fully tested within required procedures. As examples:

- New approaches to clinical trials and clinical practice need to be consistent with laws regulating clinical procedures, for example developing personalized medicine stratification in randomized clinical trials.
- Large scale patient genetic databases and interpreted results need to maintain confidentiality, when the genetic information itself can be a source of identification.
- Data and analysis may indicate a propensity to certain cancers, but the decision to use preventive strategies or not may have legal and health implications.
- New approaches to cancer treatment should be carefully handled and in a strictly ethical framework, due to the risk of inadvertently violating standards, which additionally might lead to public stigmatization and other consequences.

1.2 Laboratory, Clinical, Data and Educational Resources

1.2.1 Global Molecular and Cellular Measurement Technologies

A wide range of measurement technologies has been developed and used in cancer research and applied methodology (e.g. in tumour and tissue diagnostics) at the genomics and proteomics levels (Fisher 2007). Some are particularly suitable for adaptation to systems analysis and applications in cancer genetic variation studies (ICGC 2010); an example is DNA expression chips, invented by Lehrach et al. (1990) and subsequently greatly improved.

Cancer-specific genomic (Sysbiomed-cancer 2010) and proteomic (Fisher 2007) measurements can be used to guide and organize other areas and to support technology development programmes focussed on systems applications. Key cancer-relevant genomic technologies include: combined bisulphite restriction analysis (COBRA), sequencing, methylation analysis by DNA immunoprecipitation (MeDIP) and sequencing (MeDip-seq), chromatin immunoprecipitation (ChIP-chip) for histones, and 2nd generation bisulphite sequencing. Key proteomic and transcriptomics technologies include: DNA arrays, quantitative RT-qPCR, Xtag technology, and deep sequencing. New processes based on nanotechnology are being developed (Johnson et al. 2008).

Many measurements, which are dynamic, quantitative and space and time resolved, are being developed and improved. Especially rapid progress is being made in areas such as microscopy and imaging. The progress in these technologies is accompanied by related standards and protocols for rapidly exchanging data and information between different levels of cellular information, and for combining the data resources with theoretical models.

1.2.2 Cell Lines, Tissue Samples, Model Organisms, Biobanks

The diversity of cancer-relevant cell lines, tissue samples, model organisms, and biobanks highlights the need for systems approaches in order to apply these resources efficiently to critical questions of research and treatment. The relevance and appropriateness of human and model organisms, cell lines and tissues to cancer studies must be taken into account when considering quantitative validation and application. In particular, the requirements placed on those resources and their related databases for categorization and selection for modelling need to be specified in terms of relevance to investigations. Key areas include: intercellular/intracellular signalling, cellular microenvironment, metastases, cell cycle, cell death, cancer immunology, stem cells and differentiation, vascular/interstitial biology, DNA damage/repair, virus interaction, genetic factors, and carcinogenic progression.

Tests on cellular and *in vivo* animal models (including implants/transplants) constitute the final steps of the pre-clinical phase for assessing any drug tested for antineoplastic effect, and are essential prerequisites to the jump to human applications (Cell lines 2010). An *in silico* model provides inter-operability of methodologies for streamlining the test phase and optimizing the verification phase of drug design. For example, Klingmüller et al. (2006) developed standard operating procedures for the preparation and cultivation of primary mouse hepatocytes, which allowed reliable monitoring of the dynamic induction of signalling pathways.

In developing *in silico* modelling approaches (Sysbiomed-cancer 2010), it is often preferable to use high quality data from an artificial cell type, rather than mediocre data from a more realistic source. Good candidates are human-derived HEK293 and HeLa cell lines, with suitable growth characteristics, outstanding transfectability, and ease of performing RNAi experiments and homologous recombination. Unfortunately, since they are aneuploid and artificial, these single cell types are not appropriate for studying every type of response. Cell lines in the NCI60 (2010) panel, those from patients and differentiated iPS (induced pluripotent stem) cells prove superior in many areas.

Tissues and biopsies from patients would at first seem to provide the most relevant data, but patient-derived tumour cells and cell lines are often highly mixed, owing to the various stages of evolution of different and differentiated cells in the tumour and in metastases. These cells reflect the individual molecular pathology of patients and the stage of treatment and medication. Analyses become time consuming, with lower throughput and problems of availability and reproducibility. Even within individual types of cancer, a number of cell lines are available, all of which

may furnish important information. For example, a large number of non-small cell lung adenocarcinoma cell lines include Hop62, H650, HCC4006, HCC827, EKVX, HCC2935 and A549. (Tomshine et al. 2009).

As sources of cellular and patient derived material, the relevance of biobanks depends critically on the quality of the methodology and its standardization. In contrast to using cells directly from patient biopsies, Yuille et al. (2008) note that existing biobanks are well-organized resources comprising biological samples and associated information that are accessible to scientific investigation, where millions of samples with related data are held in different types of collections. Therefore, the full implementation of a systems approach requires careful integration of biobank resources into existing modelling and drug testing procedures, involving careful and consistent choices of cells, cell lines, animal models and patient tissues.

1.2.3 Expression and Genetic Variation Databases for Cancer Research

Major initiatives are in progress to gather wide ranges of germ-line (GEN2PHEN 2010) and cancer-relevant somatic genetic variation data. Somatic initiatives include the International Cancer Genome Consortium (ICGC 2010) following on from The Cancer Genome Atlas (TCGA 2010), based on genetic sequencing technology (Sanger Genetics 2010) and on expression profiling (Parkinson et al. 2009). This type of genetic variation data needs to be taken and restructured to provide optimal input for modelling cancer and its response to drugs.

To further the integration of clinical and medical resources, the BioSapiens-WP109 (2010) project has established methods for the analysis of functional consequences of cancer-associated oncogene mutations. This integration involved close collaboration with the (CGP) Cancer Genome Project (2010) of the Wellcome Trust Sanger Institute (WTSI 2010). The CGP uses the human genome sequence and high throughput mutation detection techniques to identify somatically acquired sequence variants/mutations and hence identify genes critical in the development of human cancers. These projects involve major collaborations with the USA project TCGA (2010), the Cancer Genome Atlas, of the NCI (2010) and NHGRI (2010).

The TCGA (2010) project includes an integrated network of clinical sites, core resources, specialized genome characterization and sequencing centres, working together to select genes and regions in order to drive high-throughput cancer genome sequencing. The TCGA and ICGC teams apply resources from caBIG (2010), the Cancer Biomedical Information Grid. Resources from caBIG include common data elements, metadata, and middleware to facilitate interactions among distributed databases.

1.2.4 Education and Research Infrastructures

Education is a key component in the success of any research activity involving systems approaches to cancer. A balance needs to be found between interdisciplinarity

and narrower specialist knowledge. Ideally, future systems biologists would need to be expert in wet lab work and computer modelling, but many educational programmes do not attempt this highly ambitious goal. Similarly, clinicians and biologists should have some familiarity with each other's disciplines and with modelling. Many systems-oriented educational programmes concentrate on special areas of expertise, with an admixture of broader interdisciplinary introductory courses, often linked to research projects and educational curricula. Examples of systems-oriented PhD programmes are to be found at the Massachusetts Institute of Technology (MIT-CSB 2010) and the University of Oxford (Oxford-SB 2010). The MIT programme integrates coursework and research opportunities in biology, engineering, mathematics, microsystems, and computer science with interdisciplinary courses in computational and systems biology developed for this programme. Integrative cancer research is pursued at the MIT-Koch (2011) centre. Coursework is also made available worldwide via MIT-opencourseware (2011), including systems biology courses. Graduates of the programme are taught how to develop original methods, make discoveries, and establish new paradigms. At Oxford, the Systems Biology Programme is associated with the Oxford Centre for Integrative Systems Biology (OCISB 2010) and with several core departments: Biochemistry, Chemistry, the Mathematical Institute, the Computing Laboratory, and the Weatherall Institute for Molecular Medicine (itself founded to support translational research). This close alliance between several major interdisciplinary centres and departments gives students access to world-leading research in life and physical sciences.

Cancer research institutions and their infrastructures, including central data repository and analysis capabilities, when linked to major educational resources, provide an integrated means of implementing systems approaches. For example, the Spanish National Cancer Institute (CNIO Training 2010) offers an extensive theoretical and practical programme, at both pre- and postdoctoral levels, as well as training in molecular pathology and familial cancer for medical residents and for those who hold a master's degree in other biomedical areas. The training programme also includes practical laboratory courses for undergraduate students and future technicians. This range of training in systems biology, in conjunction with direct exposure to cancer research and clinical application, equips and qualifies young researchers to make important individual contributions, and to work effectively in interdisciplinary teams.

1.3 Bioinformatics and Systems Biology Analysis

1.3.1 Mathematical Tools in Cancer Signalling

Mathematical models, both analytical and computational, give detailed insights into the dynamics of cancer-relevant pathways and systems, as well as providing mathematical solutions in several key areas, including genetic mutation, cancer tumours

and carcinogenesis, malignant growth and invasion, and therapy (Nagl 2006). The topic of dynamic modelling by analytical means alone is comprehensively treated by Johnston et al. (2007), for cell population dynamics in colorectal cancer. Once the cancer has been initiated, a succession of genetic mutations or epigenetic changes leads to the overcoming of homeostasis in the colonic crypt, and subsequently to unbounded growth. The authors consider the dynamics of a single colorectal crypt by using a compartmental approach which accounts for populations of stem cells, differentiated cells, and transit cells. They are thereby enabled to model increases in cell renewal leading to the growth of cancers.

Other theoretical approaches, including deterministic and stochastic models (Ullah and Wolkenhauer 2007), have played a significant role in hypothesis generation. Many aspects of cancer and carcinogenesis require multiscale modelling of intracellular modules. Genetic mutation simulations are important for understanding genetic instability (Beckman and Loeb 2005). Combinatory models are often required to analyse epidemiologic and molecular biology data.

1.3.2 Computational Tools

Sophisticated well-adapted computational tools and techniques for numerical analysis of data are available for implementing a systems approach to cancer research and applications. The numerical modelling of molecular pathways and physiological systems is described in detail by Klipp et al. (2009). Many of these models have been incorporated into toolboxes for research and clinical applications (ENFIN 2010). A key tool is Reactome (2010), a curated resource of core pathways and reactions in the biology of humans. Information is cross-referenced to the NCBI Entrez Gene, Ensembl and UniProt databases, the UCSC and HapMap Genome Browsers, the KEGG Compound and ChEBI small molecule databases, PubMed, and GO. Tools for data analysis include Skypainter and Biomart. An example is given by Vera et al. (2008), in which the amplification and responsiveness of the JAK2-STAT5 pathway using a kinetic model is investigated.

Databases play a central role in cancer research at the cellular level (caBIG 2010; Cancer Genome Project 2010; DTU-CBS 2010; EBI 2010; EMBRACE 2010). The Cancer Genome Project database names include: cancer gene census, mutated genes causally implicated in human cancer, catalogue of somatic mutations in cancer, CGP resequencing studies, somatic mutations from systematic large scale resequencing of genes in human cancers, resequencing of known cancer genes and other analyses of human cancer cell lines, and copy number analysis in cancer.

Data is available on copy number and loss of heterozygosity in cancer cell lines and primary tumours. Clinical, medical and epidemiological databases and records (PDQ 2010) comprise areas such as adult and paediatric treatment, supportive and palliative care, screening/detection, prevention, genetics, complementary and alternative medicine. There are also a number of important databases and

facilities applicable to cancer research which span the area between biological and clinical data. An example is the International ImMunoGeneTics information system (IMGT® 2000) for immunology, which comprises resources such as nucleotide sequences of IG and TR, sequences of the human MHC, oligonucleotides (primers) of IG and TR, nomenclature for IG and TR genes and 3D structures. See also Immunogrid (2010).

1.3.3 The Hallmarks of Cancer Revisited Through Modelling

Models are available for several key cellular pathways (Reactome 2010) and processes of central importance for understanding the hallmarks of cancer. This analysis is a key element in developing the concept of systems medicine (Auffray et al. 2009) applied to cancer research. The modelling is based on new laboratory data integrated with existing databases, including those related to various cancer stages. The process of modelling these characteristics (Aebersold et al. 2009) includes parallel signal transduction; nonlinear effects of regulatory feedbacks; pathway cross-talk and non-stationary biochemical processes; determinations for normal and cancerous cells; cell-context specific molecular interaction maps in cancer (Cancer Interactomes) and cellular network-based contexts.

As detailed examples, the COSBICS (2010) project team analysed two important signalling systems, the Ras/Raf/MEK/ERK and JAK/STAT pathways. For mitogen-activated protein kinase (MAPK) signalling pathways in cancer, Dhillon et al. (2007) demonstrate that cancer can be perceived as a disease of communication between and within cells. Cancerous mutations in MAPK pathways mostly affect Ras and B-Raf in the extracellular signal-regulated kinase pathway. The balance and integration between signals may widely vary in different tumours, but are important for the outcome and the sensitivity to drug therapy. This principle has been analysed and described by Kim et al. (2007) who noted that the Wnt and the extracellular signal regulated-kinase (ERK) pathways are both involved in the pathogenesis of various types of cancers. Cellular pathways are also analysed and mapped by functional interaction proteomics (Kriegsheim et al. 2008).

1.3.4 Analysis of Cell Death Pathways in Cancer: The Role of Collaborative and Interdisciplinary Research

Cell cycle deregulation and apoptosis inhibition are widely recognised as primary determinants in the induction and progression of cancer. The two main characteristics of all neoplastic cells are abnormal proliferation and aneuploidy, easily recognizable as direct consequences of cell cycle deregulation. The DIAMONDS (2010) team has developed a systems analysis of the cell cycle. This project has created

a firm basis for a high-throughput functional analysis of findings and hypotheses. In this context, Hainaut and Wiman (2007) have described the central role of p53, which triggers apoptosis in cancer.

Cell death pathways constitute key and highly complicated cancer mechanisms. The comprehensive modelling needed to implement a full systems approach currently requires a major collaborative effort. A quantitative understanding of the cell death pathways involves complex conceptual models of cell-fate decision, requiring an understanding of mutant phenotypes. A further prerequisite is selection of those cancers susceptible to integrated systems approaches, *i.e.* where there are appropriate clinical measurements and resources available. For interdisciplinary integration of knowledge, a common language is required for describing pathways and integrating different kinds of models from various repositories. The APO-SYS (2010) project directly addresses apoptosis with a systems biology approach involving 22 collaborating research institutions around Europe. The team implements an integrated approach of experimental biology, data mining, mathematical modelling, biostatistics, systems engineering and molecular medicine, in order to investigate lethal signal transduction pathways leading to apoptotic or non-apoptotic (necrotic, autophagic, mitotic) cell death. This has applications for breast and other cancers, including analysis of clinical trials.

A description of collaborative research programmes in cancer and related areas is presented by Marcus (2008). Further updates are available from Kyriakopoulou (2009) and most recently by Kyriakopoulou (2010) in the Workshop on Systems Biology to Systems Medicine. Outlines of all European Commission (EC) funded projects are accessible via Cordis Projects (2010). In the EC's Framework Programme for Research (FP7 2010), there were 42 collaborative projects funded in systems biology or systems medicine, funded or under negotiation; several of these projects focus on cancer and/or related processes. Projects such as Cancersys (2010) target the signalling aspects of liver cancer. Other projects have important elements relevant to systems approaches to cancer research: ACGT (2010), Angiotargeting (2010), Attack (2010), BioSapiens (2010), BioSim (2010), Brecosm (2010), Combio (2010), Cosbics (2010), Diamonds (2010), DNA Repair (2010), EMBRACE (2010), Enfin (2010), ESBIC-D (2010), Lymphangiogenomics (2010), Mismatch-2model (2010), Mutp53 (2010), Sybilla (2009), Trireme (2010), Tumour-Host Genomics (2010), Unicellsys (2010), Valapodyn (2010).

Most recently, the EC funded three new cancer systems biology projects (Kyriakopoulou 2010), of the same large scale as Apo-sys. The cancers covered (with project acronyms and references) are: colorectal (Syscol 2011), embryonic (Asset 2011), and liver (Modhep 2011). These new large projects, together with the others, especially Apo-sys and Cancersys, will create a world-leading programme in implementing systems analysis methods in cancer, to be closely coupled with clinical work. Information about the entire project system, from calls to proposal to negotiation and operation, is available from the European Commission official and comprehensive website CORDIS (2010), and discussed in Marcus (2008, Chap. 10) The USA ICBP (2010) Integrative Cancer Biology Programme of the NCI (2010)

also supports major interdisciplinary and collaborative research programmes, especially via its dedicated centres and the caBIG (2010) programme.

1.3.5 Approaches to Cancer Progression Outcomes

Cancers of the liver or breast may serve as illustrative models for a full systems analysis, leading to advances in understanding and, ideally, development of improved patient treatment. Such an approach is essential in dealing with various aspects of cancer progression, especially in the presence of drug treatment (Hengstler et al. 2006). Conditional mouse tumour models have been established that allow the expression of specific oncogenes controlling tumour growth to be switched off, resulting in useful insights for oncogene-blocking therapies. The initiation and progression of recurrent tumours is triggered/regulated by oncogenes different from those involved in the onset of the primary tumours, possibly explaining why clinical inhibitors of a broader spectrum of protein kinases (so-called 'dirty inhibitors') may be superior to highly specific ones.

The Cancersys (2010) project uses a multi-scale model for two major signalling pathways (beta-catenin and ras) involved in hepatocellular carcinoma. The model integrates different levels, ranging from the cellular level to the tissue and organ level, often based on major research programmes devoted to understanding the function of healthy tissue (Hepatosys 2010). The project focuses on studying near term human applications in model organisms. These methodologies are supported by new developments in proteomics (Abeloff et al. 2008). By defining the collective protein-protein interactions in a cancer cell, functional relationships between disease-promoting genes are examined to provide novel candidates for intervention, and to investigate neoplastic initiation and progression. These efforts are complemented by direct genetic variation analysis designed to identify disease-promoting genes that are suitable subjects for systems biology analyses.

Systems approaches help to develop innovative therapeutic strategies aimed at modulating the immune system (De Duve 2009; Sybilla 2009), by linking to biomarker simulation, genomic and environmental effects. Other investigations involve the cancer-ageing link, with a focus on particular cancers, including comparisons of mouse models, cell lines and human samples (Cancer Research UK 2006; Mukherjee 2005).

1.3.6 Modelling at the Physiological and Tumour Level

A focused evolutionary approach to cancer progression may result in new paradigms in cancer treatment (Gatenby 2009). A goal in cancer therapy, similar to that of antimicrobial treatments, is that of killing as many tumour cells as possible. The assumption is that this will, at best, cure the disease, and at least keep the patient

alive as long as possible. However, the principles for successful cancer therapy might lie not in magic bullets (as for antibiotics in microbiology) but in adapting the evolutionary dynamics of applied ecology. Support for this idea comes from *in vivo* experiments, computer simulations and mathematical models of tumoral evolutionary dynamics. Efforts to eliminate cancers may actually hasten the emergence of resistance and tumour recurrence, thus reducing a patient's chances of survival. This controversial approach lends itself to more systematic and case-by-case analysis to indicate the most pragmatic approaches in individual cases.

Models of multicellular tumours have become fairly advanced. Detailed descriptions of multiscale modelling have been made of colorectal cancer by Johnston et al. (2007), and of solid tumour growth by Maini and Gatenby (2006), Nagl (2006), Byrne et al. (2008) and Owen et al. (2009). The Center for the Development of a Virtual Tumor (CViT 2010) has assembled and linked a wide range of models, including tumour growth, vascular growth, and multiscale modelling of specific cancers via a cancer virtual tumour platform. The CViT collaboration is developing a generic module-based toolkit for modelling and simulating selected cancer types of interest, such as breast, brain, and melanoma, following a complex systems approach. Combined with biomedical data, this toolkit has significant value for cancer research as it allows the study of cancer initiation and such critically linked progression features as invasion, angiogenesis and metastasis.

1.4 Diagnosis and Treatment Applications

1.4.1 Diagnostic and Prognostic Cancer Biomarkers

Reliable biomarkers are crucial for correct diagnosis and prognostic determination. Biomarkers are essential for detecting the presence of cancer, and evaluating its progression and treatment responses. They rely on annotated high-quality clinical samples and on well-structured biobanks, ideally coupled with updated cancer registries. Using biomarkers successfully depends upon measurements in tissues and body fluids to assemble a profile of gene expression proteins and other products. An essential first step is the identification, standardization and validation of effective biomarkers. Biological and modelling platforms are needed to assist in the processes of identification and prioritization of potential biomarkers. It is hoped that very small colonies of cancer cells can be detected at an early stage with new nanodetection technologies by means of their secretions of proteins and other substances into the urine, blood or breath, i.e. 'early stage biomarkers'.

However, identification and interpretation of biomarkers is complex, as a large number of false positives dominate the field (Harris 2005). McShane et al. (2005) have produced a consensus on how to design and present biomarker studies. The development of well-characterized tissue microarrays has expedited the analysis of biomarkers and pathways. Many new techniques are available, including extraction

of DNA for comparative genome hybridization and extraction of RNA. This allows reverse transcription and quantitative polymerase chain reaction (rt-PCR) assays to derive gene profiles (Paik et al. 2004). Use of biomarkers to predict which patients will benefit from increasingly complex and expensive therapies is potentially of high utility. Patients with poor prognosis can be selected for special therapeutic approaches. Biomarkers applicable in multi-cancer situations are being identified, which are good targets for modelling validation, where pathway analysis reveals important functional categories of tumorigenesis (Tseng et al. 2009). Such an approach is needed to achieve the goal of personalized medicine, with the selection of the correct therapy tailored to the molecular pathways in the tumour.

As always, there is the challenge of identification of primary biomarkers in biopsies, and of secondary biomarkers in more accessible body fluids. A further challenge is to determine organ-specific blood markers, for which systems approaches will be essential. Pharmacokinetic and pharmacodynamic simulations need to be applied to biomarkers, to determine concentrations and timings for detection. The application of bioinformatics is also essential for understanding biomarkers detection, for example by determining binding of biomarkers to probes. A systems approach to the use of biomarkers in the diagnostic and therapeutic setting should optimize their use in the clinical setting.

1.4.2 Cancer Drug Development

Consistent pharmacological and medical data are required for appropriate drug development. The quantitative simulation of pathways greatly enhances the identification and characterization of drug targets and effects (Chassagnole et al. 2006). Complex cell cycle models (Physiomics 2010) are used to develop anti-cancer therapeutics; these can predict the cytotoxicity of various compounds (without model adaptation to any specific cancer cell types) and may be matched with experimentally determined data. Understanding of drug effects can be greatly improved by using models where biomarkers are quantified, involving vertically integrated multiscale (VIMS) models (Physiomics Virtual Tumour 2010). This connection allows the possibility of simulating combination therapies. Modelling integrates clinical measurements and correlated responses for better examination of anti-cancer drug effects, in order to develop tests at the cellular level and later in the preclinical animal model. It provides an important preliminary step before advancing to early-stage clinical trials. Trial results, through a reverse translation process, can in turn yield precious data on such topics as tumour resistance to (the selected) therapy. Optimata (2010) is also developing relevant models.

Bioinformatics tools have long been essential in the drug development process, and become more powerful when applied in a systems approach context, facilitating elucidation of the relationships between cancer drugs and disease genes (mutations and polymorphisms) and exploration of defective proteins as targets for drug development (Arnesen et al. 2008).

Biomarker-driven drug development becomes more possible with an integrated approach (Carden et al. 2009), which includes:

- cancer-specific diagnostic and prognostic markers
- patient stratification by genotyping
- use of predictive markers for efficacy
- surrogate 'markers' (end-points) for long-term drug efficacy
- predictive tumour genotyping for efficacy (responders/non-responders and safety).

Compared with conventional chemotherapy, rationally designed molecularly targeted agents may be more likely to exert antitumoral activity in selected tumour subgroups when driven by the oncogenic signals targeted by these compounds and a different side-effect profile.

1.4.3 Cancer Chronotherapy

Eriguchi et al. (2003) described cancer chronotherapy as a novel and logical therapy in which anti-cancer drugs are administered with optimal timing according to circadian rhythms of anti-cancer action and adverse effects on normal cells. Advances in chronobiology have identified the suprachiasmatic nucleus (SCN) as both the centre of biological rhythms and the area in which clock genes act to generate and coordinate them. These findings have led to the development of cancer chronotherapy. Clinically, patients with advanced gastrointestinal cancer have been treated by chronomodulated chemotherapy with good response. For colorectal cancer patients with unresectable liver metastases, chronotherapy with folinic acid has been reported to allow complete surgical resection of liver metastases, resulting in 39–50% 5-year survival. A systems approach to cancer chronotherapy has been developed by the BioSim (2010) project, in the form of detailed models for calculating the optimum timing for drug administration during cancer treatment: the results have been verified in clinical tests.

1.4.4 Clinical Applications of Systems Biology Approaches

The process of clinical decision-making is multidisciplinary in its nature, arising out of complex diagnostic procedures and the multiple options represented by various therapeutic strategies, including chemotherapy, radiation therapy and surgery. To complement the standard inputs from imaging, biomarkers, biopsies and pathology diagnoses, new and more reliable computational decision-making tools for cancer and its treatment are being developed, such as models for key systems of cells, tissues, body systems and the cancer patient.

One implementation (Mills 2009) is to translate understanding of the genetic and epigenetic defects that underlie the initiation and progression of cancer, into improved patient management, by linking cancer genetics, molecular diagnostics, genomics, proteomics and signal transduction to molecular therapeutics and individualized patient care. Targeting the underlying defects, and particularly signal transduction pathways in cancer cells, would ideally result in effective non-toxic molecular therapeutics which would improve the outcome in cancer patients. This process can be implemented by pharmacokinetic and pharmacodynamic modelling (Physiomics 2010), chronotherapy simulations, quantitative drug dose modelling, and combined therapy modelling. The integration of information from imaging (from cellular to organ level) and pathology provides additional data from which to construct the simulation environment.

Personalized cancer medicine (Jain 2009) has now been made possible by technological innovations coupled with physiological and pathology-enhanced determinations. In particular, genetic variation measurements and analysis, coupled with modelling tools, are now creating much greater possibilities of identifying and targeting treatment for both sub-populations and individuals. Auffray et al. (2009) refer to the combined advances in genetics and new measurement technologies which facilitate the development of personalized medicine; systems medicine, too, is facilitated by the modelling capabilities of systems biology, and their application to medical problems. The increased efficiency of DNA sequencing, in combination with extensive genetics measurements by the International Cancer Genome Consortium (ICGC 2010) and other genomics programmes, opens up the possibility of analysing a large number of individual genomes and transcriptomes. Complete reference proteomes and metabolomes are within reach, thanks to powerful analytical techniques based on chromatography, mass spectrometry and nuclear magnetic resonance. Simulations of the role of individual variations put personalized medicine on a much firmer basis, after using bioinformatics analyses and pathways simulations of the role of genetic variations. The formulation of treatment strategies needs, however, to be embedded in clinical trials, as well as tailored to individual patients.

Another key concept is the management of complexity in (chronic) cancer care. Not only is cancer itself a highly complicated multistage disease, but it also involves a wide range of complications due to the cancer itself and to side effects of treatments. Examples include the complexity represented by progression from inflammation related to chronic obstructive pulmonary diseases (COPD) to lung cancer, or the coexistence with cancer of diseases such as COPD, heart failure, and diabetes. Such an approach may pave a pragmatic way to wider applications of systems medicine.

1.4.5 Cancer Robustness and Therapy Strategies

Robustness (Kitano 2004) is a concept crucial to cancer progression, and it opens new perspectives in treatment. Cancer is robust because it continues to progress against our natural body defences and against external clinical intervention. Robust-

ness against perturbations is a property of complex systems. Its understanding in the context of cancer is crucial, and has major implications for choices of treatment strategies, since robustness in some areas goes along with fragility in other areas. Targeting these fragilities can lead to major insights and new strategies, including novel targets for drug development and applications. In particular, both Kitano (2004) and Gatenby (2009) argue that attacking a cancer tumour on the simple criterion of reducing its mass may well be a misplaced strategy. New concepts such as these provide new paradigms for developing less toxic and more effective cancer treatments, ideally resulting in improved patient outcomes.

1.5 Perspectives and Conclusions

1.5.1 Perspectives

Perspectives for the success of systems approaches depend on the development of specific supporting research and infrastructure. Review articles and workshops discuss the requirements and integrated perspectives for the field (Aebersold et al. 2009; Sysbiomed-cancer 2010; Auffray et al. 2009; Makarow 2008; FCSB 2008). Infrastructure requirements for systems biology in general are discussed by Cassman and Brunak (2007), who cite the need for clinical and data resources and modelling tools for their analysis.

Future systems approaches to the complexity represented by cancer detection and treatment will involve, for example, detection via nanotechnological methods (Heath et al. 2009), biomarkers, environmental factors, circadian rhythms, immune system, cancer vaccines, etc. The objectives of established and newly-started integrated and interdisciplinary projects already provide perspectives. Major initiatives are also starting in the application of physics methods (http://web.mit.edu/ki/research/centers/psoc.html) to cancer research.

Synthetic biology techniques and technologies (Anderson et al. 2006, 2007; Rowe 2009) offer prospects of major advances, especially when integrated into a systems and bioinformatics analysis approach (Yao 2002) involving applications to testing models, diagnosis, and drug development. This technical advance is occurring at the same time as a major increase in the understanding of atomic level processes studied through nanomeasurement and nanotechnology methods. Mismatch2model (2010) directly observes and performs quantitative systems modelling of DNA mismatch repair and its role in the maintenance of genomic stability and cancer avoidance.

1.5.2 Conclusion

Systems approaches have a major and highly favourable role to play in cancer research and clinical applications. Systems approaches, many of which are already

well established, offer real promise for rapid and major improvements in our understanding of cancer and its treatment.

References

Abeloff et al (2008) Abeloff's clinical oncology, 4th edn. Elsevier, Churchill Livingston, London

Aebersold R, Auffray C, Baney E, Barillot E, Brazma A, Brett C, Brunak S, Butte A, Califano A, Celis J, Cufer T, Ferrell J, Galas D, Gallahan D, Gatenby R, Goldbeter A, Hace N, Henney A, Hood L, Iyengar R, Jackson V, Kallioniemi O, Klingmuller U, Kolar P, Kolch W, Kyriakopoulou C, Laplace F, Lehrach H, Marcus F, Matrisian L, Nolan G, Pelkmans L, Potti A, Sander C, Seljak M, Singer D, Sorger P, Stunnenberg H, Superti-Furga G, Uhlen M, Vidal M, Weinstein J, Wigle D, Williams M, Wolkenhauer O, Zhivotovsky B, Zinovyev A, Zupan B (2009) Report on EU-USA workshop: how systems biology can advance cancer research (27 Oct 2008). Mol Oncol 3(1):9–17

Aherne GW, McDonald E, Workman P (2002) Finding the needle in the haystack: why high-throughput screening is good for your health. Breast Cancer Res 4:148–154

Anderson JC, Clarke EJ, Arkin AP, Voigt CA (2006) Environmentally controlled invasion of cancer cells by engineered bacteria. J Mol Biol 355(4):619–627

Anderson JC, Voigt CA, Arkin AP (2007) Environmental signal integration by a modular AND gate. Mol Syst Biol 3:133

ANGIOTARGETING (2010) Tumour angiogenesis research. http://www.uib.no/rg/trancancer/projects/angiotargeting. Accessed 18 Oct 2010

APO-SYS (2010) Apoptosis systems biology applied to cancer and AIDS. http://www.apo-sys.eu/. Accessed 18 Oct 2010

Arnesen T et al (2008) The protein acetyltransferase ARD1: a novel cancer drug target? Curr Cancer Drug Targets 8(7):545–553

Asset (2011) Analysis and striking the sensitivities of embryonal tumours. http://www.ucd.ie/sbi/research/areasofresearch/sbicollaborativeprojects/asset/. Accessed 30 Jan 2011

ATTACK (2010) Adoptive engineered T cell targeting to activate cancer killing. http://www.attack-cancer.org. Accessed 18 Oct 2010

Auffray C, Chen Z, Hood L (2009) Systems medicine: the future of medical genomics and healthcare. Genome Med 2009, 1:2. doi:10.1186/gm2

Beckman RA, Loeb LA (2005) Review—genetic instability in cancer: theory and experiment. Semin Cancer Biol 15:423–435

BioSapiens (2010) A European virtual institute for genome annotation. http://www.biosapiens.info. Accessed 18 Oct 2010. See also BioSapiens (2005) A European network for integrated genome annotation. Eur J Hum Genet 13:994–997

BioSapiens-WP 109 (2010) BioSapiens work package 109—cancer. http://www.biosapiens.info/page.php?page=package&pack=109&new=1. Accessed 18 Oct 2010

BioSim (2010) Biosiumulation—a new tool in drug development. http://biosim-network.eu/. Accessed 18 Oct 2010

BRECOSM (2010) Identification of molecular pathways that regulate the organ-specific metastasis of breast cancer. http://www.healthcompetence.eu/converis/publicweb/project/1693;jsessionid=397ae485b508dd6ec6d1b1106f19?show=Person. Accessed 18 Oct 2010

Byrne HM, Leeuwen IMM van, Owen MR, Alarcon T, Maini PK (2008) Multiscale modelling of solid tumour growth. In: Selected topics in cancer modeling. Genesis, evolution, immune competition, & therapy. Birkhauser, Boston, pp 449–473. ISBN 978-0-8176-4712-4

caBIG (2010) Cancer biomedical information grid (caBIG™) of the NCI (2010). https://cabig.nci.nih.gov/. Accessed 18 Oct 2010

Cancer Facts (2010) Cancer facts and figures 2009. http://www.cancer.org/docroot/PRO/content/PRO_1_1_Cancer_Statistics_2009_Presentation.asp. Accessed 18 Oct 2010
Cancer Genome Project (2010) The WTSI (2010) Cancer genome project. http://www.sanger.ac.uk/genetics/CGP/. Accessed 18 Oct 2010
Cancer Research UK (2006) "Virtual cancer patient" predicts how breast cancer patients respond to treatment. http://info.cancerresearchuk.org/news/archive/pressrelease/2006-10-10-virtual-cancer-patient-predicts-how-breast-cancer-patients-respond-to-treatment
Cancersys (2010) Cancersys project. http://www.ifado.de/cancersys/. Accessed 18 Oct 2010
Carden C et al (2009) From darkness to light with biomarkers in early clinical trials of cancer drugs. Clin Pharmacol Ther 85(2):131–133. Issn 0009-9236
Cassidy J, Bissett D, Spence RAJ (2002) Oxford handbook of oncology. Oxford University Press, Oxford
Cassman M, Brunak S (2007) The US-EC workshop on infrastructure needs for systems biology. http://ec.europa.eu/research/biotechnology/ec-us/index_en.html. Accessed 18 Oct 2010
Cellis J (2008) Editorial. Mol Oncol 2(1):1
Cell lines (2010) Cell lines from high-risk breast tissue. http://clinicaltrials.gov/ct2/show/NCT00032201. Accessed 18 Oct 2010
Cesario et al (2004) Non-small cell lung cancer: from cytotoxic systemic chemotherapy to molecularly targeted therapy. Curr Med Chem Anticancer Agents 4(3):231–245
Chan JK (2001) The new world health organization classification of lymphomas: the past, the present and the future. Hematol Oncol 19:129–150
Chassagnole C et al (2006) Using mammalian cell cycle simulation to interpret differential kinase inhibition in anti-tumour pharmaceutical development. Biosystems 83(2–3):91–97
CNIO Training (2010) Spanish national cancer institute. http://www.cnio.es/ing/cursos/. Accessed 18 Oct 2010
Colotta F, Allavena P, Sica A, Garlanda C, Mantovani A (2009) Cancer-related inflammation, the seventh hallmark of cancer: links to genetic instability. Carcinogenesis 30(7):1073–1081. Epub 2009 May 25
COMBIO (2010) An integrative approach to cellular signalling and control processes: bringing computational biology to the bench. http://combio.crg.es. Accessed 18 Oct 2010
CORDIS (2010) Community Research and Development Information Service (CORDIS) provides information on all EU-supported R&D activities. http://cordis.europa.eu/home_en.html. Accessed 18 Oct 2010
CORDIS Projects (2010) Cordis project reference systems. http://cordis.europa.eu/fp7/projects_en.html. Accessed 18 Oct 2010
COSBICS (2010) Computational systems biology of cell signalling. http://www.sbi.informatik.uni-rostock.de/cosbics/. Accessed 18 Oct 2010
CVIT (2010) The center for the development of a virtual tumour. https://www.cvit.org. Accessed 18 Oct 2010
Daskalaki A (2009) Handbook of research on systems biology applications in medicine. ISBN: 978-1-60566-076-9. doi: 10.4018/978-1-60566-076-9. IGI Global, Pennsylvania, pp 1–982
De Duve (2009) De Duve Institute. http://www.deduveinstitute.be/cancer_immunology.php
Deisboeck TS, Stamatakos G (2010) Multiscale cancer modelling. CRC press. ISBN 9781439814406.
Dhillon AS, Hagan S, Rath O, Kolch W (2007) MAPK kinase signalling pathways in cancer. Oncogene 26(22):3279–3290
DIAMONDS (2010) Dedicated integration and modelling of novel data and prior knowledge to enable systems biology. http://www.sbcellcycle.org. Accessed 18 Oct 2010
DNA Repair (2010) DNA damage response and repair mechanisms. http://www.dna-repair.nl. Accessed 18 Oct 2010
Druker BJ (2002) ST1571 (Gleevec) as a paradigm for cancer therapy. Trends Mol Med 8(Suppl):S14–S18
DTU-CBS Databases (2010) Danish technical university, centre for biological sequence analysis. http://www.cbs.dtu.dk/databases/. Accessed 18 Oct 2010

EBI (2010) European bioinformatics institute. http://www.ebi.ac.uk/. Accessed 18 Oct 2010
EMBRACE (2010) A European model for bioinformatics research and community education—bioinformatics grid. http://www.embracegrid.info. Accessed 18 Oct 2010
ENFIN (2010) Enabling systems biology PP6 (2007) project. http://www.enfin.org. Accessed 18 Oct 2010
ESBIC-D (2010) European systems biology initiative combating complex diseases. http://www.healthcompetence.eu/converis/publicweb/project/2128/. Accessed 18 Oct 2010
Eriguchi M, Levi F, Hisa T,Yanagie H, Nonaka Y, Takeda Y (2003) Chronotherapy for cancer. Biomed Pharmacother 57:92s–95s
Faratian D et al (2009). Systems biology reveals new strategies for personalizing cancer medicine and confirms the role of PTEN in resistance to trastuzumab. Cancer Res 69(16):6713–6720. http://www.csa.com. Accessed 18 Oct 2010
FCSB (2008) FCSB 2008 first future challenge for systems biology. http://systems-biology.org/conference/report/2008-calendar-1/000005.html
Fisher PB (2007) Cancer genomics and proteomics: methods and protocols. Humana Press, New York
FP7 (2010) EC framework programme for research. http://cordis.europa.eu/fp7/home_en.html. Accessed 18 Oct 2010
Fulda S (2009) Apoptosis pathways and their therapeutic exploitation in pancreatic cancer. J Cell Mol Med Jul;13(7):1221–7. Epub 2009 Mar 27
Gatenby R (2009) A change of strategy in the war on cancer. Nature 459:508–509
GEN2PHEN (2010) Genotype to phenotype databases. http://www.gen2phen.org. Accessed 18 Oct 2010
Godoy P et al (2009) Extracellular matrix modulates sensitivity of hepatocytes to fibroblastoid dedifferentiation and transforming growth factor beta-induced apoptosis. Hepatology 49(6):2031–2043
Gogvadze V, Orrenius S, Zhivotovsky B (2009) Mitochondria as targets for chemotherapy. Apoptosis 14(4):624–640
Hache H et al (2009) GeNGe: systematic generation of gene regulatory networks. Bioinformatics 25(9):1205–1207
Hahn WC, Weinberg RA (2002) Modelling the molecular circuity of cancer. Nat Rev Cancer 2(5):331–341
Hainaut P, Wiman KG (eds) (2007) 25 years of p53 research. Springer, Dordrecht
Hanahan D, Weinberg RA (2000) The hallmarks of cancer. Cell 100(1):57–70
Harris AL (2005) Editorial—REporting recommendations for tumour MARKer prognostic studies (REMARK). Br J Cancer 93:385–386. doi:10.1038/sj.bjc.6602730 http://www.bjcancer.com. Published online. Accessed 16 Aug 2005
Heath JR, Davis ME, Hood L (2009) Nanomedicine targets cancer. Sci Am 300(2):44–51
Hengstler JG et al (2006) Oncogene-blocking therapies: new insights from conditional mouse tumour models. Curr Cancer Drug Targets 6(7):603–612
HepatoSys (2010) German network systems biology hepatocyte programme. http://www.systembiologie.de/de/index.html. Accessed 18 Oct 2010
Hoffmann J et al (2008) Improved cellular pharmacokinetics and pharmacodynamics underlie the wide anticancer activity of sagopilone. Cancer Res 68(13):5301–5308
Hornberg JJ et al (2006) Cancer: a systems biology disease. Biosystems 83(2–3):81–90. http://www.csa.com. Accessed 18 Oct 2010
ICBP (2010) Integrative cancer biology programme of the NCI (2010) http://icbp.nci.nih.gov/. Accessed 18 Oct 2010
ICGC (2010) International cancer genome consortium. http://www.icgc.org. Accessed 18 Oct 2010
ICSB (2010) International conferences on systems biology. http://www.icsb2010.org.uk/. Accessed 18 Oct 2010
IMGT® (2010) The international ImMunoGeneTics information system. http://www.imgt.org. Accessed 18 Oct 2010

IMMUNOGRID (2010) The European virtual human immune system project. http://www.immunomics.eu/index.php?option=com_content&view=article&id=5%3Aimmunogrid-project-brief-description&catid=2%3Aimmunogrid&Itemid=10&lang=en. Accessed 18 Oct 2010

Jain KK (2009) Textbook of personalized medicine. Springer, New York

Jemal A (2009) Cancer statistics. CA Cancer J Clin 59:225–249. doi:10.3322/caac.20006

Johnson CJ, Zhukovsky N, Cass AE, Nagy JM (2008) Proteomics, nanotechnology and molecular diagnostics. Proteomics 8:715–730

Johnston MD, Edwards CM, Bodmer WF, Maini PK, Chapman SJ (2007) Mathematical modelling of cell population dynamics in the colonic crypt and in colorectal cancer. PNAS 104(10):4008–4013. See also http://www.pnas.org/cgi/reprint/104/10/4008.pdf. Accessed 18 Oct 2010

Khalil IG, Hill C (2005) Systems biology for cancer. Curr Opin Oncol 17(1):44–48. http://www.csa.com. Accessed 18 Oct 2010

Kim D, Rath O, Kolch W, Cho KH (2007) A hidden oncogenic positive feedback loop caused by crosstalk between Wnt and ERK Pathways. Oncogene 26(31):4571–4579

Kitano H (2004) Opinion: cancer as a robust system: implications for anticancer therapy. Nat Rev Cancer 4:227–235. doi:10.1038/nrc1300

Klingmüller U et al (2006) Primary mouse hepatocytes for systems biology approaches: a standardized in vitro system for modeling of signal transduction pathways. IEE Proc Syst Biol Nov;153(6):433–447

Klipp E, Liebermeister W, Herwig, WC, Kowald A, Lehrach H, Herwig R (2009) Systems biology—a textbook. Wiley-VCH, Weinheim

Knox SS (2010) From 'omics' to complex disease: a systems biology approach to gene-environment interactions in cancer. Cancer Cell Int 10:11–11. http://www.csa.com. Accessed 18 Oct 2010

Kreeger PK, Lauffenburger DA (2010) Cancer systems biology: a network modeling perspective. Carcinogenesis Jan;31(1):2–8. Epub 2009 Oct 27

Kriegsheim A, Preisinger C, Kolch W (2008) Mapping of signalling pathways by functional interaction proteomics. Methods Mol Biol 484:177–192

Krebs HA (1953) 1953 nobel prize in medicine or physiology. http://nobelprize.org/nobel_prizes/medicine/laureates/1953/index.html. Accessed 18 Oct 2010

Kyriakopoulou C (ed) (2009) The book of life. Published European commission. ftp://ftp.cordis.europa.eu/pub/fp7/docs/fungen-book_en.pdf

Kyriakopoulou C (ed) (2010) Workshop on from systems biology to systems medicine. European Commission, Brussels. http://ec.europa.eu/research/health/pdf/systems-medicine-booklet_en.pdf. Accessed 18 Oct 2010

Lehrach H et al (1990) Hybridization fingerprinting in genome mapping and sequencing. Genome analysis, vol 1: genetic and physical mapping. Cold Spring Harbor Laboratory Press, New York, pp 39–81

Liu ET (2003) Classification of cancers by expression profiling. Curr Opin Genet Dev 13(1):97–103

Lymphangiogenomics (2010) Genome-wide discovery and functional analysis of novel genes in lymphangiogenesis. http://www.lymphomic.org/. Accessed 18 Oct 2010

Maini PK, Gatenby RA (2006) Some mathematical modelling challenges and approaches in cancer. In: Nagl S (ed) Cancer bioinformatics: from therapy design to treatment. Wiley, New York, pp 95–107

Makarow M et al (2008) Advancing systems biology for medical applications. http://www.esf.org. Accessed 18 Oct 2010

Mani K (2010) Systems biology and personalized medicine in cancer. Curr Pharmacogenomics Personalized Med (Formerly Current Pharmacog) 8(1):64–72(9)

Marcus FB (2008) Bioinformatics and systems biology: collaborative research and resources. Springer, Berlin

Materi W, Wishart DS (2007) Computational systems biology in cancer: modeling methods and applications. Gene Regul Syst Biol Sep 17;1:91–110

McShane LM et al (2005) REporting recommendations for tumour MARKer prognostic studies (REMARK). Br J Cancer 93:387–391
Mills GB (2009) M.D. Anderson cancer center. http://www.mdanderson.org/education-and-research/departments-programs-and-labs/departments-and-divisions/systems-biology/chair-message.html. Accessed 18 Oct 2010
Mismatch2model (2010) Characterization and quantitative modelling of DNA mismatch repair and its role in the maintenance of genomic stability and cancer avoidance. http://www.mm2m.eu/index.html. Accessed 18 Oct 2010
MIT-CSB (2010) Computational and systems biology at MIT. http://csbi.mit.edu/education/core.html. Accessed 18 Oct 2010
MIT-Koch (2011) Koch institute for integrative cancer research. MIT. http://web.mit.edu/ki/. Accessed 5 Jan 2011
MIT-opencourseware (2011) MIT open courseware. http://ocw.mit.edu. Accessed 4 Jan 2011
Modhep (2011) Systems biology of liver cancer: an integrative genomic-epigenomic approach. http://cordis.europa.eu/fetch?CALLER=PROJ_EN&ACTION=D&DOC=16&CAT=PROJ&QUERY=01238acac6d1:af70:07e5f5e3&RCN=97663. Accessed 30 Jan 2011
Mukherjee A et al (2005) Virtual cancer patient: simulated biomathematical model for treatment personalisation in metastatic breast cancer (MBC). http://www.optimata.com/src/VIRTUAL%20CANCER%20PATIENT(1).pdf
Mutp53 (2010) Mutant p53 as a target for cancer treatment. http://www.healthcompetence.eu/converis/publicweb/project/1667. Accessed 18 Oct 2010
Nagl S (ed) (2006) Cancer bioinformatics: from therapy design to treatment. Wiley, Chichester
NCI (2010) National cancer institute of the national institutes of health. http://www.cancer.gov/. Accessed 18 Oct 2010
NCI60 (2010) Discovery services. http://dtp.nci.nih.gov/docs/misc/common_files/cell_list.html. Accessed 18 Oct 2010
NHGRI (2010) National human genome research institute of the NIH. http://www.genome.gov/. Accessed 18 Oct 2010
OCISB (2010) Oxford centre for integrative systems biology. http://www.sysbio.ox.ac.uk/. Accessed 18 Oct 2010
Optimata (2010) http://www.optimata.com. Accessed 18 Oct 2010
Owen MR, Alarcon T, Maini PK, Byrne HM (2009) Angiogenesis and vascular remodelling in normal and cancerous tissues. J Math Biol 58(4–5):689–721
Oxford-SB (2010) Oxford systems biology doctoral training centre. http://www.sysbiodtc.ox.ac.uk/. Accessed 18 Oct 2010
Paik S et al (2004) A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 351:2817–2826
Parkinson H, Brazma A et al (2009) ArrayExpress update—from an archive of functional genomics experiments to the atlas of gene expression. Nucleic Acids Res 37(Database issue):D868–D872. Epub 2008 Nov 10
PDQ (2010) PDQ-NCI's comprehensive cancer database. http://www.cancer.gov/cancertopics/pdq/cancerdatabase. Accessed 18 Oct 2010
Physiomics (2010) Physiomics plc, 2010. http://www.physiomics-plc.com. Accessed 18 Oct 2010
Physiomics Virtual Tumour (2010) Physiomics virtual tumour. http://www.physiomics-plc.com/results-update-%E2%80%93-virtual-tumour/. Accessed 18 Oct 2010
Price ND, Foltz G, Madan A, Hood L, Tian Q (2008) Systems biology and cancer stem cells. J Cell Mol Med 12(1):97–110. http://www.csa.com. Accessed 18 Oct 2010
Ptitsyn AA, Wei MM, Thamm DH (2008). Systems biology approach to identification of biomarkers for metastatic progression in cancer. BMC Bioinform [computer file] 9(Suppl 9):S8. http://www.csa.com. Accessed 18 Oct 2010
Reactome (2010) Reactome—a curated knowledgebase of biological pathways. http://reactome.org/. Accessed 18 Oct 2010

Rosenfeld S, Kapetanovic I (2008) Systems biology and cancer prevention: all options on the table. Gene Regul Syst Biol Oct 10;2:307–19
Rowe A (2009) Experimental drug makes the immune system revolt against cancer. http://www.wired.com/wiredscience/tag/synthetic-biology
Sanga S, Sinak J, Frieboes B, Ferrari M, Freuhauf J, Cristini V (2006) Mathematical modeling of cancer progression and response to chemotherapy. Expert Rev. Anticancer Ther 6(10):1361–1376
Sanger Genetics (2010) Genomics and genetics at the sanger institute. http://www.sanger.ac.uk/research/areas/humangenetics/. Accessed 18 Oct 2010
Seigneuric R, Riel NAW van, Starmans MHW, Erk A van (2009) Systems biology applied to cancer research. pp 339–353 in Daskalaki (2009)
Sen F, Vega F, Medeiros LJ (2002) Molecular genetic methods in the diagnosis of hematologic neoplasms. Semin Diagn Pathol 19:72–93
Singer S (1999) New diagnostic modalities in soft tissue sarcoma. Semin Surg Oncol 17:11–22.
Sybilla (2009) Systems biology of T-cell activation in health and disease. http://www.sybilla-t-cell.de/
Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL (1987) Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science 235(4785):177–182
Sysbiomed-cancer (2010) Sysbiomed workshop on systems biology and cancer. http://www.healthcompetence.eu/converis/publicweb/project/1123. Accessed 18 Oct 2010
Syscol (2011) Systems biology of colorectal cancer. http://www.genexplain.com/syscol. Accessed 30 Jan 2011
TCGA (2010) The cancer genome atlas. http://cancergenome.nih.gov. Accessed 18 Oct 2010
Tomshine JC, Severson SR, Wigle DA et al (2009) Cell proliferation and epidermal growth factor signalling in non-small cell lung adenocarcinoma cell lines are dependent on Rin1. J Biol Chem 284(39):26331–26339. Epub 2009 Jul 1
Trireme (2010) Systems-level, multi-layer understanding of cellular responses to ionizing radiation. http://www.healthcompetence.eu/converis/publicweb/project/2877;jsessionid=2e31223356d266c8188d0e9edbef. Accessed 18 Oct 2010
Tseng GC et al (2009). Investigating multi-cancer biomarkers and their cross-predictability in the expression profiles of multiple cancer types. Biomark Insights 4:57–79
Tumour-Host Genomics (2010) FP6(2010) project on tumour host genomics. http://research.med.helsinki.fi/tumorhostgenomics. Accessed 18 Oct 2010
Ullah M, Wolkenhauer O (2007) Family tree of Markov models in systems biology. IET Syst Biol 1(4):247–254
Unicellsys (2010) Eukaryotic unicellular organism biology—systems biology of the control of cell growth and proliferation. http://www.unicellsys.eu/. Accessed 18 Oct 2010
Vera J et al (2008) A systems biology approach to analyse amplification in the JAK2-STAT5 signalling pathway. BMC Syst Biol 2:38
VALAPODYN (2010) Validated predictive dynamic model of complex intracellular pathways related to the cell death and survival. http://www.valapodyn.eu/. Accessed 1 May 2010
Wang E (2010) Cancer systems biology. Chapman & Hall/CRC Mathematical & Computational Biology, Boca Raton
Weinberg RA (2007) The biology of cancer. Garland Science, New York
WTSI (2010) The wellcome trust sanger institute. http://www.sanger.ac.uk/. Accessed 18 Oct 2010
Yao T (2002) Bioinformatics for the genomic sciences and towards systems biology. Japanese activities in the post-genome era. Prog Biophys Mol Biol 80(1–2):23–42
Yuille M et al (2008) Biobanking for Europe. Briefings in bioinformatics 9(1):14–24

Chapter 2
Cancer: Clinical Background and Key Challenges

Antonio Llombart-Bosch, Ulrik Ringborg, Sergio Rutella and Julio E. Celis

Abstract This chapter is aimed at a wide audience ranging from biologists to medical students and cancer specialists. It provides a comprehensive overview of systems approaches to the pathology and treatment of cancer. In particular, it addresses diagnosis and therapy by interconnecting various aspects of cancer at both the molecular and clinical level, and contrasts the unifying features of malignancies with the daunting diversity of cancer types, stages, and evolutionary processes during treatment. The importance is emphasized of both prevention and innovative treatments in reducing the cancer burden, and of early detection as the link between these two major areas. It sets the stage for analysis of cancer by means of systems biology, bioinformatics, and systems medicine. These methods involve the processing of cytological, histological, and imaging data, combined with genetic and expression profiling. The application of systems approaches to cancer-related clinical practice and research is discussed. The necessity is demonstrated for signalling pathways analysis to be fully integrated into grading and clinical staging of cancers, as well as into the process of discovering novel targets and biomarkers for diagnosis and prognosis. Key challenges and limitations are outlined for systems approaches to cancer, and areas are indicated where research needs to be focused in the future. Finally, pointers are provided to the paths that must be followed in order to move from a carefully controlled biological investigation, to approaches and technologies that will eventually accelerate the translation of new discoveries into prevention and clinical applications.

2.1 Introduction

Cancer is one of the major health issues affecting our society, and is forecast to worsen globally as the population ages. The WHO statistics from 2008 (Boyle and Levin 2008) estimate an increase of new cancer cases from 12.4 million to 20 million world-wide by 2030. Over the same period, the number of deaths from cancer will increase from 8.3 million to 12.9 million, and the number of patients living with cancer from 28 to 82 million, with similar trends in Europe (Ferlay et al. 2010a). In 2008, there were 1.7 million deaths due to cancer, a number that corresponds

A. Llombart-Bosch (✉)
Department of Pathology, School of Medicine, University of Valencia, Valencia, Spain
e-mail: antonio.llombart@uv.es

A. Cesario, F. B. Marcus (eds.), *Cancer Systems Biology, Bioinformatics and Medicine,*
DOI 10.1007/978-94-007-1567-7_2,

to more than three deaths every minute; while the number of new cases was 3.2 million. As the population ages, the burden is expected to increase even further. Moreover, as a result of more effective treatments for cancer, its prevalence as a chronic disease will sharply increase, particularly in countries such as the EU where life expectancy is already high. Cancer is one of the main chronic diseases, a fact that translates into a substantial extra demand on health care systems, because of the required surveillance and recurrent treatment of both the disease and observed side-effects.

Worldwide, lung cancer is the most frequent form of the disease and causes the majority of deaths. Among men, the commonest types of cancer are lung, prostate, stomach and colorectal; among women, breast, cervix, uteri, and colorectal. The principal cause of death is cancer of the lung, followed by stomach and liver (Ferlay et al. 2010b). There are many differences in occurrence between the more developed countries as compared to the less developed ones; for example, breast, prostate, lung, and colorectal cancer are the most frequent malignancies in developed countries, while in less developed ones liver, cervical, and oesophageal cancers are more prevalent (Kamangar et al. 2006; Boyle and Levin 2008).

A number of factors affect incidence rates in different countries, the most important of which are demographic aspects, external environment, lifestyle, and economic status. Differences in genetic risk factors can also contribute to dissimilarity and disparity in the results of diagnostic activities, and variations in the recording of cancer statistics make comparison between countries very difficult (Parkin 2004).

The cure rate of some cancers has increased significantly (for a review, see DeVita et al. 2008). For example, patients with some forms of leukaemias, lymphomas, and paediatric tumours have been treated successfully with combinations of antitumoural agents, while the increased cure rate for patients with cutaneous malignant melanoma is mainly owing to early detection (Cohn-Cedermark et al. 2000; Balch et al. 2003; Aitken et al. 2010). Early detection has improved the cure rate for breast cancer, with additional curative effects from both radiation therapy and antitumoural agents (Fletcher and Elmore 2003; Berry et al. 2005; Clarke et al. 2005, 2008; Cuppone et al. 2008; Madarnas et al. 2008; Dowsett et al. 2010; Richards 2009). The average mortality reduction for all forms of cancer is, however, modest. In the follow-up of the European Code against Cancer, a 9% reduction was achieved over a period of 15 years (Boyle et al. 2003b).

Significant relief of the cancer problem can only be achieved by concerted actions aimed at improving prevention and therapeutic strategies. Moreover, early detection is of fundamental importance both for prevention and improving cancer treatments. Prevention initiatives might focus on either high-risk individuals or total populations. Thanks to molecular genetics and epidemiological studies, a number of risk factors, both inherited and lifestyle-acquired, have been identified. Advances in understanding the molecular pathways involved in tumour initiation and progression have provided unique opportunities for research on the correlation between molecular biomarkers and risk factors. Approximately one-third of cancers are considered to be preventable. Successful prevention, however, largely depends on change in lifestyles, a challenging problem for behavioural sciences. Examples

of risk factors are tobacco smoking (Doll et al. 2004; Pleasance et al. 2010); Human papilloma virus (HPV) infections (Cogliano et al. 2005); alcohol abuse (Boffetta and Hashibe 2006); UV radiation, obesity and insufficient physical activity (Boyle et al. 2003a; Boyle and Levin 2008).

Early detection is often used synonymously with secondary prevention. For breast, cervical, and colorectal cancer population screening is considered to be a satisfactorily evidence-based method (Boyle and Levin 2008). An important area of modern cancer research is the identification of premalignant (precursor) lesions likely to progress to invasive cancers; this is crucial to the development of innovative approaches to prevention. For prediction, appropriate molecular markers are required, since premalignant lesions represent a heterogeneous group intermingled with benign lesions unlikely to develop malignant behaviour.

Today's treatment strategies aim to develop personalized cancer medicine based on the understanding of the molecular mechanisms underlying the disease. To achieve this goal, predictive biomarkers of primary and metastatic disease need to be identified, and sensitive and accurate methods have to be developed to monitor response to treatment and the occurrence of side-effects. Moreover, knowledge of the mechanisms underlying sensitivity and resistance to anticancer agents will constitute an important aspect of the endeavour, given that treatment failures are often the result of existing or acquired drug resistance. A close coordination of preclinical with clinical research is required in order to gain insight into which of the molecular pathways are disturbed and drive tumour growth. Novel targets for therapy will thereby be generated, which should foster the development of new anticancer agents. We expect, however, that major advances can only be made in the short term by choosing the appropriate combination of already existing drugs based on extensive diagnostic information.

The purpose of this chapter is to provide a pathological and clinical overview of cancer, as well as to pinpoint major challenges that need to be addressed if we are to fulfil the goal of bringing individualized cancer care closer to reality. Initially, we cover tumour nomenclature, grading and staging, and various technological approaches to categorizing morphology. The various methods and strategies for treatment are then outlined. With this background, several major cancers are then treated in detail, including various further classifications and modifications which have been developed for each individual type, leading to implications for treatment. Finally, these descriptions are linked to a systems biology and medicine context, demonstrating both the need for and the advantages of these approaches.

2.2 Pathology Integration in Cancer Biology Systems

For many years, the study of morphology, as applied to biology and medicine, has played a major role in basic and clinical science, helping to establish and confirm the diagnosis and prognosis of disease. Histopathology, with the aid of old and new ancillary techniques, remains a key protagonist in medicine today. Nevertheless,

new complementary classifications based on gene expression profiles enhance or more accurately predict survival in the case of certain neoplasms (Gravendeel et al. 2009; Aparicio and Huntsman 2010).

Over 150 years have elapsed since Rudolf Virchow postulated the key approach '*on the structural basis of diseases and their anatomical location as a consequence of an altered response to any external or internal injury produced to the cell or the tissue*.' Even though the biological concept of disease has changed to a great extent over the years, the anatomo-clinical approach for diagnosing diseases remains the basis of medicine (Llombart-Bosch 2001; Costa 2009).

This concept is particularly valid in oncology, where the notion of malignancy is associated with a loss of control in normal cell growth, tissue and cell differentiation. Malignancy is a biological concept, according to which the anarchic proliferation rate of the newly transformed tissue overgrows the normal, and extends into distant areas of the organism (metastasis), often producing the death of the patient in what is known as the 'natural history of cancer'. In this situation, the morphological (i.e. an image) diagnosis, provides objective support to the clinical diagnosis.

Histopathology and cytology, based upon the microscopic examination of tissues and cells, continue to play a seminal role, and are reliable methods for diagnosing patients with cancer. The hematoxylin-eosin (H&E) staining of paraffin-embedded tissues is considered the gold standard for the histological diagnosis of cancer (Rosai 2007). In addition, a number of new technological approaches based both on the microscopical analysis of cells and tissues, such as electron microscopy, enzyme histochemistry, immunohistochemistry (IHC), fluorescence in situ hybridizations (FISH) etc., yield enhanced information on the initiation, promotion, and progression of cancer cells, both in human and experimental models. These technologies are described in detail in Sect. 2.3.

The findings obtained with these methods have been supplemented by the application of molecular biology approaches to the study of cells in normal and pathological conditions. Thus, molecular pathology, in the process of complementing conventional histopathology, raises new challenges for considering cancer not only as a disease of the organ-tissue-cell system, but also as a consequence of genetic disease involving several complex biological systems (Hahn and Weinberg 2002; Celis 2008; Abeloff et al. 2008).

Nevertheless, concepts such as benign or malignant, as well as tumour typing, continue to be connected to the conventional notion of pathology, as defined by use of a microscope. Thus, morphology continues to be a reliable tool for the diagnosis of a neoplasm, as well as for controlling tissue quality in molecular biological and oncological research (Rosai 2001; Llombart-Bosch 2001; Beckman 2006).

2.2.1 Definition of a Neoplasm

A neoplasm consists of a new growth of cells, initiated within a tissue by substituting for normal cells and causing the emergence of a mass (tumour) in which the

cells exhibit high growth rate, loss of differentiation, self-autocrine feedback, loss of cell death self-control, ability for angiogenic modulation, and metastatic capacity. The degree of failure in these biological controls, which conditions the benign or malignant nature of the neoplasm, threatens the death of the host. Evolution towards malignancy may be rendered irreversible by genomic instability and the acquisition of new biological properties, as a result of continuous genetic and epigenetic remodelling (Hanahan and Weinberg 2000; Weinberg 2007).

2.2.2 Tumour Nomenclature

The nomenclature of tumours is based either on cell origin (histogenesis) or on microscopical similarity with normal tissue (histology); both categorizations may sometimes overlap.

Two types of tissue compose solid tumours: neoformed tissue (*parenchyma*), whether *carcinoma* in epithelial context (tissue composed of cells that line the cavities and surfaces of the body) or *sarcoma* in mesenchymal (connective tissue, bone, cartilage, and circulatory and lymphatic); and the *stroma* component, which provides the support cells and the vessels. Haematological neoplasia (*leukaemia*) lacks stroma while growing within the bone marrow and the blood compartment. The stroma component is an active partner in tumour behaviour and seems to be induced by the malignant cells in tumours, providing a decisive role in invasion and metastasis. Interplay between both components varies from case to case, giving a fleshy or dense consistency. The amount of the collagenous involvement causes increased density or *desmoplasia*.

Tumour nomenclature is based upon the type of parenchymal cells. Two principal types of tumours occur, benign and malignant, although a number of intermediate clinical possibilities must also be considered, such as semi-malignant, pseudo-malignant, and of questionable malignancy. The most important features used to characterize a tumour are its predictable clinical behaviour and its microscopical appearance. Benign neoplasms are harmless and slow-growing, whereas malignant lesions exhibit rapid proliferation, invade adjacent tissue, and metastasize.

In general, the suffix *-oma* follows the originating cell type, independently of grade (benign or malignant). Thus, the name for an epithelial benign tumour is *adenoma* when originating in glandular tissue (gastro-intestinal tract, breast, kidney or liver, etc.), and *papilloma* if the cells arise from the non-secretory epithelial surfaces (skin, respiratory mucosa, lower urinary tract). An additional number of terms complement this generic histological terminology, based on the presence of cysts, micropapillae, or mixed components. The terms for benign mesenchymal tumours are based on the cell type: *fibroma* is used when fibroblasts constitute the tumour, *lipoma* in the case of adipocytes, and *osteoma* for osteoblasts. Tumours may present a variable extent of interstitial components such as myxoid, angiomatous, collagen, or elastic fibres. Mixed epithelial-mesenchymal tumours are named fibropapilloma, adenofibroma and so on. A *polyp* is a composite tumour containing both the

neoformed epithelial parenchyma and the stroma; it may be pediculated or sessile. The term polyp has no implication for clinical outcome.

The generic name for a malignant epithelial neoplasm is *carcinoma*. When a dual component is present, either glandular or ductal, the term is *adenocarcinoma*. A tumour originating in a specific cell type, such as terminal lobules of the breast, is named *lobular carcinoma*. When derived from squamous or keratin producing epithelia, they are called *squamous* or *epidermoid carcinomas*. Neoplasms with hybrid features of glandular and squamous participation are known as *adenosquamous carcinoma*. Histological varieties of glandular differentiation such as cords, trabecula, tubes, papillae, or cysts, provide a variety of histological subtypes. Neoplasms may also be named by the organ from which they originate: *hepatocellular carcinoma* is a generic name for a malignant tumour originating from hepatocytes, and *mesothelioma* for tumours of mesothelial origin.

The size of the cell is also used in some poorly differentiated neoplasms such as *small-cell* or *large-cell carcinoma* of the lung. Cell secretory activity provides additional varieties with names such as *mucinous, colloid, serous, apocrine*, or *neuroendocrine* for carcinomas in diverse anatomical locations such as breast, ovary, or GI tract. Malignant epithelial cells may suffer metaplasia, conditioning the so-called *metaplastic carcinoma*, in which the carcinomatous and sarcomatous components appear together.

Malignant tumours of mesenchymal origin are designated sarcomas; several varieties of sarcomas are defined based upon their histological resemblance to the normal cell counterpart: *liposarcoma* resembles adipocytes, *fibrosarcoma* mimics fibroblasts, *leiomyosarcoma* resembles smooth muscle, *angiosarcoma* is similar to blood vessels.

The cell line of differentiation is not sufficient in several malignant mesenchymal tumours to define their true nature, thus in these cases the terminology *pleomorphic sarcoma, myxoid sarcoma, small round cell sarcoma*, or *spindle cell sarcoma* is employed, based exclusively on the cell configuration, cell size, and type of stroma. Composite tumours may be found within the same neoplasms (*dedifferentiated sarcomas*). Epithelial-like components occur in some sarcomas (*synovial sarcoma, epithelioid sarcoma, alveolar sarcoma*) making the differential diagnosis with carcinomas difficult.

The nomenclature for embryonal (occasionally ectopic) neoplasms (germ-cell tumours) of the testis or ovary is mainly dependent on tissue similarity to the cell of origin, while exhibiting malignant behaviour, and includes *seminoma, embryonal carcinoma, yolk sac tumour, granulosa cell tumour. Teratoma, hamartoma*, and *choristoma* are terms associated with embryonal cell growth, showing one, two, or all three germ-cell layers (ectoderm, endoderm, and mesoderm) in a variable degree of differentiation or state of maturation. No particular clinical outcome is implied by these terms for the patient. For example, teratoma may be malignant or benign depending on the cells present and their level of maturity.

Tumours of haematopoietic tissue are named either *leukaemia* or *lymphoma*. *Lymphoma* is frequently associated with haematogenous extension (leukemic). *Leukaemia* acquires a tumour component when seeding into a solid organ. Many

types are classified according to their cell of origin: cancer stem cell (lymphocytes, myelocytes, granulocytes, monocytes, erythrocytes) and their degree of maturation. For further details, see the WHO tumour classification (Chan 2001).

Tumours of the nervous system are often named from their similarity to the cells of origin. Other terms focus on histological similarities, as is the case for *glioma*, when initiated from the glial cells, *medulloblastoma* and *neuroblastoma* for primitive stem neuroblasts or embryonal neuroblasts. A tumour derived from Schwann cell is called *schwannoma* when benign, and *malignant peripheral nerve sheath tumour* when malignant. As with haematological tumours, a large number of terms are employed for the several varieties existing within these tumours.

The term *blastoma* is used for tumours composed of germ cells of a highly malignant behaviour. The term is preceded by the name of the potential cancer stem cell that may be the origin of the neoplasm: *nephroblastoma* for embryonal tumour of the kidney (also called Wilm's tumour); *hepatoblastoma* for a similar embryonal neoplasm of the liver. Malignantly transformed germinal cells of the ovary are referred to by the term *disgerminoma*. Highly malignant neoplasms such as grade III (see below) gliomas are denominated *glioblastoma*.

2.2.3 Tumour Grading

Nomenclature provides only a partial view of the biology and clinical outcome of neoplasms. Additional information is therefore necessary for a clear characterization of each tumour type. To this end, several grading systems have been proposed (Rosai and Ackerman 1996). Tumour grading, from **benign** to **malignant**, is based primarily on: degree of differentiation of the tumour cells, nuclear features, rate of growth (mitotic activity), stromal response (angiogenesis, inflammatory reaction). The grading criteria vary greatly for different neoplasms and the correlation between histology and clinical behaviour is in some cases imperfect. Accuracy in grading is difficult because of the heterogeneity existing in some neoplasms as well as being dependent on site-related events.

2.2.3.1 Benign Tumours

These lesions are made up of mature, differentiated cells mimicking the tissue from which they originate and resembling the normal tissue, but showing a lower grade of architectural organization: stroma is poor, and vessels scarce. Nuclear pleomorphism is almost absent or very rare, while mitosis may be limited, and there is a lack of chromosomal abnormalities. In most cases the nuclei have a diploid DNA content. The tumour is well delimited and is frequently encapsulated or polypoid. Such lesions are occasionally multiple and synchronous (uterine *leiomyomas*, breast *fibroadenomas*, GI *polyps*). Some reach large sizes (*fibromas, leiomyomas, ovarial enteral cysts, prostatic adenomas*). Exceptionally, they are functional and hormone

producing (*pancreatic insular adenomas*); nevertheless, they may cause severe clinical problems and even a fatal outcome. Basically, a benign tumour is not a metastasizing neoplasm, but benign metastasizing tumours have been described in solid or in endocrine organs (thyroid gland, kidney, ovary, testes) and bone (giant cell tumour). In order to distinguish a benign from a malignant well-differentiated neoplasm, meticulous attention must be paid to the presence of any capsular or vascular infiltration which would indicate the capacity to metastasize. Tumours with potential metastatic capacity have also been defined as neoplasms with uncertain biological potential.

2.2.3.2 Malignant Neoplasms

These lesions are characterized by a lesser degree of structural and architectural organization, providing a poor histological configuration when compared to their normal counterpart. Nevertheless, the degree of differentiation may vary from well to poor, from undifferentiated to dedifferentiated neoplasia. The latter two terms are used indiscriminately, even though they represent two diverse biological situations: the first means a lack of differentiation, while the second expresses a loss. Lack of differentiation is one of the major structural patterns of malignancy, since the tissue mimics its embryonal counterpart. The existence of 'cancer committed stem cells' as progenitors for tumours indicates that the arrest of tumour maturation in a given stage is probably the cause of loss of architectural organization. Additionally, the stroma plays a role in the loss of differentiation (Kalluri and Zeisberg 2006).

Neoplastic cells induce angiogenesis by secreting angiogenic factors, and also induce fibroblastic proliferation, interfering with the normal reparatory process associated with tumour growth (Bergers and Benjamin 2003). An inflammatory reaction may participate in the process, imparting additional complexity to the tumour phenotype.

The cellular and nuclear pattern is the second attribute that characterizes a malignancy. The cytological criteria of a malignant tumour are well known and used for diagnosis, either by tumour scraping (cervical cytology) or by aspiration (fine needle aspiration cytology). Histology also provides adequate structural criteria. The loss of cytoplasm and nuclear structures varies greatly from tumour to tumour. Cells may be larger or smaller than normal cells, or they may resemble embryonic cells (loss of the nuclear to cytoplasmic ratio) with large nuclei and nucleoli, occasionally multiple and with an irregular contour. The distribution and configuration of the chromatin varies widely from cell to cell (hyperchromatism). Total loss of cell configuration is described as *anaplasia*: the neoplasm lacks any structure that resembles its normal histological counterpart. Anaplasia is associated with higher malignancy, but may also be caused by therapy (chemo- or radiotherapy). In addition to undifferentiation and pleomorphism, some tumours present abundant cellularity with dominance of the parenchyma over the stroma. Cell rich neoplasms, mainly in sarcomas, are associated with a poor prognosis.

Invasive growth is another characteristic of malignant neoplasms. In carcinomas, the normal surrounding tissue is infiltrated by neoplastic cells which disrupt the

basal membranes and extend into the neighbouring stroma, producing new structural patterns such as cords, trabecula, tubules or glands. The malignant cells may appear isolated within the neoformed stroma or produce tiny nests. Infiltration is a microscopic event and conditions the capacity of some tumours to local relapse. The lack of defined spatial limits is associated with the absence of a capsule as happens in benign neoplasms. Some slow-growing carcinomas may induce a pseudocapsular structure mimicking a benign tumour. Particular attention has to be paid to these tumours (*follicular carcinoma of the thyroid* is an example) because they may be confused with an adenoma. Thus, the limits of invasion have to be confirmed in order to assure tumour-free margins for treatment. To this end, multiple histological sections of the neoplasia are necessary, and are made possible by colouring the excision margins with Indian ink (which resists discoloration when treated with alcohol-xylene dehydration), embedding in paraffin, and staining with H&E.

Tumour invasion may affect the neighbouring tissues and vessels, thus originating local relapses. The invasion of the vessels involves regional (lymph nodes) or distant metastasis. A particular type of invasion is the so-called *skip metastasis*, corresponding to tumour implants in the skin adjacent to a primary neoplasm due to their local extension. This type of metastasis occurs mainly in melanomas or sarcomas.

Sarcomas possess highly invasive growth potential and imprecise margins, lacking anatomical boundaries that necessitate large compartmental resections to avoid local relapses. The absence of invasion of the margins is a mandatory requisite, to assure the absence of relapse in the preserved limb. New therapeutic methodologies for malignant tumours have created the possibility of preserving organs or limbs, avoiding previously necessary organ resections.

Necrosis is another feature present mainly in malignant tumours. Tumour progression is dependent on a balance between cell proliferation, differentiation, and death. Necrosis is the usual end for many tumour cells, independently of *apoptosis* (programmed cell death). This variety of non-programmed destruction of cells is expressed by a loss of the architectural and cytological (*pyknosis, karyorrhexis, karyolysis*) structures in a focal or large area within the tumour.

Most of these necrotic foci adopt a geographical distribution depending on the amount of neovascularisation. Only cells located at less than 200 μm from a vessel receive sufficient oxygen and nutrients to stay alive. Fast-growing carcinomas and sarcomas present an imbalance between angiogenesis and cell proliferation, and therefore suffer extensive areas of necrosis. In contrast, slow-growing epithelial or mesenchymal tumours maintain a good angiogenic capacity and so the degree of ischemia is more limited. This is also the case for benign tumours, which rarely suffer necrosis.

2.2.3.3 Miscellaneous and Borderline Lesions

In addition to benign and malignant neoplasms, a number of biological possibilities arise out of the clinical behaviour of some neoformations. *Pseudomalignant*

tumours are neoformations that express many of the attributes of malignancy such as cell pleomorphism, loss of architectural patterns, and mitosis. Nevertheless, their clinical behaviour is benign. Endocrine tumours such as *pheochromocytoma* or *paraganglioma* may present these peculiarities. The *pleomorphic adenoma* of the salivary glands also belongs to this category. In the skin, a *keratoacanthoma* may resemble a squamous epidermoid carcinoma, but with benign behaviour and may even regress spontaneously. Other examples are pigmented tumours in the skin such a *cellular blue nevi* or *Spitz nevi*. Some soft tissue neoplasms display this peculiarity (*pseudosarcomatous fasciitis*).

Semimalignant tumours are neoplasms with a peculiar biology. They express local aggressivity like malignant neoplasms and have cell atypia and mitosis, infiltrating neighbouring tissues; nevertheless, they lack metastatic capacity, even if they may relapse locally. *Basal cell carcinoma* of the skin is a good example. *Dermatofibrosarcoma protuberans* can also be included in this category, as can *phyllodes tumours of the breast*, both of which rarely metastasize.

Currently there is a tendency to consider these neoplasms as *low-grade tumours* (tumours with questionable malignancy), meaning proliferations with scant aggressivity, but which may occasionally relapse or even metastasize. Examples of these are atypical lipomatous tumour or the *angiectatic pleomorphic tumour* in soft tissue. They should not be confused with another category of neoplasia: *borderline lesions,* a term denoting the inability of the pathologist to define the possible clinical outcome of neoplasms. The number of lesions classified within the context of *borderline* is progressively increasing, and represents a substantial grey area between benign and malignant processes. Examples include borderline pigmented lesions of the skin, or borderline lesions in mucinous and serous tumours of the ovary, as well as lesions present in the breast or in the prostate.

2.2.4 Growth Rate of a Tumour

Growth rate constitutes one of the main biological features marking the distinction between benign and malignant tumours. Within the latter category, slow-growing tumours are distinguished from those that possess a rapid rate of proliferation and therefore a higher biologically aggressive behaviour.

The study of tumour cell kinetics has clarified the mechanisms by which a tumour proliferates. This rate is also dependent on the degree of differentiation (cell maturation) and programmed cell death (cell loss). Thus, it would be helpful to know not only how many cells replicate and go into mitosis, but also how many are within the cell cycle by maturation or remain outside the cycle (cancer progenitor stem cells). In addition, the growth rate is also dependent on other factors such as cell doubling time or the amount of the replicative pool. Finally, tumour growth is balanced between the predominance of cell proliferation over cell loss. Thus, in tumours with a relatively high growth fraction, the disparity is large; resulting in more rapid growth than in tumours in which cell production exceeds cell loss by only a small margin.

The number of cells in mitosis or in cell cycle provides two ways to measure the growth rate within a neoplasm. Histologically, both features are easily and objectively assessable. Mitotic counts have been performed for many years as the most reliable way to measure the proliferation activity of a given neoplasm. It is well known that benign tumours lack or have few mitoses, while carcinomas and sarcomas display high mitotic count together with abnormal mitosis, aberrant chromosomes and loss of ploidy (aneuploidy). There is a good correlation between the mitotic cell count in a tumour and prognosis. Breast carcinoma is a good example of this correlation with the so called Bloom index (Nottingham index) (Elston and Ellis 1998). In soft tissue sarcomas, all grading systems (American system, French Federation system) include the mitotic count as a reliable parameter to measure the histological grade of malignancy.

A second method is to evaluate the number of cycling cells. Several immunohistochemical techniques offer reliable ways to measure their number and establish correlations with the mitotic count and prognosis. Today, the most popular antibody used in histology is Ki67 (MIB1) for assessing the proliferative rate of a tumour. Numerous publications confirm its value (Ueda et al. 1989; Hoos et al. 2001; Meara et al. 2007; Lopez-Guerrero et al. 2011).

2.2.5 Dysplasia and Carcinoma in situ

The process of cancer promotion and progression involves a number of genetic and molecular events that are variably expressed with a number of structural changes in the phenotype of the cell and tissue (Ponten 2001). The loss of cytological and architectural organization is consistent with the term *dysplasia* (altered form) and can be considered a step in the transition from normal to cancer cells. The presence of dysplasia in a tissue does not necessarily mean a precancerous lesion or *in situ* cancer. Many of the dysplastic alterations may reverse or are secondary to an adapted cell response to a metabolic injury, inflammation, or tissue repair. Thus, the distinction between what represents a harmless dysplasia and a precancerous lesion exceeds the limits of histology and occupies an imprecise grey area. German pathologists of the early twentieth century used to consider this the realm of '*Persönlich einstellung*' (personal appreciation), because at that time histological diagnosis was based exclusively on the experience of the expert. One of the principal advances in modern genetics and molecular biology is to have provided objectivity in differentiating these types of lesion, in which histopathology has a limited capacity for definition.

Most lesions presenting dysplasia occur in the epithelia, either stratified or glandular, and are associated with *metaplasia* (transformation of a tissue originating in a given germ layer to another of the same origin) or *hyperplasia* (increase in the number of cell layers within an epithelia, but preserving their normal phenotype). Hyperplasia, metaplasia, and dysplasia are biological events that merge in both early precancerous lesions and carcinoma *in situ* or invasive cancer. Examples of these lesions can often be seen in resected specimens of the bronchial mucosa as-

sociated with carcinoma of the lung (Franklin et al. 2004). Similarly, the presence of intestinal metaplasia of the gastric mucosa is consistent with a precancerous lesion (Whitehead 1994), while the processes of hyperplasia-metaplasia and dysplasia characterize GI polyps and are also related to the successive genetic rearrangements that precede carcinoma (Shia et al. 2003).

In situ carcinoma occurs in numerous stratified glandular epithelial mucosa and in the skin. Several examples have been described that characterize this entity. The most frequently quoted models are uterine cervix for squamous epithelia, and the breast for glandular epithelia, perhaps because they are commonly occurring lesions of a type easily defined by cytology or histology. This form of carcinoma depends primarily on two components that preserve the capacity of the transformed cells to maintain the organization and integrity of the epithelia, namely the continuity of the basal membrane that isolates the epithelia from the supporting connective tissue, and maintenance of cell to cell attachment, owing to the activation of several cytoplasmic and membranous adhesion molecules. The loss of these functions leads to early invasion and infiltration (microinvasion).

Additionally, the tumour induces angiogenesis which provides a seminal capacity for growth, invasion, and metastasis. The tumour vessel formation mimics vascular embryogenesis, promoting several processes. These mechanisms are very complex and may occur together inside the same tumour and their surrounding stroma. A number of possible processes have been recognized:

- *Neo-angiogenesis*: new vessels grow by branching from pre-existing vessels (mainly capillaries). The process occurs through *vascular sprouting* or by *intussusceptions* in which interstitial columns of tissue are incorporated into the lumen of newly formed or pre-existing vessels.
- *Vasculogenesis*: *de novo* production of vessels originating from undifferentiated precursors (mesenchymal pluripotential cells with angioblastic capacity) forming an initial tubular network. At this stage the endothelial cells mature and integrate closely with smooth muscle cells, pericytes, and the surrounding matrix.
- *Angiogenic remodelling*: the initial network is modified by pruning and vessel enlargement to form interconnecting branching figures, characteristic of a more mature vasculature.
- *Lymphangiogenesis*: the endothelial vessels proliferate, producing new lymph vessels, either by angiogenesis or a vasculogenesis-like mechanism, induced by lymphatic endothelial growth factors and their receptors.
- *Vascular co-option*: groups of avascular tumour cells co-opt with pre-existing host vessels and initiate as well-vascularised small tumours.
- *Mosaic vessels*: the tumour cells come in contact with a lumen, together with neoformed endothelial cells, producing an interface or mosaic on the surface of the intratumoural capillaries.
- *Vascular mimicry*: the tumour cells transform themselves into a pseudoendothelium, mimicking new vessels that are incorporated into the vascular network.
- *Capillary drop-out*: regression induced in recently developed microvessels by anti-angiogenic drugs, such as Avastin (bevacizumab).

2.2.6 Metastasis

Metastasis is an especially complex process which occurs through a series of sequential steps in which tumour cells first migrate from the primary tumour, penetrate blood vessels, circulate within the bloodstream, and after migration, finally colonize distant sites, reproducing the disease. Metastases are tumour implants discontinuous with the primary tumour. New data indicates that the mechanisms controlling metastasis are regulated independently of the primary tumour growth and are due to a sequence of events mediated by different classes of metastatic genes (Meyer and Hart 1998; Chiang and Massague 2008; Nguyen et al. 2009). Approximately 30% of cancer patients harbouring solid tumours present metastases at diagnosis, this being the most common cause of death, even if the patients respond transiently to conventional therapy.

The dissemination of cancers may occur through several pathways: by direct extension and seeding into neighbouring organs, surfaces, or body cavities, or by either lymphatic or haematogenous vascular spread. Transplacental tumour extension has also been reported, even though it may be very unusual. Direct implants of tumour cells caused by surgery may rarely occur.

The term *local metastasis* refers to lymphatic invasion or direct spread of tumour cells into the surrounding interstitial tissue and/or space, as in the case of the peritoneum, pleurae, SNC liquoral ventricles, etc. In the breast, local invasion of the subcutaneous tissue and dermis produces a skin retraction (*peau d'orange*). *Metastasis in continuity* affects mainly the natural ducts such as the bronchial tree for lung cancer, or urethra in carcinoma of the bladder, due to the seeding of neoplastic cells in proximity of the original neoplasia. Children's medulloblastoma in the Central Nervous System (CNS) may extend directly into the lateral and medial ventricles, while high-grade glioblastoma produces massive local invasive growth, but does not usually metastasize outside the CNS. A particular type of local invasion is *Paget's disease of the nipple*, in which an intraductal breast carcinoma originating in the main collector ducts of the gland progressively substitutes, as isolated or small cell nests, for the normal superficial stratified cutaneous epithelia.

Tumour extension into the natural cavities affecting the serosa is frequent in the abdominal cavity and pleura, secondary to serous or mucinous carcinoma of the ovary or GI tract, lung carcinoma, and malignant mesothelioma. Tumour implants may be microscopical in size or form large nodules (*carcinomatosis peritonei/pleurae*). *Primary peritoneal carcinomatosis (pseudomyxoma peritonei)* is almost always secondary to a mucinous carcinoma of the appendix or the ovary, while in the pleural spaces, the origin of focal or diffuse *pleuritis carcinomatosa* is mainly secondary to carcinomas of the lung or the breast.

Peritoneal implants caused by borderline mucinous carcinoma of the ovary are not necessarily cases of true metastasis, in that the tumour deposits only superficially and does not invade the serosa, prognosis being favourable in this case. Detection of pleural or peritoneal tumour extension is, however, an indication of poor clinical outcome.

Lymphatic nodes may be colonized by tumour cells (mainly carcinomas and melanomas, more rarely sarcomas), a fact that signifies the regional extension of the tumour. The early detection of nodal metastasis is an important aim in cancer assessment. Contrast radiological lymphography has been used for years to detect the presence of metastasis in nodes. Other imaging technologies such as Positron Image Tomography—PET and Computed Tomography—CT/PET, have greatly improved the detection of small metastasis in nodes, since they are able to detect and depict with high precision areas of relatively intense metabolism as actively replicating tumour cells form in their growth process. Even so, the histological study of the resected lymph nodes is still required for the purpose of precisely determining the "pathological staging" of the tumour (pTNM, see Sect. 2.2.7). Small deposits of tumour cells have been called 'clandestine metastasis' and their prognostic and therapeutic significance is today of high oncological value. The presence of isolated tumour cells, freely located in the cortical sinus of a node, has no clinical significance, unless the tumour increases in size to 2 mm or more, in which case it is considered as a positive metastasis node.

The *sentinel lymph node* is the node anatomically located in closest proximity to the cancer, and provides excellent information on the extension of the neoplasia into the regional territory. This is of great value in breast carcinoma, sparing unnecessary surgery on the residual axillary nodes while the sentinel nodes continue to present negative. Other tumours may benefit from this approach (melanoma, gastric and colon cancer, prostate). In addition, the capsular invasion by tumour cells and their extra nodal extension are associated with a negative prognosis. Furthermore, histological study of all nodes remaining after surgery offers the possibility of gaining additional information about the presence or absence of metastasis, which will condition both therapy and clinical outcome.

2.2.7 Tumour Staging

Several types of local, regional, and distant metastasis are characterized by *tumour staging*, which is based upon the presence or absence of any various types of tumour extensions together with the size of the neoplasm. Staging of the tumour is an important procedure before deciding on treatment. The widely used classification and staging system is the Tumour Node Metastasis (TNM). Data for the vast majority of solid neoplasms is integrated in the framework of the TNM system. The T, N, and M parameters and their combination define the stage of disease and represent a powerful criterion in the therapeutic decision making process. The presence of distant metastasis puts the tumour at the most advanced stage, irrespective of the T and N status.

Two versions of TNM exist: the one developed by UICC, the International Union Against Cancer -http://www.uicc.org—(Sobin et al. 2009) and the other by AJCC, the American Joint Committee on Cancer—http://www.cancerstaging.org- (Edge et al. 2010).

- **T** indicates the size of the primary tumour and its behaviour towards surrounding structures (adjacent/in contact versus infiltrating).
- **N** represents the involvement of regional lymph nodes.
- **M** the presence of metastases (loco-regional and/or distant).

The clinical staging (producing the cTNM status and the cStage), performed at the moment of the initial diagnosis, depends essentially on imaging techniques (Computed Tomography (CT- and PET-CT); Nuclear Magnetic Resonance (NMR), Ultrasound (US), radionuclide scan, etc.) and outlines, defines the presence and characteristics of the local and loco-regional disease (including the T and N status), and the presence and characteristics of loco-regional and distant metastases. Over time, the process of revision of the TNM staging system is a forum of continuous discussion and validation. The pathological staging (producing the pTNM status and the pStage), assessed if surgery is performed, is essentially based on the final pathology. The T and N statuses are unequivocally characterised by histology. The M status can be confirmed cytologically or histologically if metastases are simply biopsied or surgically removed. The M status is recorded at the time of clinical staging as well as pathological staging, even if a final pathology is not obtained on neoplastic tissue originating from the detected metastases, because of the fact that these are not surgically removed (in general terms, in fact there is no surgical indication for the removal of local and distant metastase, except in selected cases).

An unusual situation arises where a tumour has been judged to be resectable owing to its locally advanced stage, as opposed to a cancer judged to be inoperable because of distant spread. Whereas chemotherapy and/or radiation is typically administered as 'adjuvant' primary treatment for unresectable tumours. In such cases, after primary therapy has been administered, the tumour status is re-assessed in what is known as "clinical re-staging" and a yTNM plus yStage is produced. This provides a basis for clinicians to evaluate whether the situation is better (down-staged), the same (unchanged), or worse (progression) and to make decisions as to further treatment options. If surgery is indicated, a pTNM and pStage will be determined.

Since staging is fundamental for the choice of local, loco-regional, or systemic treatment, it is important to improve staging methods. In the future, by using molecular biomarkers, it will be possible to predict local and metastatic disease with much higher precision. It may even be possible to predict different metastatic phenotypes of the primary tumour, such as regional and systemic metastases. Promising developments of both molecular pathology/cytology and imaging, which will incorporate molecular biomarker research outcomes into clinical technologies, should soon have an important impact on clinical staging procedures.

The presence of distant metastasis is associated with vascular spread and represents Stage IV of the disease (the most advanced). Patient survival at this stage is generally poor, with therapy being based on a palliative approach. Haematogenous spread occurs when the tumour cells irrupt into the vessel, circulate within the blood, adhere to the endothelia, transverse the wall vessel, and invade, colonizing a distant tissue. The site of the metastasis depends on the anatomical configuration

and the local circulatory network in the distant organ, into which the tumour grafts and reproduces a new growth with some similarities to the original. Here again, the cancer cells induce their own new vessels by secreting vascular growth factors (Kerbel 2008). There are two main theories postulating the mechanism for metastatic organ distribution. The 'mechanical theory' maintains that the spreading of metastasis depends primarily on the number of vessels present in the tissue: tissues with dense vascularisation are better predisposed to receive metastasis. This is the case for the liver, lung, CNS, or bone marrow, but not in others such as the spleen. There is evidence for a direct relationship between venous blood drainage and the anatomical location of the metastasis. Basically, some neoplasms show a tendency to venous invasion; an example is the paravertebral plexus in the prostate or renal veins in kidney carcinoma. Colon cancer extends not only to the local regional lymph-nodes, but can also affect the liver as the most frequent first distant location. This is also the case in gastric cancer. Nguyen et al. 2009, and others before them (Mehlen and Puisieux 2006), have argued for another possible mechanism to explain the metastatic process known as the 'root and seed' process. This implies the existence of tumour cell specificity for dissemination and organ-specific colonization. The organ specificity of the metastatic cells is determined by a particular infiltrative and colonizing capacity, gained after its dissemination from a primary tumour. For a given type of cancer, these events occur within particular temporal kinetics and in a unique organ site. The varied latency period for metastasis occurring in certain tumours suggests a need for a specific tumour progression allowing the cells to adapt to the microenvironments of the particular organ. The acquisition of specific pro-metastatic functions, earlier during primary tumour promotion, might enable distinct cancer types to express different timings to relapse.

Mehlen and Puisieux 2006 have provided additional evidence that metastatic potential is associated with an increased resistance to apoptosis. They postulate the concepts of *anoikis* and *amorphosis* as barriers to metastasis. *Anoikis* should be considered as cell death induced by disruption of the cell attachment and interactions between the cell-matrix complexes, whereas *amorphosis* appears to depend on tumour cell death stimulated by the loss of cytoskeletal architecture. Both processes interfere with metastatic spread.

Sanchez-Garcia 2009 in a review on 'the crossroads of oncogenesis and metastasis' analysed two known possibilities of the metastatic cascade in cancer. In the classic model of human cancer progression, the metastatic process occurs in the advanced clinical stage. However, recent studies support a second possibility that could be modulated by activation of protein expression that mediates the epithelial-to-mesenchymal transition. This would concomitantly activate both the malignant conversion and the metastatic dissemination occurring in early tumour stages. According to this model, dissemination of the initiated malignant cells could happen at any time during cancer promotion or progression and not only in the advanced stage of the disease. If this theory were accepted, cancer could be considered as a systemic disease from its early phase (*ab initio*) and should therefore be treated as such. Studies of certain ovarian and breast carcinomas (Naora and Montell 2005; Weigelt et al. 2005) support this possibility.

In conclusion, evidence suggests that some tumours bear a genetically controlled metastatic phenotype, and are prone to dissemination from the beginning of their growth, even where the tumour size is so small that it cannot be clinically detected (generalized metastasis with unknown primary). A number of metastatic genes have been postulated to be involved in this process. The detection of what could be a metastatic phenotype, using biomarkers to identify distant metastases, would aid in the selection of patients to be treated with targeted therapy.

2.2.8 Cytology and Diagnosis

Clinical cytology, used in conjunction with histology, is the most reliable tool available for the microscopical diagnosis of cancer for precancerous lesions and other pathological conditions such as atypical reactive proliferations or inflammatory processes. It has also been used for hormonal checks in women and in addition serves to orient a rapid diagnosis of internal lesions that are radiologically detectable as lumps or scars. Image diagnosis is complemented with a quite simple procedure using fine-needle aspiration (FNA) of the suspected process (Arisio et al. 1998). This is an economical and very reliable way to assess pathology diagnosis of cancer in neoplasms located in deep organs which are not easily accessible and require surgery.

The application in population-based screening programs of systematic cytology examinations using cervical exfoliation or endometrial aspiration has provided one of the more valuable methods to detect early-stage cancer or precancerous lesions. The pioneering work of George Papanicolau deserves honourable mention; in 1941 he proposed exfoliation of cells of the cervix as a reliable method for the diagnosis of cervical cancer. This simple and quite inexpensive test has probably saved more women's lives than any other diagnostic procedure in clinical practice. Cervical screening with the 'Pap-test' has continued to be an almost indispensable system for controlling cervical cancer and following up the response to therapy or the hormonal status. The application of this test has been extended to other mucosa (oral mucosae, gastric, bronchial tree, and urinary tract) and corporeal fluids (urine, sputum, liquids in pleural or peritoneal cavities).

The FNA technology was first developed at the Radiumhemmet, the Karolinska Hospital in Sweden in the early 1950 by the group of cytologists and pathologists lead by Sixten Franzén and Jozef Zajicek and extended rapidly to the USA, becoming nowadays a very popular technique in pathology laboratories all over the world. Nonetheless, it must be taken into account that exfoliative or FNA cytology shares the limitations inherent in any clinical tests: the balance between sensitivity and specificity. Generally, few cases of neoplasia are missed when diagnosed by an expert cytopathologist, which means a higher sensitivity of the technique, nevertheless a number of patients without neoplastic disease will need additional follow-up testing which means a lower specificity. The balance between sensitivity and specificity varies greatly from organ to organ, and is also dependent on the good quality

of the material obtained by the technical extraction procedure of the cells (scraping or aspiration) and on the quality of smearing cells onto the slide together with good fixation and staining. Quality control within cytological laboratories and training of personnel is a major problem in developing countries, contributing to failures that may discredit this kind of test.

Cytology material has been successfully used in numerous conventional and molecular techniques, such as immunohistochemistry (IHC), morphometry, electron microscopy and in recent years in molecular biology DNA and RNA extraction for blotting, PCR, or microarray gene cluster analysis.

2.3 Technological Approaches to Morphology and Pathology

2.3.1 Hematoxylin-Eosin (H&E) Staining in Histological Diagnosis

The rising use and value of histopathology in cancer diagnosis has for many years been based mainly upon a relatively simple and economical technology that has stood the test of time and should continue to do so in the future, provided that it remains in the hands of adequately trained pathologists. Microscopical observation of tumour slides with hematoxylin-eosin (H&E) staining offers an especially selective and sensitive means of approaching cancer diagnosis. Occasionally, however, H&E staining of paraffin-embedded tissue sections is insufficient to confirm a diagnosis of malignancy: Immunohistochemistry (IHC) staining should additionally be performed to acquire sufficient information for a final diagnosis. There is no problem with IHC staining of formalin-fixed paraffin-embedded tissue, and it does not require the extra time or major expense of more sophisticated diagnostic procedures such as electron microscopy (EM) or FISH. However, in view of the fact that some special stains have specific requirements for uncommon tissue fixation and processing, the need for special stains must be anticipated to ensure that adequate tissue samples are appropriately processed, e.g. fixation for electron microscopy or preservation of fresh tissue for cell cultures, xenografting or RNA extraction. Tissue banking, not only with paraffin blocks, but also with fresh normal and tumour tissue, is becoming mandatory in all modern, well-equipped laboratories. The EM procedure is quite expensive and very time consuming when compared to IHC, so its use is therefore restricted to specific tumour types or the search for a particular inclusion. In fact, a large number of laboratories have abandoned this diagnostic procedure, limiting it to a very few research indications.

2.3.2 Immunohistochemistry

Immunohistochemistry (IHC) is an important ancillary tool that has provided great support to the diagnosis and prognosis of tumours. IHC is an old technique, still used

in fluorescence microscopy, deriving from the pioneering work of Albert Coons in 1941, who labelled antibodies with fluorescein isocyanate. The data it yields must be considered together with other information available, such as histology, molecular findings, and clinical records. It should not constitute the sole decision criterion for a final diagnosis, but must be integrated into the decision-making process of the pathologist (Taylor and Cote 1997; Natkunam and Mason 2006).

However, transmission light microscopy offers greater advantages in evaluating morphology over fluorescence, because cells and tissues are more clearly recognized and tumour architecture is well preserved, allowing better histological correlations. This makes enzyme IHC (peroxidase, alkaline phosphatase, biotin-avidin) more useful for routine diagnosis. The sensitivity of fluorescence is higher, but is mostly limited to FISH (Fluorescence In Situ Hybridization) for procedures such as detection of chromosomal segments or gene probes (Jin and Lloyd 1997). The enzymatic procedure involves three steps in which the selected antibody is first bound to an unlabeled antibody and posteriorly to a second antibody having generic specificity for the first, and is then conjugated with biotin, providing a bridge for the subsequent binding of an avidin-biotin horseradish peroxidase complex that completes the immunochemical assembly. Nuclei are counterstained with hematoxylin.

Usually the antigenic determinants that are targets of the IHC stains are not totally tumour-specific. Therefore, it is necessary to employ more than one antibody in order to elaborate sets of primary antibodies for the identification of the subtype of a given tumour, thereby constructing an 'antigenic fingerprint' that will allow final interpretation. On this basis a number of algorithms have been constructed to facilitate the classification or differential diagnosis of histologically similar neoplasms.

Several examples of these antigenic fingerprints are discussed in Sect. 2.5, related to colorectal, breast and bronchial carcinoma. Differential diagnosis of 'small round cell tumours' including lymphomas located in soft tissue or in bone, provides a good model of integrating histology with a panel of IHC markers in molecular genetically similar neoplasms.

2.3.3 Electron Microscopy

Electron microscopy (EM) can be especially helpful when other specialized techniques do not provide a definitive diagnosis. It should not be expected to be able to differentiate benign from malignant cells, as they may not display specific patterns (Dardick and Herrera 1998; Ordonez and Mackay 1998). Samples have to be prepared with extreme care; small pieces of tissue are processed in a special fixative, embedded in epoxy resin, and thinly sectioned (0.1 mm thick), after which they are impregnated with osmium. This technique, by enabling a differential absorption of the focused electron beam traversing through the specimen, can produce an image that is directly observed on a fluorescent screen, stored on photographic film, or digitized. The image should be evaluated by a trained electron microscopist, analysing specific, but previously focused images at low magnification. Study of the cell configuration, membrane preservation, and intracellular content has to be progressively

reviewed with higher magnifications to look for specific structural organelles and other features; the type of malignant cell may be characterized, as well as cytoplasmic contents such as filaments, neurosecretory granules, secretory material, lipid droplets, or inclusions such as viruses located either in the cytoplasm or nucleus. All this information has to be comprehensively analysed, by comparing conventional histological slides of the tumour and the semi-thin slides coloured with Alcian blue.

2.3.4 Tissue Microarray (TMA)

Tissue microarray (TMA) technology is based upon the concept of obtaining high-throughput phenotyping profiles of intact tissues. The conventional investigation of fresh frozen/paraffin embedded tissues is too expensive and time consuming for analysis of hundreds and thousands of genes associated with tumours. Therefore, TMA is based on the idea of miniaturization and a high-throughput rapid and cost effective approach to the validation of molecular targets in a large number of tissue specimens at DNA, RNA, or mainly protein level with IHC (Shergill et al. 2004).

Donor blocks must be at least 1 mm thick, but an optimal thickness of 3–4 mm is recommended for better results. Recipient array blocks are prepared by pouring paraffin into moulds of about 5–10 mm thickness. The block surface is made flat and parallel to the underside of the plastic cassette by trimming off in a microtome. The TMA technique has the advantage of permitting rapid analysis of genomic alterations in a large number of specimens, and of achieving an unprecedented level of standardization with minimal destruction of original tissue blocks. A uniform staining quality is also achieved with internal positive and negative controls, and efficiency is high, saving reagents, time, and money.

Nevertheless, the technique presents a number of limitations, such as possible loss on the glass slide of samples floated off during sectioning or unmasking procedure, and the tendency of shallower samples to be used up more quickly. On some occasions inadequate tumour tissue is present in the sample, which may therefore fail to be representative of the entire lesion, owing to the tumour heterogeneity (cells may be intermingled with areas of stroma and necrosis). Even so, several studies have confirmed that two cores from each tumour give 95% accuracy (Kallioniemi et al. 2001).

The applications of TMA are numerous. Some of the possibilities include routine protocols for IHC, *in situ* RNA hybridization, or interphase FISH. Another possible application is the analysis for prevalence of genetic alteration in one or more tumour samples lacking clinicopathological information. In addition, TMA can be used to analyse molecular markers in relation to various stages of the tumour, or for prognosis studies linking molecular findings and clinical outcome.

The TMA technique, as distinct from gene microarrays, covers many samples for one antibody or *in situ* hybridization product, whereas with gene microarrays, one tissue sample analyses several probes delivering a gene expression profile of amplifications/deletions.

Frozen tumour TMA technology (Schoenberg Fejzo and Slamon 2001) for the analysis of tumour RNA, DNA, and proteins may be useful in some circumstances. It is known that paraffin embedding incurs a problem of fixation, which could partially mask some antigenic targets. This can introduce chemical modification of RNA, resulting in less optimal conditions for *in situ* analysis of DNA, RNA, or proteins. Frozen TMA for IHC and *in situ* RNA hybridization with FISH have been developed to overcome the fixation problem.

A matter of debate is whether TMA automation is necessary for routine pathology. There are arguments in favour of automatic screening, since a pathologist may need to review up to 1000 samples in 1–2 h. An argument against automation is that the procedure is based on signal intensity/spot and lack of distinction between different cell types, such as epithelial and stromal. Automation may also underestimate staining in samples with few cancer cells.

Bioinformatics-type analyses in TMA 'virtual cores' with software tools have been developed for archival IHC data in TMA (Liu et al. 2002). Analysis and storage of the large amount of data generated by the staining results is accomplished by recording directly into Microsoft Excel work sheets. Data is subsequently reorganised by a program ('TMA -DECONVOLUTER') into a format suitable for hierarchical cluster analysis. The immunoprofile of a case can be retrieved and reviewed by using the 'STAINFINDER' software. Digital images are available at <http:// genome- http://www.stanford.edu/TMA/explore.shtml>.

The National Human Genome Research Institute (NHGRI) and National Cancer Institute (NCI) have created the Tissue Array Research Program (TARP) (http://ccr.cancer.gov/tech_initiatives/tarp/) of the National Cancer Institute (NCI) with the goal of promoting TMA research and development and providing assistance in arraying unique tissue materials such as those collected from clinical trials. The TARP programme also provides training and arranges workshops and protocols concerning the TMA technology.

2.4 Treatments

In Western countries about half of cancer patients have advanced cancer at the time of diagnosis, with either advanced visible tumours or concurrent micrometastases. It therefore follows logically that treatment is mostly multimodal in nature, involving combinations of surgery, radiation therapy, and medical treatment. For example, in a Swedish investigation 84% of those patients treated by radiation therapy were also treated with surgery and/or antitumour agents (Ringborg et al. 2003). Multidisciplinarity, as reflected by diagnostic activities, is essential, as well as psychosocial oncology, rehabilitation, high-quality supportive and palliative care. The cure rate of patients in European cancer registries was 21–47% in men and 38–59% in women (Francisci et al. 2009). In general, about 1/3 of patients are cured by surgery alone. The addition of radiation therapy increases the cure rate to about 50%, while further medical treatment increases the cure rate to about 60% in countries with well-developed cancer care. These, of course, are approximate.

2.4.1 Surgical Treatment

The main aim of surgical treatment is to remove the primary tumour and loco-regional metastases. Over the years a number of surgical methods have been tested, all too often with mutilation as a consequence. During the last decades, however, our enhanced understanding of the biology underlying the various cancer types has resulted in a reduction of the invasiveness of surgical interventions (the "organ sparing" philosophy). Malignant melanoma is a good example of this trend, as the availability of better parameters for predicting recurrent disease has considerably reduced the size of resection margins (Cohn-Cedermark et al. 2000; Balch et al. 2003). Similarly, combining surgery with radiation therapy in breast cancer has enabled strategies to be changed from mastectomies to breast-conserving treatments. The outcome is the same, but the quality of life of the patients is improved (Clarke et al. 2005). For other tumour types like rectal cancer, improved surgical techniques and combined chemo-radiation plus surgery approaches have reduced the local recurrence rate and the need for extensive *demolitive surgery* (Simunovic et al. 2009). The same is true in head and neck cancer, where combined chemo-radiation therapy has replaced extensive surgery (Lango 2009).

2.4.2 Radiation Therapy

Innovations in radiation physics have significantly improved the quality of radiation therapy (Levitt et al. 2006, 2008; Verellen et al. 2007). Modern imaging technologies now permit the volume of the tumour to be defined and delineated in a more precise way that facilitates image-guided radiotherapy. The availability of high-energy, well-collimated radiation facilities, together with effective dose planning, makes it possible to save more normal tissue, while at the same time increasing the dose delivered to the tumour. Moreover, with intensity-modulated radiation therapy, the dose can be shaped to fit almost any irregular tumour mass in the body. Methods have also been developed for breathing-synchronized irradiation.

Stereotactic radiation therapy (radiosurgery) can effectively kill primary tumours and metastases and can, for selected patients, replace surgery (Baumann et al. 2009). Some cancer centres are currently evaluating treatments with light ions having different radiobiological effects as compared to x-ray photons, the source of energy most widely used (Jakel et al. 2008).

The main role of radiation therapy is to eliminate loco-regional tumour disease. This is done by irradiating the total tissue volume around the primary tumour as well as the anatomical region where *in-transit* and lymph-node metastases may grow. Total systems irradiation may be indicated in cases of sensitive disseminated disease. Radiation therapy also has an important role to play in the palliative treatment of patients. Prediction of regional metastases, *in- transit*, and/or lymph-node metastases is particularly important when they are microscopic. If these can be detected using biomarkers, molecular imaging may be used to determine the target volume for radiation.

Table 2.1 Effects of cytostatic/cytotoxic treatment

Treatment may cure
Acute leukaemia, above all in children
Choriocarcinoma
Malignant lymphoma
Testicular carcinoma
Wilm's tumour
Treatment may prolong survival and be of palliative value
Breast cancer
Colorectal cancer
Chronic leukaemia
Myeloma
Ovarian carcinoma
Sarcoma
Small-cell lung cancer
Urinary bladder cancer
Cancer of corpus uteri
Treatment has modest effects
Gastrointestinal (except colorectal) cancer
Malignant glioma
Malignant melanoma
Prostate cancer

An important area of research is the investigation of molecular mechanisms underlying sensitivity and resistance to radiation therapy, since this has a bearing both for tumours and normal tissues. The optimal dose and optimal fractionation of the radiation are crucial questions. Here a number of radiobiological factors, including molecular factors linked to DNA damage response, are of strategic importance (Sarkaria and Bristow 2008). Understanding the molecular radiobiology of ionizing radiation is expected to lead to new prediction methods regarding radiosensitivity and resistance. Currently, there is a trend to combine radiation therapy with antitumour agents, with the aim of increasing the combined effect.

2.4.3 Systemic Treatment

Medical treatment of malignancies is the youngest of the treatment modalities currently available. Even if surgery and radiation therapy jointly cure the majority of potentially curable patients, medical treatment is effective in several forms of the disease. Its role has increased considerably over time, as the main challenge is to solve the problem posed by disseminated disease, which currently have different levels of response to treatment (Table 2.1). A large number of antitumoural agents are in clinical use, with different modes of action (for a review, see Chabner and Longo 2006). Examples of *cytostatic/cytotoxic agents* are presented in Table 2.2.

Table 2.2 Examples of cytostatic/cytotoxic drugs in clinical use

Alkylating agents and platinum compounds	*Topoisomeras inhibitors*
Busulfan	Amsacrine
Cisplatin	Daunomycin
Carboplatin	Doxorubicin
Cyclophosfamide	Epirubicin
Chlorambucil	Etoposide
Dacarbazine	Idarubicin
Ifosfamide	Irinotecan
Lomustine	Mitoxantron
Melphalan	Topotecan
Oxaliplatin	
Temozolomide	
Antimetabolites	*Mitotic inhibitors*
Azatioprine	Docetaxel
Capecitabine	Estramustin
Chlorodeoxyadenosine	Paclitaxel
Cytarabine	Vinblastine
Fludarabine	Vincristine
Gemcitabine	Vinorelbine
Mercaptopurine	
Methotrexate	
Thioguanine	
Other drugs	
Bleomycin	
Actinomycin D	
Mitomycin C	

2.4.3.1 Cytostatic/Cytotoxic Agents

Alkylating agents bind preferentially to the N7-position of guanine in DNA and are effective in the resting phase of the cell cycle. Some are monofunctional alkylating agents, but the majority of the agents are bifunctional and form intra- and interstrand crosslinks in DNA. There are two main groups of alkylating agents. Nitrogen mustard gas derivatives, which include melphalan, chlorambucil, cyclophosphamide, and ifosfamide, are the most frequently used agents and are clinically active in the treatment of both haematological malignancies and solid tumours. The other group includes the nitrosureas which are represented by the lipid soluble chloroethylnitrosourea compounds carmustine (BCNU), lomustine (CCNU), and methyl-CCNU. Following intracellular degradation, the active part of the chloroethylnitrosourea binds to the O^6-position of guanine before crosslinking the DNA. The O^6 of guanine is the target for the DNA repair protein O^6 –methylguanine-DNA methyltransferase, which can modify the cytotoxic effect of chloroethylnitrosoureas (Bobola et al. 2005; Hansen et al. 2007). Antitumour effects are

observed in several types of malignancies such as lymphomas, small-cell lung cancer, melanoma, and brain tumours. Streptozotocin and chlorozotocin are variants of nitrosureas; they are monofunctional alkylating agents that are active in the treatment of endocrine pancreatic cancer. Busulfan, an alkyl alkan sulfonate, causes bifunctional DNA crosslinks and its antitumour activity has been shown in CML (chronic myelogeneous leukaemia). Procarbazine, a monofunctional akylating agent, is used in combination chemotherapy for malignant lymphoma, while dacarbazine (DTIC) is active in malignant melanoma. Temozolomide has a similar mode of action and has shown positive clinical effects in brain tumours and malignant melanoma.

Platinum compounds react with DNA, noting that the most frequent DNA lesions are intrastrand crosslinks. Cisplatin is active in several types of solid tumours, particularly in testicular carcinoma. Combination of chemotherapy with cisplatin cures 70% of patients with disseminated testicular cancer (Ehrlich et al. 2010). Carboplatin, a second-generation platinum agent, has similar antitumour effects, while oxaliplatin is a more complex molecule active in the treatment of colorectal cancer.

Topoisomerase inhibitors interfere with the rejoining of DNA strands after topoisomerase action (topoisomerase I and II). Antitumour antibiotics like the anthracyclines daunorubin and doxorubicin are specific for the action of topoisomerase II. Daunorobicin has a role in treatment of leukaemias, while doxorubicin is active in the treatment of several solid tumours. Epirubicin is a derivative of the anthracycline molecule with less cardiac toxicity as compared to doxorubicin. Idarubicin is another anthracycline with activity in CML. Etoposide and teniposide are podophyllotoxin derivatives that interact with topoisomerase II. Etoposide has a role in the treatment of lung cancer, leukaemia, lymphomas, and testicular carcinoma. Derivatives of camptotecin, such as topotecan and irinotecan, are inhibitors of topoisomerase I that have antitumour effects for several solid tumours. Mitoxantrone is used in the treatment of breast cancer and haematologic malignancies and amsacrine in the treatment of acute leukaemia.

Antimetabolites are substances that simulate normal precursors of DNA and RNA and are cell-cycle specific. Methotrexate is a folic acid analogue that inhibits dihydrofolate reductase. It is active in treatment of leukaemias and lymphomas, but also in the treatment of several solid tumours. Mercaptopurine, thioguanine, and azathioprine (Imuran, an immuno-suppressant), are purine analogues which are phosphorylated to triphosphates with inhibitory effect on DNA synthesis. They are used in the treatment of haematologic diseases. The purine nucleoside analogues chlorodeoxyadenosine and fludarabine are mainly active in myeloproliferative diseases. Cytarabine is a pyrimidine analogue active in treatment of leukaemia. 5-Fluorouracil is a fluoropyrimidine with antitumour activity in treatment of colorectal and breast cancer. Capecitabine is a pro-drug which is metabolized to 5-fluorouracil in the tumour. Gemcitabine, a cytidine analogue, inhibits ribonucleotide reductase and can also be incorporated into DNA. Gemcitabine is active in the treatment of several solid tumours, such as pancreatic, non-small-cell lung, and head-neck cancer.

Mitotic inhibitors are represented by vinca alkaloids (vinblastine, vincristine, vindesine, and vinorelbine), compounds that act by binding to tubulin. Neurotox-

icity is a common side-effect particularly after treatment with vincristine, which is active in leukaemia, lymphoma, and testicular carcinoma. Vinblastine has fewer neurotoxic side-effects and is used in combination therapy for testicular and ovarian carcinoma as well as lymphoma. Vindesine is used in the treatment of ALL (acute lymphatic leukaemia) in children, malignant melanoma, and squamous cell carcinoma. Vinorelbin is used in non-small-cell lung and breast cancer and has limited toxicity to normal tissues. Taxanes (paclitaxel and docetaxel) are mitotic inhibitors that interfere with microtubule function. Anticancer activity has been demonstrated for non-small-cell, ovarian, and breast cancer. Estramustine, composed of estradiol and non-nitrogen mustard, is cytotoxic mainly by binding to tubulin and is used in the treatment of hormone refractory prostate cancer.

Bleomycin is a mixture of glycopeptides. Its cytotoxic effects are most probably caused by strand brakes in DNA. Bleomycin is used in the treatment of malignant lymphoma, testicular cancer, and squamous cell carcinoma of the head and neck and in combination with radiation therapy for penis and anal cancer. Actinomycin D, an inhibitor of RNA synthesis, is active in colon carcinoma, Wilm's tumour, neuroblastoma, embryonic rhabdomyosarcoma, and Ewing sarcoma. Mitomycin D is a DNA crosslinker with antitumour effect in several solid tumours.

2.4.3.2 Endocrine Treatment

A special area of medical oncology is endocrine treatment of malignant tumours, as several tumour diseases are dependent on both steroid and peptide hormones. Most endocrine therapies aim at decreasing the stimulating effects of steroid hormones in malignant cells. Examples are anti-oestrogen and aromatase inhibition of breast cancer (Dowsett et al. 2010) and LHRH agonists for treatment of cancer of the breast (Sharma et al. 2008) and prostate (Albertsen 2009). Glycocorticoids may be effective in the treatment of lymphomas, and gestagenes in the treatment of breast cancer and corpus carcinoma. Tyrosine is used for treatment of thyroid carcinoma with the aim of decreasing the stimulatory effect of thyreotropin on the disease.

2.4.3.3 Targeted Treatment

During recent decades, targeted therapies have been developed on the basis of enhanced knowledge of the molecular mechanisms underlying cancer (Chabner and Longo 2006; Baselga 2006; Baselga and Swain 2009; Yamanaka and Saya 2009, and references therein). The main molecular targets used to develop anticancer drugs are cell surface receptors, signal transduction pathways, gene transcription targets, ubiquitin-proteasome/heat shock proteins, and tumour microenvironment constituents (Fig. 2.1). Examples of targeted drugs are given in Tables 2.3 and 2.4 and Fig. 2.1. Currently, there are widespread efforts to identify new targets and develop new compounds, and it is expected that about 10 new antitumoural agents

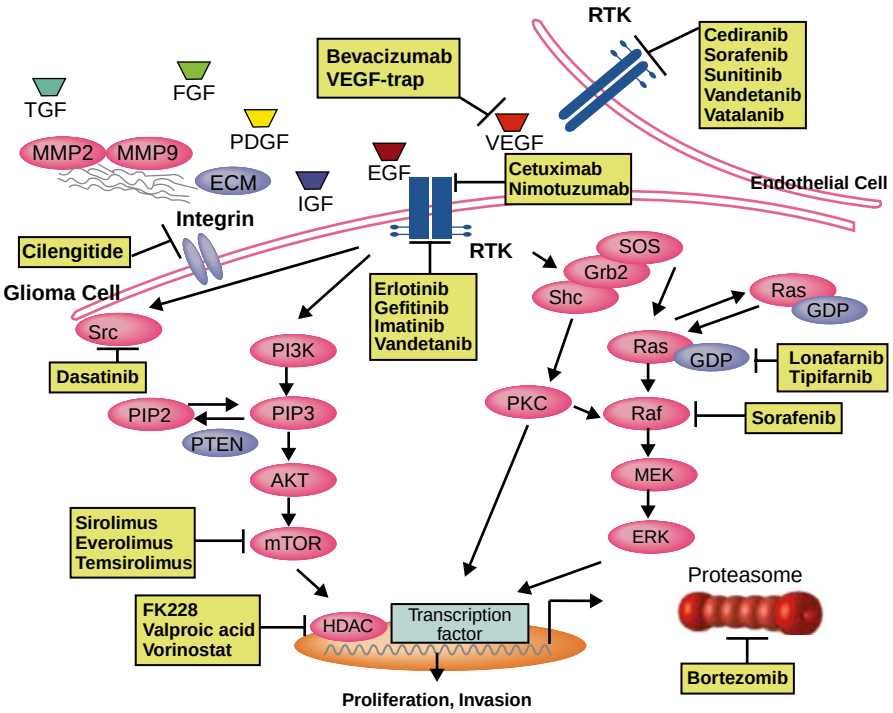


Fig. 2.1 Molecular targeted therapy for glioma. A representation of signalling pathways and therapeutic molecular targets in glioma cells. ECM = extracellular matrix; EGF = epidermal growth factor; ERK = extracellular regulated kinase; FGF = fibroblast growth factor; GDP = guanosine diphosphate; Grb2 = growth factor receptor-bound protein; GTP = guanosine triphosphate; HDAC = histone deacetylase; IGF = insulin-like growth factor; MEK = mitogen-activated protein extracellular regulated kinase; MMP = metalloproteinase; mTOR = mammalian target of rapamycin; PDGF = platelet-derived growth factor; PI3K = phosphatidylinositide-3-kinase; PIP2 = phosphatidylinositol (4,5) biphosphate; PIP3 = phophatidylinositol (3,4,5) triphosphate; PKC = protein kinase C; PTEN = phophatase and tensin homolog; RTK = receptor tyrosin kinase; Shc = Src homology 2 domain containing transforming protein; SOS = son of sevenless homolog; Src = Schmidt-Ruppin A2 viral oncogene homolog; TGF = transforming growth factor; VEGF = vascular endothelial growth factor. From Yamanaka and Saya (2009)

will be registered per year in the near future. Estimates indicate that there are 800–1000 new compounds in the pipeline for anticancer drug development.

Today, most of the targeted therapies are aimed at growth factor signalling pathways and tyrosine-kinase receptors. The main group of trans-membrane tyrosine kinase receptors correspond to the Erbb family that includes EGFR (epidermal growth factor receptor, HER1), ERBB2 (HER2), ERBB3 (HER3) and ERBB4 (HER4). The receptors have an extra-cellular domain which is a target for the ligand, a transmembrane segment, and an intracellular tyrosine kinase domain. Binding of the ligand to the receptor causes its dimerization and this in turn leads to receptor autophosphorylation and to pathway activation and a cascade of downstream events.

Table 2.3 Examples of protein kinase inhibitors

	Target	Clinical effects in
Imatinib	BCR-ABL	CML
	c-Kit	GIST
	PDGFRA	Chronic myelomonocytic leukaemia
Gefinitib	EGFR	Small-cell lung cancer
Erlotinib	EGFR	Non-small-cell lung cancer
		Pancreatic cancer
Sunitinib	VEGFR	GIST
	PDGFR	Renal cell carcinoma
	c-Kit	
Sorafenib	VEGFR	Liver carcinoma
	PDGFR	Renal cell carcinoma
Dasatinib	SRC-ABL	CML
		ALL
Lapatinib	EGFR	Breast cancer
	HER2	
Nilotinib	BCR-ABL	CML
Temsirolimus	mTOR	Renal cell carcinoma
		Mantle cell lymphoma

Table 2.4 Examples of monoclonal antibodies

	Target	Clinical effects in
Rituximab	CD20	Follicular non-Hodgkin's lymphoma
		B-cell lymphoma
		CLL
Trastuzumab	HER2	Breast cancer
Alemtuzumab	CD52	CLL
		Sézary's syndrome
		Cutaneous T-cell lymphoma
Cetuximab	EGFR	Colorectal cancer
		Head-neck carcinoma
Bevacizumab	VEGF	Colorectal cancer
		Breast cancer
		Non-small-cell lung cancer
Panitumumab	EGFR	Colorectal cancer

In many malignancies, overexpression and mutations of receptor tyrosine kinases cause pathologic molecular signalling which leads to uncontrolled cell proliferation and invasion. So far, drugs interfering with EGFR and HER2 have shown the most significant clinical effects. Targeting aberrant tyrosine kinase activities has opened new possibilities for therapies having more specific antitumour activity, and also fewer side-effects.

Imatinib, a small molecule inhibitor that binds to the site of tyrosine kinase activity, is a potent BCR-ABL inhibitor and the standard of care for first-line treatment of chronic myeloic leukaemia (CML) (Druker et al. 2001; O'Brien et al. 2003; Bacca-

rani et al. 2009). Imatinib also blocks other tyrosine kinases like c-Kit and PDGF receptors which are aberrantly expressed in GISTs (gastrointestinal stromal tumours) (van Oosterom et al. 2001; Heinrich et al. 2003; Heinrich and Corless 2004). Owing to other PDGF alterations, the antitumour activity of imatinib has also been demonstrated for chronic myelomonocytic leukaemia (Baselga and Arribas 2004).

Trastuzumab is a monoclonal antibody that binds to the extra cellular domain of HER2. Amplification of the HER2 gene occurs in about 25% of invasive primary breast cancers (Slamon et al. 1987).Treatment of HER2-positive early breast cancer patients with trastuzumab has shown significant reduction of recurrent disease (Baselga 2006; Smith et al. 2007).

Cetuximab is an anti-EGFR monoclonal antibody with clinical effect in colorectal and head and neck cancer. Gefinitib and erlotinib are small molecules that inhibit the tyrosine kinases associated with EGFR, effective against lung and pancreatic cancer. Dasatinib, a second generation of small molecules targeting tyrosine kinases, is clinically active in patients with imatinib-resistant CML. Sunitinib, which targets different tyrosine kinases associated with EGFR, PDGFR, and c-Kit, has been shown to have anti tumour activity in imatinib resistant GISTs and renal cell carcinoma. Lapatinib, on the other hand, is a second generation of tyrosine kinase inhibitors associated with EGFR and HER2 that exhibits antitumoural effect in trastuzumab-resistant breast cancer. Sorafinib targets VEGFR and PDGF and is effective for the treatment of liver cancer and renal cell carcinoma. Bortezomib is a proteasome inhibitor which is used in the treatment of multiple myeloma. Retuximab binds to CD20 with antitumour activity on follicular non-Hodgkin's lymphoma, diffuse large-cell B-cell lymphoma, and CLL (chronic lymphatic leukaemia). Alemtuzumab is another antibody that targets CD52 with clinical effects in treatment of CLL, Sézary's syndrome, and cutaneous T-cell lymphoma. Bevacizumab targets VEGF and interferes with the angiogenesis regulatory process. The antibody has been shown to have clinical effect for the treatment of colorectal carcinoma.

2.4.3.4 Other Treatment Modalities

Biological treatment includes a group of agents with natural functions in the body. Some of them have antitumour characteristics, as is the case for interferons and interleukins. Throughout the years, different types of immunotherapies have been explored, based on antibodies and cellular immunity (for a review, see Chabner and Longo 2006). From the increased knowledge about molecular defects causing tumour development, different possibilities to correct these defects have been identified in experimental systems, but so far gene therapy remains an experimental approach.

2.4.3.5 Drug Resistance

Drug resistance is becoming an increasingly important clinical problem. Some tumours like melanoma, renal cell carcinoma, and non small-cell lung cancer are

often primarily resistant, while others like myeloma and breast cancer may recur after a remission. Recurrent disease may depend on the heterogeneity of the tumour cell population with a minority of resistant cells surviving the primary treatment. Several types of resistance mechanisms have been described. Some antitumour agents are dependent on active transport through the cell membrane, and decreased uptake of the drug may lead to resistance. One special resistance mechanism is linked to the action of the 170 kDa membrane protein from the MDR-1 (multidrug resistance) gene which acts to transport molecules outside the cell, some cytostatics included (Marie et al. 1991). To be active, a number of anticancer agents require metabolization by specific enzymes. On the other hand, cytostatics may be inactivated by glutathione conjugation. Thus, changes in activation as well as inactivation may lead to drug resistance. Genetic instability of tumour cells may represent one of the main causes of acquired resistance. Gene amplification and mutations may cause alterations in the amount of target molecules, or cause qualitative changes of these molecules with decreased efficacy of the anticancer agent. Drugs causing DNA damage will be more or less effective depending upon DNA repair mechanisms. Examples are activities of 06-methylguanine-DNA-methyltransferase and nucleotide excision repair. Several antitumour agents induce apoptosis, and changes in apoptosis-regulating molecules may simultaneously confer drug resistance. This can be seen as a system-level phenomenon (see Chap. 12 and 17).

2.4.3.6 Side-Effects

An important aspect of treatment with anticancer agents is side-effects. Acute side-effects like nausea, vomiting, and fatigue occur frequently. Depending on their mode of action, different anticancer agents cause organ related side-effects. A large number of anticancer agents produce alopecia, and neurotoxicity is common following treatment with vincristine and cisplatinum. A majority of antitumour agents cause bone marrow toxicity, and gastrointestinal side-effects are common after treatment with antimetabolites. Anthracyclines can induce specific cardiotoxicity. Cisplatinum and methotrexate (Ledertrexate, an antineoplastic antimetabolite with immunosuppressant properties) are examples of drugs causing nefrotoxicity. A long-term side-effect may be secondary tumours caused mainly by alkylating agents.

2.4.4 Treatment Strategies

Combining antitumour agents with different modes of action has been successful in the treatment of lymphomas and leukaemias (DeVita et al. 2008), with strong posi-

tive effects being demonstrated in paediatric oncology (Pinkerton et al. 2007; Trigg et al. 2008). With some exceptions, e.g. in testicular carcinoma (Ehrlich et al. 2010), treatment of solid tumours poses a more difficult problem. An important concept is applying adjuvant systemic treatment at an early stage of the metastatic disease. For example, it has been shown that for breast cancer, being subject to most oncologic activities, adjuvant systemic treatment after primary surgery and radiation therapy in high-risk individuals, significantly improves survival (Clarke et al. 2005, 2008; Cuppone et al. 2008; Madarnas et al. 2008). For treatment of patients with solid tumours, the general trend is towards systemic treatment after primary surgery/radiation therapy for high-risk individuals, regarding the presence of micro-metastases. Another application of chemotherapy for solid tumours is in preoperative treatment in order to reduce the tumour volume before surgery and to treat potential metastatic disease as early as possible.

There are a large number of anticancer agents with positive clinical effects on only a fraction of patients and often for only a limited time. To improve cure rates, treatments tailored to the individual patient are required, taking into account the specific phenotypic and molecular characteristics of the tumour and the patient. This includes treatment of the cancers at an early stage. Early detection should therefore include prediction or detection of micrometastatic disease. Systemic treatment is indicated in such cases, as it will increase the probability of cure.

Personalized cancer medicine has as its goal delivery of ‘the right treatment to the right patient at the earliest possible time’. The time is now ripe to identify and validate prognostic biomarkers anticipating the risk that the patient will develop progressive disease and predictive biomarkers that point to the likely response of the tumour to a particular intervention as well as side-effects. There are three steps to biomarker discovery: identification, retrospective validation in archival material, and prospective validation in clinical trials. For this purpose we need molecular pathway-driven and adaptive clinical trials, before the potential benefit of predictive biomarkers used for stratification of patients will be evaluated in randomized clinical trials. Since one expects a high correlation between the biomarker profile and the response to a particular treatment, this approach requires trials with only a small number of patients in a specific treatment area. This strategy, along with the implementation of systems biology and systems medicine approaches, will likely help in solving the dilemma of the present approach of evidence-based cancer medicine, where large numbers of patients are needed for time-consuming trials which involve identifying small differences between treatment groups. This should lead to a much faster development and implementation of new therapies. Molecular pathology/cytology will be fundamental in predicting prognosis and treatment outcomes. The development of imaging technologies, already spectacular, will benefit in the future from molecular pathology discoveries, making the biological analyses of tumours possible with fewer surgical intervention (Fass 2008).

2.5 Major Cancers, Diagnosis, Disease-specific Supplementary Classifications, and Treatment Implications

In light of the evidence discussed so far we will now discuss the features and treatment of some of the 'big killers' at a level of detail that illustrates the complexity which needs to be ultimately addressed by systems approaches in conjunction with existing classification methods, to produce improved outcomes for patients.

2.5.1 Colorectal Cancer

Colon and rectal cancer (CRC) is a major cause of death in Western countries and the incidence is increasing rapidly worldwide due to the adoption of Western lifestyles and increased ageing of the population. About 1 out of 20 people will be affected by this disease, and in the European Community alone around 200,000 new cases per year are diagnosed. Nevertheless, in recent years, understanding of the initiation and progression of the adenoma-to-carcinoma sequence, and the genes involved in these processes, has increased enormously, opening up new preventive and therapeutic approaches. Survival has increased thanks to better surgical treatments, especially for rectal carcinoma, with a reduction in local recurrences, and to new targeted therapies with monoclonal antibodies, together with neoadjuvant radiotherapy and improved conventional chemotherapy.

In all these steps, accurate pathological reporting of the specimens plays an increasing role in two situations: the diagnostic and the post-surgical phases (Bosman 1995). The role of histopathology consists both in confirming the presence of clinically suspected intestinal adenoma or atypical adenoma to *in situ* or invasive carcinoma sequence, and is furnishing a precise description of the surgically resected specimen, including the resection margins and the nodes. This double role will influence the therapeutic approach and clinical management, and predict the survival of the patient.

The modified, original Duke's prognostic classification (Hutter and Sobin 1986; Whitehead 1994; Fenoglio-Preiser et al. 1999) is based upon pathological examination of the surgically resected specimen and the assessment of the tumour extension in the depth of the intestinal wall affecting or not the neighbouring tissues together with nodal invasion or the presence of distal metastasis. This classification has been further improved with the addition of TNM staging (Edge et al. 2010). T describes the size of the tumour and whether it has invaded nearby tissue, N describes regional lymph nodes that are involved, M describes distant metastasis (spread of cancer from one body part to another).

Duke's classification is based upon the depth of tumour invasion:

A: limited to the mucosa;
B1: limited to the *muscularis propria*, but with negative nodes;

B2: penetrates the *muscularis propria*, with negative nodes;
C1: limited to wall, but with positive nodes;
C2: extends through the wall and nodes are positive;
D: any grade, but with the presence of distal metastases.

Colorectal cancer–TNM classification (Sobin et al. 2009) includes four stages:

Stage I: Tumour invades muscularis propria, but has not spread to nearby lymph nodes.
Stage II: Tumour spreads into the subserosa and/or perirectal tissues with up to three regional lymph nodes, or directly invades adjacent tissues without lymph node involvement.
Stage III: Any depth of tumour invasion, with four or more positive lymph nodes, but without distant metastases.
Stage IV: Any depth of tumour involvement, any number of involved lymph nodes, with distant metastases.

Staging at diagnosis has been recognized as the most powerful indicator of clinical outcomes (e.g. survival rates) in patients with colorectal malignancies, which varies from 3% to 100%, depending upon the grade and stage of the tumour. An especially important aspect of staging is the determination of the N status. The introduction of sentinel-node detection has proved to be of value, similar to that in breast carcinoma. In the majority of patients, the sentinel node procedure is successful, and almost one-quarter of the clinically suspected node-negative patients have microscopic disease, which has profound implications for the outcome of the disease. However, approximately one-third of node-negative patients have recurrent disease. Another important parameter affecting prognosis and guidelines is accurate pathological staging after neoadjuvant radio- and chemotherapy prior to surgery. An added problem is the morphological prediction of response to therapy, but several approaches have been proposed with promising results (Dworak et al. 1997; Saad et al. 2006).

Histological typing of the tumour follows the accepted international histological classification of CRC proposed by the WHO (Hamilton and Aaltonen 2000). Adenocarcinoma is the basic epithelial neoplasm displaying several histological varieties (mucinous-colloid, signet-ring cell, medullary, small-cell, squamous cell, adenosquamous, and undifferentiated). Several grading systems have been proposed, but a common accepted standard is lacking. The College of American Pathologists proposed a system based upon the proportion of gland formation, with an average of more than, or less than 50%, and also 0%, thus distinguishing between well-differentiated and undifferentiated adenocarcinoma. The proportion of glands present in the tumour allows a diagnosis of well/moderate, low grade (grades 1 and 2) to poorly/undifferentiated, high grade carcinomas (grades 3 and 4). Some types are directly categorized as high grade: signet ring, small-cell, and undifferentiated carcinomas, while medullary carcinoma has a better prognosis. Neuroendocrine carcinomas, which may present a small-cell pattern or be undifferentiated with large cells, display a worse prognosis. The presence of isolated neuroendocrine cells in

a conventional adenocarcinoma does not, however, imply an adverse prognosis. In fact, patients with the worse prognosis tend to be at a more advanced stage of the disease, according to pTNM classification and pStage.

In colo-rectal cancer, completeness of the mesorectum removal during surgery allows a good assessment of the adequacy of excision and the regularity of the circumferential resection margin, while the evaluation of mesorectal completeness provides significant information on the prognosis. Patients with an incomplete mesorectum have a higher risk of recurrence and the circumferential margin involvement is one of the most powerful predictors of local recurrence. Although there has been discussion about the definition of positive margins, Nagtegaal and van Krieken (2002) have shown that in order to predict local recurrence, margins smaller than or equal to 2 mm should be regarded as involving an increased risk of local recurrence, while margins of 1 mm or less are predictive of an increased risk of developing distant metastases and therefore of shorter survival times.

A decisive factor in prognosis is the presence of nodal metastases at the time of surgical treatment. The TNM guidelines recommend examining at least 12 lymph nodes; they must be tumour-free before a case can be classified as N0. In daily practice this number is hard to attain, and the number may vary from one patient to another; in elderly patients retrieved lymph-nodes are less numerous. In addition, pre-operative neoadjuvant treatment (normally chemo-radiotherapy) influences the number and status of lymph nodes that are resected and examined. Detailed macroscopical examination of the resected specimen is an important element in determining the number of lymph nodes as well; there are in fact considerable differences in the numbers of lymph nodes retrieved by different pathologists in different institutions or even within the same laboratory (Ingoldsby and Callagy 2009).

However, data obtained from colon or colorectal cancer patients may not be applicable to rectal cancer patients. It is clear that the mean number of examined nodes in the ascending and descending colon may be almost twice as high as the number of nodes in the rectum. Very small lymph nodes that are missed using routine examination can be detected using fat clearance techniques. However, these techniques are time-consuming and expensive and might interfere with determination of the circumferential resection margin. Immunohistochemical and molecular support provide promising results, but their relevance is still not clear.

Several linear analyses (Jass 2007; Markowitz and Bertagnolli 2009) are consistent with the subdivision of all CRC developed in the general population into five major classes: Hereditary nonpolyposis colorectal cancer (HNPCC), suspected HNPCC, juvenile cases, familial tumours, and apparently sporadic cases. All these cases evolve through several genetic pathways defined by specific molecular expressions: (1) DNA microsatellite instability (MSI) status being sub-stratified as MSI-high (MSI-H), MSI-low (MSI-L) and MS stable (MSS), and (2) CpG island methylated phenotype (CIMP) divided as CIMP-high, CIMP-low and CIMP-negative (CIMP-neg). Jass (2007) proposed the presence of a morphological correlation in the five molecular subtypes:

Type 1 (CIMP-high/MSI-H/BRAF mutation),
Type 2 (CIMP-high/MSI-L or MSS/BRAF mutation),

Type 3 (CIMP-low/MSS or MSI-L/KRAS mutation),
Type 4 (CIMP-neg/MSS) and
Type 5 or Lynch syndrome (CIMP-neg/MSI-H).

These molecular states can be detected at an early evolutionary stage and are present in polyps with precancerous lesions. For instance, serrated polyps are the precursors of Types 1 and 2 of CRC, whereas Types 4 and 5 evolve through the steps of conventional adenomatous polyp, *in situ* carcinoma, and invasive carcinoma sequence. Type 3 CRC may arise within either type of polyp.

In addition, a better understanding of the molecular pathways that characterize CRC cell growth, cell cycle, apoptosis, angiogenesis, and invasion has led to the identification of novel targets for cancer therapy. Duff et al. (2006) have given an interesting overview on proteins that play a role in CRC, grouped according to their location in the cell: membrane receptor targets (epidermal growth factor receptor (EGFR), vascular endothelial growth factor receptor (VEGFR), insulin-like growth factor receptor (IGFR), platelet-derived growth factor receptor (PDGFR), tumour necrosis factor-related apoptosis-inducing ligand receptor (TRAIL-R), and c-Met, intracellular signalling targets (Ras/Raf/MAPK pathway, phosphatidylinositol- 3-kinase (PI3K)/Akt/mammalian target of rapamycin (mTOR), src kinase, and p53/Hdm2, as well as other protein kinases that control cell division.

These molecular approaches give further understanding to the sequence of already well known morphological events. This evolution is a consequence of an aberrant gain in cell proliferation that starts with somatic or germinal mutations and leads to the gene instability that typifies the progression of CRC. Thanks to these findings, medicine today has a better opportunity to prevent and clinically manage the polyp-adenoma sequence before malignant transformation (Fearon and Vogelstein 1990; Lynch and de la Chapelle 2003).

2.5.2 Breast Carcinoma

Breast neoplasms comprise a heterogeneous group of proliferative processes including distinct entities, many of which present a malignant behaviour. Breast carcinoma is the primary cause of cancer morbidity in women in developed countries, and its incidence is increasing worldwide, resulting in more than 500,000 deaths annually according to the WHO.

Pathology participates decisively in combating this process through microscopical diagnosis, which joins other image techniques such as routine mammographic screening, echography, and MRI. Needle-core biopsy (NCB) is widely used, being a well validated technique that reduces the need for diagnostic breast surgery and almost totally excludes fine needle aspiration cytology. The current increase in cure rates is based on early and generalized tumour screening, improved surgery, better radiation therapy, and more effective chemotherapy regimens. Nevertheless, breast cancer mortality remains high, not only when the diagnosis is performed at advanced stages (Stages III and IV), but also in subgroups of patients affected with

small tumours in early stages (stage I and stage II) prior to any evidence of distant metastasis.

The advances attained on knowledge of the disease have been spectacular in recent decades; mainly thanks to the fields of genetics and molecular biology. New biological modulators and monoclonal antibodies targeted to specific cell pathways disturbed in cancer, have contributed to identifying particular types of tumours that benefit from a more effective therapy. Nevertheless, progress remains limited and many questions these new bio-technological approaches have raised are still largely a matter of laboratory and clinical investigation. Closer integration by systems approaches of diverse fields of science is a pressing and necessary desideration for conquering the scourge of cancer.

For many years the grading and the staging of breast tumours (Tavassoli and Devilee 2003) has constituted the basis for therapy, follow-up, and prognosis. In this context, determining the clinical status of the patient and the pathology of the tumour (TNM) is mandatory before starting any type of treatment. These include a number of clinical parameters such as the tumour size, the nodal status, the presence or absence of distant metastasis, and the histological type and grade. These have been historically complemented with the determination of the status of oestrogen (ER) and progesterone (PgR) receptors and more recently of the epidermal growth factor receptor 2 (HER2/neu) (Payne et al. 2008; Faratian and Bartlett 2008).

Malignant progression to carcinoma of the breast follows a successive number of steps similar to what happens in other glandular epithelia. In this process, a ductal-lobular unit initiates a hyperplastic focus proceeding to a usual dysplasia, which leads to an atypical ductal or lobular dysplasia (atypical intraductal or lobular proliferation), and eventually into an *in situ carcinoma* (ductal or lobular) which may progress to invasion and metastasis into axillary nodes or distant organs. Other changes include columnar cell changes, complex sclerosing lesions, and papillary proliferations. Obviously, not all breast carcinomas necessarily follow these successive steps, and a few probably become malignant *ab initio* (from the matter) without manifesting the precancerous stages. Fortunately, a large number of women suffer a dysplasia that will never progress into carcinoma.

The NHS proposed in 2001 (Hayes and Quinn 2009) a coding system designed to simplify the screening programs as guidelines for classifying the lesions obtained with NCB:

- normal non-malignant breast (B1),
- benign (B2),
- uncertain malignant potential (B3),
- suspicious of malignancy (B4), and
- malignant (B5).

Why there is so much heterogeneity in the biological and pathological behaviour of the mammary gland is still poorly understood. This diversity affects not only the histology of the dysplastic lesions, but also the varieties of *in situ carcinoma* including their transition to infiltrating carcinoma. Below we highlight some particular breast cancer aspects of more general interest.

2.5.2.1 Histological Types: WHO Classification and Nottingham Grading

Three major histological varieties (Tavassoli and Devilee 2003) of invasive carcinomas can be distinguished: ductal carcinoma (IDC), some of which may lack a specific organization (not otherwise specified, NOS) 70%, lobular carcinoma (ILC) 8%, and combined infiltrating ductal lobular carcinoma (IDLC) 12%. In addition, there are around 10% of cases with miscellaneous phenotypes (colloid, mucinous, secretory, medullary, papillary, micropapillary, anaplastic, etc.).

The most frequent form of breast carcinoma is IDC. This category mainly comprises those easily identifiable cases with enlarged ducts filled with more or less pleomorphic cells and necrosis (comedo-necrosis is not always seen). They present a varied diversity of stroma infiltrations such as ducts, cords, papillae, or solid nests that extend into the fat tissue and invade local structures, large ducts, deep muscle, or superficial dermis. Most of the NOS varieties initiate as an IDC. The limits of the neoplasm are usually better defined than in ILC and contain grouped microcalcifications which help the mammographic detection in the early invasion stage or even as carcinoma *in situ*. Desmoplasia varies and the elastosis produces a central sclerotic core that mimics a benign radial sclerosis or sclerosing complex, lesions being visible senographically as a dense stellate lump. Their size varies and is a prognostic factor together with the presence or not of axillary nodal metastasis. Some histological varieties display better prognosis than others; tubular, papillary carcinomas are low-grade tumours, while solid invasive comedocarcinoma, micropapillary, or colloid IDC are high-grade. IDC varies widely in the expression of ER, PgR, HER2/neu, cytokeratins, EGFR, or proliferative markers such as MIB1. In addition, most cases express the E-Cadherin adhesion molecule, which helps differentiate them from the ILC negativity. New molecular microarray genetic expression analysis has led to a more advanced molecular classification with the support of IHC in paraffin slides, complementing present histological types and better adapted to the available targeted therapies (Reis-Filho et al. 2006; Badve and Nakshatri 2009).

ILC shows peculiar clinical characteristics in its lack of microcalcifications, the absence of distinct borders, and the difficulty in visualizing it with conventional mammography. Histological analysis may reveal remaining *in situ* areas with dilated lobules filled with small cells of bland nuclear appearance, associated with infiltration in files of single cells (Indian files) caused by the loss of expression of certain adhesion molecules such as E-Cadherin. There are subtypes that display major nuclear polymorphism and solid, alveolar, or mixed architecture with desmoplasia. Hormone receptors (ER and PgR) are positive in 90% of cases, while HER/neu2-positive cases are low (about 8%) and mostly correspond to the pleomorphic subtype.

In terms of prognosis, ILC presents a better clinical outcome compared to IDC in early years, but this trend reverses later on, 6 years or more after diagnosis, achieving IDC mortality rates. Local recurrences and contralateral association are higher when compared to IDC, but total mastectomy is not justified, and lumpectomy is the choice for small tumours.

Clinical and histological data may be used to stratify the carcinoma in order to determine its prognosis and therapy. The pTNM and WHO Stage classifications are based upon tumour size, axillary nodes status, and the presence of distal metastasis, in conjunction with histologic type and grade (Bloom and Richardson 1957 modified by Elston and Ellis 1998), known as the Nottingham grade. Computation of all this information stratifies carcinomas into low, intermediate, and high-grade, with an apparent different clinical outcome (relapse-free disease, disease-free survival, and overall survival) dependent on the response to therapy which combine neoadjuvant and/or adjuvant chemotherapy, hormonal inhibitors, surgery, and high-energy beam radiotherapy, in regimens, depending on the patient's age, the disease stage, and hormonal status (Eden et al. 2004).

There are, however, shortcomings in the results; they are not always as clear-cut as might be expected. Divergences between the clinical behaviour of tumours considered as low grade, but which relapse early or present shortened overall survival, are in contrast with high-grade neoplasms that display an unexpected favourable clinical outcome. Additionally, carcinomas with analogous grade and stage may respond differently to a similar therapeutical protocol. Mortality from breast cancer has tended to decline; an increase in survival of over 10% since the late nineties is attributable in part to early detection by population screening and the implementation of hormonal and adjuvant chemotherapy. However, the disease still remains a major cause of morbidity and mortality, and global survival rates are not satisfactory. Further improvements are expected, given newly available diagnostics and targeted therapeutical agents.

2.5.2.2 Immunohistochemistry and a New Molecular Classification

In recent years, microarray gene expression measurements on fresh tissue RNA, analysed by hierarchical clustering, have offered new possibilities for genetic profiling of breast carcinomas (Perou et al. 2000; O'Shaughnessy 2006; Reis-Filho et al. 2006). This technology has provided new sub-classifications, dividing breast carcinomas into groups that facilitate more precise prognostic and therapeutical approaches. Microarray technology is, however, impractical or difficult to implement in daily routine, partly because of current high costs (Desmedt et al. 2008; Correa Geyer and Reis-Filho 2009).

Nevertheless, these advances have provided seminal information leading to more accessible and cheaper methods, such as IHC in paraffin embedded tissues combined with tissue microarray technology. Using these two techniques, large series of tumours can be tested in single slides with a particular antibody, not only retrospectively, but also in prospective studies. Carcinomas of the breast have been reclassified not only according to their histology, but also by the positive or negative expression of a number of marker proteins that possess clinical, prognostic, or therapeutic relevance. This classification has now been validated by retrospective clinical analysis, and new studies, currently underway, will combine microarray gene expression analysis with the methodology, hope-

fully resulting in a better and more comprehensive view of breast cancer biology (Tang et al. 2009).

Four major types of breast carcinoma are recognized to date (McCafferty et al. 2009). These types are known as: *Luminal A, Luminal B, Basal-like*, and *HER2/neu*; all identified using a four-marker immunopanel based on hormone receptor status: oestrogen receptor (ER) status, progesterone receptor (PR) status, HER2/neu, and Ki-67 proliferation index. At least two to four more subtypes have been proposed. The proposed addition of CK 5/6 and EGFR allows the sub-classification of the Basal-like subtypes in a triple negative and in core basal phenotypes. Moreover, an Apocrine phenotype (based upon AR status and GCDP-15), which is close to the HER2/neu and 'Claudin 1 low' (stem-cell like) phenotypes (Pinero-Madrona et al. 2008), has been proposed. Although this immunohistochemical-molecular classification has attracted wide interest, its clinical validation is still in progress.

The *Luminal A subtype,* which expresses ER and PR positivity, is HER2/neu negative, and displays a low proliferative index (KI-67). This is the most common breast tumour mimicking normal luminal cells (positivity for luminal low-molecular-weight cytokeratins 8/18) involving genes associated with an active ER pathway. It corresponds histologically to low grade carcinomas, mainly ductal, tubular, cribriform, and lobular, following the WHO classification, and therefore presents low clinical stages and favourable prognosis.

The *Luminal B subtype* is the second most frequent breast tumour. They express ER but the PR status is low or negative. The lesions are HER2/neu negative and show a high Ki-67 proliferative index. They consist of derivates of luminal cells (positivity for low molecular-weight cytokeratins 8/18), with activated ER gene pathways and p53 mutations. The histological counterparts are mainly high grade ductal, NOS, and micropapillary carcinomas. Clinical outcome and prognosis is worse than for tumours of luminal type A, but the B subtype responds well to chemotherapy with Taxotere Adriamycin Ciclophosphamide (TAC) or Fluorouracil Adriamycin Cyclophosphamide (FAC), and to hormonal control. The clinical stages may be more advanced (stages II and III).

Basal-like subtypes comprise a low number of tumours (around 15% of breast carcinomas correspond to this category), but they can be subdivided in groups that have prognostic implications (Dent et al. 2007; Cheang et al. 2008). All basal-like carcinomas have the characteristics of positivity for basal high-molecular-weight cytokeratins and specific myoepithelial cell markers (CK5/6, CK17, Caveolin1, Calponin1, P63), as well as lack of ER, PR, and HER2/neu expression (triple negative). The lesions exhibit a high Ki-67 proliferation index and harbour both p53 mutations and DNA repair defects. At present, there is controversy regarding these groups of tumours, because not all triple-negative types are genetically basal-like and not all basal-like genetically confirmed tumours display triple-negative features (Rakha et al. 2006). Moreover, a subgroup of basal-like carcinomas expresses EGFR and C-KIT positivity. The latter have a worse prognosis when compared to the already known unfavourable clinical outcome and poor response to therapy of the subgroup. The basal-like category also covers medullary, adenoid cystic, and metaplastic carcinoma, and includes a small subgroup of high grade NOS in the

WHO classification. The category as a whole has more numerous BRCA1 germ line mutation carriers. Some authors prefer to consider as 'unclassified' those tumours with the triple negative features with negativity for CK5/6 and EGFR should be considered as unclassified tumours (Tang et al. 2009).

The *HER2/neu subtype* comprises carcinomas with definite positivity for immunostaining with this antibody (clone DAKO, 3+) and confirmation by FISH or CRIST analysis. These tumours may belong to the luminal B type, but the majority correspond to the category of ER and PR negative tumours with a high Ki-67 positivity and occasional low CK 5/6 expression. They are very aggressive high-grade ductal NOS carcinomas; but respond well to HER2 tyrosine kinase inhibitors (trastuzumab).

The *Apocrine type* is very infrequently diagnosed. Focal apocrine features are found in the majority of ductal not-otherwise-specified (NOS) carcinomas; the term is limited to exceptionally pure histological apocrine carcinomas. Apocrine metaplasia is very usual in benign ductal dysplasia and less frequent in *in-situ* carcinoma. Clinically, apocrine carcinomas correspond to high grade tumours, are negative for ER and PR, and present AR positivity together with intense, but focal GCDP-15 and occasionally HER2/neu 3+. Their genetic profile has recently been partially identified (Celis et al. 2007, 2008). However, it is not clear if this group, as is also the case for '*Claudin1 low stem cell like carcinoma*' (Pinero-Madrona et al. 2008), constitute particular clinical entities or should be included within any of the above indicated categories.

Normal cell breast-like type carcinoma has been considered by some authors as another specific entity (Sorlie et al. 2006). However, the characteristics of this tumour, detected via unsupervised hierarchical clustering analysis by the Stanford group (Perou et al. 2000), are unclear because it mimics normal epithelial cells; its histological and clinical significance has still to be determined (Brenton et al. 2005; Tang et al. 2009).

Breast cancer control has progressed dramatically in the past few years, mainly because detection of clinical stage I disease has been improved by modern imaging and screening campaigns, thus reducing cases of more advanced phases involving regional tumour spread and distant metastasis. Simultaneously, a better understanding of the morphological and molecular mechanisms underlying the disease has resulted in additional diagnostic tools and better tailored therapy approaches. Incorporation of this knowledge into systems biology is an example of the new modern approaches to enhancing prevention, early diagnosis, and cure of this disease.

2.5.2.3 Tumorectomy vs. Mastectomy: Free Margins and Sentinel Lymph Nodes

Breast cancer is now considered as a systemic disease and not merely a localized problem. Hence, the approach to its control has changed, with surgery limited to lumpectomies with tumour excisions instead of large amputations (radical mastectomy with complete dissection of axillary lymph nodes), and increased use of

adjuvant chemotherapy and targeted drugs including hormonal inhibitors. This new trend allows surgery to be geared to preserving the mammary gland in the case of small localized tumours. The axillary nodes too may be preserved if the sentinel lymph-node is free of metastasis.

The pathologist plays a seminal role in these new therapeutic directions. Keeping margins free from carcinoma invasion is essential, if local tumour relapse is to be avoided. To this end, it is necessary to excise at least 10 mm of tumour-free margins containing normal tissue surrounding the neoplasm (Lopez-Guerrero et al. 2006). Several procedures have been proposed for determining the extent of the tumour-free margins in the resected surgical specimen; the usual procedure involves colouring the resected specimen with permanent ink that resists embedding in paraffin and does not discolour in the histological section. The practice of free-margin tumorectomies does not preclude the need to complement the area around the excised tumour with local irradiation therapy: the incidence of tumour relapse decreases from 10% to 3–4% in irradiated patients (Lopez-Guerrero et al. 2006).

Sentinel-node biopsy is a standard procedure in operable breast cancer for clinically negative axillae, replacing axillary node dissection in the staging of breast carcinoma. Its value has been confirmed by numerous randomized studies demonstrating improved quality of life and reduced morbidity in patients that undergo this procedure.

Nevertheless, several problems have been raised regarding this new procedure, both of a technical and oncological nature. A technical problem is whether the pathologist should perform a diagnosis of frozen samples, with or without the support of IHC. Another is the variability of histological techniques employed, leading to diversity of results emanating from the different laboratories. Although the American College of Pathology guidelines do not recommend routine IHC as mandatory for diagnosis, many pathologists use it (pan-cytokeratins cover over 90% of epithelial cells, but also non epithelial cells such as the normal reticular cell) and consequently not only metastasis (>2 mm in size) but also micrometastases are detected (0.2–2 mm or isolated or grouped free tumour cells located in the cortical sinus). An oncological problem concerns the determination of micrometastases.

Another controversy concerns the substitution of conventional microscopy diagnosis traditionally performed by the pathologist, with molecular assaying techniques (RT-PCR) based on GeneSearch and OSNA by Sysmex, both using cytokeratin 19. This molecular procedure requires lyzing the node and therefore causes loss of tissue. Nevertheless, preliminary results indicate that the diagnostic capacity of this assay is at least similar or superior to intra-operatory imprint cytology or frozen sections, even with the help of IHC.

2.5.3 Lung Cancer

Bronchial carcinoma (lung cancer) is today the most common neoplasia in the world (12.6% of all new cancers) with a male-female gender ratio of 2.7:1 affecting

populations of both developed and developing countries. While the incidence in the developed countries remains stable or even decreases, in developing countries it is increasing. The primary cause of lung cancer in 90% of patients is smoking, and it is estimated that carcinoma will develop in 10–15% of all smokers. Environmental factors may play a role, as well as genetic predisposition in a multistep carcinogenic process, but stopping cigarette smoking is the most effective and least expensive means of reducing the risk.

Lung cancer is the deadliest of human neoplasiae, and yet no early diagnosis is currently available. Staging is still mainly based on histopathological and clinical criteria, which have a limited capacity to predict relapses and survival. In recent years, a major effort to improve the control of lung cancer has been carried out by introducing molecular profiling to typify different groups of bronchial carcinomas, and to provide more accurate predictions of the outcome after treatment, particularly with new targeted therapies. A good example is the EGFR mutations and amplifications that identify patients with non-small-cell lung cancer who may respond well to EGFR tyrosine kinase inhibitors.

Four major histological groups have to be considered, aside from a small number of rarer tumours that display unusual behaviour. The main groups are: *squamous carcinoma; adenocarcinoma; large-cell carcinoma; small-cell carcinoma*. In addition, combined types exist such as adeno-squamous, neuroendocrine (carcinoids), sarcomatoid, and some other more infrequent carcinomas. Mesenchymal and lymphoid neoplasms, together with dysgenetic pulmonary blastoma, complete the picture of this family of malignant lung tumours, excluding pleural mesothelioma (Travis et al. 2004).

For the processes of clinical staging and therapeutic indications, tumours are divided in two major categories based upon cell size. ‘Small-cell Lung Cancer’ (SCLC) and ‘Large-cell Lung Cancer’ (LCLG) or ‘Non Small-cell lung Cancer’ (NSCLC). The last comprising squamous carcinoma, adenocarcinoma, large-cell carcinoma, and adeno-squamous carcinoma, each with different clinic outcomes and distinct available therapeutic approaches.

Small-cell lung cancer (SCLC) was previously known as ‘oat cell carcinoma’. It corresponds to an anaplastic epithelial neoplasm, microscopically composed of small cells of round and/or spindle contour, containing one nucleus with dense chromatin and scanty poorly-defined cytoplasm. It is not infrequently combined with composite structures (about 10% of the tumour extent is necessary for it to be considered a mixed type) such as spindle-shaped, squamous, adenomatous, large, or even giant cells. Mitotic activity is high and necrosis may be massive, producing a diagnostic dilemma when the sample is small. IHC support is necessary to distinguish SCLC from lymphomas or metastatic small-round-cell-tumours (SRCT) (see Sect. 2.5.4), found in other anatomical locations. The SCLC tumour expresses neuroendocrine antigens (chromogranin and synaptophysin) and TTF-1, containing neurosecretory granules which are detectable with EM. The p53 gene is very frequently mutated, with a type of mutation related to cigarette smoking, mainly in women. A large number of genetic rearrangements have been described, that not only differentiates from classic neuroendocrine carcinoma, but also from NSCLC.

There is no detectable *in-situ* phase. Histogenesis involves a pluripotential stem cell of the bronchial tree with neuroectodermal differentiation and neuroendocrine expression. The tumour is not however considered as a true member of the family of neuroendocrine carcinomas.

Due to its high malignancy and generally adverse but unpredictable clinical outcome, the neoplasia is not graded following the TNM system, but is clinically considered as being either at a limited or an advanced stage of the disease. The advanced stage is associated with the presence of distant metastasis. Clinical symptom diagnosis of SCLC are generally reflected in disseminated disease with metastasis in liver, brain, or bone marrow. Original neoplasms located in the hilar or parahilar area may be asymptomatic and remain occult, even where the mediastinal node is involved, or where distant metastasis exists.

Non small-cell lung cancer (NSCLC) comprises a large number of histological varieties of bronchial carcinomas. For clinical purposes, three major histologic subtypes are considered: squamous-cell carcinoma, adenocarcinoma, and large-cell carcinoma. Smoking causes all types of lung cancer, but is more strongly linked with squamous-cell carcinoma, while adenocarcinoma is more frequent in patients who have not smoked. Staging of NSCLC is based on the TNM system. Before treatment, the tumour size, lymph-node status, and the possible presence of metastases must be determined. Lung cancer often spreads to the nodes in the hilum and mediastinum. The combination of PET and X-ray CT scans appears to have great sensitivity and specificity, and the use of both is recommended as part of the clinical evaluation before making any therapeutical decision.

We now focus on the histology of the three major carcinomas and on recent advances in the molecular study of the origin and biology of squamous-cell carcinoma and adenocarcinoma (Pass et al. 2000).

Squamous cell carcinoma is more frequent in men (44%) than in women (25%). Microscopical investigation shows it to possess large cells with diverse amounts of keratinization and pearl formation. The relative amount of squamous maturation serves to stratify well-differentiated tumours (with abundant keratinization) and undifferentiated tumours (with intercellular bridges and focal cytoplasmic keratin only in occasional large cells). Cytologic atypia with highly hyperchromatic nuclei, mitosis, and necrosis are hallmarks of this type of tumour. Several subtypes are considered: clear-cell, small-cell, basaloid, and alveolar carcinomas, as well as the adenocarcinoma mixed phenotype. IHC supports the diagnosis with positivity for low- and high-weight molecular keratins. In addition, EMA (epithelial membrane antigen) and CEA (carcinoembryonic antigen) stain focally isolated cells. TTF1 (thyroid transcriptions factor1) is positive only in some tumours.

Adenocarcinoma, which is becoming more frequent and surpassing the incidence of squamous carcinoma, occurs currently at a rate of almost 80% of NSCLC (Thun et al. 2006), a fact that is explicable due to changes in smoking behaviour. Its anatomical location varies within the lung, most usually as a peripheral tumour close to the pleura or with mesothelioma-like pleural extension, followed by central bronchial and endobronchial siting. The tumour sometimes adopts a pneumonia-like infiltration with nodular foci in the basal lobes owing to a bronchioloalveolar

extension. Scar carcinoma with desmoplasia is quite rare. In histology, glandular differentiation is the dominant pattern, with or without mucin secretion and acinar, papillary, or bronchioloalveolar associated phenotypes. Some adenocarcinomas adopt a solid configuration in which mucin production is lost or limited to isolated fine droplets within the cell. During histological grading, in cases of mixed histology, the least differentiated grade has to be measured. The bronchioloalveolar pattern is consistent with a grade 1, while solid adenocarcinoma is grade 3.

For the differential diagnosis with a metastatic carcinoma in the lung, IHC may be of interest. CAM 5.2, EMA, CEA, and CK7 are very frequently positive, but provide little additional value for the differential diagnosis of a metastasis. TTF1 positivity (75% of cases) gives support to the bronchial origin because other adenocarcinomas, excluding thyroid, are negative.

Large-cell carcinoma accounts for 9% of all lung cancers. This carcinoma combines old terminology: large cell anaplastic carcinoma and large-cell undifferentiated carcinoma, to which should be added the large-cell neuroendocrine carcinoma (LCNEC) and some rare combined forms (lymphoepithelioma-like, basaloid, and rhabdoid phenotype). The stage at diagnosis is similar to other NSCLC. Anatomically, they present large nodules generally located at the periphery of the lung, infiltrating pleura and even the chest wall or large bronchi.

Their histology encompasses a variety of microscopical patterns that contain large undifferentiated cells with large polygonal cytoplasm and prominent nuclei exhibiting numerous mitosis, but lacking squamous, adenoid, or microcellular differentiation. Large-cell neuroendocrine carcinomas (LCNEC) may present alone or be combined with other microscopical components such as adenocarcinoma, squamous, or even giant cells. They stain for chromogranin and synaptophysin, but also for TTF-1. Histological prognostic criteria for LCNEC are controversial, with apparently better outcome than with conventional large-cell carcinomas. Depending upon the staging, the 5-year survival of localized tumours with <3 cm resected NSCLC reaches almost 100% (T1). The local IIA decreases to 55%, while locally advanced (IIB, IIIA, IIIB) shows 39%, 23% and 3%, while IIIB is quite similar to advanced stages (stage IV) (Spira and Ettinger 2004).

Several alterations have been found in histologically apparently normal specimens of bronchial epithelium from smokers. Changes such as hyperplasia and metaplasia have been considered as a slightly abnormal epithelium and are regarded as early changes. The lesion affects normal bronchial mucosa composed of stratified cylindrical epithelia with cilia, and leads to a hyperplasia with or without metaplastic changes of squamous type. These modifications are not necessarily cancer precursors and may regress spontaneously. Nevertheless, in smokers the epithelial metaplasia may progress into dysplasia and subsequently to *in situ* squamous carcinoma. These microscopic lesions are usually multicentric in the bronchial tree and are frequently found in resected lungs of smokers, cohabiting with invasive squamous carcinoma. The molecular variations detected in dysplasia are regarded as occurring at an intermediate stage, whereas those found *in situ* or in invasive carcinoma are consistent with late changes. This multiple-step genetic rearrangement present in most lung cancers varies in its progression during the preneoplastic process.

In fact, the origin of lung cancer depends on a number of interactions between the environment and host genetic susceptibility, including changes in deregulated signalling pathways, which are potential targets for new therapeutical approaches. New techniques for genomic, transcriptomics, epigenetic, and proteomic profiling (Patz et al. 2007; Esteller 2008; Herbst et al. 2008) have improved the clinical approach to several histological varieties of NSCLC, identifying particular molecular markers of individual sensitivity, prognosis, and response to treatment. Good examples are the EGFR and VEGF inhibitors (erlotinib and bevacizumab), which have improved the clinical outcome in these patients (Shepherd et al. 2005; Sandler et al. 2006).

2.5.4 Small Round Cell Tumours (SRCT)

This category comprises a number of malignancies consisting predominantly of small round cells or round-spindle cells, independently of their origin or anatomical setting. Their histological patterns are very similar, thus a differential diagnosis becomes necessary in order to provide clinical information for therapy and prognosis. Among these, the most frequent is Ewing's Sarcoma (ES) and its family of tumours (ESFT) (including peripheral neuroectodermal tumour—PNET). Other frequent SRCT malignancies are neuroblastoma, rhabdomyosarcoma, and lymphoma (non-Hodgkin). A more extensive categorization might include tumours such as microcellular anaplastic osteosarcoma, myxoid chondrosarcoma, small-cell carcinoma of the lung, and small-cell neuroendocrine carcinoma (e.g. Merkel's tumour of the skin). Clinically, SRCT occurs in varied anatomical locations: not only bone and soft tissue, but also solid organs (ovary, testes, kidney, lung, meninges) and skin. There are no gender differences and it may occur at any age, although children and young adults are affected more frequently. Recently developed ancillary techniques such as cytogenetics, molecular biology, FISH, DNA or RNA microarray, and TMA, are offering new diagnostic and prognostic possibilities. Nevertheless, immunohistochemistry and electron microscopy continue to play an important role in the characterization of SRCTs.

Histologically, several variants of ESFT have been described, combining images of conventional or classical ES with other varieties such as atypical ES with large cells, ES with neuroectodermal pattern, and ES with endothelial features (Llombart-Bosch et al. 1996, 2009). Moreover, peripheral neuroepithelioma of soft tissue belongs to this group of neoplasms and may mimic a conventional undifferentiated neuroblastoma. Although neuroblastoma in adults is a rare event, it may still occur (Hasegawa et al. 2001); the differential diagnosis is mainly based upon immunohistochemical and molecular genetic findings. Adamantinoma-like and desmoplastic ES are other unusual variants of this family (Folpe et al. 2005). Genetic confirmation is essential for their identification.

Several immunohistochemical techniques are necessary for diagnosis. Antibody HNK 1 (CD 57), neuron-specific enolase, S-100, NF-70, synaptophysin, chromo-

Table 2.5 Chromosomal and genetic rearrangements in ES/pPNET tumours

Tumour	Translocation	Fusion gene	Frequency (%)
ES/pPNET	t(11;22)(q24;q12)	EWS/FLI-1	85
ES/pPNET	t(21;22)(q22;q12)	EWS/ERG	10
ES/pPNET	t(7;22)(p22;q12)	EWS/ETV1	<1
ES/pPNET	t(17;22)(q12;q12)	EWS/EA1AF	<1
ES/pPNET	t(2;22)(q33;q12)	EWS/FEV	<1

granin and PG-9.5, lend support to a neuroectodermal lineage (Navarro et al. 2007). Cytokeratin positivity (AE1/3) has been seen in several cases. The cell surface protein product p30/32 MIC-2 (Fellinger et al. 1991) CD99 is expressed in nearly 99% of cells in these tumours, independent of the histological subtype, but is absent in neuroblastoma, while positive in other SRCT unrelated to ESFT, such as B-lymphomas, rhabdomyosarcomas, synovial sarcomas (Weidner and Tjoe 1994; Llombart-Bosch et al. 2009). Caveolin 1 has also been confirmed as an excellent marker for this family of tumours (Llombart-Bosch et al. 2009).

Moreover, it has been demonstrated (Nilsson et al. 1999) that the 68 kDa fusion protein derived from the EWS/FLI1 hybrid gene can be specifically detected by Western blotting using a polyclonal antibody to the C-terminal of FLI1 on fresh tissue as well as paraffin-embedded ES. Eighty percent of the tumours exhibited a positive reaction for the FLI1 antibody, mainly with a nuclear location, but negative in neuroblastoma (Llombart-Bosch and Navarro 2001; Folpe et al. 2005).

Combining histology with immunohistochemistry confirms the structural heterogeneity of ESFT, which varies from conventional ES to atypical ES (including the large-cell variants) to PNET, with numerous Homer-Wright rosettes. Intermediate types may be found within a single tumour, reinforcing this heterogeneous microscopic pattern, such as the presence of vascular lakes with endothelial-like cells. The combination of staining with four antibodies, CD99, HNK1, FLI1, and Caveolin1, provides 100% of positivities in genetically confirmed tumours (Llombart-Bosch et al. 2009) aside from their histology.

Electron microscopy provides further support to the diagnosis. The cells are characterized according to their homogeneity. Large amounts of well-preserved glycogen are seen, and cell contacts show desmosomes. The detection of neurosecretion does not alone exclude ESFT; exclusion is confirmed by other cytoplasmic inclusions such as myofilaments (rhabdomyosarcoma) or interstitial deposits of osteoid material (small-cell osteosarcoma) (Llombart-Bosch et al. 1996).

Accuracy in diagnosis additionally requires confirmation by cytogenetic and molecular biology, which should indicate the chromosomal and genetic rearrangements detected in this group of tumours (Table 2.5). Moreover, the EWS gene presents fusion products with transcription factors: ATF (clear cell sarcoma of soft tissue), WT1 (desmoplastic small round cell tumour), and TEC (myxoid chondrosarcoma).

The detection of a balanced translocation in the ES/pPNET tumours, t(11;22)(q24;q12) shown in Table 2.5, turned out to be a formidable diagnostic marker pro-

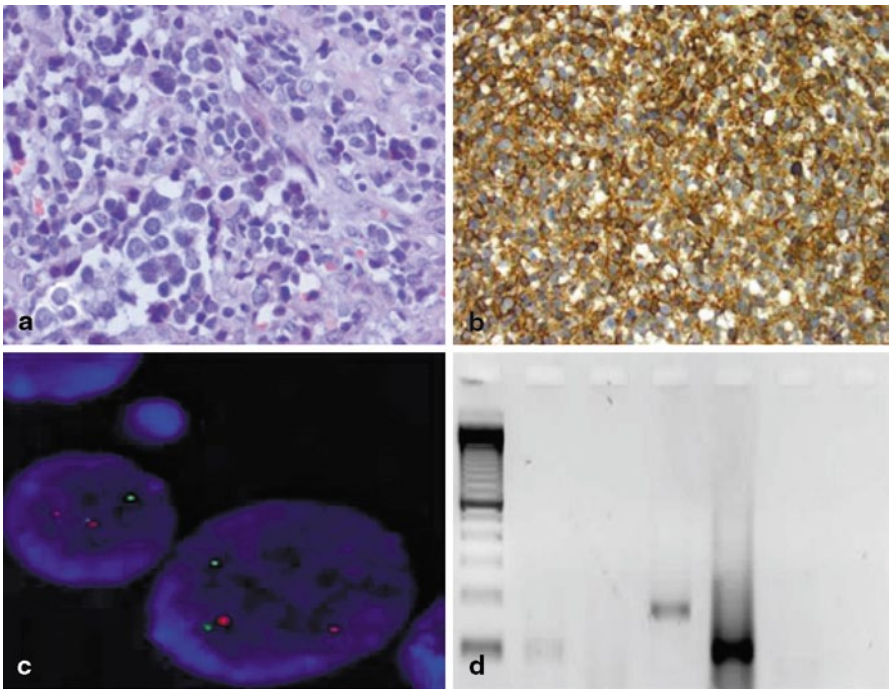


Fig. 2.2 **a** Small round cell tumour Ewing sarcoma type H/E 40X; **b** Membranous CD99 positivity +++ 40X; **c** FISH EWSR1 positive translocation (split signal) 100X; **d** RT-PCR positive (EWS/FLI1) gene fusion

viding a new phenotypic trait in this group of SRCT. More recently, new chromosomal translocations have been observed, independent of the morphological subtype of tumour. These, in decreasing frequency, are as follows: t(21;22)(q22;q12) in approximately 10% of cases; t(7;22)(q22;q12) in approximately 1% of tumours; t(17;22)(q12;q12) also in less than 1% of tumours, and finally a very rare translocation t(2;22)(q33;q12) described only in three ES/pPNET tumours. Thus, this family of tumours is characterized by at least five variants of translocations, in which the locus of the chromosome 22q12 is affected (Delattre 2008).

Nonetheless, approximately 5% of ES/PNET tumours with histological consistency and clinical evidence, using RT-PCR, remain negative for such types of transcript, with the result that a small number of SRCTs still remain outside any genetical phenotyping. New tools are evolving to facilitate the discrimination of most of these breakpoints. For example, the complementary technique of FISH employs specific probes flanking the ES breakpoint regions, not just metaphase chromosomes, but interphase nuclei as well (Bridge et al. 2006; Machado et al. 2009) (see Fig. 2.2a–d).

Gene expression profiling using cDNA microarrays, which allows simultaneous analysis of multiple markers, has been used to categorize different types of SRCT, especially when complemented with artificial neural networks (ANNs) (Khan et al.

2001; Kauer et al. 2009; Zambelli et al. 2010). Gene-expression signatures associated with specific variants of these tumours have thereby been identified: ESFT versus neuroblastoma, rhabdomyosarcoma, and Burkitt lymphoma.

These findings confirm the number of additional mutations that have occurred in ES, which add further insights into the malignant transformation of the initiating cell. These new mutations are not necessarily tumour specific, but may account for the clinical and histological variability of these neoplasms and even show prognostic value. Among them, trisomy 8 appears in 50% of ES and trisomy 12 appears in 20% of tumours. Moreover, an unbalanced translocation t(1;16) has been described in several tumours with ES phenotype. Furthermore, several oncogenes, such as ras, CMYC, MDM2 and tumour suppressor genes (p53, p16, pRb) have been analysed in ES tumours. Alterations expressed in oncogenes are not representative, while suppressor genes may play a major role. The frequency of p53 mutation is low (approximately 10% of ES) and pRb is not inactivated (Parham et al. 1999). In contrast, p16 homozygous deletions have been described in one third of ES without chromosomal aberrations of the 9q21 locus for the p16 gene (Kovar et al. 1997). In addition, our group has performed a molecular analysis of the 9p21 locus and p53 genes in this family of tumours using cell lines and original neoplasms (Lopez-Guerrero et al. 2001). Hence, the molecular alteration in either or both the pRB and p53 pathways seems to constitute a multi-step process with equivalent cellular effects, in which the EWS-ETS gene fusion seems to be the initiating mechanism.

A search for prognostic factors is necessary, since only one third of children with non-metastatic disease and 10% of metastatic patients survive the neoplasm. Several multivariate analyses show that the clinical prognosis provides the most valid criteria at present: those criteria are male sex, age greater than 12, fever at presentation, anaemia, high lactate dehydrogenase, axial location, and older chemotherapy regimes associated with adverse outcome (Cotterill et al. 2000). The determination of neuroectodermal expression in ES provides no differences in overall survival or disease-free survival, but histological atypical variants show a worse clinical outcome when compared to classical ES (Terrier et al. 1995; Parham et al. 1999; Llombart-Bosch et al. 2009).

Several assays have been undertaken to ascertain the molecular changes that could indicate further clinical prognostic markers in ES. In the analysis of the EWS/ETS gene fusion types, it was found that the most common fusion type is EWS/FLI-1 exon 7/6 (type 1), whereas in a third of cases the FLI-1 exon 5 is included in the transcript joined to the EWS exon 7 (type 2 fusion seen in 30% of cases). More occasionally, the fusions result in inclusion of the EWS exon 9 or 10 (10%) with FLI-1 exon 4 or 6 (10%) (Ladanyi 1995). Other studies (Zoubek et al. 1996) suggest a better clinical outcome for patients with localized ES, carrying mutation type 1. Thus, the EWS/ETS gene fusion type could serve as a prognostic indicator in both localized and metastatic disease (de Alava et al. 1998). Other molecular factors could also play an important role from the prognostic point of view. These genes include the cell cycle regulators p53 and p16 and the recently demonstrated Ki67 in localized disease, whereby those cases expressing Ki67 in more than 5% of nuclei of tumour cells had the worst behaviour independently of the type of treat-

ment (Lopez-Guerrero 2010). In addition, Zambelli et al. (2010) have proposed that lectin galactoside-binding soluble 3 binding protein (LGALS3BP) is a novel and reliable prognostic indicator for ES/PNET patients showing associated high mRNA expression levels of HINT1, STOML2 and c.MYC.

2.5.5 Leukaemias and Lymphomas

Molecular genetics has been at the forefront of research into cancer pathogenesis. The identification of recurrent chromosomal translocations has provided insights into the molecular events leading to leukaemic transformation. The classification of leukaemias has been, until recently, based on morphology and immunophenotype. The recognition of the correlation between distinctive morphologies and specific translocations mainly occurring in *de novo* leukaemias has had an indisputable impact on leukaemia classification, leading to the introduction of a subset of "acute myeloid leukaemias (AML) with recurrent genetic abnormalities" in the new WHO classification of acute leukaemias. The majority of the translocation events produce a fusion gene that encodes an aberrant protein, in which the '5 end of one translocation partner encodes the N-terminal protein sequence of the fusion protein, and the 3' end of the other translocation partner encodes the C-terminal protein sequence of the fusion protein. The fusion genes produced as a result of translocation events or the mutated genes are transcriptional regulatory proteins with altered properties of transcriptional activation or repression. It is now realized that many leukaemia cases that appear to be cytogenetically normal have point mutations or deletions in genes encoding key regulatory proteins, such as the fms-like tyrosine kinase-3 (FLT3) or the CCAAT/enhancer binding protein-α (C/EBPα).

More than one genetic hit is usually necessary for the development of leukaemia. The concept that multiple genetic defects are involved in leukaemogenesis is supported by the discovery of frequent FLT3 mutations in leukaemias with recurrent translocations. The breakthroughs in understanding the molecular genetics of leukaemia have had a direct impact on clinical treatment. In this respect, chronic myeloid leukaemia (CML) has been the paradigm for the translation of basic research to clinical treatment. CML was the first leukaemia to be associated with a recurrent translocation, t(9;22)(q34;q11) (the Philadelphia chromosome) and the first leukaemia for which the product of the translocation, BCR-ABL, was characterized. In addition, a specific molecular inhibitor, imatinib (Gleevec), was designed for CML and was successfully used for patient treatment. The Philadelphia (Ph) chromosome is also the most frequent recurring translocation in adult acute lymphoblastic leukaemia (ALL) occurring in 15–30% of patients, and is also present in 5% of paediatric B-cell ALL. In both clinical scenarios, it is an adverse prognostic factor. The most common breakpoint region within the BCR gene, the major breakpoint cluster region (M-bcr), results in a fusion protein of 210 kD, referred to as $p210^{bcr\text{-}abl}$. A minor breakpoint, the m-bcr, results in a truncated fusion protein of 190 kD ($p190^{bcr\text{-}abl}$). Importantly, $p210^{bcr\text{-}abl}$ is much more common in CML, whereas

p190 $^{bcr\text{-}abl}$ is present in 80–90% of paediatric Ph+ ALL and 50% of adult Ph+ ALL. BCR-ABL has leukaemogenic properties as a constitutive tyrosine kinase that activates multiple downstream signal transduction intermediates, including ras, PLCγ and PI3 kinase, leading to proliferation and resistance to apoptosis.

Conceivably, similar mechanisms are operative in Ph+ ALL, a disease whose treatment is highly problematic. Remissions tend to be short-lived and stem cell transplantation is the most effective way of attaining durable control of the disease. The 2-phenylaminopyrimidine derivative, imatinib, is an ABL-specific tyrosine kinase inhibitor that constrains the proliferation of CML cell lines by inhibiting BCR-ABL kinase activity. The drug is administered orally and is generally well tolerated. In a multicenter phase II trial, imatinib at 400 mg/d induced a complete haematological response in 95% of patients and a major cytogenetic response in 60% (Kantarjian et al. 2002). After a median follow-up of 18 months, 95% of the patients were alive and CML had progressed to accelerated or blast crisis in 11% of patients. This has represented a dramatic breakthrough in the treatment of CML, a disease context where initial therapies were aimed at controlling the elevated white blood cell count, reducing the symptoms of concomitant splenomegaly, and treating metabolic complications arising from profound marrow proliferation, such as hyperuricaemia and gout.

Unfortunately, the issue of resistance to imatinib is beginning to emerge. Of patients who start imatinib in the early chronic, late chronic, and accelerated phases of CML, 12%, 32%, and 62% respectively develop resistance mutations within two years of commencing treatment, these being attributable to either a single amino acid substitution in the ATP-binding region of BCR-ABL or, occasionally, to progressive BCR-ABL gene amplification. New BCR-ABL tyrosine kinase inhibitors are currently being evaluated in clinical trials: the improved-potency, selective Abl inhibitor nilotinib, and the highly potent dual Src/Abl inhibitor dasatinib. Despite their greater activity, there is some concern arising from *in vitro* studies of dasatinib that these drugs too will turn out to be unable to clear the leukaemic stem cells (Copland et al. 2006). Clinical trials will show whether these second-generation tyrosine kinase inhibitors should be used alone or in combination as first-line therapy for newly diagnosed CML.

Another elegant example of the interaction between molecular advances and clinical treatment is the case of acute promyelocytic leukaemia (APL). The t(15;17)(q22;q21) translocation in APL is associated with the characteristic morphology of hypergranular blast cells with frequent Auer rods or the microgranular variant. An initial report from China (Huang et al. 1988) indicated that APL could be treated successfully with all-trans retinoic acid (ATRA). This observation preceded the discovery that the t(15;17) translocation involved the retinoic acid-α gene on chromosome 17. In the t(15;17)(q22;q21), the most common translocation associated with APL, the 5' portion of the fusion protein is encoded by the PML (promyelocytic leukaemia) gene from 15q22, and the 3' portion is encoded by the RARα gene from 17q21. The wild-type RARα is a nuclear receptor acting as a transcription factor and binding to retinoic acid response elements (RARE) in the promoter of many genes, including those implicated in myeloid differentiation (granulocyte col-

ony-stimulating factor [G-CSF], cell-surface adhesion molecules [CD18, CD11b], regulators of apoptosis [Bcl-2], and several transcription factors). In the absence of retinoic acid, the wild-type RARα binds to corepressor proteins and histone deacetylates, resulting in transcriptional repression. Wild-type PML protein is localized in subnuclear oncogenic domains, called nuclear bodies, and may act as a tumour-suppressor protein, although it does not bind DNA directly. In APL, the aberrant fusion protein PML-RARα is delocalized from the nuclear bodies to a microspeckled nuclear pattern and acts in a dominant negative manner, competing with wild-type RARα for binding to the RAREs in the absence of ligand. However, pharmacological concentrations of retinoic acid are required to convert PML-RARα into a transcriptional activator. These observations have provided the rational basis for the efficacy of ATRA treatment in patients with APL to induce the differentiation of leukemic promyelocytes.

The t(8;21) translocation is present in approximately 15% of patients with AML and involves the RUNX1 (AML1) gene which is located on chromosome 21q22.3. The fusion partner of RUNX1 in t(8;21) is named eight-twenty-one (ETO) and is a transcriptional regulator. The murine counterpart of RUNX1 was first described as part of the core binding factors, which are essential for haematopoietic development, as indicated by gene deletion experiments in mice. RUNX1 is a transcriptional activator regulating lymphoid genes such as B-cell tyrosine kinase, T-cell receptor α and β, interleukin (IL)-3, and granulocyte proteins. The RUNX1-ETO fusion protein binds to the same DNA-binding site as RUNX1 and acts as a dominant negative inhibitor of wild-type RUNX1 but also as an active transcriptional repressor. Targets of RUNX1-ETO repression are presumed to be genes important for granulocytic differentiation and tumour suppressors such as p14ARF and NF1.

More recently, development of the technology of microarray analysis has led to a better understanding of the global changes in gene expression that occur as a result of leukaemic transformation and has allowed the subtyping of acute leukaemias. When patients with AML are grouped on the basis of gene expression signatures, most clusters correspond to the common recurrent translocations or known gene mutations. Valk et al. 2004 identified 16 groups when analysing blood or bone marrow from 285 patients with AML. Patients with recurrent translocations (t(8;21), t(15;17) and inv(16)), and thus with cytogenetically defined disease subsets, formed clear clusters, thus validating the significance of the gene expression patterns and suggesting that microarray analysis may allow subclassification of leukaemias into meaningful groups with unique prognosis and pathogenesis.

Lymphomas are a heterogeneous group of malignant diseases of the lymphoid system, whose classification remains a confusing and controversial issue. They comprise the *non-Hodgkin lymphomas* (NHL) and *Hodgkin lymphoma* (HL). In 1994, the Revised European American Lymphoma (REAL) classification was proposed and listed "real disease entities recognized and diagnosed in daily practice". More recently, the WHO classification has provided a comprehensive definition of lymphomas by morphology, immunophenotype, genetics, and clinical information. The NHL are a diverse collection of lymphoid neoplasms with varied pathology, cell of origin, natural history, and response to treatment. The histological diagno-

sis of NHL is among the most difficult tasks that surgical pathologists are asked to undertake. The molecular genetic lesions of pathogenic importance in selected forms of NHL have recently been unravelled via the molecular analysis of structural chromosomal abnormalities that alter critical genes regulating growth and/or differentiation.

There is a worldwide epidemic of NHL and its rise has been faster than that of all other malignancies except lung cancer in women, melanoma, and prostate cancer. More than 60,000 new cases per year will be diagnosed in the United States in the 2000s. The majority of patients with NHL presents with painless lymphoadenopathy, more commonly in the cervical or supraclavicular regions, but extranodal disease, mainly of the gastrointestinal tract, can be detected at presentation in up to 40% of patients. Systemic symptoms are associated with advanced stages of disease and portend a poor prognosis. Actually, there is no substitute for tissue diagnosis, although PET scanning using ^{18}F-fluorodeoxyglucose has been used as a diagnostic technique for disseminated disease and to assess treatment response (Haioun et al. 2005). Disease prognosis is highly variable, as might be expected from the broad spectrum of NHL subtypes, and biological differences have been described between young and old patients, translating into a greater mortality in some series of elderly patients compared with younger cohorts. The type of therapy is generally based on pathology and its intensity is based on both pathology and stage of disease. Independent prognostic determinants include advanced stage, tumour bulk as reflected by size and LDH, and number of extranodal sites of involvement.

Recent advances in clinical scoring systems and in molecular and phenotypic markers have improved our ability to predict therapeutic responses. In this respect, immunotherapy approaches are rapidly advancing and include non-specific immunostimulation with interferons, passive therapy with anti-lymphoid antibodies such as rituximab, radioimmunotherapy, patient-specific autologous anti-idiotype vaccines, and novel cellular immunotherapy modalities. Rituximab is a chimeric IgG anti-CD20 monoclonal antibody composed by fusing a light and heavy chain-variable domain of a murine IgG with a human IgG light and heavy chain-constant region. Rituximab has rapidly become an effective component as a single agent or in combination with chemotherapy in the treatment of all types of NHL, producing an undisputed survival advantage over chemotherapy alone. New monoclonal antibodies directed against CD20 and other B-cell antigens are under investigation, including epratuzumab, a chimeric anti-CD22 antibody.

Multiple myeloma (MM) is a highly treatable but incurable neoplastic plasma-cell dyscrasia characterized by the clinical pentad of anaemia, a monoclonal protein in the serum and/or urine, abnormal radiographs and bone pain, hyperuricaemia, and renal insufficiency or failure. Until recently, the higher response rates seen with regimens that combine multiple agents as initial therapy (alkylators, anthracyclines, corticosteroids, and interferon) had not resulted in improved survival rates. Over the last five years, we have witnessed enormous progress in fundamental and therapeutic research in MM. The current preferred therapies are all in the "novel" category. In particular, a new class of immune modulatory drugs has dramatically changed the previous scenario. The recognition in 1999 of the activity of thalidomide against

MM and the subsequent development of lenalinomide and bortezomib, has made MM treatment more promising and rewarding. Thalidomide is believed to inhibit angiogenesis in MM but also to target the surrounding stroma and cytokines and to affect NK cells. When used as a single agent, thalidomide induces response rates of 25% in previously untreated patients and is now considered a standard therapy for MM. Bortezomib is the first drug in its class of proteasome inhibitors. Bortezomib selectively and reversibly inhibits the proteasome, an intracellular complex that degrades ubiquitinated proteins and plays a key role in cell cycle regulation, protein degradation, and gene expression. Single-agent response rates in relapsed/refractory MM range from 28% to 38% and the median duration of response is 8 months (Weber et al. 2003). The long-term outcome of treatment with novel drugs is presently not known, because of the short duration of follow-up. The protagonists of the current treatment armamentarium embrace a holistic total-therapy approach and combine multiple agents. In the near future, we may be able to ascertain biological differences among disease subsets and direct specific forms of therapies to their biology. Cytogenetic testing is an integral element to establish prognosis and a treatment plan for newly diagnosed MM. Nearly all MM patients have cytogenetic abnormalities diagnosed by FISH, but abnormal karyotypes are seen in only 18–30% of cases. Molecular classification systems have been proposed based on gene expression profiling but these are deemed not to be ready for general clinical application, as distinct from cytogenetic classification systems that are easily applied to the clinic at present. Nearly 85% of newly diagnosed MM patients have gene expression-defined good risk features and fare so well that the prospect of cure has become a reality. By using high-density oligonucleotide microarrays and hierarchical clustering analysis, four distinct subgroups of MM (MM1, MM2, MM3 and MM4) have been identified (Zhan et al. 2003). Of interest, clinical variables associated with poor prognosis, including abnormal karyotype and high serum β2-microglobulin levels, were most prevalent in MM4. Also, over-expression of genes involved in DNA metabolism and cell cycle control was primarily observed in MM4. Whether this information may be incorporated into novel prognostic algorithms and be used to tailor current treatment strategies, remains to be addressed.

2.6 Systems Biology of Cancer: Key Challenges for the Future

As mentioned earlier, recent advances in molecular and cellular cancer biology, as well as the explosion of novel technologies within genomics, transcriptomics, proteomics, and functional genomics, promise to have a major impact on clinical practice. These developments are likely to change the way in which diseases will be diagnosed, treated, and monitored in the future (Celis et al. 2005). Major areas of research that will benefit from these developments include the identification of molecular biomarkers for non-invasive early diagnosis, subclassification based on clinical outcome, prediction of prognosis and response to treatment, as well as

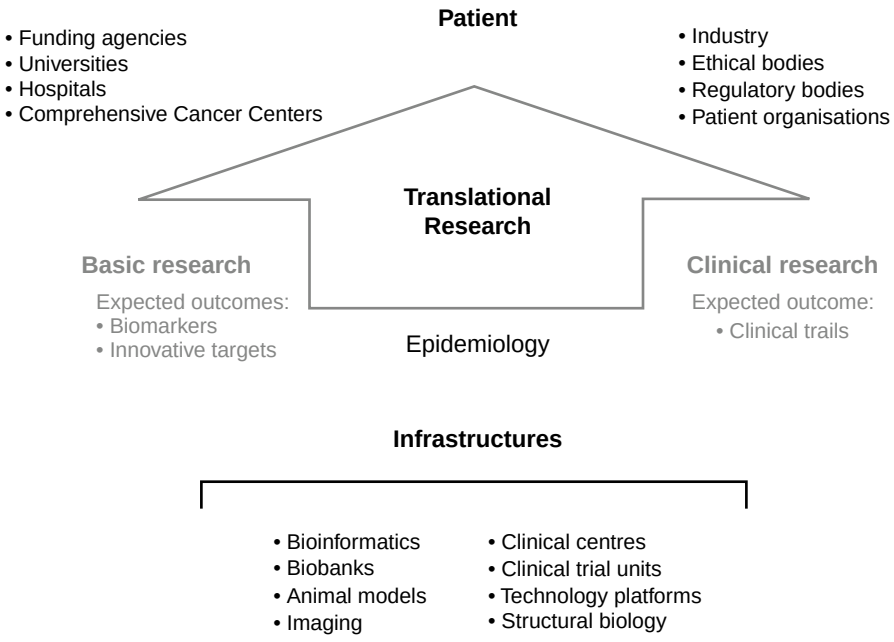


Fig. 2.3 Graphical illustration of stakeholders and support infrastructures in discovery-driven translational cancer research

determination of drug efficacy and toxicity (Zhang et al. 2007; Ransohoff 2009; Alymani et al. 2010 and references therein). Also, the identification of novel therapeutic targets—identified on the basis of systems biology approaches to the analysis of pathways that are affected in cancer cells (Aebersold et al. 2009; Laubenbacher et al. 2009; Kreeger and Lauffenburger 2010)—will be another area that will greatly benefit from high-throughput 'omic' technologies (Sara et al. 2010).

Research efforts in these areas, however, must be supported by the development of bioinformatics and modelling tools for integrating and mining data, as well as by proper technological and clinical infrastructures. In the long run, this integrated approach should lead to a better understanding of the biology underlying cancer cells, which in turn will lead to a more effective translation of basic discoveries into new diagnostics and therapies (Fig. 2.3). The daunting heterogeneity of human cancer, in terms of cellular phenotypes, genetic make-up, molecular profiling, and clinical behaviour is however posing major challenges that must be addressed if we are to fulfil the dream of bringing personalized medicine and individualized cancer care closer to reality. Tumours usually contain malignant cells showing different degrees of differentiation (Celis et al. 2003; Celis et al. 2007) as well as other cell types, which together compose the 'tumour microenvironment' (Celis et al. 2004; Tlsty and Coussens 2006; Witz 2009). The heterogeneity-related problems in characterization and treatment have been partially addressed using techniques that allow the dissection of a defined set of purified cell populations (Espina et al. 2007), but these

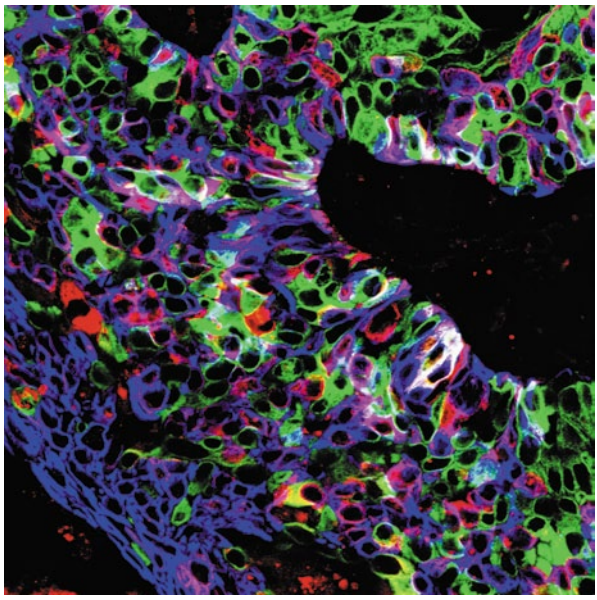


Fig. 2.4 Triple IHC stained of a breast carcinoma *in situ* reacted with antibodies against CK's 8, 15, and 19. Photograph kindly provided by Jose Moreira

technologies cannot solve the problem altogether, as heterogeneity can be observed even in a small number of cells within a given lesion, as illustrated in Fig. 2.4. In this particular case a breast carcinoma *in situ* has been stained with antibodies against cytokeratins 8, 15, and 19. As seen in Fig. 2.4, the heterogeneity of the epithelial cells in terms of phenotype is such that it would be very difficult to interpret expression data generated from it, unless one had access to speedy procedures for validation at the cellular level; for example, using a large battery of specific antibodies (Tlsty and Coussens 2006). The fact that only a few cells in this pre-cancerous lesion may harbour the malignant phenotype underlines the complexity of the problem and emphasizes the need to develop strategies to identify biomarkers that specifically predict the prognosis of each cell type expressing a given phenotype. In addition, we must increase our efforts to identify cancer stem cells (which generate cellular heterogeneity), as these will be the focus for developing new targeted therapies (Nirmalanandhan and Sittampalam 2009; Watt and Driskell 2010).

It is also urgent to address the problem of clinical relevance when selecting the source of samples used to derive new biomarkers and targets (Celis et al. 2005). The use of well-annotated and accessible clinically relevant samples to generate new data is important, as the final outcome will very much depend on the quality and relevance of the data. Accordingly, we are increasingly moving from the study of cultured cells to the analysis of freshly collected cells, tissue samples, and biofluids, but one of the main challenges one faces is how best to apply the powerful 'omics' technologies to the study of clinically relevant samples in a well-defined clinical and pathological framework (Celis et al. 2003). There is still a significant gap between mechanistic research based on cellular model systems, and their potential in clinical applications.

Finally, we must also acknowledge the value of long-term research and provide the appropriate legal and ethical framework to encourage collaboration among all the stakeholders in the cancer continuum. Bridging the gap between basic and clinical research, facilitating the engagement of industry, establishing new infrastructures, as well creating innovative clinical trials, are among the items that require urgent action (Fig. 2.3). The aim of cancer research is to improve the life expectancy and quality of life of patients and we must make every effort to coordinate current activities in order to achieve this goal.

Acknowledgements We are indebted to Laila Fischer for expert secretarial assistance. This work was supported by the IVO Cancer Institute and the EuroBoNeT consortium, a network of excellence granted by the European Commission for studying the pathology and genetics of bone tumours (to ALLB), the Stockholm Cancer Society (to UR), and the Danish Cancer Society, the Danish Medical Research Council, and the John and Birthe Meyer Foundation (to JEC).

References

Abeloff MD, Armitage JO, Niederhuber JE, Kastan MB, Gillies McKenna W (2008) Abeloff's clinical oncology, 4th edn. Churchill Livingstone, UK

Aebersold R, Auffray C, Baney E, Barillot E, Brazma A, Brett C, Brunak S, Butte A, Califano A, Celis J, Cufer T, Ferrell J, Galas D, Gallahan D, Gatenby R, Goldbeter A, Hace N, Henney A, Hood L, Iyengar R, Jackson V, Kallioniemi O, Klingmuller U, Kolar P, Kolch W, Kyriakopoulou C, Laplace F, Lehrach H, Marcus F, Matrisian L, Nolan G, Pelkmans L, Potti A, Sander C, Seljak M, Singer D, Sorger P, Stunnenberg H, Superti-Furga G, Uhlen M, Vidal M, Weinstein J, Wigle D, Williams M, Wolkenhauer O, Zhivotovsky B, Zinovyev A, Zupan B (2009) Report on EU-USA workshop: how systems biology can advance cancer research (27 October 2008). Mol Oncol 3(1):9–17

Aitken JF, Elwood M, Baade PD, Youl P, English D (2010) Clinical whole-body skin examination reduces the incidence of thick melanomas. Int J Cancer 126(2):450–458

Albertsen P (2009) Androgen deprivation in prostate cancer—step by step. N Engl J Med 360(24):2572–2574

Alymani NA, Smith MD, Williams DJ, Petty RD (2010) Predictive biomarkers for personalised anti-cancer drug use: discovery to clinical implementation. Eur J Cancer 46(5):869–879

Aparicio SA, Huntsman DG (2010) Does massively parallel DNA resequencing signify the end of histopathology as we know it? J Pathol 220(2):307–315

Arisio R, Cuccorese C, Accinelli G, Mano MP, Bordon R, Fessia L (1998) Role of fine-needle aspiration biopsy in breast lesions: analysis of a series of 4110 cases. Diagn Cytopathol 18(6):462–467

Baccarani M, Cortes J, Pane F, Niederwieser D, Saglio G, Apperley J, Cervantes F, Deininger M, Gratwohl A, Guilhot F, Hochhaus A, Horowitz M, Hughes T, Kantarjian H, Larson R, Radich J, Simonsson B, Silver RT, Goldman J, Hehlmann R (2009) Chronic myeloid leukemia: an update of concepts and management recommendations of European LeukemiaNet. J Clin Oncol 27(35):6041–6051

Badve S, Nakshatri H (2009) Oestrogen-receptor-positive breast cancer: towards bridging histopathological and molecular classifications. J Clin Pathol 62(1):6–12

Balch C, Houghton A, Sober A, Soong S (2003) Cutaneous melanoma, 4th edn. Quality Medical, St Louis

Baselga J (2006) Targeting tyrosine kinases in cancer: the second wave. Science 312(5777):1175–1178

Baselga J, Arribas J (2004) Treating cancer's kinase 'addiction'. Nat Med 10(8):786–787
Baselga J, Swain SM (2009) Novel anticancer targets: revisiting ERBB2 and discovering ERBB3. Nat Rev Cancer 9(7):463–475
Baumann P, Nyman J, Hoyer M, Wennberg B, Gagliardi G, Lax I, Drugge N, Ekberg L, Friesland S, Johansson KA, Lund JA, Morhed E, Nilsson K, Levin N, Paludan M, Sederholm C, Traberg A, Wittgren L, Lewensohn R (2009) Outcome in a prospective phase II trial of medically inoperable stage I non-small-cell lung cancer patients treated with stereotactic body radiotherapy. J Clin Oncol 27(20):3290–3296
Beckman M (2006) Tumor complexity prompts caution about sequencing. J Natl Cancer Inst 98(24):1758–1759
Bergers G, Benjamin LE (2003) Tumorigenesis and the angiogenic switch. Nat Rev Cancer 3(6):401–410
Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M, Mandelblatt JS, Yakovlev AY, Habbema JD, Feuer EJ (2005) Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med 353(17):1784–1792
Bloom HJ, Richardson WW (1957) Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. Br J Cancer 11(3):359–377
Bobola MS, Silber JR, Ellenbogen RG, Geyer JR, Blank A, Goff RD (2005) O6-methylguanine-DNA methyltransferase, O6-benzylguanine, and resistance to clinical alkylators in pediatric primary brain tumor cell lines. Clin Cancer Res 11(7):2747–2755
Boffetta P, Hashibe M (2006) Alcohol and cancer. Lancet Oncol 7(2):149–156
Bosman FT (1995) Prognostic value of pathological characteristics of colorectal cancer. Eur J Cancer 31A(7–8):1216–1221
Boyle P, Autier P, Bartelink H, Baselga J, Boffetta P, Burn J, Burns HJ, Christensen L, Denis L, Dicato M, Diehl V, Doll R, Franceschi S, Gillis CR, Gray N, Griciute L, Hackshaw A, Kasler M, Kogevinas M, Kvinnsland S, La VC, Levi F, McVie JG, Maisonneuve P, Martin-Moreno JM, Bishop JN, Oleari F, Perrin P, Quinn M, Richards M, Ringborg U, Scully C, Siracka E, Storm H, Tubiana M, Tursz T, Veronesi U, Wald N, Weber W, Zaridze DG, Zatonski W, Zur HH (2003a) European code against cancer and scientific justification: third version. Ann Oncol 14(7):973–1005
Boyle P, d'Onofrio A, Maisonneuve P, Severi G, Robertson C, Tubiana M, Veronesi U (2003b) Measuring progress against cancer in Europe: has the 15% decline targeted for 2000 come about? Ann Oncol 14(8):1312–1325
Boyle P, Levin B (2008) World cancer report. http://www.iarc.fr/en/publications/pdfs-online/wcr/2008/index.php
Brenton JD, Carey LA, Ahmed AA, Caldas C (2005) Molecular classification and molecular forecasting of breast cancer: ready for clinical application? J Clin Oncol 23(29):7350–7360
Bridge RS, Rajaram V, Dehner LP, Pfeifer JD, Perry A (2006) Molecular diagnosis of Ewing sarcoma/primitive neuroectodermal tumor in routinely processed tissue: a comparison of two FISH strategies and RT-PCR in malignant round cell tumors. Mod Pathol 19(1):1–8
Celis JE (2008) Editorial. Mol Oncol 2(1):1–1
Celis JE, Gromov P (2003) Proteomics in translational cancer research: toward an integrated approach. Cancer Cell 3(1):9–15
Celis JE, Gromov P, Cabezon T, Moreira JM, Ambartsumian N, Sandelin K, Rank F, Gromova I (2004) Proteomic characterization of the interstitial fluid perfusing the breast tumor microenvironment: a novel resource for biomarker and therapeutic target discovery. Mol Cell Proteomics 3(4):327–344
Celis JE, Gromov P, Cabezon T, Moreira JM, Friis E, Jirstrom K, Llombart-Bosch A, Timmermans-Wielenga V, Rank F, Gromova I (2008) 15-prostaglandin dehydrogenase expression alone or in combination with ACSM1 defines a subgroup of the apocrine molecular subtype of breast carcinoma. Mol Cell Proteomics 7(10):1795–1809
Celis JE, Gromov P, Gromova I, Moreira JM, Cabezon T, Ambartsumian N, Grigorian M, Lukanidin E, thor Straten P, Guldberg P, Bartkova J, Bartek J, Lukas J, Lukas C, Lykkesfeldt A, Jaattela M, Roepstorff P, Bolund L, Orntoft T, Brunner N, Overgaard J, Sandelin K, Blichert-Toft

M, Mouridsen H, Rank FE (2003) Integrating proteomic and functional genomic technologies in discovery-driven translational breast cancer research. Mol Cell Proteomics 2(6):369–377
Celis JE, Gromova I, Cabezon T, Gromov P, Shen T, Timmermans-Wielenga V, Rank F, Moreira JM (2007) Identification of a subset of breast carcinomas characterized by expression of cytokeratin 15: relationship between CK15+ progenitor/amplified cells and pre-malignant lesions and invasive disease. Mol Oncol 1(3):321–349
Celis JE, Moreira JM, Gromova I, Cabezon T, Ralfkiaer U, Guldberg P, Straten PT, Mouridsen H, Friis E, Holm D, Rank F, Gromov P (2005) Towards discovery-driven translational research in breast cancer. FEBS J 272(1):2–15
Chabner BA, Longo DL (2006) Cancer chemotherapy and biotherapy. Principles and practice, 4th edn. Lippincott Williams & Wilkins, Baltimore
Chan JK (2001) The new world health organization classification of lymphomas: the past, the present and the future. Hematol Oncol 19(4):129–150
Cheang MC, Voduc D, Bajdik C, Leung S, McKinney S, Chia SK, Perou CM, Nielsen TO (2008) Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. Clin Cancer Res 14(5):1368–1376
Chiang AC, Massague J (2008) Molecular basis of metastasis. N Engl J Med 359(26):2814–2823
Clarke M, Coates AS, Darby SC, Davies C, Gelber RD, Godwin J, Goldhirsch A, Gray R, Peto R, Pritchard KI, Wood WC (2008) Adjuvant chemotherapy in oestrogen-receptor-poor breast cancer: patient-level meta-analysis of randomised trials. Lancet 371(9606):29–40
Clarke M, Collins R, Darby S, Davies C, Elphinstone P, Evans E, Godwin J, Gray R, Hicks C, James S, MacKinnon E, McGale P, McHugh T, Peto R, Taylor C, Wang Y (2005) Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomised trials. Lancet 366(9503):2087–2106
Cogliano V, Baan R, Straif K, Grosse Y, Secretan B, El Ghissassi F (2005) Carcinogenicity of human papillomaviruses. Lancet Oncol 6(4):204
Cohn-Cedermark G, Rutqvist LE, Andersson R, Breivald M, Ingvar C, Johansson H, Jonsson PE, Krysander L, Lindholm C, Ringborg U (2000) Long term results of a randomized study by the Swedish Melanoma Study Group on 2-cm versus 5-cm resection margins for patients with cutaneous melanoma with a tumor thickness of 0.8–2.0 mm. Cancer 89(7):1495–1501
Copland M, Hamilton A, Elrick LJ, Baird JW, Allan EK, Jordanides N, Barow M, Mountford JC, Holyoake TL (2006) Dasatinib (BMS-354825) targets an earlier progenitor population than imatinib in primary CML but does not eliminate the quiescent fraction. Blood 107(11):4532–4539
Correa Geyer F, Reis-Filho JS (2009) Microarray-based gene expression profiling as a clinical tool for breast cancer management: are we there yet? Int J Surg Pathol 17(4):285–302
Costa J (2009) Systems approach to the practice of pathology: a new role for the pathologist. Arch Pathol Lab Med 133(4):524–526
Cotterill SJ, Ahrens S, Paulussen M, Jurgens HF, Voute PA, Gadner H, Craft AW (2000) Prognostic factors in Ewing's tumor of bone: analysis of 975 patients from the European Intergroup Cooperative Ewing's Sarcoma Study Group. J Clin Oncol 18(17):3108–3114
Cuppone F, Bria E, Carlini P, Milella M, Felici A, Sperduti I, Nistico C, Terzoli E, Cognetti F, Giannarelli D (2008) Taxanes as primary chemotherapy for early breast cancer: meta-analysis of randomized trials. Cancer 113(2):238–246
Dardick I, Herrera GA (1998) Diagnostic electron microscopy of neoplasms. Hum Pathol 29(12):1335–1338
Alava E de, Kawai A, Healey JH, Fligman I, Meyers PA, Huvos AG, Gerald WL, Jhanwar SC, Argani P, Antonescu CR, Pardo-Mindan FJ, Ginsberg J, Womer R, Lawlor ER, Wunder J, Andrulis I, Sorensen PH, Barr FG, Ladanyi M (1998) EWS-FLI1 fusion transcript structure is an independent determinant of prognosis in Ewing's sarcoma. J Clin Oncol 16(4):1248–1255
Delattre O (2008) Ewing's tumours, genetic and cellular aspects. Pathol Biol (Paris) 56(5):257–259
Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, Lickley LA, Rawlinson E, Sun P, Narod SA (2007) Triple-negative breast cancer: clinical features and patterns of recurrence. Clin Cancer Res 13(15):4429–4434

Desmedt C, Ruiz-Garcia E, Andre F (2008) Gene expression predictors in breast cancer: current status, limitations and perspectives. Eur J Cancer 44(18):2714–2720

DeVita VT Jr, Hellman S, Rosenberg SA (2008) Cancer: principles and practice of oncology. Lippincott Williams and Wilkins, Baltimore

Doll R, Peto R, Boreham J, Sutherland I (2004) Mortality in relation to smoking: 50 years' observations on male British doctors. BMJ 328(7455):1519

Dowsett M, Cuzick J, Ingle J, Coates A, Forbes J, Bliss J, Buyse M, Baum M, Buzdar A, Colleoni M, Coombes C, Snowdon C, Gnant M, Jakesz R, Kaufmann M, Boccardo F, Godwin J, Davies C, Peto R (2010) Meta-analysis of breast cancer outcomes in adjuvant trials of aromatase inhibitors versus tamoxifen. J Clin Oncol 28(3):509–518

Druker BJ, Talpaz M, Resta DJ, Peng B, Buchdunger E, Ford JM, Lydon NB, Kantarjian H, Capdeville R, Ohno-Jones S, Sawyers CL (2001) Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. N Engl J Med 344(14):1031–1037

Duff SE, Jeziorska M, Rosa DD, Kumar S, Haboubi N, Sherlock D, O'Dwyer ST, Jayson GC (2006) Vascular endothelial growth factors and receptors in colorectal cancer: implications for anti-angiogenic therapy. Eur J Cancer 42(1):112–117

Dworak O, Keilholz L, Hoffmann A (1997) Pathological features of rectal cancer after preoperative radiochemotherapy. Int J Colorectal Dis 12(1):19–23

Eden P, Ritz C, Rose C, Ferno M, Peterson C (2004) "Good Old" clinical markers have similar power in breast cancer prognosis as microarray gene expression profilers. Eur J Cancer 40(12):1837–1841

Edge SB, Byrd DR, Compton CC, Fritz AG, Greene FL, Trotti A (2010) AJCC cancer staging manual, 7th edn. Springer, Berlin

Ehrlich Y, Brames MJ, Beck SD, Foster RS, Einhorn LH (2010) Long-term follow-up of Cisplatin combination chemotherapy in patients with disseminated nonseminomatous germ cell tumors: is a postchemotherapy retroperitoneal lymph node dissection needed after complete remission? J Clin Oncol 28(4):531–536

Elston C, Ellis I (1998) Systemic pathology: the breast. Elsevier, Churchill Livingstone, London

Espina V, Heiby M, Pierobon M, Liotta LA (2007) Laser capture microdissection technology. Expert Rev Mol Diagn 7(5):647–657

Esteller M (2008) Epigenetics in cancer. N Engl J Med 358(11):1148–1159

Faratian D, Bartlett J (2008) Predictive markers in breast cancer—the future. Histopathology 52(1):91–98

Fass L (2008) Imaging and cancer: a review. Mol Oncol 2(2):115–152

Fearon ER, Vogelstein B (1990) A genetic model for colorectal tumorigenesis. Cell 61(5):759–767

Fellinger EJ, Garin-Chesa P, Su SL, DeAngelis P, Lane JM, Rettig WJ (1991) Biochemical and genetic characterization of the HBA71 Ewing's sarcoma cell surface antigen. Cancer Res 51(1):336–340

Fenoglio-Preiser CM, Noffsinger AE, Stemmermann GN (1999) Gastrointestinal pathology: an atlas and text. Lippincott Williams & Wilkins, Baltimore

Ferlay J, Parkin DM, Steliarova-Foucher E (2010a) Estimates of cancer incidence and mortality in Europe in 2008. Eur J Cancer 46(4):765–781

Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM (2010b) Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 127(12):2893–2917

Fletcher SW, Elmore JG (2003) Clinical practice. Mammographic screening for breast cancer. N Engl J Med 348(17):1672–1680

Folpe AL, Goldblum JR, Rubin BP, Shehata BM, Liu W, Dei Tos AP, Weiss SW (2005) Morphologic and immunophenotypic diversity in Ewing family tumors: a study of 66 genetically confirmed cases. Am J Surg Pathol 29(8):1025–1033

Francisci S, Capocaccia R, Grande E, Santaquilani M, Simonetti A, Allemani C, Gatta G, Sant M, Zigon G, Bray F, Janssen-Heijnen M (2009) The cure of cancer: a European perspective. Eur J Cancer 45(6):1067–1079

Franklin W, Muller KM, Wistuba II, Sozzi G, Geisinger K, Brambilla E, Lam S, Gazdar A, Hirsch FR (2004) Squamous dysplasia and carcinoma in situ. Pathology and genetics of tumours of the lung, pleura, thymus and heart. WHO—IARC press, Lyon

Gravendeel LA, Kouwenhoven MC, Gevaert O, Rooi JJ de, Stubbs AP, Duijm JE, Daemen A, Bleeker FE, Bralten LB, Kloosterhof NK, De MB, Eilers PH, Spek PJ van der, Kros JM, Sillevis Smitt PA, Bent MJ van den, French PJ (2009) Intrinsic gene expression profiles of gliomas are a better predictor of survival than histology. Cancer Res 69(23):9065–9072

Hahn WC, Weinberg RA (2002) Modelling the molecular circuitry of cancer. Nat Rev Cancer 2(5):331–341

Haioun C, Itti E, Rahmouni A, Brice P, Rain JD, Belhadj K, Gaulard P, Garderet L, Lepage E, Reyes F, Meignan M (2005) [18F]fluoro-2-deoxy-D-glucose positron emission tomography (FDG-PET) in aggressive lymphoma: an early prognostic tool for predicting patient outcome. Blood 106(4):1376–1381

Hamilton SR, Aaltonen LA (2000) Pathology & genetics of tumours of the digestive system. WHO—IARCPress, Lyon

Hanahan D, Weinberg RA (2000) The hallmarks of cancer. Cell 100(1):57–70

Hansen RJ, Nagasubramanian R, Delaney SM, Samson LD, Dolan ME (2007) Role of O6-methylguanine-DNA methyltransferase in protecting from alkylating agent-induced toxicity and mutations in mice. Carcinogenesis 28(5):1111–1116

Hasegawa T, Hirose T, Ayala AG, Ito S, Tomaru U, Matsuno Y, Shimoda T, Hirohashi S (2001) Adult neuroblastoma of the retroperitoneum and abdomen: clinicopathologic distinction from primitive neuroectodermal tumor. Am J Surg Pathol 25(7):918–924

Hayes BD, Quinn CM (2009) Pathology of B3 lesions of the breast. Diagnostic Histopathol 15(10):459–469

Heinrich MC, Corless CL (2004) Targeting mutant kinases in gastrointestinal stromal tumors: a paradigm for molecular therapy of other sarcomas. Cancer Treat Res 120:129–150

Heinrich MC, Corless CL, Demetri GD, Blanke CD, Mehren von M, Joensuu H, McGreevey LS, Chen CJ, Van den Abbeele AD, Druker BJ, Kiese B, Eisenberg B, Roberts PJ, Singer S, Fletcher CD, Silberman S, Dimitrijevic S, Fletcher JA (2003) Kinase mutations and imatinib response in patients with metastatic gastrointestinal stromal tumor. J Clin Oncol 21(23):4342–4349

Herbst RS, Heymach JV, Lippman SM (2008) Lung cancer. N Engl J Med 359(13):1367–1380

Hoos A, Stojadinovic A, Mastorides S, Urist MJ, Polsky D, Di Como CJ, Brennan MF, Cordon-Cardo C (2001) High Ki-67 proliferative index predicts disease specific survival in patients with high-risk soft tissue sarcomas. Cancer 92(4):869–874

Huang ME, Ye YC, Chen SR, Chai JR, Lu JX, Zhoa L, Gu LJ, Wang ZY (1988) Use of all-trans retinoic acid in the treatment of acute promyelocytic leukemia. Blood 72(2):567–572

Hutter RV, Sobin LH (1986) A universal staging system for cancer of the colon and rectum. Let there be light. Arch Pathol Lab Med 110(5):367–368

Ingoldsby H, Callagy G (2009) Pathology of minimal metastatic disease in sentinel lymph nodes in breast cancer. Diagn Histopathol 15(10):470–477

Jakel O, Karger CP, Debus J (2008) The future of heavy ion radiotherapy. Med Phys 35(12):5653–5663

Jass JR (2007) Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. Histopathology 50(1):113–130

Jin L, Lloyd RV (1997) In situ hybridization: methods and applications. J Clin Lab Anal 11(1):2–9

Kallioniemi OP, Wagner U, Kononen J, Sauter G (2001) Tissue microarray technology for high-throughput molecular profiling of cancer. Hum Mol Genet 10(7):657–662

Kalluri R, Zeisberg M (2006) Fibroblasts in cancer. Nat Rev Cancer 6(5):392–401

Kamangar F, Dores GM, Anderson WF (2006) Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. J Clin Oncol 24(14):2137–2150

Kantarjian H, Sawyers C, Hochhaus A, Guilhot F, Schiffer C, Gambacorti-Passerini C, Niederwieser D, Resta D, Capdeville R, Zoellner U, Talpaz M, Druker B, Goldman J, O'Brien SG, Russell N, Fischer T, Ottmann O, Cony-Makhoul P, Facon T, Stone R, Miller C, Tallman M,

Brown R, Schuster M, Loughran T, Gratwohl A, Mandelli F, Saglio G, Lazzarino M, Russo D, Baccarani M, Morra E (2002) Hematologic and cytogenetic responses to imatinib mesylate in chronic myelogenous leukemia. N Engl J Med 346(9):645–652

Kauer M, Ban J, Kofler R, Walker B, Davis S, Meltzer P, Kovar H (2009) A molecular function map of Ewing's sarcoma. PLoS One 4(4):e5415

Kerbel RS (2008) Tumor angiogenesis. N Engl J Med 358(19):2039–2049

Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu CR, Peterson C, Meltzer PS (2001) Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nat Med 7(6):673–679

Kovar H, Jug G, Aryee DN, Zoubek A, Ambros P, Gruber B, Windhager R, Gadner H (1997) Among genes involved in the RB dependent cell cycle regulatory cascade, the p16 tumor suppressor gene is frequently lost in the Ewing family of tumors. Oncogene 15(18):2225–2232

Kreeger PK, Lauffenburger DA (2010) Cancer systems biology: a network modeling perspective. Carcinogenesis 31(1):2–8

Ladanyi M (1995) The emerging molecular genetics of sarcoma translocations. Diagn Mol Pathol 4(3):162–173

Lango MN (2009) Multimodal treatment for head and neck cancer. Surg Clin North Am 89(1):43–52

Laubenbacher R, Hower V, Jarrah A, Torti SV, Shulaev V, Mendes P, Torti FM, Akman S (2009) A systems biology view of cancer. Biochim Biophys Acta 1796(2):129–139

Levitt SH, Perez CA, Hui S, Purdy JA (2008) Evolution of computerized radiotherapy in radiation oncology: potential problems and solutions. Int J Radiat Oncol Biol Phys 70(4):978–986

Levitt SH, Purdy JA, Perez CA, Vijayakumar S (2006) Technical basis of radiation therapy. Practical clinical applications, 4th edn. Lippincott Williams & Wilkins, Baltimore

Liu CL, Prapong W, Natkunam Y, Alizadeh A, Montgomery K, Gilks CB, Rijn M van de (2002) Software tools for high-throughput analysis and archiving of immunohistochemistry staining data obtained with tissue microarrays. Am J Pathol 161(5):1557–1565

Llombart-Bosch A (2001) De la anatomia patologica estructural a la patologia molecular. Discurso de Recepcion. Real Academia de Medicina de la Comunidad Valenciana. RAMCV Press, Valencia, Spain

Llombart-Bosch A, Contesso G, Peydro-Olaya A (1996) Histology, immunohistochemistry, and electron microscopy of small round cell tumors of bone. Semin Diagn Pathol 13(3):153–170

Llombart-Bosch A, Machado I, Navarro S, Bertoni F, Bacchini P, Alberghini M, Karzeladze A, Savelov N, Petrov S, varado-Cabrero I, Mihaila D, Terrier P, Lopez-Guerrero JA, Picci P (2009) Histological heterogeneity of Ewing's sarcoma/PNET: an immunohistochemical analysis of 415 genetically confirmed cases with clinical support. Virchows Arch 455(5):397–411

Llombart-Bosch A, Navarro S (2001) Immunohistochemical detection of EWS and FLI-1 proteinss in Ewing sarcoma and primitive neuroectodermal tumors: comparative analysis with CD99 (MIC-2) expression. Appl Immunohistochem Mol Morphol 9(3):255–260

Lopez-Guerrero JA, Machado I, Scotlandi K, Noguera R, Pellin A, Navarro S, Serra M, Calabuig-Farinas S, Picci P, Llombart-Bosch A (2011) Clinicopathological significance of cell cycle regulation markers in a large series of genetically confirmed Ewing's sarcoma family of tumors. Int J Cancer 128(5):1139–1150

Lopez-Guerrero JA, Llombart-Cussac A, Noguera R, Navarro S, Pellin A, Almenar S, Vazquez-Alvadalejo C, Llombart-Bosch A (2006) HER2 amplification in recurrent breast cancer following breast-conserving therapy correlates with distant metastasis and poor survival. Int J Cancer 118(7):1743–1749

Lopez-Guerrero JA, Pellin A, Noguera R, Carda C, Llombart-Bosch A (2001) Molecular analysis of the 9p21 locus and p53 genes in Ewing family tumors. Lab Invest 81(6):803–814

Lynch HT, la Chapelle A de (2003) Hereditary colorectal cancer. N Engl J Med 348(10):919–932

Machado I, Noguera R, Pellin A, Lopez-Guerrero JA, Piqueras M, Navarro S, Llombart-Bosch A (2009) Molecular diagnosis of Ewing sarcoma family of tumors: a comparative analysis of 560 cases with FISH and RT-PCR. Diagn Mol Pathol 18(4):189–199

Madarnas Y, Trudeau M, Franek JA, McCready D, Pritchard KI, Messersmith H (2008) Adjuvant/neoadjuvant trastuzumab therapy in women with HER-2/neu-overexpressing breast cancer: a systematic review. Cancer Treat Rev 34(6):539–557

Marie JP, Zittoun R, Sikic BI (1991) Multidrug resistance (mdr1) gene expression in adult acute leukemias: correlations with treatment outcome and in vitro drug sensitivity. Blood 78(3):586–592

Markowitz SD, Bertagnolli MM (2009) Molecular origins of cancer: molecular basis of colorectal cancer. N Engl J Med 361(25):2449–2460

McCafferty MPJ, Healy NA, Kerin MJ (2009) Breast cancer subtypes and molecular biomarkers. Diagn Histopathol 15(10):485–489

Meara RS, Cangiarella J, Simsir A, Horton D, Eltoum I, Chhieng DC (2007) Prediction of aggressiveness of gastrointestinal stromal tumours based on immunostaining with bcl-2, Ki-67 and p53. Cytopathology 18(5):283–289

Mehlen P, Puisieux A (2006) Metastasis: a question of life or death. Nat Rev Cancer 6(6):449–458

Meyer T, Hart IR (1998) Mechanisms of tumour metastasis. Eur J Cancer 34(2):214–221

Nagtegaal ID, Krieken JH van (2002) The role of pathologists in the quality control of diagnosis and treatment of rectal cancer-an overview. Eur J Cancer 38(7):964–972

Naora H, Montell DJ (2005) Ovarian cancer metastasis: integrating insights from disparate model organisms. Nat Rev Cancer 5(5):355–366

Natkunam Y, Mason DY (2006) Prognostic immunohistologic markers in human tumors: why are so few used in clinical practice? Lab Invest 86(8):742–747

Navarro S, Giraudo P, Karseladze AI, Smirnov A, Petrovichev N, Savelov N, Varado-Cabrero I, Llombart-Bosch A (2007) Immunophenotypic profile of biomarkers related to anti-apoptotic and neural development pathways in the Ewing's family of tumors (EFT) and their therapeutic implications. Anticancer Res 27(4B):2457–2463

Nguyen DX, Bos PD, Massague J (2009) Metastasis: from dissemination to organ-specific colonization. Nat Rev Cancer 9(4):274–284

Nilsson G, Wang M, Wejde J, Kreicbergs A, Larsson O (1999) Detection of EWS/FLI-1 by immunostaining. An adjunctive tool in diagnosis of Ewing's sarcoma and primitive neuroectodermal tumour on cytological samples and paraffin-embedded archival material. Sarcoma 3(1):25–32

Nirmalanandhan VS, Sittampalam GS (2009) Stem cells in drug discovery, tissue engineering, and regenerative medicine: emerging opportunities and challenges. J Biomol Screen 14(7):755–768

O'Brien SG, Guilhot F, Larson RA, Gathmann I, Baccarani M, Cervantes F, Cornelissen JJ, Fischer T, Hochhaus A, Hughes T, Lechner K, Nielsen JL, Rousselot P, Reiffers J, Saglio G, Shepherd J, Simonsson B, Gratwohl A, Goldman JM, Kantarjian H, Taylor K, Verhoef G, Bolton AE, Capdeville R, Druker BJ (2003) Imatinib compared with interferon and low-dose cytarabine for newly diagnosed chronic-phase chronic myeloid leukemia. N Engl J Med 348(11):994–1004

O'Shaughnessy JA (2006) Molecular signatures predict outcomes of breast cancer. N Engl J Med 355(6):615–617

Ordonez NG, Mackay B (1998) Electron microscopy in tumor diagnosis: indications for its use in the immunohistochemical era. Hum Pathol 29(12):1403–1411

Parham DM, Hijazi Y, Steinberg SM, Meyer WH, Horowitz M, Tzen CY, Wexler LH, Tsokos M (1999) Neuroectodermal differentiation in Ewing's sarcoma family of tumors does not predict tumor behavior. Hum Pathol 30(8):911–918

Parkin DM (2004) International variation. Oncogene 23(38):6329–6340

Pass HI, Mitchell JB, Johnson DH, Turrisi AT, Minna JD (2000) Lung cancer: principles and practice, 2nd edn. Lippincott Williams & Wilkins, Baltimore

Patz EF Jr, Campa MJ, Gottlin EB, Kusmartseva I, Guan XR, Herndon JE (2007) Panel of serum biomarkers for the diagnosis of lung cancer. J Clin Oncol 25(35):5578–5583

Payne SJ, Bowen RL, Jones JL, Wells CA (2008) Predictive markers in breast cancer—the present. Histopathology 52(1):82–90

Perou CM, Sorlie T, Eisen MB, Rijn M van de, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D (2000) Molecular portraits of human breast tumours. Nature 406(6797):747–752

Pinero-Madrona A, Polo-Garcia L, Onso-Romero JL, Salinas-Ramos J, Canteras-Jordana M, Sola-Perez J, Galindo-Fernandez PJ, Illana-Moreno J, Bermejo-Lopez J, Navarrete-Montoya A, Parrilla-Paricio P (2008) Immunohistochemical characterisation of breast cancer: towards a new clasification? Cir Esp 84(3):138–145

Pinkerton R, Matthay K, Shankar AG (2007) Evidence-based pediatric oncology, 2nd edn. Blackwell, Oxford

Pleasance ED, Stephens PJ, O'Meara S, McBride DJ, Meynert A, Jones D, Lin ML, Beare D, Lau KW, Greenman C, Varela I, Nik-Zainal S, Davies HR, Ordonez GR, Mudie LJ, Latimer C, Edkins S, Stebbings L, Chen L, Jia M, Leroy C, Marshall J, Menzies A, Butler A, Teague JW, Mangion J, Sun YA, McLaughlin SF, Peckham HE, Tsung EF, Costa GL, Lee CC, Minna JD, Gazdar A, Birney E, Rhodes MD, McKernan KJ, Stratton MR, Futreal PA, Campbell PJ (2010) A small-cell lung cancer genome with complex signatures of tobacco exposure. Nature 463(7278):184–190

Ponten J (2001) Cell biology of precancer. Eur J Cancer 37(Suppl 8):97–113

Rakha EA, Putti TC, bd El-Rehim DM, Paish C, Green AR, Powe DG, Lee AH, Robertson JF, Ellis IO (2006) Morphological and immunophenotypic analysis of breast carcinomas with basal and myoepithelial differentiation. J Pathol 208(4):495–506

Ransohoff DF (2009) Promises and limitations of biomarkers. Recent Results Cancer Res 181:55–59

Reis-Filho JS, Westbury C, Pierga JY (2006) The impact of expression profiling on prognostic and predictive testing in breast cancer. J Clin Pathol 59(3):225–231

Richards MA (2009) The size of the prize for earlier diagnosis of cancer in England. Br J Cancer 101(Suppl 2):S125–S129

Ringborg U, Bergqvist D, Brorsson B, Cavallin-Stahl E, Ceberg J, Einhorn N, Frodin JE, Jarhult J, Lamnevik G, Lindholm C, Littbrand B, Norlund A, Nylen U, Rosen M, Svensson H, Moller TR (2003) The Swedish Council on Technology Assessment in Health Care (SBU) systematic overview of radiotherapy for cancer including a prospective survey of radiotherapy practice in Sweden 2001—summary and conclusions. Acta Oncol 42(5–6):357–365

Rosai J (2001) The continuing role of morphology in the molecular age. Mod Pathol 14(3):258–260

Rosai J (2007) Why microscopy will remain a cornerstone of surgical pathology. Lab Invest 87(5):403–408

Rosai J, Ackerman LV (1996) Ackerman's surgical pathology. CRC Press, Boca Raton

Saad RS, Kordunsky L, Liu YL, Denning KL, Kandil HA, Silverman JF (2006) Lymphatic microvessel density as prognostic marker in colorectal cancer. Mod Pathol 19(10):1317–1323

Sanchez-Garcia I (2009) The crossroads of oncogenesis and metastasis. N Engl J Med 360(3):297–299

Sandler A, Gray R, Perry MC, Brahmer J, Schiller JH, Dowlati A, Lilenbaum R, Johnson DH (2006) Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer. N Engl J Med 355(24):2542–2550

Sara H, Kallioniemi O, Nees M (2010) A decade of cancer gene profiling: from molecular portraits to molecular function. Methods Mol Biol 576:61–87

Sarkaria JN, Bristow RG (2008) Overview of cancer molecular radiobiology. Cancer Treat Res 139:117–133

Schoenberg Fejzo M, Slamon DJ (2001) Frozen tumor tissue microarray technology for analysis of tumor RNA, DNA, and proteins. Am J Pathol 159(5):1645–1650

Sharma R, Hamilton A, Beith J (2008) LHRH agonists for adjuvant therapy of early breast cancer in premenopausal women. Cochrane Database Syst Rev (4):CD004562

Shepherd FA, Rodrigues PJ, Ciuleanu T, Tan EH, Hirsh V, Thongprasert S, Campos D, Maoleekoonpiroj S, Smylie M, Martins R, Kooten M van, Dediu M, Findlay B, Tu D, Johnston D, Bezjak A, Clark G, Santabarbara P, Seymour L (2005) Erlotinib in previously treated non-small-cell lung cancer. N Engl J Med 353(2):123–132

Shergill IS, Shergill NK, Arya M, Patel HR (2004) Tissue microarrays: a current medical research tool. Curr Med Res Opin 20(5):707–712

Shia J, Ellis NA, Paty PB, Nash GM, Qin J, Offit K, Zhang XM, Markowitz AJ, Nafa K, Guillem JG, Wong WD, Gerald WL, Klimstra DS (2003) Value of histopathology in predicting microsatellite instability in hereditary nonpolyposis colorectal cancer and sporadic colorectal cancer. Am J Surg Pathol 27(11):1407–1417

Simunovic M, Smith AJ, Heald RJ (2009) Rectal cancer surgery and regional lymph nodes. J Surg Oncol 99(4):256–259

Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL (1987) Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science 235(4785):177–182

Smith I, Procter M, Gelber RD, Guillaume S, Feyereislova A, Dowsett M, Goldhirsch A, Untch M, Mariani G, Baselga J, Kaufmann M, Cameron D, Bell R, Bergh J, Coleman R, Wardley A, Harbeck N, Lopez RI, Mallmann P, Gelmon K, Wilcken N, Wist E, Sanchez RP, Piccart-Gebhart MJ (2007) 2-year follow-up of trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer: a randomised controlled trial. Lancet 369(9555):29–36

Sobin LH, Gospodarowicz MK, Wittekind C (2009) TNM classification of malignant tumours, 7th edn. Wiley-Blackwell, Oxford

Sorlie T, Perou CM, Fan C, Geisler S, Aas T, Nobel A, Anker G, Akslen LA, Botstein D, Borresen-Dale AL, Lonning PE (2006) Gene expression profiles do not consistently predict the clinical treatment response in locally advanced breast cancer. Mol Cancer Ther 5(11):2914–2918

Spira A, Ettinger DS (2004) Multidisciplinary management of lung cancer. N Engl J Med 350(4):379–392

Tang P, Skinner KA, Hicks DG (2009) Molecular classification of breast carcinomas by immunohistochemical analysis: are we ready? Diagn Mol Pathol 18(3):125–132

Tavassoli FA, Devilee P (2003) Pathology & genetics: tumours of the breast and female genital organs. IARC press—WHO, Lyon

Taylor CR, Cote RJ (1997) Immunohistochemical markers of prognostic value in surgical pathology. Histol Histopathol 12(4):1039–1055

Terrier P, Henry-Amar M, Triche TJ, Horowitz ME, Terrier-Lacombe MJ, Miser JS, Kinsella TJ, Contesso G, Llombart-Bosch A (1995) Is neuro-ectodermal differentiation of Ewing's sarcoma of bone associated with an unfavourable prognosis? Eur J Cancer 31A(3):307–314

Thun MJ, Henley SJ, Burns D, Jemal A, Shanks TG, Calle EE (2006) Lung cancer death rates in lifelong nonsmokers. J Natl Cancer Inst 98(10):691–699

Tlsty TD, Coussens LM (2006) Tumor stroma and regulation of cancer development. Annu Rev Pathol 1:119–150

Travis WD, Brambilla E, Muller-Hermelink HK, Harris CC (2004) Pathology & genetics of tumours of the lung, pleura, thymus and heart. IARC press—WHO, Lyon

Trigg ME, Sather HN, Reaman GH, Tubergen DG, Steinherz PG, Gaynon PS, Uckun FM, Hammond GD (2008) Ten-year survival of children with acute lymphoblastic leukemia: a report from the Children's Oncology Group. Leuk Lymphoma 49(6):1142–1154

Ueda T, Aozasa K, Tsujimoto M, Ohsawa M, Uchida A, Aoki Y, Ono K, Matsumoto K (1989) Prognostic significance of Ki-67 reactivity in soft tissue sarcomas. Cancer 63(8):1607–1611

Valk PJ, Verhaak RG, Beijen MA, Erpelinck CA, Barjesteh van Waalwijk van Doorn-Khosrovani, Boer JM, Beverloo HB, Moorhouse MJ, Spek PJ van der, Lowenberg B, Delwel R (2004) Prognostically useful gene-expression profiles in acute myeloid leukemia. N Engl J Med 350(16):1617–1628

Oosterom AT van, Judson I, Verweij J, Stroobants S, di Donato PE, Dimitrijevic S, Martens M, Webb A, Sciot R, Van GM, Silberman S, Nielsen OS (2001) Safety and efficacy of imatinib (STI571) in metastatic gastrointestinal stromal tumours: a phase I study. Lancet 358(9291):1421–1423

Verellen D, Ridder MD, Linthout N, Tournel K, Soete G, Storme G (2007) Innovations in image-guided radiotherapy. Nat Rev Cancer 7(12):949–960

Watt FM, Driskell RR (2010) The therapeutic potential of stem cells. Philos Trans R Soc Lond B Biol Sci 365(1537):155–163

Weber D, Rankin K, Gavino M, Delasalle K, Alexanian R (2003) Thalidomide alone or with dexamethasone for previously untreated multiple myeloma. J Clin Oncol 21(1):16–19

Weidner N, Tjoe J (1994) Immunohistochemical profile of monoclonal antibody O13: antibody that recognizes glycoprotein p30/32MIC2 and is useful in diagnosing Ewing's sarcoma and peripheral neuroepithelioma. Am J Surg Pathol 18(5):486–494

Weigelt B, Peterse JL, van't Veer LJ (2005) Breast cancer metastasis: markers and models. Nat Rev Cancer 5(8):591–602

Weinberg RA (2007) Biology of cancer. Garland Science, London

Whitehead R (1994) Gastrointestinal and oesohageal pathology. Churchill Livingstone, London

Witz IP (2009) The tumor microenvironment: the making of a paradigm. Cancer Microenviron 2(Suppl 1):9–17

Yamanaka R, Saya H (2009) Molecularly targeted therapies for glioma. Ann Neurol 66(6):717–729

Zambelli D, Zuntini M, Nardi F, Manara MC, Serra M, Landuzzi L, Lollini PL, Ferrari S, Alberghini M, Llombart-Bosch A, Piccolo E, Iacobelli S, Picci P, Scotlandi K (2010) Biological indicators of prognosis in Ewing's sarcoma: an emerging role for lectin galactoside-binding soluble 3 binding protein (LGALS3BP). Int J Cancer 126(1):41–52

Zhan F, Tian E, Bumm K, Smith R, Barlogie B, Shaughnessy J Jr (2003) Gene expression profiling of human plasma cell differentiation and classification of multiple myeloma based on similarities to distinct stages of late-stage B-cell development. Blood 101(3):1128–1140

Zhang X, Li L, Wei D, Yap Y, Chen F (2007) Moving cancer diagnostics from bench to bedside. Trends Biotechnol 25(4):166–173

Zoubek A, Dockhorn-Dworniczak B, Delattre O, Christiansen H, Niggli F, Gatterer-Menz I, Smith TL, Jurgens H, Gadner H, Kovar H (1996) Does expression of different EWS chimeric transcripts define clinically distinct risk groups of Ewing tumor patients? J Clin Oncol 14(4):1245–1251

Part II
Laboratory, Clinical, Data and Educational Resources

Chapter 3
Global Molecular and Cellular Measurement Technologies

Bodo M. H. Lange, Michal R. Schweiger and Hans Lehrach

Abstract Measuring technologies in the field of molecular biology and cellular biology have developed rapidly over the recent years, most obviously in the sequencing field where capacity and throughput have increased by several orders of magnitude. This has been a major factor in systems biology research, which thrives on technologies facilitating efficient and relatively economical genome-wide readout on DNA, mRNA, protein, and metabolome level. With data generation increasing exponentially, we are faced with new challenges of transforming this data into useful models that help to predict the outcome of genomic aberrations and to develop novel diagnostic and therapeutic strategies. There is currently a technological and digital transition from many array-based assays to second-generation sequencing approaches that analyse gene expression, genotype, single nucleotide polymorphisms and methylation patterns. Sequencing technologies are developing rapidly, as are preparatory enrichment techniques, which include the amplification of sample signal or target enrichment to reduce sample complexity. Proteomics and functional assays have also been much advanced as a result of technological progress in mass spectrometry or automatic microscopy and image analysis. Nevertheless, we are still far away from routinely measuring whole proteome data, because of the complexity of different transcripts and post-translational modifications. In spite of this, numerous new or improved analytical techniques embedded in integrated systems approach frameworks will potentially generate clinical usefulness.

Abbreviations

BRET	bioluminescence resonance energy transfer
BS	bisulfite
CDK	cyclin dependent kinases
CGH	comparative genome hybridization
CGP	cancer genome project
CNVs	copy number variations
DIGE	differential-in-gel-electrophoresis
ds	double stranded

B. M. H. Lange (✉)
Department of Vertebrate Genomics, Max-Planck Institute for molecular Genetics, Ihnestrasse 73, 14195, Berlin, Deutschland
e-mail: Lange_b@molgen.mpg.de

A. Cesario, F. B. Marcus (eds.), *Cancer Systems Biology, Bioinformatics and Medicine,*
DOI 10.1007/978-94-007-1567-7_3,

esiRNA	endoribonuclease-derived short interference RNAs
FISH	fluorescence in situ hybridizations
FFPE	formalin-fixed and paraffin-embedded
FRET	fluorescence resonance energy transfer
GC	gas chromatography
HPR	Human Protein Atlas
HRG	Histidine rich glycoprotein
ICGC	international cancer genome consortia
IEF	isoelectric focusing
IHC	immuno-histo chemistry
IMAC	immobilized metal affinity chromatography
InDel	insertions/deletions
IPG	immobilized pH gradient
ICAT	isotope encoded affinity tags
LC-MS	liquid chromatography MS
LUMIER	luminescence-based mammalian interactome mapping
MBP	methyl-binding protein
MeDIP	methylated DNA immuno-precipitation
MGS	microarray-based genomic selection
MS	mass spectrometry
MSCC	Methyl-Seq and methyl-sensitive cut counting
NGS	next-generation sequencing
NMR	nuclear magnetic resonance
NSL	non-small lung cancer
ORF	open reading frame
PAC	phosphoramidate chemistry
PGP	personal genome project
PLA	proximity ligation assay
PPI	protein-protein interaction
PSA	prostate specific antigen
PTM	posttranslational modification
QQQ	triple quadrupole
RISC	RNA-induced silencing complex
RNAi	RNA interferences
RPPA	reverse phase protein array
RRBS	reduced representation BS sequencing
SBGN	Systems Biology Graphical Notation
SBML	Systems Biology Markup Language
SNPs	single nucleotide polymorphisms
SILAC	stable-isotope labelling by amino acids in cell culture
TAP	tandem affinity purification
TCGA	The Cancer Genome Atlas
Y2H	yeast-two-hybrid
4sU	4-thiouridine

3.1 Introduction—The Need for Systems Biology Predictive Models

Knowledge about human biology has been considerably expanded in recent years. Systematic global datasets covering DNA, RNA, proteins, metabolite, and many other aspects of phenotypes now make it possible to characterize the biology of a disease process in individual patients to a far greater extent than could have been achieved only a few years ago. While it took 10 years and about a billion US dollars to complete the sequence of the human genome, we are now able to generate that much information on one next generation sequencing (NGS) instrument in 10 days (Illumina, Solid announcements). We expect to be able within a few years to generate the sequence of the genome of an individual patient in a few minutes (PacBio announcement) at a cost of 1000 US$ or less, representing reductions in cost and time of roughly 6 orders of magnitude.

Similar, though perhaps less dramatic, improvements have been made in many other technology areas, providing a basis for combining the information on pathways gained in decades of worldwide biomedical research, with a genome (and proteome) scale analysis of the disease process in individual patients, so as to establish models able to predict in detail the individual response of patients to different therapies.

The first chemotherapies developed were based on a disruption of cell homeostasis, involving alkylating agents or antimetabolites. It is only in recent years that targeted chemotherapies are under development and in clinical use. These drugs aim at the inhibition or activation of specific cellular target proteins, most frequently protein kinases. Parallel with this development, there emerged a recognition that only a subset of patients significantly benefit from those chemotherapies. Treatment of the others is pointless; they suffer the side-effects without benefitting from the treatment. The ability to identify in advance the subpopulation of patients most likely to respond to a particular therapy is summarized in the concept of personalized cancer therapy (Fig. 3.1). This fundamental underlying principle is the enormous heterogeneity among tumours, even among tumours of the same class. Using the newest high-throughput technologies, we have learned that a wide spectrum of different genetic alterations, from mutations to copy number and structural variations, are found in each tumour. This complexity is even aggravated by epigenetic variations seen in many tumour entities.

Recent improvements in high-throughput technologies for the generation of global data sets covering the genome, transcriptome, proteome and metabolome have made it possible for the first time to draw phenotype-genotype conclusions and to optimally stratify patients for the most appropriate cancer treatments. Efficient handling of these large datasets requires specific bioinformatics tools which primarily aim at directly identifying the correlation between experimental and clinical data. In a second step, the data are integrated in large computer models of cellular pathways. Adjustment of these cellular models to complex diseases enables the establishment of cancer computer models, which in turn can be used as predic-

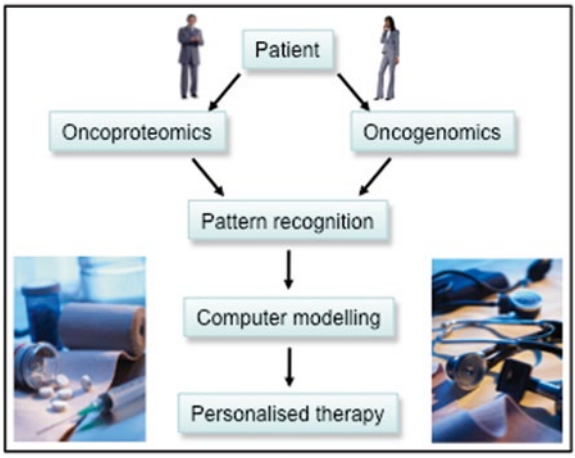


Fig. 3.1 Outline for the process of personalized therapy: Patients are diagnosed with a tumour which is thoroughly characterized using modern molecular biological and high throughput technologies. Personalized therapy is planned using computer models on the basis of a reliable prediction of response

tive models for the calculation of optimal chemotherapeutics combinations. It is becoming increasingly clear that not only are alterations in single genes essential for tumour pathogenesis, but also modifications in pathways. Computer models can be used to calculate the influence of each single gene on the complete pathways.

An approach in this direction has been taken by the Max-Planck Institute for Molecular Genetics, the comprehensive cancer centre of the Charité Hospital, Harvard Medical School and three small companies (Alacris, Theranostics and CollabRx). They aim to characterize tumour genomes and transcriptomes, as well as patient genomes, in order to construct predictive models of the effects and side-effects of specific drugs and drug combinations on specific patients. As an example of this strategy, results of modelling the dose-dependent effects and side-effects of a specific drug combination on a specific patient with metastatic melanoma are shown below (Fig. 3.2). This model is based on deep sequencing (40×coverage of the tumour genome, 30×coverage of blood DNA, 30×coverage of the exome), ~300 million reads from a tumour, a cellular control (melanocyte), as well as a cancer stem-cell preparation of the same tumour (CD133plus, CD133minus).

In order to establish such individual cancer models and to optimize the treatment for individual patients, many critical points have to be considered. This chapter gives an overview of technologies and strategies which can contribute to this goal, starting from the recording of medical history and the generation of high-throughput genetic and proteomic data, and proceeding to the integration of this data in predictive computer models.

3.2 Sample Preparation

With the combination of advanced Next Generation Sequencing (NGS) technologies and clinical tumour material one is certain to identify additional novel genes or DNA regions that contribute to tumorigenesis. Since genetic and epigenetic (such

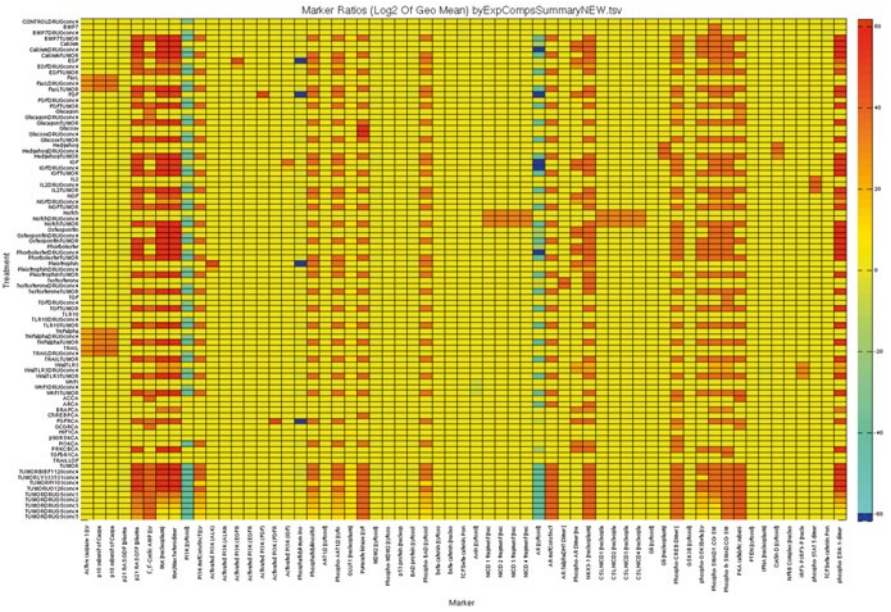


Fig. 3.2 The modelling of mutations represented by individual markers (x-axis) and of the response of particular different treatments (y-axis) is displayed in a heat-map identifying potential efficient potent drug cocktails and side effects of treatment in a particular genetic background

as DNA methylation) marks are chemically stable and relatively easy to detect, they are attractive biomarkers in oncology. Specialized protocols permit the extraction and conversion of DNA from formalin-fixed and paraffin-embedded (FFPE) tissue samples which are routinely stored at pathology departments (Bian et al. 2001). However, an FFPE preparation is incompatible with many downstream molecular biology techniques, whereas NGS technologies are applicable to small samples of FFPE material dating back over 20 years (Schweiger et al. 2009). With these experiments in mind, it is highly likely that bisulphite converted FFPE DNA used for DNA methylation analyses can also be employed for NGS analyses. The use of FFPE DNA material opens up access to a variety of population stratifiers for clinical trials and enables routine diagnostic work-ups of patients. DNA analyses can be performed on small numbers of cells obtained by laser capture micro-dissection, as well as on DNA extracted from diverse body fluids such as blood, urine or sputum (Kerjean et al. 2001). Combinations of all these methods open up a broad field of clinically or molecular-biologically relevant questions, such as the problem of tumour resistance to therapy, the different growth rates and progression of tumours, properties required for metastatic spreading or tumour evolution from single tumour stem cells.

3.3 Analysis of the Genome

3.3.1 DNA Microarrays

Global cellular changes can be measured using microarray technologies. In standard microarrays, nucleotide probes are immobilized on solid surfaces. Single-stranded DNA or cDNA samples of interest are distributed over the array and complementary fragments can hybridize targets to the probes. Unbound fractions are washed off. Probe-target hybridizations are usually detected and quantified by detection of fluorophore-, silver-, or chemiluminescence-labelled targets. A wide variety of DNA microarrays offer the possibilities of measuring changes in gene expression, genotyping single nucleotide polymorphisms (SNPs), re-sequencing DNA regions or determining the copy number of DNA segments (Ragoussis 2009). Gene expression analysis in cells or tissues is used for determining specific cell stages or obtaining functional readouts of the protein, obtained with over-expression or gene knock-down methods.

Another important area for gene expression arrays is the stratification of tumours with regard to distinct biological properties or individual prognoses. The expression levels of thousands of distinct genes are measured in tissue preparations. Subsequent bioinformatics correlation analyses make it possible to identify a subset of genes, called signature-genes, which are characteristic for a specific tumour phenotype, e.g. drug resistance, progression or recurrence. Pioneering work has been carried out on breast cancers to distinguish those which become metastatic from those that will remain locally restricted. Using gene expression arrays it has been possible to predict the clinical course with more than 90% accuracy. Whereas women used to be treated with chemotherapy irrespective of the tumour characteristics, the distinction between aggressive and less aggressive forms now spares many women, of whom up to 85% will not develop metastasis (van de Vijver et al. 2002), unnecessary treatment and side-effects (Williams et al. 2006).

Other fields of application of arrays are comparative genome hybridizations, called Array-CGHs. These are based on making the test DNA and a control DNA compete to hybridize to a target DNA immobilized on an array. This technique is commonly used to detect microdeletions and duplications in patients with congenital abnormalities. Array-CGHs are also used to detect copy number alterations in cancer samples. Comparisons with NGS data show, as was expected, good overlap between the two methods.

3.3.2 Next Generation Sequencing (NGS)

The sequencing of the human genome represents one of the greatest scientific achievements in the history of mankind. Advancements in automated capillary electrophoresis led to the launching in 1990 of the human genome project (HUGO)

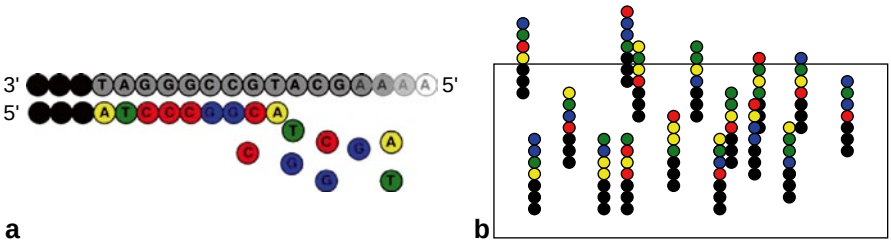


Fig. 3.3 Principle of next generation sequencing (NGS) technologies: In comparison to (**a**) conventional sequencing processes (e.g. Sanger) where only one DNA strand is sequenced at a time, for NGS (**b**) millions of DNA strands are spatially immobilized and sequenced in parallel

intended to decipher the entire human genome. In an international effort, over 1000 scientists sequenced the genome, the project was completed and announced in 2001 (Lander et al. 2001). Continually growing demand for high-throughput technologies led to the subsequent development of new NGS technologies, making it possible to sequence the whole human genome within a few days, and offering exciting prospects for a systematic genome analysis. The worldwide personal genome project (PGP), and the 1000 genomes project aim at sequencing over 1000 individual genomes to gain insight into genomic variability.

3.3.2.1 Sequencing Techniques

The development of NGS technologies (454 (Roche), Genome Analyzer (Illumina), SOLiD (Applied Biosystems, AB)) has initiated a real revolution in genomics analyses. With these technologies, an enormous parallel analysis of genomic DNA has become possible in a time-frame of a few days. Key features of these technologies are the spatial immobilization of millions of short DNA fragments followed by a massively parallel sequencing process (Fig. 3.3).

Fluorescence markers incorporated into the DNA fragments either by ligation (SOLiD) or polymerase activity (Illumina) during the sequencing process, or by a light signal emitted from luciferase activity coupled to the incorporation of nucleotides (454FLX), are detected by high resolution cameras. They provide digital information on DNA sequences of the fragments which are assembled and aligned to reference genomes using bioinformatics tools. Digital information is the basis for re-sequencing approaches as well as quantification modules for gene expression analyses or chromatin immuno-precipitation experiments. The parallel sequencing of millions of DNA molecules is especially useful for sequencing heterogeneous material, such as cancer tissues.

While NGS technologies have enormously increased the sequencing throughput, the several steps of enrichment, amplification and labelling still cause the performance to be relatively time and cost-intensive. Costs are, however, dropping rapidly for most second-generation sequencing systems (http://www.illumina.com, http://

www.454.com, http://www.solid.appliedbiosystems.com, http://www.helicosbio.com), down to costs of a few thousand US dollars per genome. This was the figure announced by Complete Genomics (http://www.completegenomics.com), a recent start-up company developing an optimized automated second-generation sequencing approach to be run in service mode.

In contrast, future third generation sequencing approaches, which rely on detecting the binding of the nucleotidetriphosphate to the polymerase in real time (Pacific Biosciences, Visigen), nanopore (*e.g.* Oxford Nanopore) and scanning probe sequencing approaches (Blow 2008; Clarke et al. 2009; Greenleaf and Block 2006; Rusk 2009; Sugiyama et al. 2006), are directed towards sequencing of single DNA molecules without any prior amplification or labelling. PacBio and Visigen use optical techniques to detect the triphosphate about to be incorporated. In the procedure developed by Pacific Biosciences, multiple zero-mode waveguide structures on a chip define minute volumes containing single polymerase molecules. DNA sequences are read out by a series of desoxynucleotide triphosphates labelled with different fluorescent dyes and illuminated during the incorporation step. During incorporation of the triphosphate analogue, the fluorescent label is cleaved off together with the pyrophosphate group, allowing the next incorporation step. In addition to determining the sequence, this procedure has also been shown to be capable of detecting base modifications in the DNA, because of its influence on the kinetics of incorporation. The protocol currently under development at Visigen is conceptually quite similar. In this case however, the binding of the fluorescence labelled desoxynucleotide triphosphate is detected by energy transfer from a quantum dot on the polymerase to the fluorescence label on the triphosphate group. The basic principle of nanopore sequencing is that a DNA strand or a cleaved nucleotide is passed through a nanopore and induces changes in the current applied (Clarke et al. 2009). The use of electrical currents for nucleotide identification is promising for discrimination of all four nucleotides as well as for the identification of methylated cytosines. This suggests that during one sequencing process, all five nucleotides, A, T, C, G and methylation of cytosine in the five position (5mC), could be distinguished, and no additional manipulation of DNA would be required for the construction of methylation patterns. In another approach under development, that of scanning probe sequencing, the DNA molecule is immobilized and the scanning instrument records the nucleotides (Sugiyama et al. 2006).

3.3.2.2 Total Versus Enriched DNA Sequencing

Sequencing of entire genomes is an important application of NGS. Although all types of genetic polymorphisms can be identified using whole-genome re-sequencing approaches, this method is still too cost-intensive to be conducted routinely. Instead, many research and diagnostic goals might be achieved by sequencing only a fraction of the genome (Fig. 3.4).

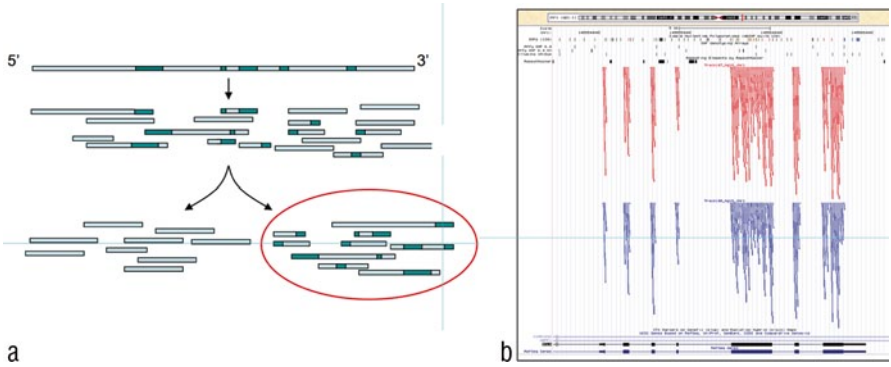


Fig. 3.4 a Schematics of the DNA capturing process: After DNA fragmentation, regions of interest (*dark green*) are enriched using different DNA capturing approaches, e.g. hybridization approaches or multiplex PCR reactions. **b** UCSC visualization of enriched and sequenced DNA fragments. *Bottom*: Exon/intron structure of the gene of interest with thick bars representing exons. *Red* and *blue* represent two different samples with the height of each peak being proportional to the amount of DNA fragments sequenced at this localization

For disease analysis, specific sets of genes involved in the pathomechanism or implicated by whole genome association studies are of major interest. However, there might be limitations on sequencing capacity to only re-sequence all protein-coding regions ('exome') which encompass approximately 1% of the genome. Several targeted sequence enrichment techniques to reduce DNA sequence complexities have been established. In particular, microarray-based genomic selection (MGS), multiplex exon capture or bead-based enrichment methods are already commercially available and used for targeted sequencing approaches. The main differences are the amount of input DNA, the ease of performance and whether they are based on hybridization or synthesis. In view of the fact that sequencing capacities per run continue to increase and are soon predicted to reach throughputs of up to 300 gigabases per run (SOLiD announcement), greater ease in processing multiplex multiple samples will become increasingly important.

The array-based enrichment of DNA segments is based on a hybridization of genomic DNA to oligonucleotides immobilized on arrays. After extensive washing steps, bound DNA can be eluted and used for sequencing. This procedure results in a large reduction of background DNA and a good enrichment factor. However, relatively large amounts of input DNA are currently required and the procedure cannot be easily automated for the analysis of large sample sizes. Solution-based enrichment technology (Agilent) uses biotinylated RNA 'baits' to fish targets out of a pond of DNA fragments. The RNA baits are synthesized by reverse transcription of a library of DNA-oligonucleotides, which have themselves been synthesized and released from an array. The enrichment process is possible with small amounts of DNA (so far 500 ng have been tested) and can be parallelized. Enrichment factors of 94% of the theoretically achievable values are within range, calculated as:

Calculation of enrichment factors (EF) for human:
EF = reads on target/(size of enrichment region × all aligned seq. reads/3×10^9)

For a target-region independent factor, the maximum possible enrichment factor for the target region is calculated and the percentage of the real EF determined.

Another possibility for an enrichment of specific DNA regions is a rolling-circle amplification strategy. For this, primers with arms complementary to the genomic DNA are hybridized to the region of interest and rolling circle amplifications are performed. Sequencing adapters are subsequently ligated to the amplified DNA and NGS is performed. Following a similar direction, multiplex PCRs are designed. However, since PCR substrates and products negatively influence the power of the multiplex reactions, a spatial separation of the PCR reactions in an emulsion of buffer droplets in oil is required. Again, amplified material can be used for NGS. So far this technology is optimized for 55,000 genomic regions. A potentially quite powerful approach for the selective amplification of relatively small numbers of DNA (or RNA) segments has been developed by Raindance Technologies (http://www.raindancetechnologies.com/). It relies on the use of microfluidics to generate small droplets containing the nucleic acid to be amplified; each droplet is mixed with another droplet containing one out of thousands of different primer pairs defining the sequences to be amplified. The sequences in this emulsion can then be amplified by carrying out emulsion PCR, and the amplified DNAs can in turn be sequenced by different second (or third) generation sequencing techniques. No matter which method is used, an enrichment of DNA regions results in a significant reduction of required sequencing capacities, permitting the throughput of analyses to be significantly increased.

3.3.2.3 Copy Number Variations (CNVs)

Genome rearrangements resulting in aberrant transcriptional events are common features in human cancer. Not only point mutations, but also extended genome rearrangements are thus implicated in tumorigenesis, including translocations, inversions, small insertions/deletions (InDels) and copy number variations (CNVs). InDels are most often defined as deletions or insertions below one thousand base pairs (1 kb) of DNA, whereas CNVs comprise alterations larger than 1 kb of DNA. Recent analyses by genome-wide approaches have revealed the importance of structural genomic variations in health and disease; moreover, genetic association studies have implicated CNVs in cancer. Insight into the connection of changes in CNVs to several diseases has boosted the development of new technologies investigating CNV rearrangements. Initial methods have involved microscopic examination of chromosome bandings, PCR, FISH, and microarrays; more recently, NGS technologies have been utilized for these analyses. These approaches offer important advantages over conventional methods such as microarrays or array compara-

tive genomic hybridization. In addition to quantitative information, they provide data about qualitative mechanisms, e.g. balanced rearrangements such as reciprocal translocations and inversions, which might otherwise have been overlooked. Moreover, since the sequencing is based on digital modes, they are able to detect variants present in a subpopulation of cells. Given the short read-lengths of most NGS technologies, paired-end sequencing approaches have been developed whereby short DNA segments, which are many hundreds of base pairs apart, are brought together and sequenced. Along with the technology advances, new algorithms have been developed which identify 'chimeric' reads. These are fragments which incorporate DNA segments from different parts of the genome and need to be split before correct alignments can be determined.

3.3.2.4 DNA Methylation: Bisulfite Conversion

The nucleotide sequence is the primary level of genetic information and the basic principle of genetic inheritance. Another level of complexity in genomic DNA arises from epigenetic variations of DNA segments which underlie the inheritance of phenotypes from generation to generation as well as from cell to cell during cell division. In humans, cytosine methylation was the first mark discovered. It is required both for the regulation of gene expression and for silencing transposons and other repetitive sequences (Beck and Rakyan 2008). The chemical modification occurs predominantly via a covalent attachment of a methyl group to the C5 position of the cytosine ring (5mC) in CpG dinucleotides. The structure of cytosine is thereby altered without changing its base-pairing properties. Altered methylation patterns have been reported in a diverse array of complex human diseases such as cancer, systemic autoimmune and psychiatric diseases as well as in monogenic epigenetic diseases (Feinberg 2007). Research conducted mainly on cancer epigenetics indicates that cytosine methylations are among the earliest events in tumorigenesis. The first biomarkers were developed on the basis of these modifications (Banerjee and Verma 2009).

For the detection of methylated nucleotides, the DNA is marked using a 'bisulfite (BS) conversion' reaction or by methylation-specific restriction analyses. In BS conversion, genomic DNA is treated with sodium bisulfite under denaturing conditions. Cytosine residues are deaminated and converted to uracil, leaving methylated cytosine moieties unaffected. During the following amplification reactions, uracil is converted to thymine, and subsequent technologies rely on SNPs detection methods for distinguishing between cytosines and thymines which result from bisulfite conversions. Pitfalls in all technologies may lead to false positive base identifications resulting from incomplete conversion reactions, degraded DNA caused by harsh conversion conditions, and methylations in pseudogenes. For human genomes, up to 1% or accordingly 28 million CpG sites are estimated. To investigate all of these by conventional PCR amplification and sequencing strategies would be extremely time and cost-intensive. Only with the aid of NGS technologies do whole human genome m5C patterns become feasible. Preliminary human genome-wide epigen-

etic maps after bisulfite treatment have been constructed and show that more than 93% of all CpGs can be targeted (Lister et al. 2009).

However, the sequencing capacities and costs required for whole genome analyses are still relatively high. Thus, limited analyses of parts of the genome are more practical for gaining insight into methylation patterns in mammals, especially if large numbers of samples need to be analysed. Initial approaches to reducing genome complexity have been performed by BS conversion of genomic DNA, PCR-amplification of target regions, and sequencing of all fragments in one NGS run (Korshunova et al. 2008; Taylor et al. 2007). Another approach in aiming for a reduction of sequencing capacities involves a combination of targeted enrichment of specific DNA regions, BS conversion and NGS. In the technology known as the bisulfite padlock probe (BSPP), genomic DNA was bisulfite-treated and 10,000 independent regions were interrogated (Ball et al. 2009). Reduced-representation BS sequencing (RRBS) exploits digestion of the genomic DNA at CCGG sites with a methylation-insensitive restriction enzyme followed by size selection by gel electrophoresis. Apart from BS treatments, DNA methylation sites can also be searched for using methylation-sensitive (Hpall) and methylation-insensitive (Mspl) restriction enzymes; these methods are known as Methyl-Seq and methyl-sensitive cut counting (MSCC). With this approach a maximum of 1.4 million sites can be interrogated, and after approximately 20 million sequencing reads, 66% of the sites have been covered with at least one read. A drawback of these methods is that any region showing at least one read in the Hpall-digest is currently called ‘unmethylated’, with the result that the quantitative methylation state of the individual region is lost and partial methylation or imprinted loci remain unidentified.

In addition to the direct approaches described above, indirect approaches examine genome-wide methylation profiles, providing information about methylated regions of approximately 100–200 bp length. These approaches basically rely on precipitations of DNA fragments containing methylated cytosines (5mC) using an anti-5mC antibody or methyl-binding proteins (MBP) and are thus termed methylation-dependent immunoprecipitation (MeDIP) and methyl-binding protein (MBP) assays (Cross et al. 1994; Keshet et al. 2006; Weber et al. 2005). The use of NGS (MeDIP-seq, MBP-seq) instead of custom-designed arrays to identify precipitated DNA fragments (Fig. 3.5) provides unbiased genome-wide information about methylated regions (Down et al. 2008).

This implies that all DNA fragments can be identified, and not just pre-selected regions which are immobilized on the array. The completeness of the data is especially advantageous in generating methylation profiles outside of CpG-islands and promoter regions, for example in gene bodies where DNA methylation changes have recently been shown to occur (Ball et al. 2009; Rakyan et al. 2008). Taken together, the number of different NGS epigenetic technologies is large, and each has its own advantages and disadvantages. The selection of the right technology for the research question investigated is a key feature in making the most of the enormous potential NGS bears for basic and clinical research directions of research.

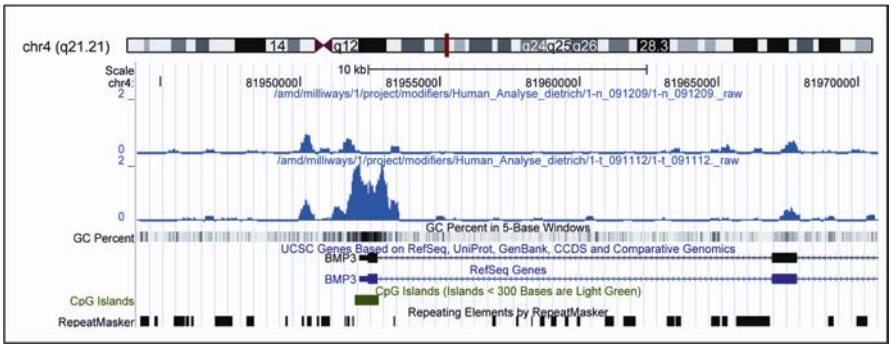


Fig. 3.5 Visualization of a MeDIP-seq region after precipitation of DNA isolated from normal (*top*) and tumour (*bottom*) colon tissue. *Blue* peaks indicate the amount of precipitated DNA fragments at these regions and peak heights are proportional to the amounts of reads sequenced. *Green*: CpG island

3.3.2.5 Transcriptome Analyses

First candidate-gene approaches used Northern blot techniques to analyse expression levels of single genes. Later, the development of microarray techniques allowed the simultaneous examination of thousands of genes in one experiment. Due to their high power and relatively low costs, microarrays are now commonly used in laboratories. However, because of the pre-selection of examined genes these methods do not permit a detection of novel transcripts. As an alternative, NGS technologies are used for the analysis of transcriptomes (Morozova et al. 2009; Sultan et al. 2008). With this technology not only can gene expression profiles be established, but also alternative splicing and translocations can be detected. Using long reads from 454 or PacBio technologies, splice patterns can be directly detected, providing information which is otherwise difficult to reconstruct from short-read datasets.

For the sequencing of RNA species, mRNA is reverse transcribed into cDNA, fragmented, and adapters ligated before the sequencing is performed. Protocols for sequencing on all three commonly used NGS platforms are available. The data can be used not only for expression profiling, but also for the detection of new transcripts and the discovery of new splice variants. If the sequencing effort is sufficiently large, it can also server for mutational profiling. Transcriptome sequencing has been performed on a wide variety of organisms, from *Arabidopsis thaliana* and *Drosophila melanogaster* to human cell lines and tissues. Most of the experiments for gene expression profiles are performed at the total cellular RNA level. A disadvantage of performing these experiments at an endpoint is that temporal information, which is especially required for modelling the kinetics of cellular processes, is lost. This problem can be addressed by using nascent RNA which is identified by metabolic labelling of newly generated RNA. Elegant approaches have been developed which use an incorporation of 4-thiouridine (4sU) into nascent RNA followed by a thiol-specific biotinylation and magnetic separation of nascent RNA. Thanks

to this method, kinetic pictures are emerging of RNA metabolism. Combining NGS technologies with several samples in one sequencing run, throughput can be increased and detailed time curves can be established.

The use of NGS technologies can also make it possible for those small RNA molecules called miRNAs to be easily enriched, detected and analysed. MicroRNAs have recently arisen as crucial regulators of development and cell fate determination, both essential elements in cancer progression. In this field, NGS technologies are extremely powerful for the discovery of novel miRNA genes on a genome-wide scale. The combination of both sequencing approaches—transcriptome sequencing and miRNA sequencing—enables the prediction of miRNA-target RNA pairs, which adds a functional viewpoint to the discovery of new and already known miRNAs.

3.4 Proteomics

After the completion of several genome-sequencing projects for man and other organisms, systematic analysis of gene function by functional genomics techniques has become central to further progress in understanding complex biological processes. Proteomics, the systematic analysis of the proteins present in specific cells or tissues, is essential for our understanding of the regulation of cellular signalling and homeostasis. However, because of the complexity of protein variants created, for example, through alternative splicing and post-translational modifications, the systematic analysis of entire proteomes constitutes a challenge for even the most advanced measurement technologies. Nevertheless, advances have been made that now permit efficient sub-fractionation of the proteome and multi-parameter quantitative and time resolved analysis of protein interactions and function.

3.4.1 Two-dimensional Gel Electrophoresis

Two-dimensional (2D) gel electrophoresis is the method of choice for the separation of complex mixtures of proteins (in the order of 1000 to 10,000) discriminated by their charge and molecular weight. In a first step, proteins are separated according to their isoelectric point by isoelectric focusing (IEF) using either an immobilized pH gradient (IPG) or carrier-ampholyte-based IEF. In a second step, proteins are separated according to their electrophoretic mobility by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). 2D electrophoresis has been successfully employed for the identification of biomarkers because of its excellent capacity to detect and quantify changes within complex protein mixtures (Reymond and Schlegel 2007; Kulasingam and Diamandis 2008). An example is the use of prostate-specific antigen (PSA) as a prognostic marker for prostate cancer (Stamey et al. 1987; Charrier et al. 2001). 2D electrophoresis can further be enhanced through

differential-in-gel-electrophoresis (DIGE) employing differential labelling, for example of normal versus tumour tissue. This provides fluorescence-based quantitative read-out of relative differences of protein expression. Subsequent protein identification is carried out by mass spectrometry (MS) for biomarker characterization, as for example in hepatocellular carcinoma (Lee et al. 2005). However, the sensitivity of 2D gel-electrophoresis is limited by variations of protein separation, gel polymerisation differences, reproducibility in staining procedures and difficulty in separating hydrophobic proteins that are of limited solubility. Consequently, the quantitative analysis of unique protein spots through software and bioinformatics tools can be hampered by artefacts both on the level of image analysis and statistical evaluation of results (Biron et al. 2006). Nevertheless, 2D gel electrophoresis, in combination with advanced differential labelling techniques and MS, is a powerful technique for biomarker discovery and for systems biology approaches, but is currently too complex for routine diagnostic application to patient tissues in cancer research.

3.4.2 Mass Spectrometry

Mass spectrometry (MS) has greatly contributed to the characterization of cancer-related pathways and the identification of biomarkers. MS in this area has focused on

a. whole proteome analysis,
b. post-translational modifications (PTM) detection,
c. quantitative readout of the effect of modulating environmental and signalling cues,
d. protein-protein interaction (PPI) identification for pathway analysis, and
e. the identification of cancer specific metabolites

Proteomic studies can be divided into discovery approaches that aim to identify novel effectors and biomarkers, and quantification and validation approaches that analyse known molecules quantitatively (Mann 2006; Zhou and Veenstra 2007; Gentleman and Huber 2007; Nishizuka and Spurrier 2008; Dang et al. 2009).

Proteomic discovery approaches rely on MS/MS instruments that provide high mass accuracy (MALDI-TOF/TOF, qTOF, and LTQ Orbitrap instruments), see (Yates et al. 2006). These techniques are also suitable for relative quantification of the identified molecules, for instance, if combined with stable isotope labelling techniques. In contrast, quantification and validation approaches are conducted on mass-selective instruments, such as triple quadrupole (QQQ) mass spectrometers. From a complex mixture, these instruments are able to select, fragment and quantify a number of pre-chosen peptides or metabolites (Hopfgartner et al. 2004; Picotti et al. 2009).

Because quantitative measurements are essential for systems biology approaches, multiple quantitative MS strategies have been developed that facilitate measurement of relative and absolute changes of protein levels, the analysis of protein kinetics, and better discrimination between false and true positive hits in complex

protein mixtures (Mann 2006). Examples of quantification strategies that involve stable isotope labels in discovery approaches are isotope encoded affinity tags (ICAT) (Gygi et al. 1999) or stable-isotope labelling by amino acids in cell culture (SILAC) (Mann 2006). These work by spiking protein mixtures with exogenous labelled prototypic peptides for targeted approaches (Kuster et al. 2005). However, because of the high costs and low flexibility of stable isotope incorporation, researchers are increasingly turning to label-free peptide quantifications (Zhu et al. 2010). Depending on the technological platform, changes in chromatographic ion intensity, or spectral counting of identified proteins, is translated into quantitative information.

To increase sensitivity for applying MS to biomarker discovery, validation and screening, several approaches to reducing the complexity of samples have been suggested. These attempts range from selective depletion of major proteins fractions with marginal information content (i.e. albumin in plasma), the enrichment of sub-proteomes such as the glycoproteome, or the selective enrichment of target molecules by stable isotope standards with capture by anti-peptide antibodies (Whiteaker et al. 2010; Schiess et al. 2009). Indeed, the application of a combination of these techniques has improved the detection sensitivity down to the physiologically relevant pg/ml range of protein concentration. In spite of great technical improvements with regard to sensitivity and separation techniques, neither the number of cancer related markers nor their predictive power have greatly increased. Although latest studies report quantification limits in the lower attomolar range (Kuzyk et al. 2009), the major limitation for biomarker discovery and validation in body fluids remains the sensitivity of MS techniques. In plasma, for instance, albumin and clinically relevant biomarkers such as the prostate specific antigen (PSA) are separate by more than 10 orders of magnitude in concentration (Anderson and Anderson 2002). However, it is likely that some of the current limitations of biomarkers with respect to reliability of the clinical diagnosis and predictions will be overcome by combining several non-unique markers to enhance the likelihood of predictions.

3.4.3 Quantitative Protein Arrays

Protein and antibody arrays have become informative tools for systems biology approaches and for discovery of biomarkers, by obtaining direct quantitative or relative quantitative readout of protein or protein phosphorylation levels. Two main approaches (Fig. 3.6) are used: the microarray immunoassay and reverse phase protein arrays (RPPA) (Espina et al. 2009; Korf et al. 2008; Korf et al. 2009). The RPPA technology involves the dispersion of very small (in the order of 1 nl) tumour lysates on, for example, nitrocellulose-coated glass slides, detection of tumour markers with specific primary antibodies; and finally detection of the primary antibody with a fluorescently labelled secondary antibody. In contrast, microarray immunoassays involve the spotting of different antibodies onto glass

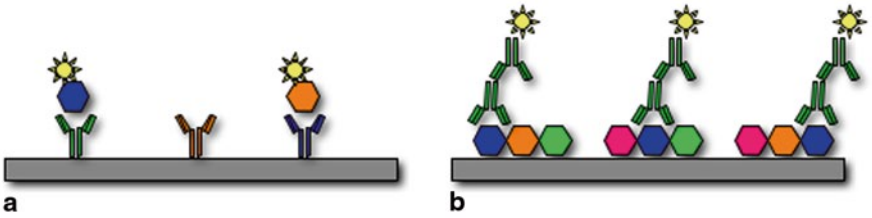


Fig. 3.6 Simplified scheme of the two main protein arrays for quantitative readout: **a** The microarray immuno array carries different antibodies on e.g. glass slides capturing fluorescently labelled proteins from a protein mixture with direct quantitative detection of the fluorescence. **b** The RPPA carries different spots of protein mixtures on e.g. nitrocellulose-coated glass slides. Specific antigens are identified with primary antibodies followed by detection of the primary antibody with a fluorescently labelled secondary antibody

slides for the capture of fluorescently labelled proteins from cell or tissue extracts and direct quantitative detection of the fluorescently labelled target protein. For systems biology modelling, quantitative readout of both protein levels and phosphorylation status using RPPAs have been used for exploring ERBB signalling in combination with other genomics technologies such as mRNA profiling and siRNA approaches (Sahin and Wiemann 2009). On the clinical level, several potential diagnostic markers have been described, such as phospho-AKT for prostate cancer progression (Paweletz et al. 2001) and glioma progression (Jiang et al. 2006). It is conceivable that such approaches will be used more extensively in the future for measuring therapeutic response (Espina et al. 2009; Rapkiewicz et al. 2007). However, these technologies are limited by lack of availability of specific antibodies against protein targets or their phosphorylated variants, and by the need for better standardization techniques in tissue collection and processing. Even with increased availability of good antibody probes, the protein array is likely to provide information limited to defined pathways or at best to large groups of proteins (e.g. kinases, phosphatases). Further improvements and target coverage of this methodology are very necessary, since mRNA and protein profiles are at best complementary. Recently it was shown that about only a third of about 1000 gene products correlate in a comparative study between mRNA and protein expression in 23 cell lines (Gry et al. 2009).

3.4.4 Immunohistochemistry

Protein profiling is instrumental in determining protein expression levels or post-translational modification status in cancer cells and tissues. Laser capture microdissection methods can prepare homogeneous populations of intact cells from sections of heterogeneous tissue and facilitate the proteomic, mutational or mRNA-expression analysis of cancer tissues and cancer cells (Mustafa et al. 2008). Despite these advances, classical cytological examination made by experienced patholo-

gists still provides the most important information on the tissue and disease status (see Chap. 2). On a systems approach level, information on large cohorts of patients and tissues is required. The Human Protein Atlas (HPR) project http://www.proteinatlas.org/ has carried out cytological analysis as well as protein expression and localization studies, in a systematic approach using mono-specific antibodies generated against protein fragments or large peptides (Ponten et al. 2008). These probes are employed for IHC on tissue culture cells, normal and disease related tissues, and provide an exhaustive readout of both protein expression and localization (on a tissue and sub-cellular level), as well as a valuable correlative database on both tissue and cellular level between proteins and diseases. These data complement those mRNA expression profiles (http://www.oncomine.org; http://www.ebi.ac.uk/gxa/) that are used to analyse differences of gene expression in normal or diseased tissue/cells on a genome-wide level.

3.4.5 Phosphoproteome

Protein phosphorylation controls most cellular signalling pathways, and deregulation of kinases and phosphatases has been linked to cancer pathology. Consequently, improved analysis of protein phosphorylation in cancer cells and tissues is critical for our understanding of cancer progression on a systems level. A series of methods have been developed or refined over recent years that employ MS, protein/phospho-specific antibody arrays and 2D gel electrophoresis (White 2008). In order to enhance detection sensitivity, these are combined with enrichment technologies for phosphopeptides such as phosphoramidate chemistry (PAC), immobilized metal affinity chromatography (IMAC), and titanium dioxide prior to analysis or phosphospecific protein stains in the case of gel electrophoresis or mass spectrometry (Bodenmiller et al. 2007). A combination of some of these approaches has been applied to analyse the dynamics of growth-factor stimulation of cellular signalling. For example, EGFR-dependent phosphotyrosine signalling was characterized by using isotope labelling of amino acids and enrichment of tyrosine-phosphorylated protein using immunoprecipitation (Blagoev et al. 2004). In a second study, the effects of increased human epidermal growth factor receptor 2 (HER2) expression in the context of EGF and HRG stimulation showed differential stimulation of cell proliferation or cell migration pathways (Wolf-Yadlin et al. 2006). For our understanding of the precise signalling event, it is important to identify actual phosphorylation sites and relevant peptides within individual proteins and whole pathways. Astonishing progress has been made towards this aim and databases are now available listing actual phospho-sites, see http://www.phosphopep.org/, identified in systematic studies of different cell or tissues. More narrowly defined targets are now available for kinase/phosphatase substrate interaction studies. This information is complemented through bioinformatically calculated phosphorylation sites (http://www.phosphosite.org/).

3.4.6 Metabolome

A quantitative analysis of the metabolic complement in cells or tissues can provide readout of biochemical and physiological activity, representing the potential endpoints of pathways perturbed through deregulatory events or mechanisms inherent in cancer cells (Griffin and Shockcor 2004). Nuclear magnetic resonance (NMR) spectroscopy, gas chromatography (GC), and liquid chromatography MS (LC-MS) have been the main tools in investigating the metabolic changes of biofluids, cells, tissues or organism. While NMR analysis is relatively rapid and works over a broad range of metabolites, MS analysis is several-fold more sensitive but is not equally applicable to all types of metabolites; these differ widely in their capacity to be ionized, a necessary prerequisite for MS analysis (Griffin and Shockcor 2004). Despite technical advances in metabolite identification and separation to improve sensitivity, a challenge remains to obtain a clear distinction between cancer and healthy cells or tissue, and to identify cancer specific biomarkers (Van and Veenstra 2009). A recent analysis of prostate cancer identified six metabolites that are present at higher levels in the cancer tissues using LC/GC-MS (Sreekumar et al. 2009). In particular, sarcosine levels correlated with prostate cancer progression, suggesting that metabolite levels might have a more direct role in contributing to this process. This is further evidence for the interdependencies of complex metabolic feedback loops and cellular signalling, such as has been reported for the tumour suppressor p53 that both responds to metabolic changes but also influences cellular metabolism (Vousden and Ryan 2009). In the future, metabolite markers, in combination with other clinically known markers (as for example PSA), might provide more reliable diagnostic or predictive tools for certain tumour types. Additional studies in the field of metabolomics will be critical for our understanding and modelling of the effect and efficiency of anti-cancer drugs and combinatorial drug application therapies such as cassette dosing (Chung and Griffiths 2007; Smith et al. 2007).

3.5 Functional Studies

Functional studies in basic and in clinical research have so far predominately focused on single-gene analysis and genetic interaction studies. However, to understand complex genetic diseases and networks, a systems biology approach needs to be applied. This can be achieved through genome-wide cellular loss-of-function studies (e.g. through genetic screens or RNA interference experiments) and by combining high-throughput studies, single-gene/protein studies, and modelling approaches. In this context, the genetic background of the cell or organism is critical to understanding the mechanism of the developing phenotypes and cancer progression. Hence, defining the context of gene/protein function on the molecular, organism and environmental levels is becoming more and more important in order to understand and intervene in the progression of disease (Fig. 3.7).

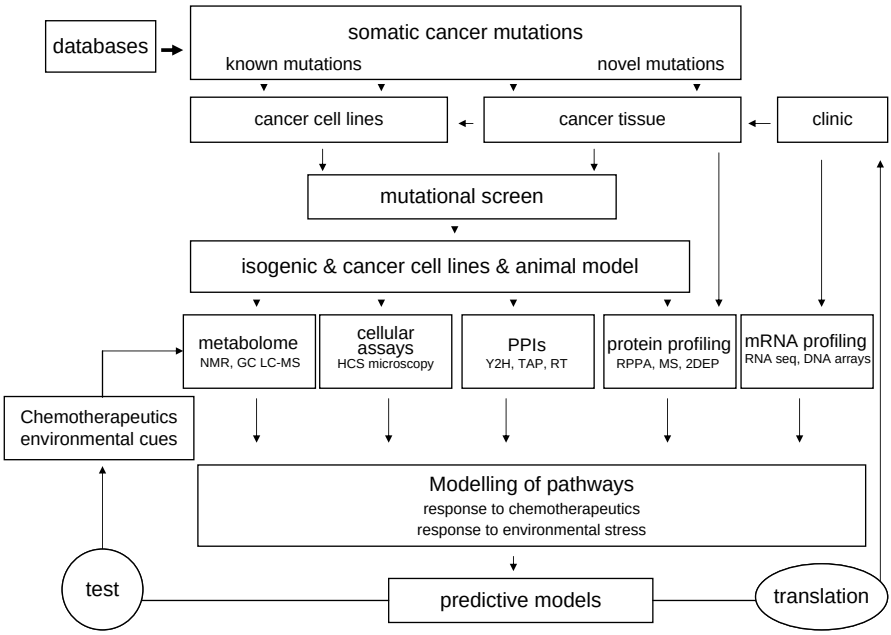


Fig. 3.7 Schematic representation of a systems approach exploring the consequence of cancer mutations on a cellular and organism level. Quantitative data of cellular function, PPI, mRNA & protein profiles, and metabolome data are analysed for pathway modelling. Predictive models are generated to predict the functional consequence of mutations or drug treatment on a systems level, and are tested by modulation of signalling events

3.5.1 RNA Interferences (RNAi)

This technology has revolutionized the exploration of gene function and pathway perturbation for a wide variety of organisms on a whole genome level. Introducing double-stranded (ds) RNA into cells initiates, via the RNA-Induced Silencing Complex (RISC), the degradation of the mature mRNA and or blocks translation, followed by depletion of wild type protein. To exploit this mechanism, several approaches were developed for introducing dsRNAs into cells or organisms, by either bathing cells or organisms in long dsRNA (*Drosophila melanogaster* cells, *Caenorhabditis elegans* worms); by feeding *C. elegans* with bacteria expressing dsRNAs; by viral transduction of hairpin expression constructs; or by transfection of endoribonuclease-derived short-interference RNAs (esiRNAs) into mammalian cells. On this basis, several genome-wide libraries and approaches have been established (Boutros and Ahringer 2008; Kittler et al. 2004; Poulin et al. 2004). This is the flexibility of the RNAi approach, and its feasibility for virtually any type of functional readout through biochemical, microscopy and profiling methodology, that makes this technique so powerful. The combination of these assays with automatic microscopes and robotic support for cellular assays has yielded multi-

parameter phenotypic readouts (also described as high-content analysis) and driven research forward. Major developments and improvements have been made in the field of commercial and not-for-profit image processing software that now allow quantitative and time-resolved readout of complex phenotypic parameters in high-throughput fashion (Boutros et al. 2006; Carpenter et al. 2006; Jones et al. 2009). Results from RNAi screens have facilitated component and pathway analysis on a genome-wide level, for example in cell cycle regulation (Bjorklund et al. 2006), cell division (Kittler et al. 2004), and the Wnt-wingless signalling pathway (DasGupta et al. 2005). Future applications of the RNAi technique on a clinical level might involve therapeutics approaches. These would still need to overcome major hurdles of optimal delivery or tissue specific-expression in diseased tissue (Dykxhoorn and Lieberman 2005).

3.5.2 Model Organisms

Several model organisms such as mice, zebrafish, *C. elegans*, *D. melanogaster* and yeast have played a major role in exploring complex multigenic disorders such as cancer (Spradling et al. 2006). These models have provided information on tissue-specific or developmental aspects of gene-gene and gene-environment interactions, gene/protein function, protein-protein interactions, regulatory networks, and mRNA profiles. Genetic screening has been a major tool to investigate loss-of-function or genetic interactions including the study of cell environment, stem-cell niche and metastasis. A major vertebrate cancer model has been the mouse: it allows efficient genetic manipulation, for example through site-specific recombination and conditional tissue-specific gene expression and silencing (Frese and Tuveson 2007). The model has made many important contributions. One recent example in the field of pancreatic cancer research suggests improved ways for drug delivery (Olive et al. 2009). The mouse has also contributed, as a xenograft model, to our knowledge on tumour progression and environmental factors that influence cancer growth. However, the validity of some of these models has been questioned, for example with respect to the physiological and metabolic variations between mouse and man, and the number and sequence of genetic changes required to achieve cell transformation and metastasis (Frese and Tuveson 2007; Wagner 2004). In addition to the mouse model, zebrafish and other chordates (Feitsma and Cuppen 2008), and non-vertebrate models (*Drosophila, C. elegans*) have been further developed over recent years, and successfully employed to study, for example, cell migration or metastasis (Januschke and Gonzalez 2008; Brumby and Richardson 2005; Poulin et al. 2004). Yeast has contributed greatly to our understanding of cyclin dependent kinases in cell cycle regulation (Hayles and Nurse 2001) and has proved effective in the identifying of drug target screens by haploinsufficiency profiling (Giaever et al. 2004). Such profiling can identify drug targets essential for cell viability, by lowering the gene dosage. This is achieved by replacing one of the two gene copies in diploid yeast through a selectable marker and a molecular barcode. After quantitative read-

out of cell-growth phenotypes, the molecular barcode permits the efficient identification of specific target genes through oligonucleotide arrays or, more recently, through second-generation sequencing approaches (Chan et al. 2009).

3.5.3 Determining Drug and Compound Action

Systems biology approaches now aim at developing predictive models that can be applied to select the most efficient drug treatment for a particular disease or to attempt to predict and reduce drug side-effects. Drug development for cancer treatment has become over the last decade more and more expensive and in some sense, less successful overall (Rothenberg et al. 2003). This fact has spiked a controversial debate on the economic advisability and clinical benefit of drug treatment which sometimes achieves only a relatively small increase in overall patient survival for enormous associated costs (Schmidt 2009). However, systems approaches are now making major inroads in the field of drug discovery and defining mechanisms of drug action. These approaches are now starting to take into account the genetic background of both cancer (somatic) and patient (germ-line), metabolic response, and global protein and mRNA levels. While the genetic landscape of cancer mutations is increasingly better defined, it is now essential to elucidate the functional consequence of these somatic genetic alterations, and to translate this knowledge into targeted therapeutic approaches. Such an approach has been taken, for example, by Sos et al. (2009) who predicted that KRAS mutations in non-small-cell lung cancer (NSCLC) cells confer enhanced Hsp90 dependency, and thus higher susceptibility to a Hsp90 inhibitor. In other approaches, isogenic or knock-in cell lines have been used that differed in their mutational status or presence/absence of EGFR, KRAS, BRAF, and PIK3CA (Di Nicolantonio et al. 2008; Torrance et al. 2001). Such experiments have led to the identification of new compounds and suggested new therapeutic approaches that match particular genetic profiles in individual tumours (McDermott and Settleman 2009). These or similar type of approaches open new avenues for developing therapies using cancer cell collections (Sharma et al. 2010); exploiting the fact that somatic lesions in cancer often create dependencies on activated oncogenes (Sos et al. 2009).

Uncontrolled proliferation of cancer cells is often associated with amplification of the regulatory subunits of cyclin-dependent kinases (CDK) that control multiple steps in cell cycle progression. Consequently, CDKs have been frequently chosen as potential drug targets for cancer therapy. Such compounds may, however, be toxic in normal cells, as a result of down-regulation of essential cellular functions. Nevertheless, recent evidence points to the fact that sensitivity to CDK inhibitors might depend on the individual tumour type, thereby offering new opportunities for the development of individual cancer therapies (Lapenna and Giordano 2009). While these approaches contribute towards understanding of the action of compounds on a pathway and organism level, other more direct approaches are required to identify compound target interactions. With the increased sensitivity and quantitative

approaches provided by MS, chemical proteomics will contribute both to our understanding of drug action and towards the finding of more specific or more potent inhibitors (Bantscheff et al. 2007; Rix and Superti-Furga 2009).

3.5.4 Protein-Protein and Protein-DNA Interactions

Clearly, network information for systems biology modelling needs to include also quantitative and dynamic information on translational and post-translational regulation; binding constants; protein-protein interactions and protein localization (Tucker et al. 2001; Stelzl and Wanker 2006). The analysis of protein-protein interactions has become possible with the availability of open reading frames (ORF) in high-throughput vector systems on a genome-wide level. Model organisms such as yeast have provided genome-wide data sets, for protein-protein interactions based on the yeast-two-hybrid (Y2H) assay (Uetz et al. 2000) and tandem affinity purification (TAP) (Gavin et al. 2006). A large human protein-protein interaction network has been defined (Stelzl et al. 2005; Rual et al. 2005). The advantage of the methodology is the high throughput and the detection of relatively transient protein-protein interactions. However, this assay has problems in detecting membrane-bound protein-protein interactions that might possibly be imported only to a limited extent into the nucleus, as required for the Y2H assay, which has nevertheless, become a standard for the identification of binary protein-protein interaction networks. In contrast, the TAP method identifies protein complexes and relatively more stable complexes, and has been used, for example, to define a functional map of the TNF alpha/NF-κB signal transduction pathway in human cultured cells (Bouwmeester et al. 2004), thereby identifying several new modulators of this pathway. While these approaches provide a two-dimensional network map, additional dynamic and quantitative data are required for systems biology modelling. Such integrated approaches have been carried out using quantitative MS, mRNA profiling and PPI data to model, for example, galactose utilization (Ideker et al. 2001) or cell-cycle regulation in yeast (de Lichtenberg et al. 2005). More dynamic data are also generated by fluorescence-based technologies employing energy resonance transfer based on bioluminescence (BRET) or fluorescence (FRET) in living cells (Ciruela 2008). In addition, protein-protein interactions can be read out dynamically and in high-throughput systems by luminescence in mammalian cells (luminescence-based mammalian interactome mapping = LUMIER). Using this assay, new components were linked to the TGF-beta pathway by screening luciferase-tagged proteins co-expressed with FLAG-tagged proteins, followed by immunoprecipitation and analysis of luciferase activity in a plate reader (Barrios-Rodiles et al. 2005). Last but not least, protein-array technologies have been used for successful investigations of protein-protein interactions in high-throughput fashion (Kung and Snyder 2006). Protein-array technology has also been successfully applied to the analysis of protein-DNA interactions, for example identifying ERK2 as a transcriptional repressor of interferon signalling (Hu et al. 2009).

An alternative and novel approach to the detection of proteins and protein interactions is provided by proximity ligation assays (PLA) (Soderberg et al. 2008), based on the selective ligation of two or more DNA tags on specific binding agents (antibodies, aptamers) in the presence of a template strand added in excess, which can then be selectively amplified by PCR or rolling circle amplification. Since ligation products will only form if the tags are held in close proximity (hence proximity ligation), this approach can be used to detect specific proteins or protein complexes with enormous sensitivity.

3.6 Overall Determining Factors and Future Outlook

Measuring technologies in the field of molecular biology and cellular biology have developed rapidly over the recent years, most conspicuously in the sequencing field where the capacity and throughput has increased by several orders of magnitude. This has been a driving factor in systems biology research which thrives on technologies that facilitate genome-wide readout on DNA, mRNA, protein and metabolome levels. Because data generation has increased exponentially, we are confronted by new challenges in transforming these data into useful models to help predict the outcome of cancer-related genomic aberrations, and to develop novel diagnostic and therapeutic strategies. The challenges are enormous and require completely new types of infrastructure, such as storage that dynamically adapts to volumes of data not seen hitherto. International quality and format standardization is essential for comparing and utilizing data in different modelling approaches. Some of the technology required for high-throughput analysis is only affordable by large institutions. While national and international consortia that share such equipment and data are already established, optimization of networking could be of great benefit to the entire field. Another challenge will be to translate the newly gained knowledge and models into day-to-day medical applications for the ultimate benefit of patients, providing them a personalized systems medicine. Systems biology technologies and modelling will contribute on the level of target detection, diagnostic and therapeutic approaches to diseases. Progress has already begun in these areas, as outlined at the beginning of this chapter. It is now possible to completely analyse the genome and transcriptome of each individual patient, and to generate an exhaustive analysis of the products of transcription. This data, together with clinical, metabolomic and proteomic data, can be used to establish predictive models of effects and side-effects of specific drug combinations in individual patients. Proteomics techniques have also improved rapidly, but analytical methodology and equipment have not yet achieved the capacity to carry out routine whole-proteome analysis of higher organisms. Progress has however been made in the selective detection of specific proteins, based on various principles (quantitative MS techniques, RPPA, PLA). It is to be hoped that many of the new or improved analytical techniques will go on to attain further clinical usefulness.

The mostly two-dimensional structure of current network models is a limitation that needs to be overcome with new graphical models, improved bioinformatic approaches, and standardization. Approaches to standardization are progressing with the Systems Biology Graphical Notation (http://www.sbgn.org) or Systems Biology Markup Language (http://sbml.org/). New models will need to link multivariate phenotypic information with time-resolved data taking into account individual characteristics. Such an integrative approach will benefit more efficiently from current complex experimental data, as obtained by modern molecular investigation, such as the Treat1000 project (http://www.treat1000.org), and International Cancer Genome Consortia (ICGC: http://www.icgc.org), the Cancer Genome Project (CGP: http://www.sanger.ac.uk/genetics/CGP), and The Cancer Genome Atlas (TCGA: cancergenome.nih.gov). These projects have been established with the aim of bringing the benefits of genomic medicine to cancer care. The combination of sequencing and modelling might in the future provide information swiftly enough to allow the sequencing of tumour materials to be analysed in a computer model during the course of a surgical operation, for immediate support to the therapeutic decision-making process.

Acknowledgments We thank our colleagues for comments on text and discussions, in particular M. Ralser for his extensive comments and suggestions on the MS part of the manuscript and C. Wierling for contributing Fig. 3.2. Our work is funded through: **BL:** NGFN IG Cellular Systems Biology, IG Neuronet, IG Mutanom; **MS:** IG Mutanom, IG Intestinal Modifiers; **HL:** Max Planck Society

References

Anderson NL, Anderson NG (2002) The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics 1:845–867

Ball MP, Li JB, Gao Y, Lee JH, LeProust EM, Park IH, Xie B, Daley GQ, Church GM (2009) Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nat Biotechnol 27:361–368

Banerjee HN, Verma M (2009) Epigenetic mechanisms in cancer. Biomark Med 3:14

Bantscheff M, Eberhard D, Abraham Y, Bastuck S, Boesche M, Hobson S, Mathieson T, Perrin J, Raida M, Rau C et al (2007) Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. Nat Biotechnol 25:1035–1044

Barrios-Rodiles M, Brown KR, Ozdamar B, Bose R, Liu Z, Donovan RS, Shinjo F, Liu Y, Dembowy J, Taylor IW et al (2005) High-throughput mapping of a dynamic signaling network in mammalian cells. Science 307:1621–1625

Beck S, Rakyan VK (2008) The methylome: approaches for global DNA methylation profiling. Trends Genet 24:231–237

Bian YS, Yan P, Osterheld MC, Fontolliet C, Benhattar J (2001) Promoter methylation analysis on microdissected paraffin-embedded tissues using bisulfite treatment and PCR-SSCP. Biotechniques 30:66–72

Biron DG, Brun C, Lefevre T, Lebarbenchon C, Loxdale HD, Chevenet F, Brizard JP, Thomas F (2006) The pitfalls of proteomics experiments without the correct use of bioinformatics tools. Proteomics 6:5577–5596

Bjorklund M, Taipale M, Varjosalo M, Saharinen J, Lahdenpera J, Taipale J (2006) Identification of pathways regulating cell size and cell-cycle progression by RNAi. Nature 439:1009–1013
Blagoev B, Ong SE, Kratchmarova I, Mann M (2004) Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. Nat Biotechnol 22:1139–1145
Blow N (2008) DNA sequencing: generation next-next. Nat Methods 5:267–274
Bodenmiller B, Mueller LN, Mueller M, Domon B, Aebersold R (2007) Reproducible isolation of distinct, overlapping segments of the phosphoproteome. Nat Methods 4:231–237
Boutros M, Ahringer J (2008) The art and design of genetic screens: RNA interference. Nat Rev Genet 9:554–566
Boutros M, Bras LP, Huber W (2006) Analysis of cell-based RNAi screens. Genome Biol 7:R66
Bouwmeester T, Bauch A, Ruffner H, Angrand PO, Bergamini G, Croughton K, Cruciat C, Eberhard D, Gagneur J, Ghidelli S et al (2004) A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway. Nat Cell Biol 6:97–105
Brumby AM, Richardson HE (2005) Using Drosophila melanogaster to map human cancer pathways. Nat Rev Cancer 5:626–639
Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, Guertin DA, Chang JH, Lindquist RA, Moffat J et al (2006) CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol 7:R100
Chan JN, Nislow C, Emili A (2009) Recent advances and method development for drug target identification. Trends Pharmacol Sci 31:82–88
Charrier JP, Tournel C, Michel S, Comby S, Jolivet-Reynaud C, Passagot J, Dalbon P, Chautard, D, Jolivet M (2001) Differential diagnosis of prostate cancer and benign prostate hyperplasia using two-dimensional electrophoresis. Electrophoresis 22:1861–1866
Chung YL, Griffiths JR (2007) Using metabolomics to monitor anticancer drugs. Ernst Schering Found Symp Proc (4):55–78
Ciruela F (2008) Fluorescence-based methods in the study of protein-protein interactions in living cells. Curr Opin Biotechnol 19:338–343
Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H (2009) Continuous base identification for single-molecule nanopore DNA sequencing. Nat Nanotechnol 4:265–270
Cross SH, Charlton JA, Nan X, Bird AP (1994) Purification of CpG islands using a methylated DNA binding column. Nat Genet 6:236–244
Dang L, White DW, Gross S, Bennett BD, Bittinger MA, Driggers EM, Fantin VR, Jang HG, Jin S, Keenan MC, Marks KM, Prins RM, Ward PS, Yen KE, Liau LM, Rabinowitz JD, Cantley LC, Thompson CB, Vander Heiden MG, Su SM (2009) Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. Nature 462:739–744
DasGupta R, Kaykas A, Moon RT, Perrimon N (2005) Functional genomic analysis of the Wnt-wingless signaling pathway. Science 308:826–833
de Lichtenberg U, Jensen LJ, Brunak S, Bork P (2005) Dynamic complex formation during the yeast cell cycle. Science 307:724–727
Di Nicolantonio F, Arena S, Gallicchio M, Zecchin D, Martini M, Flonta SE, Stella GM, Lamba S, Cancelliere C, Russo M et al (2008) Replacement of normal with mutant alleles in the genome of normal human cells unveils mutation-specific drug responses. Proc Natl Acad Sci U S A 105:20864–20869
Down TA, Rakyan VK, Turner DJ, Flicek P, Li H, Kulesha E, Graf S, Johnson N, Herrero J, Tomazou EM et al (2008) A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. Nat Biotechnol 26:779–785
Dykxhoorn DM, Lieberman J (2005) The silent revolution: RNA interference as basic biology, research tool, and therapeutic. Annu Rev Med 56:401–423
Espina V, Liotta LA, Petricoin EF 3rd (2009) Reverse-phase protein microarrays for theranostics and patient tailored therapy. Methods Mol Biol 520:89–105
Feinberg AP (2007) Phenotypic plasticity and the epigenetics of human disease. Nature 447 433–440
Feitsma H, Cuppen E (2008) Zebrafish as a cancer model. Mol Cancer Res 6:685–694
Frese KK, Tuveson DA (2007) Maximizing mouse cancer models. Nat Rev Cancer 7:645–658

Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dumpelfeld B et al (2006) Proteome survey reveals modularity of the yeast cell machinery. Nature 440:631–636

Gentleman R, Huber W (2007) Making the most of high-throughput protein-interaction data. Genome Biol 8:112

Giaever G, Flaherty P, Kumm J, Proctor M, Nislow C, Jaramillo DF, Chu AM, Jordan MI, Arkin AP, Davis RW (2004) Chemogenomic profiling: identifying the functional interactions of small molecules in yeast. Proc Natl Acad Sci U S A 101:793–798

Greenleaf WJ, Block SM (2006) Single-molecule, motion-based DNA sequencing using RNA polymerase. Science 31:801

Griffin JL, Shockcor JP (2004) Metabolic profiles of cancer cells. Nat Rev Cancer 4:551–561

Gry M, Rimini R, Stromberg S, Asplund A, Ponten F, Uhlen M, Nilsson P (2009) Correlations between RNA and protein expression profiles in 23 human cell lines. BMC Genomics 10:365

Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol 17:994–999

Hayles J, Nurse P (2001) A journey into space. Nat Rev Mol Cell Biol 2:647–656

Hopfgartner G, Varesio E, Tschappat V, Grivet C, Bourgogne E, Leuthold LA (2004) Triple quadrupole linear ion trap mass spectrometer for the analysis of small molecules and macromolecules. J Mass Spectrom 39:845–855

Hu S, Xie Z, Onishi A, Yu X, Jiang L, Lin J, Rho HS, Woodard C, Wang H, Jeong JS et al (2009) Profiling the human protein-DNA interactome reveals ERK2 as a transcriptional repressor of interferon signaling. Cell 139:610–622

Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science 292:929–934

Januschke J, Gonzalez C (2008) Drosophila asymmetric division, polarity and cancer. Oncogene 27:6994–7002

Jiang R, Mircean C, Shmulevich I Cogdell D, Jia Y, Tabus I, Aldape K, Sawaya R, Bruner JM, Fuller GN et al (2006) Pathway alterations during glioma progression revealed by reverse phase protein lysate arrays. Proteomics 6:2964–2971

Jones TR, Carpenter AE, Lamprecht MR, Moffat J, Silver SJ, Grenier JK, Castoreno AB, Eggert US, Root DE, Golland P et al (2009) Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning. Proc Natl Acad Sci U S A 106:1826–1831

Kerjean A, Vieillefond A, Thiounn N, Sibony M, Jeanpierre M, Jouannet P (2001) Bisulfite genomic sequencing of microdissected cells. Nucleic Acids Res 29:E106–E106

Keshet I, Schlesinger Y, Farkash S, Rand E, Hecht M, Segal E, Pikarski E, Young RA, Niveleau A, Cedar H et al (2006) Evidence for an instructive mechanism of de novo methylation in cancer cells. Nat Genet 38:149–153

Kittler R, Putz G, Pelletier L, Poser I, Heninger AK, Drechsel D, Fischer S, Konstantinova I, Habermann B, Grabner H et al (2004) An endoribonuclease-prepared siRNA screen in human cells identifies genes essential for cell division. Nature 432:1036–1040

Korf U, Henjes F, Schmidt C, Tresch A, Mannsperger H, Lobke C, Beissbarth T, Poustka A (2008) Antibody microarrays as an experimental platform for the analysis of signal transduction networks. Adv Biochem Eng Biotechnol 110:153–175

Korf U, Löbke C, Sahin O, Haller F, Sültmann H, Arlt D, Poustka A (2009) Reverse-phase protein arrays for application-orientated cancer research. Proteomics—Clinic Appl 3:1140–1150

Korshunova Y, Maloney RK, Lakey N, Citek RW, Bacher B, Budiman A, Ordway JM, McCombie WR, Leon J, Jeddeloh JA et al (2008) Massively parallel bisulphite pyrosequencing reveals the molecular complexity of breast cancer-associated cytosine-methylation patterns obtained from tissue and serum DNA. Genome Res 18:19–29

Kulasingam V, Diamandis EP (2008). Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. Nat Clin Pract Oncol 5:588–599

Kung LA, Snyder M (2006) Proteome chips for whole-organism assays. Nat Rev Mol Cell Biol 7:617–622
Kuster B, Schirle M, Mallick P, Aebersold R (2005) Scoring proteomes with proteotypic peptide probes. Nat Rev Mol Cell Biol 6:577–583
Kuzyk MA, Smith D, Yang J, Cross TJ, Jackson AM, Hardie DB, Anderson NL, Borchers CH (2009) Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma. Mol Cell Proteomics 8:1860–1877
Lapenna S, Giordano A (2009) Cell cycle kinases as therapeutic targets for cancer. Nat Rev Drug Discov 8:547–566
Lander ES et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921
Lee IN, Chen CH, Sheu JC, Lee HS, Huang GT, Yu CY, Lu FJ, Chow LP (2005) Identification of human hepatocellular carcinoma-related biomarkers by two-dimensional difference gel electrophoresis and mass spectrometry. J Proteome Res 4:2062–2069
Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM et al (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462:315–322
Mann M (2006) Functional and quantitative proteomics using SILAC. Nat Rev Mol Cell Biol 7:952–958
McDermott U, Settleman J (2009) Personalized cancer therapy with selective kinase inhibitors: an emerging paradigm in medical oncology. J Clin Oncol 27:5650–5659
Morozova O, Hirst M, Marra MA (2009) Applications of new sequencing technologies for transcriptome analysis. Annu Rev Genomics Hum Genet 10:135–151
Mustafa D, Kros JM, Luider T (2008) Combining laser capture microdissection and proteomics techniques. Methods Mol Biol 428:159–178
Nishizuka S, Spurrier B (2008) Experimental validation for quantitative protein network models. Curr Opin Biotechnol 19:41–49
Olive KP, Jacobetz MA, Davidson CJ, Gopinathan A, McIntyre D, Honess D, Madhu B, Goldgraben MA, Caldwell ME, Allard D et al (2009) Inhibition of Hedgehog signaling enhances delivery of chemotherapy in a mouse model of pancreatic cancer. Science 324:1457–1461
Paweletz CP, Charboneau L, Bichsel VE, Simone NL, Chen T, Gillespie JW, Emmert-Buck MR, Roth MJ, Petricoin IE, Liotta LA (2001) Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. Oncogene 20:1981–1989
Picotti P, Bodenmiller B, Mueller LN, Domon B, Aebersold R (2009) Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics. Cell 138:795–806
Ponten F, Jirstrom K, Uhlen M (2008) The human protein atlas—a tool for pathology. J Pathol 216:387–393
Poulin G, Nandakumar R, Ahringer J (2004) Genome-wide RNAi screens in Caenorhabditis elegans: impact on cancer research. Oncogene 23:8340–8345
Ragoussis J (2009) Genotyping technologies for genetic research. Annu Rev Genomics Hum Genet 10:117–133
Rakyan VK, Down TA, Thorne NP, Flicek P, Kulesha E, Graf S, Tomazou EM, Backdahl L, Johnson N, Herberth M et al (2008) An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). Genome Res 18:1518–1529
Rapkiewicz A, Espina V, Zujewski JA, Lebowitz PF, Filie A, Wulfkuhle J, Camphausen K, Petricoin EF 3rd, Liotta LA, Abati A (2007) The needle in the haystack: application of breast fine-needle aspirate samples to quantitative protein microarray technology. Cancer 111:173–184
Reymond MA, Schlegel W (2007) Proteomics in cancer. Adv Clin Chem 44:103–142
Rix U, Superti-Furga G (2009) Target profiling of small molecules by chemical proteomics. Nat Chem Biol 5:616–624
Rothenberg ML, Carbone DP, Johnson DH (2003) Improving the evaluation of new cancer treatments: challenges and opportunities. Nat Rev Cancer 3:303–309

Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N et al (2005) Towards a proteome-scale map of the human protein-protein interaction network. Nature 437:1173–1178

Rusk N (2009) Cheap third-generation sequencing. Nat Methods 6:244

Sahin O, Wiemann S (2009) Functional genomics and proteomics approaches to study the ERBB network in cancer. FEBS Lett 583:1766–1771

Schiess R, Wollscheid B, Aebersold R (2009) Targeted proteomic strategy for clinical biomarker discovery. Mol Oncol 3:33–44

Schmidt C (2009) Costly cancer drugs trigger proposals to modify clinical trial design. J Natl Cancer Inst 101:1662–1664

Schweiger MR, Kerick M, Timmermann B, Albrecht MW, Borodina T, Parkhomchuk D, Zatloukal K, Lehrach H (2009) Genome-wide massively parallel sequencing of formaldehyde fixed-paraffin embedded (FFPE) tumor tissues for copy-number- and mutation-analysis. PLoS ONE 4:e5548

Sharma SV, Haber DA, Settleman J (2010) Cell line-based platforms to evaluate the therapeutic efficacy of candidate anticancer agents. Nat Rev Cancer 10:241–253

Smith NF, Raynaud FI, Workman P (2007) The application of cassette dosing for pharmacokinetic screening in small-molecule cancer drug discovery. Mol Cancer Ther 6:428–440

Soderberg O, Leuchowius KJ, Gullberg M, Jarvius M, Weibrecht I, Larsson LG, Landegren U (2008) Characterizing proteins and their interactions in cells and tissues using the in situ proximity ligation assay. Methods 45:227–232

Sos ML, Michel K, Zander T, Weiss J, Frommolt P, Peifer M, Li D, Ullrich R, Koker M, Fischer F et al (2009) Predicting drug susceptibility of non-small cell lung cancers based on genetic lesions. J Clin Invest 119:1727–1740

Spradling A, Ganetsky B, Hieter P, Johnston M, Olson M, Orr-Weaver T, Rossant J, Sanchez A, Waterston R (2006) New roles for model genetic organisms in understanding and treating human disease: report from the 2006 Genetics Society of America meeting. Genetics 172:2025–2032

Sreekumar A, Poisson LM, Rajendiran TM, Khan AP, Cao Q, Yu J, Laxman B, Mehra R, Lonigro RJ, Li Y et al (2009) Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. Nature 457:910–914

Stamey TA, Yang N, Hay AR, McNeal JE, Freiha FS, Redwine E (1987) Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate. N Engl J Med 317:909–916

Stelzl U, Wanker EE (2006) The value of high quality protein-protein interaction networks for systems biology. Curr Opin Chem Biol 10:551–558

Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S et al (2005) A human protein-protein interaction network: a resource for annotating the proteome. Cell 122:957–968

Sugiyama S, Yoshino T, Tsukamoto K, Sasou M, Kuwazaki S, Takahashi H, Suetsugu Y, Narukawa J, Yamamoto K, Ohtani T (2006) Application of scanning probe microscopy to genetic analysis. Jpn J Appl Phys 45:2305–2309

Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, Seifert M, Borodina T, Soldatov A, Parkhomchuk D et al (2008) A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science 321:956–960

Taylor KH, Kramer RS, Davis JW, Guo J, Duff DJ, Xu D, Caldwell CW, Shi H (2007) Ultradeep bisulfite sequencing analysis of DNA methylation patterns in multiple gene promoters by 454 sequencing. Cancer Res 67:8511–8518

Torrance CJ, Agrawal V, Vogelstein B, Kinzler KW (2001) Use of isogenic human cancer cells for high-throughput screening and drug discovery. Nat Biotechnol 19:940–945

Tucker CL, Gera JF, Uetz P (2001) Towards an understanding of complex protein networks. Trends Cell Biol 11:102–106

Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P et al (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature 403:623–627

Vijver MJ van de, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ et al (2002) A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 347:1999–2009

Van QN, Veenstra TD (2009) How close is the bench to the bedside? Metabolic profiling in cancer research. Genome Med 1:5

Vousden KH, Ryan KM (2009) p53 and metabolism. Nat Rev Cancer 9:691–700

Wagner KU (2004) Models of breast cancer: quo vadis, animal modeling? Breast Cancer Res 6:31–38

Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, Schubeler D (2005) Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet 37:853–862

Whiteaker JR, Zhao L, Anderson L, Paulovich AG (2010) An automated and multiplexed method for high throughput peptide immunoaffinity enrichment and multiple reaction monitoring mass spectrometry-based quantification of protein biomarkers. Mol Cell Proteomics 9:184–196

White FM (2008) Quantitative phosphoproteomic analysis of signaling network dynamics. Curr Opin Biotechnol 19:404–409

Williams C, Brunskill S, Altman D, Briggs A, Campbell H, Clarke M, Glanville J, Gray A, Harris A, Johnston K, Lodge M (2006) Cost-effectiveness of using prognostic information to select women with breast cancer for adjuvant systemic therapy. Health Technol Assess 10:iii–iv, ix–xi, 1–204

Wolf-Yadlin A, Kumar N, Zhang Y, Hautaniemi S, Zaman M, Kim HD, Grantcharova V, Lauffenburger DA, White FM (2006) Effects of HER2 overexpression on cell signaling networks governing proliferation and migration. Mol Syst Biol 2:54

Yates JR, Cociorva D, Liao L, Zabrouskov V (2006) Performance of a linear ion trap-orbitrap hybrid for peptide analysis. Anal Chem 78:493–500

Zhou M, Veenstra TD (2007) Proteomic analysis of protein complexes. Proteomics 7:2688–2697

Zhu W, Smith JW, Huang CM (2010) Mass spectrometry-based label-free quantitative proteomics. J Biomed Biotech 840518

Chapter 4
Cell Lines, Tissue Samples, Model Organisms, and Biobanks: Infrastructure and Tools for Cancer Systems Biology

Sandra Tomaszek and Dennis A. Wigle

Abstract Despite significant advances in the understanding of cancer, we have seen hitherto only limited translation into improvements in diagnosis and treatment of patients. Further advances in targeted therapies are dependent on well-characterized cell lines for drug evaluation and testing as a preclinical entry point into the pharmaceutical development pipeline. Model organism systems to investigate cancer as a living, breathing organism functioning in three-dimensional space will increase in importance for testing novel therapeutics. Cell lines and xenograft systems have significant limitations. High-quality annotated banked tumour tissue is of fundamental importance for the investigation of cancer biomarkers, molecular pathways, and networks. These components are key parts of systems biology strategies for understanding and individualizing the treatment of human cancer.

4.1 Introduction

Decades of focused cancer research have demonstrated the oncogenic process to be frustratingly complex. Despite many triumphs in scientific and clinical understanding, we still do not comprehend the formation of most solid tumours at a basic level. This has hampered improvements in detection, diagnosis, and treatment strategies. The sequencing of the human genome was widely anticipated for the contributions it would make toward understanding human evolution, the causation of disease, and the interplay between the environment and heredity in defining the human condition. The subsequent expected pace of discovery has been slow, and translation into benefit for the clinical management of cancer patients has not yet come to fruition. The current landslide of genomic-based technologies for the molecular detection and diagnosis of cancer has yet to be clinically applied. Our fundamental understanding of the biology of cancer remains poor. Other than for a handful of notable exceptions, the rate of development and application of novel therapeutics has not appreciably changed in the post-genomic era.

Despite these facts, dramatic changes in clinical cancer management are beginning to appear on the horizon as a consequence of human genome sequencing and

S. Tomaszek (✉)
Department of Thoracic Surgery, University Hospital Zurich, Raemistr. 101, 8091 Zurich, Switzerland
e-mail: sandra.tomaszek@usz.ch

A. Cesario, F. B. Marcus (eds.), *Cancer Systems Biology, Bioinformatics and Medicine,*
DOI 10.1007/978-94-007-1567-7_4,

associated technology development. Molecular sub-staging for many tumour types is approaching clinical reality, with information from mutation analysis of specific genes being incorporated into clinical decision-making. The pipeline of novel chemotherapeutics is full of promising new classes of agents with the potential for use in a patient-specific manner based on molecular sub-staging. It is an exciting time for translational and clinical cancer research.

One of the major bottlenecks for advances in treating patients with cancer lies in the appropriate modelling of the disease. Many of the advances that have been made in biological understanding and novel therapeutics are the result of model systems to test hypotheses in a high-throughput manner prior to any application in human clinical trials. As the power of molecular diagnostics continues to increase, the importance of available tissue banks with well-annotated clinical information for testing biomarker hypotheses becomes paramount for incorporating novel biomarkers into clinical trials and clinical practice. Advances in targeted therapies are dependent on well-characterized cell lines for drug evaluation and testing as a preclinical entry point into the pharmaceutical development pipeline. Model organism systems to investigate cancer as a living, breathing organism functioning in three-dimensional space will increase in importance for testing novel therapeutics, given the known limitations of the available cell lines and xenograft systems. All of these components are key parts of the systems biology strategy for understanding and individualizing the treatment of human cancer.

4.2 Human Cell Lines

Human cell lines derived from normal and cancerous tissue have to date played a key role in our understanding of the biology of cancer. Many of the 'usual suspects' in molecular oncology and their functional role in cancer such as K-ras, p53, and others, were originally discovered using cell-line approaches (Crawford et al. 1981; Der et al. 1982). In general, cell lines can be established from human normal tissue or blood, adult or embryonic stem cells, and human tumour or diseased tissue. A number of different types of cell lines are available, including primary cell cultures, propagated secondary cultures and immortalized cell lines.

Primary cell cultures are those established from fresh tissue taken either by percutaneous biopsy or at the time of surgical resection. They are typically representative of cellular phenotypes *in vivo* depending on the culture conditions employed, expressing characteristics which may not be observed in propagated secondary cultures. They are, however, limited in their ability to produce multiple generations, as they frequently only survive for one or a limited number of passages before dying or changing dramatically in phenotype. Many of these culture types are very difficult to initiate and propagate, and are frequently dependent on some form of cell separation, either by fluorescence-activated cell sorting (FACS) or differential culture in order to obtain the cell types of interest.

On the other hand, secondary cultures for propagation through multiple passages divide and grow for an extended amount of time, typically 50–100 generations when provided with adequate culture conditions. Such cells, along with immortalized lines, make up the bulk of the cancer cell lines available through many sources, such as the NCI-60 panel described below. These lines have been the workhorse of thousands of publications investigating cancer mechanisms and novel therapeutics. They fail, however to represent tumour-stromal interactions, how a cancer functions in three dimensional space, metastatic potential, and other such processes that are limited to *in vivo* or complex *in vitro* surrogate models. Many have also questioned the degree to which tumour cell lines maintain their spectrum of mutations and other genomic alterations over time in culture conditions. That being said, cell lines can be incredibly useful for systems involving high-throughput screening where feasible endpoints are being observed, such as cell growth or death.

Immortalized cells, in contrast, will typically continue to grow and divide indefinitely *in vitro*. Immortal cell lines can be created by induction of oncogenes or loss of tumour suppressor genes. HeLa cells represent a classic example of an immortalized cell line, and are one of the oldest and most commonly used human cell lines. Originally derived from cervical cancer cells taken from a patient named Henrietta Lacks, who eventually died of her cancer on October 4, 1951, the cell line continues today to be a robust and durable *in vitro* model (Masters 2002). The cells resemble human epithelial cells obtained from cervical carcinoma transformed by human papilloma virus 18.

Another prominent example is the Jurkat cell line, an immortalized line of T lymphocyte cells that are used to study acute T-cell leukaemia and T-cell signalling. The line was originally established in the late 1970s from the peripheral blood of a 14-year-old boy with T cell leukaemia (Schneider et al. 1977).

Many common cell lines are freely available (at cost of shipping and maintenance) from non-profit registries such as the American Type Culture Collection (ATCC) (https://www.atcc.org) or The European Collection of Cell Cultures (ECACC) (http://www.hpacultures.org.uk). These organizations provide cell lines along with clinical background information on the patients from whom the tissue has been obtained. However, limitations of the cell lines provided are that they rely on accuracy of clinical and histological information provided by the investigators, in many cases lacking follow-up clinical information on the patients; frequently the information is not updated according to changing current diagnostic guidelines.

4.2.1 The NCI-60 Human Cancer Cell Line Panel

In the late 1980s the US National Cancer Institute (NCI) assembled 60 human cancer cell lines representing different tumour types including leukaemia, non-small cell lung cancer, melanoma, central nervous system, renal, colon, ovarian, breast and prostate cancer (Grever 1992; Shoemaker 2006). This panel of immortalized cancer cell lines is the most extensively characterized cancer cell line set to date

(e.g. Roeschke et al. 2003; Shoemaker et al. 1991; Arguello et al. 1996; McLemore et al. 1987; Weinstein et al. 1997). The panel was initially established as an *in vitro* drug-discovery tool. This effort to screen molecules with anti-proliferative activity on a large scale has developed into a powerful tool to identify the relationship between anticancer drug cytotoxicity and molecular cell characteristics, namely the NCI-60 anticancer drug screen (http://www.dtp.nci.nih.gov). Approximately 85,000 compounds have been tested for *in vitro* anticancer activity since the initiation of the NCI-60 project. Furthermore, because of the extensive pharmacological characterization of NCI-60 cell lines, the lines have been frequently used as test samples for novel technologies. The availability of advanced molecular methods has enabled molecular specifics of these cancer cell lines to be characterized (Ross et al. 2000; Scherf et al. 2000). Gene and protein expression patterns as well as copy number variation have been integrated with the pharmacological characteristics of the cell lines, validating the impact of these interactions on the chemosensitivity of cancer cells (Dan et al. 2002; Wallqvist et al. 2002; Bussey et al. 2006; Wang et al. 2006; Shankavaram et al. 2007; Ring et al. 2008; Nissen et al. 2009; Holbeck et al. 2010). Park and colleagues further analysed the relationship between the activation status of protein networks and mutations in NCI-60 cell lines, and identified protein clusters which are associated with chemosensitivity to standard anticancer agents. The functional status of the predicted protein associations was assessed and confirmed in a whole-genome small interfering-RNA library synthetic lethality screen (Park et al. 2010).

Furthermore, with increasing knowledge of the implication of microRNAs in the pathogenesis of cancer as they regulate gene expression post-transcriptionally, microRNAs were investigated in NCI-60 cell lines (Shell et al. 2007; Gaur et al. 2007). Gaur and colleagues compared the microRNA expression in NCI-60 cells to that in normal tissues, identifying specific microRNAs which correlated with the proliferation indices of the NCI-60 cell lines (Gaur et al. 2007). Bignell and colleagues analysed 746 cancer cell lines (NCI-60 and many other cell lines used in studies of cancer biology or cancer drug sensitivity) with respect to copy number variations and derived structural signatures distinguishing between homozygous deletions over recessive cancer genes and fragile sites (Bignell et al. 2010). They identified 2428 somatic homozygous deletions within the 746 cancer cell lines, overlying 11% of protein-coding genes. This kind of complete large-scale genome sequencing will facilitate the identification of driver mutations for further functional studies. Similarly, researchers at the Wellcome Trust Sanger Institute (http://www.sanger.ac.uk), one of the leading centres of the human genome project (International Human Genome Sequencing Consortium 2004), identified new candidate cancer genes in 598 human cancer cell lines (Mattison et al. 2010) through comparative genomic hybridization. In the NCI-60 panel, the sequence analysis of 24 known cancer genes and an assessment of 4 of the 24 genes for homozygous deletions has been reported (Ikediobi et al. 2006). No less than 137 oncogenic mutations were identified in 14 (APC, BRAF, CDKN2, CTNNB1, HRAS, KRAS, NRAS, SMAD4, PIK3CA, PTEN, RB1, STK11, TP53, and VHL) of the 24 genes. All cell lines had at least one mutation among the cancer genes examined, with

most lines (73%) having multiple mutations. Identification of mutated cancer genes in the NCI-60 panel, in combination with pharmacological and molecular profiles of the cells, will clearly enhance the use of the NCI-60 cell lines for molecularly targeted screens.

4.3 Model Organisms

Although the past few decades have seen great strides in cancer research, the molecular pathogenesis of most solid tumours from many tissue types remains largely undefined. Most of what we know about the molecular steps involved in cancer formation comes from defined genetic manipulations in the mouse and other model organisms (here the words model and modelling clearly refer to the use of living organisms, not numerical models). In many types of cancer, the lack of defined models has hampered our understanding of disease progression and potential therapeutic strategies. Such models are essential tools for facilitating the development of new therapies.

As an example, we will focus on models of lung cancer to illustrate issues around the modelling of cancer. Lung cancer continues to be the leading cause of cancer-related death worldwide (Kerr 2001; Cancer Facts 2009). Despite aggressive local and systemic therapies (Johnston et al. 2001), the majority of patients succumb to progressive metastatic disease. The molecular steps involved in the pathogenesis of lung cancer unfortunately remain not defined and elusive. Non-small-cell lung cancer, a subset of lung cancer, is characterized by its aggressive biology and heterogeneity in clinical outcome. Humans are one of only a few species susceptible to developing spontaneous lung cancer. Lung tumours in domestic animals are periodically observed by veterinarians, but Livingood's histologic description 100 years ago of a papillary tumour in a mouse (Livingood 1986) initiated the idea of using animals as experimental model systems. Currently, several types of animal models have been developed for experimental lung cancer research. These include transgenic mouse models, chemically induced lung tumours, and human tumour xenografts.

The biology of cancer is rapidly emerging as one of the most difficult systems biology problems. The myriad of genetic alterations and their phenotypic outputs creates an exceptionally complex picture to analyse and dissect through a reductionist viewpoint. Cancer models that accurately reflect these changes are difficult to generate. It is unlikely that a single model system can faithfully reflect the whole process of cancer development and progression, and as a consequence this requires us to interpret results from model systems with caution. However, appropriate use of available model systems, with an appropriate understanding of their limitations, provides a valuable and necessary tool for study of malignant transformation.

Many aspects of experimental cancer research require the use of animal model systems to reflect the true system context of oncogenesis *in vivo*. Tumour-host interactions including immunologic effects, vascular and stromal effects, and host-re-

lated pharmacologic and pharmacokinetic effects, are examples of systems that are generally poorly modelled *in vitro*. Several studies have shown that lung tumours developed in mice or rats are quite similar in histological morphology and molecular characteristics to human lung cancer (Malkinson 1992; Howard et al. 1999). In general, the spontaneous or chemically induced tumour models that are either idiopathic or arise following a carcinogenic stimulus (Corbett et al. 1975, 1984) most closely mimic the true clinical situation. Unfortunately, these tumours are usually measurable only late in their course, their metastatic pattern is not uniform, and their response to therapy is generally poor.

Transplanted animal tumour models and human tumour xenografts are widely used in experimental therapeutics. Since malignant cells or tissue are directly inoculated into the host animal, effects on early events, such as initiation and carcinogenesis cannot be studied with these models. However, tumour growth, invasion and metastasis are amenable for investigation, since tumour development uniformly follows inoculation with predictable growth and metastatic patterns. Testing of new therapeutic approaches and imaging strategies is particularly well suited to these models.

4.3.1 Transgenic Mouse Models

Transgenic technology has allowed for the development of mouse models for lung cancer. The mouse is the only genetically tractable model organism with lungs similar in structure and function to humans, and the only model organism that develops cancers of similar histopathologies to those seen clinically. The ability to target regulatory genes to the lungs in a cell-specific fashion is feasible with modern gene transfer technologies. These genetically engineered mouse lung cancer models can be exploited to define the molecular events that contribute to the pathogenesis of this disease.

Transgenic mouse technology has proved extremely useful for the creation of models of tumour development, cloning immortalized cellular sub-populations, and testing experimental therapeutic approaches (Adams and Cory 1991; Fowlis and Balmain 1993; Thomas and Balkwill 1995). The ability to integrate a gene of interest into the genome of an animal provides a novel approach for cancer investigation. Gene transfection can be achieved with microinjection (Brinster et al. 1985; Gordon and Ruddle 1983), retroviral infection, or embryonic stem cell transfer (Jaenisch 1980; Jaenisch et al. 1981; Jähner and Jaenisch 1980; Soriano and Jaenisch 1986). Transgenic mice are excellent models for studying the consequences of oncogene expression in animals, the effect of oncogenes on growth and differentiation, and their potential for cellular transformation.

Transgenic mice also provide an *in vivo* preclinical model for gene therapy and gene transfer. An example of this technique as applied to drug development is the introduction of the multiple drug resistance (mdr-1) gene into transgenic animals (Galski et al. 1989). The mdr-1 gene, which is expressed in marrow stem cells, protects cancer cells from damage by extruding cytotoxic chemotherapeutic agents

from the cell and confers *in vivo* resistance to drug toxicity in the whole animal. Such animal models have the potential for identifying agents, or combinations of agents, which are nontoxic to the animal but inhibit the function of the mdr-1 gene or its product, and may also potentially reverse the resistance phenotype.

Transgenic models capable of inducing lung cancer have also been developed. When mutated K-ras, p53 or SV40 T antigen are used as transgenes and integrated into the host genome, lung tumours develop in mice soon after birth and result in early death of the animal. These genes may be non-specifically expressed throughout the body or linked to lung-specific promoters, so that their expression is selectively targeted to non-ciliated Clara cells or alveolar type II pneumocytes (Suda et al. 1987; Maronpot et al. 1991; Wikenheiser et al. 1992; Sandmöller et al. 1995). Although these animals have been used to a limited extent to investigate the molecular events involved in the progression of lung cancer, the rapid and early onset of cancer makes investigation of the early events involved in cancer development difficult (Zhao et al. 2000).

The field of transgenic technology has now evolved to allow investigators more control over specific transgenes. Bitransgenic systems are the most effective gene regulatory systems for transgenic mice, with the tetracycline-based regulatory system (Shockett and Schatz 1996) being that most commonly used. This system, which is under the control of elements responsive to tetracycline or its analogues, has at least two advantages over conventional transgenic mice. First, the transgene can, in principle, be turned on at any time, and thus resembles a somatic mutation. Second, regulated loss of expression (turning off the transgene) can be used to determine whether the transgene is required to maintain growth and proliferation of the tumour. A transgenic mouse model of lung adenocarcinoma with expression of a mutant active K-ras transgene has been developed by using this regulatory transgenic technology (Fisher et al. 2001). Tumours rapidly regress as a result of apoptosis when doxycycline, a tetracycline analogue, is withdrawn, demonstrating the role of K-ras in driving lung tumorigenesis. Several other lung cancer mouse models have also been developed with conditional activation of oncogenic K-ras (Meuwissen et al. 2001; Jackson et al. 2001; Johnson et al. 2001). The use of regulatory transgenic systems such as these is a valuable tool for identifying targets for future drug development.

Further recent developments include the description of transgenic models of small-cell lung cancer (SCLC), which is a highly aggressive human tumour with a high mortality rate. Its molecular pathogenesis remains poorly understood. Using conditional inactivation of Rb1 and Trp53 in mouse lung epithelial cells, Meuwissen et al. (2003) created a mouse model of neuroendocrine malignancy. The majority of the tumours generated spread through the lung and gave rise to extrapulmonary metastases resembling SCLC. Ongoing investigation into the combinatorial effects of transgenic manipulation of multiple oncogenes and tumour suppressor genes such as this will no doubt continue to enhance our understanding of cancer biology. One of the issues with a number of known oncogenes and/or tumour suppressors is that they are embryonically lethal when deleted in the mouse. As a consequence, the study of tissue-specific pathways of tumorigenesis involving

these genes is impossible. Although explored in only limited fashion hitherto, the potential of tissue-specific deletions using the cre-loxP system has great potential for the understanding of tissue-specific tumour pathways. Many of these avenues remain to be explored.

4.3.2 Chemically Induced Models

Humans are constantly exposed to potentially harmful mixtures of chemicals and physical agents from the environment. The laboratory environment allows controlled administration of such toxins to animals. Mice that are prone to develop spontaneous lung tumours are also often susceptible to chemically induced lung cancer (Jackson et al. 2001). If a newborn inbred strain A/J mouse is given a single intraperitoneal injection of ethyl carbamate (urethane) at a dose of more than 0.5 mg/g body weight, it will develop dozens of benign lung adenomas within a few months (Shimkin and Stoner 1975). Some of these induced tumours eventually progress to adenocarcinomas that are histopathologically indistinguishable from human adenocarcinoma (Malkinson 1992). Many chemicals and environmental agents have been tested for carcinogenic activity using this tumorigenic response of the mouse lung as an indicator of toxicity.

Strain A mice have also been extensively used as a murine lung tumour bioassay to assess carcinogenic activity of chemicals, including urethane, benzopyrene, metals, aflatoxin, and constituents of tobacco smoke such as polyaromatic hydrocarbons and nitrosamines (Shimkin et al. 1975; Kim and Lee 1996; Stoner 1991). These agents can act as initiators and/or promoters of pulmonary tumorigenesis by accelerating tumour onset and increasing tumour multiplicity. In addition to chemicals, both radiation and viruses can induce lung tumours in mice (Rapp and Todaro 1980). Although induction of lung tumours in such models is highly reproducible (Malkinson 1989), all chemically induced lung tumours for some unknown reason exhibit relatively low metastatic potential. Despite the usefulness of carcinogen-induced lung cancer models, major disadvantages remain. They are time-consuming and, more importantly, they yield a variety of different histological tumour cell types with variable natural histories that might not be directly relevant to human lung cancer.

4.3.3 Human Lung Tumour Xenografts

The success of human tumour xenografting into immunocompromised rodents, and the ability to maintain the histologic and biologic identity of tumour cells through successive passages *in vivo*, revolutionized many aspects of cancer research, including drug development (Povlsen and Rygaard 1971). Since the immunogenicity of human neoplasms causes their destruction when implanted into immunocompetent

species, experimental hosts need to be immunocompromised. Irradiation, thymectomy, splenectomy and corticosteroids were initially used to blunt acquired immunity. With the breeding of hairless nude mouse mutants (nu/nu homozygotes), severe combined immunodeficient (SCID) mice and Rowett nude rats, these laboratory animals have now become the most common recipients of human tumours in experimental therapeutics.

Subcutaneous implantation is the predominant site for transplantation of human tumour material into the nude mouse, since the procedure is simple and the site is readily accessible. This also allows for straightforward monitoring of tumour growth. Although subcutaneous xenograft models can predict clinical efficacy (Mattern et al. 1988; Boven et al. 1992; Steel et al. 1983), these models have significant limitations, which include:

1. A low tumour take rate for fresh clinical specimens, with the percentage varying widely depending on the type of cancer (Mattern et al. 1988);
2. The atypical tissue compartment where tumour growth occurs. This raises the question of how the microenvironment of the subcutaneous space might influence study results.
3. The lack of consistent invasion and metastasis is perhaps the greatest limitation of the model (Mattern et al. 1988; Fidler 1986), because these properties of cancer are most closely linked to clinical outcome.
4. Since tumour-bearing animals may succumb to local tumour effects, such as infection from skin ulceration, survival is also not a feasible endpoint for assessing drug efficacy in these animals.

Due to these limitations, orthotopic models have been pursued and developed where human tumours are implanted directly into the appropriate organ or tissue of origin in the laboratory animal. The advantages of these models, such as improved tumour take and enhanced invasive and metastatic properties, are now well established (Fidler 1986, 1991; Fidler et al. 1990). Orthotopic implantation permits the expression of the metastatic phenotype of a variety of tumours; for example, colon carcinoma cells grown in the caecal wall, bladder carcinoma in the bladder, renal cell carcinoma cells under the renal capsule, and melanomas implanted subdermally. These models all typically yield metastases at much higher frequency than when grown subcutaneously (Manzotti et al. 1993; Kerbel et al. 1991). In contrast to subcutaneous implantation models, orthotopic models demonstrate that non-small-cell lung cancer (NSCLC) cell lines implanted into the thoracic cavity of nude mice are almost always fatal (McLemore et al. 1988). Orthotopic implantation appears to result in more aggressive tumour biology and shorter animal survival. This suggests that the local environment for *in situ* growth may reflect clinical reality more closely than heterotopic tumour implantation. An organ-specific site presumably provides tumour cells with the most appropriate milieu for studying local growth and metastasis. These manifestations support Paget's hypothesis that metastasis is not a random phenomenon. Rather, he concluded, malignant cells have special affinity for growth in the environment of certain organs—the familiar seed and soil theory (Paget 1989).

Orthotopic lung cancer models have been developed using endobronchial, intrathoracic or intravenous instillation of tumour cell suspensions (Wang and Hoffman 1992; McLemore et al. 1987; Howard et al. 1991) and surgical implantation of fresh, histologically intact tumour tissue (An et al. 1996; Rashidi et al. 2000). The first orthotopic model of human lung cancer was developed by McLemore et al. who implanted human lung cancer cell lines and enzymatically dissociated human lung tumours in the right lung of nude mice via an endobronchial injection (McLemore et al. 1987). The tumours had increased growth and invasiveness within the lung, as compared to the same tumours inoculated subcutaneously. However, most of the tumours remained within the right lung, with only 3% showing distant spread to lymph nodes, liver or spleen. McLemore et al. subsequently developed a second model by percutaneously injecting lung tumour cells via an intrathoracic route into the pleural space (McLemore et al. 1988). The model gave high tumour take-rates, with reproducible growth and a mortality endpoint as a result of local disease progression; however, very few metastases were observed. This approach appears to have major drawbacks, which may result in seeding lung cancer cells into the pleural space rather than within the pulmonary parenchyma or bronchi where clinical human lung cancer originates.

Similarly, endobronchial implantation has been used to grow non-small-cell (A549, NCI-H460, and NCI-H125) and small-cell (NCI-H345) lung carcinoma lines in irradiated nude rats (Howard et al. 1991). In these models, metastases to the mediastinal lymph nodes are frequently seen, but the incidence of systemic metastasis is low. In order to develop a model capable of metastasizing both regionally and systemically from a primary bronchial site, fresh tumour fragments derived from orthotopic lung tumours were implanted and grown from the H460 cell line. This H460 nude rat model has a 100% tumour take-rate in the lung with a rapid and reproducible growth rate up to approximately four grams over a 32–35-day period. It also metastasizes at a consistent rate to both regional mediastinal lymph nodes and distant systemic sites, including bone, brain and kidney. This is the first human lung cancer model to show extensive systemic metastasis from an orthotopic primary site (Howard et al. 1999).

Several other intrathoracic human lung cancer models have been described, all using immunocompromised mice. One is the traditional intravenous model in which the lung is colonized with tumour cells via the pulmonary circulation after tail vein injection (Kuo et al. 1992, 1993). In the second, the tumour grows in a subpleural location from fragments sewn onto the surface of the left lung through a thoracotomy (Wang et al. 1992). Recently, a SCID lung cancer model has also been described that develops lymphatic metastasis following percutaneous injection of cancer cells into the mouse lung (Miyoshi et al. 2000). However, none of these models grow from a primary endobronchial site and none develop a consistent metastatic pattern in extrathoracic locations.

In addition to using human cancer cell lines or their derived tumours for orthotopic implantation, fresh, histologically intact, human lung tumour tissue or tissue from metastatic lesions can be orthotopically implanted. Such models putatively maintain intact critical stromal epithelial relationships, although the source of most,

if not all, stromal tissue appears to be from the host rather than the original human xenograft (van Weerden and Romijn 2000). Very few such lung cancer models have been developed, in part because of technical obstacles and the generally poor growth of human lung cancer tissue in immunocompromised animals.

4.3.4 Lung Cancer Models in Cancer Drug Development

Despite advances in basic cancer biology and *in silico* modelling, animal models, especially human tumour xenografts, remain pivotal to cancer drug discovery and development. The value of a model depends on its validity, selectivity, predictability, reproducibility and cost (DeVita and Schein 1973; Zubrod 1972). Initially, lung tumour xenografts were designed with the intention of permitting patient-specific chemotherapy. By learning the drug responsiveness of a particular xenograft, treatment of the patient from whom the transplanted material originated could be individualized. Unfortunately, variations in take rate, the weeks to several months required for the transplants to grow, and the expense of maintaining xenografts, make this strategy generally untenable in the clinical setting.

Early drug screening systems utilizing the L1210 or P388 mouse leukaemia models represented a compound oriented strategy. Any anticancer agent for clinical development had to prove itself in the murine leukaemia/lymphoma models before further *in vivo* animal model development in a solid tumour panel. This resulted in a low yield of agents active against other solid tumour types. In order to develop screening systems with greater predictive power for the clinic, the U.S. National Cancer Institute (NCI) started to shift from a compound oriented screen toward disease-oriented screens. NCI employs xenografts as an integral part of its drug discovery screening strategy (Khleif 1997). Drugs toxic to human cancer cell lines *in vitro* are tested on xenografts as a secondary screen. The *in vitro* studies permit high throughput screening, and the cell lines found sensitive to a particular drug are used to choose appropriate xenografts for further testing. Lung tumour transplants often reflect the chemo-sensitivity of their tumours of origin. The growth of SCLC xenografts is typically inhibited by cisplatin, etoposide, cyclophosphamide, doxorubicin, and vincristine, while NSCLC grafts are much less responsive to those agents (Shoemaker et al. 1988). Other animal tumour models can be selected to demonstrate a specific cytotoxic effect of the drug or biological agent. Primary lung tumours in mice can be used for screening effective single drugs and drug combinations prior to clinical testing. For example, cisplatin, administered by itself and in combination with indomethacin, decreases the size of NNK-induced carcinomas (Belinsky et al. 1993). Although subcutaneous xenograft models such as the Lewis lung cancer system has been widely employed as an *in vivo* drug screen, the more complicated orthotopic models may be better suited for preclinical studies. Since orthotopic rodent tumours mimic biological aspects of clinical cancer (e.g. disease progression and metastasis) much more effectively than subcutaneous rodent tumours, orthotopic tumours are also likely to provide more relevant pharmacokinetic

and pharmacodynamic information than subcutaneous tumours (Mulvin et al. 1992). Subcutaneous xenograft models have a long history in the pharmaceutical industry, and they are indisputably straightforward to use: however, their record of accurately predicting the efficacy of anticancer agents in the clinic has been questionable.

A range of methods can be used to evaluate drug effects on tumours in animal models. Tumour size and tumour weight or volume changes are simple and easily reproducible parameters in subcutaneous xenograft models, but are more difficult, except at necropsy, in most orthotopic models. Morphological changes and alterations in tumour immunogenicity or invasiveness are additional markers of response. Survival, perhaps the ultimate parameter, is a valid endpoint only if clinically relevant tumour progression is responsible for the animal's demise. Advances in CT and PET-based imaging at the micro level, such that they can be applied to model organisms, have enhanced the potential of utilizing other imaging-based endpoints, like those that would be applied clinically. A number of orthotopic nude mouse and nude rat models have been developed as *in vivo* preclinical screens for novel anticancer therapies that target invasion, metastasis and angiogenesis (Davies et al. 1993; Russell et al. 1991; Schuster et al. 1993; Furukawa et al. 1993). A specific concern in studying anticancer agents with animal models derived from human cell lines is the degree of heterogeneity involved in the sample (Manzotti et al. 1993; Price 1994). In other words, does serial passage of cell lines over months and years select out and propagate specific clonal elements of a tumour? Studies have shown that the molecular characteristics of both breast and lung cancer cell lines closely match their original human tumour (Gazdar et al. 1998; Wistuba et al. 1999). However, other important characteristics, such as cytokine production or patterns of gene expression, may be lost or muted through serial passaging. Two potential solutions come to mind; either constructing model systems from fresh clinical tumour specimens and passaging the tumours serially as tumour lines, or creating multiple models representing all of the lung cancer histologies, thereby minimizing heterogeneity issues as much as possible.

4.3.5 Models for the Study of Lung Cancer Metastasis

The most remarkable feature of human lung cancer is tumour metastasis. It has been estimated that approximately 60% of cancer patients harbour overt or sub-clinical metastases at diagnosis, and it is the general consensus that the poor prognosis of lung cancer reflects the aggressive biological nature of the disease. In particular, metastasis to mediastinal lymph nodes or distant organs produces poor prognosis in lung cancer. Unfortunately, very little is known about how lung cancer cells propagate distant metastases, and the identification of molecules with a crucial role in the distant spread of lung cancer cells has been hampered by the absence of appropriate experimental model systems.

Intravasation and extravasation are two major steps for tumour cells to metastasize distantly. Entry of tumour cells into the circulation is the critical first step in

the metastatic cascade, and although it has been assayed in various ways (Glaves 1986; Butler and Gullino 1975; Liotta et al. 1974), it has not been observed directly. Novel approaches that rely on the ability to specifically "mark" the tumour cell are promising. For example, one can engineer tumour cells to express the green fluorescence protein for *in vivo* fluorescence imaging. In order to understand the metastatic pattern of NSCLC, Yang M et al. developed a green fluorescent protein (GFP) expressed in human lung cancer cell line H460-GFP. The GFP-expressing lung cancer was visualized to metastasize widely throughout the skeleton when implanted orthotopically in nude mice (Rashidi et al. 2000). This makes possible direct observation of tumour growth and metastasis, as well as tumour angiogenesis and gene expression. This assay is able to reveal the microscopic stages of tumour growth and metastatic seeding, superior to the previous transfection of lacZ to detect micrometastases (Boven et al. 1992; Lin et al. 1990), as real-time visualization of micrometastases even down to the single-cell level becomes feasible.

In contrast to utilizing orthotopic implantation to enhance metastatic potential in lung cancer, the alternative approach of *in vivo* selection of metastatic tumour cell variants has also been applied. There is now wide acceptance that many malignant tumours contain heterogeneous subpopulations of cells with different potential for invasion and metastasis (Fidler and Hart 1982; Heppner 1984; Nicolson 1984, 1987), and that metastasis results from the survival and proliferation of specialized subpopulations of cells that pre-exist within parental tumours (Fidler and Kripke 1977). Whether these populations are true 'cancer stem cells' remains controversial. The isolation of cell populations (from heterogeneous human tumours) that differ from the parent neoplasm in metastatic capacity provides a powerful tool with which to study those intrinsic properties that distinguish metastatic from nonmetastatic cells (Morikawa et al. 1988; Naito et al. 1986; Dinney et al. 1995).

Efforts have recently been made to develop metastatic lung cancer cell variants through *in vivo* propagation and selection. New cell line variants, H460-LNM35 and H460SM were established through *in vivo* propagation of tumour cells derived from H460 tumour or lymph node metastases (Kozaki et al. 2000; Blackhall et al. 2004). Selected variants of these tumour cells differ in their ability to metastasize compared to the parent cell line. This may provide a means of producing a highly metastatic orthotopic lung cancer model by direct cell implantation. Other opportunities involve the production of cell lines from transgenic models such as the K-ras hit-and-run allele, examining the potential of tumour engraftment in the absence of immune compromise. Selecting and enriching for metastatic variants constitutes a useful model for the discovery and mechanistic evaluation of genes potentially involved in metastasis of human lung cancer.

4.3.6 Model Organisms: New Systems for Modelling Cancer

Although the mouse and rat have traditionally been used for *in vivo* modelling of cancer, a number of model systems are on the horizon that may impact the ge-

netic dissection of tumour mechanisms, and facilitate high-throughput screens for drug discovery. Model organisms such as the yeast *Saccharomyces Cerevisiae*, the nematode *Caenorhabditis Elegans*, and the fruit fly *Drosophila Melanogaster,* have been very productive in the genetic dissection of pathways in fundamental biologic and organogenic processes. Unfortunately none of these systems develops cancer as we know it in human systems. In contrast, the model organism zebrafish *Danio rerio* does develop tumours with a variety of histologic subtypes that are similar to those present in humans. Fish have a long history of use in cancer toxicology studies because of this propensity to develop cancer. Because of considerable progress in Zebrafish genetics and genomics over the past few years, the Zebrafish system has provided many useful tools for studying basic biological processes. These tools include forward genetic screens, transgenic models, specific gene disruptions and small-molecule screens. By combining carcinogenesis assays, genetic analyses and small-molecule screening techniques, the Zebrafish is emerging as a powerful system for identifying novel cancer genes and for cancer drug discovery (Stern and Zon 2003). Some of the advantages of Zebrafish include ease and low cost of housing, large numbers of embryos produced from matings, ease of mutagenesis, and external, transparent embryos in which cleavage divisions, gastrulation, morphogenesis and organogenesis occur within 24 h. Because of these advantages, the Zebrafish has become a force to reckon with in vertebrate developmental genetics. It is a potentially powerful cancer model organism.

In *D. Melanogaster*, many decades of research have produced tremendous insight into basic understanding of growth, apoptosis, differentiation, and developmental patterning. Although abnormal proliferation has been described and exploited, it has only been recently that the potential of the "fruit fly" as a cancer model has begun to be exploited. Genetic techniques available in *D. Melanogaster* can facilitate the analysis of the behaviour of cells with distinct mutations, allowing the study of cell interactions and oncogenic cooperation. As recently described by Wu et al. (2010), cooperation between the oncogenic protein Ras(V12) in *Drosophila* eye-antennal discs and loss-of-function mutations in the conserved tumour suppressor scribbled (scrib) gene gives rise to metastatic tumours that display many characteristics observed in human cancers. This interaction between Ras(V12) and scrib(-) clones appeared to involve JNK signalling propagation and JNK-induced up-regulation of JAK/STAT-activating cytokines. The development of Ras(V12) tumours could also be triggered by tissue damage, a stress condition that activates JNK signalling, suggesting that similar cooperative mechanisms could have a role in the development of human cancers. Recent descriptions at the 2010 American Association of Cancer Research meeting by Cagan et al. of drug screening in tumours of the *D. Melanogaster* eye also suggest this model may allow the adaptation of the powerful genetics underpinning *D. Melanogaster* for enhancing drug and cancer mechanistic screens. In summary, many different models of lung and other cancers models are available, but unfortunately none accurately reflects all aspects of human disease observed clinically. Each has its own advantages and disadvantages that should be understood and evaluated prior to their use in addressing specific questions. In selecting the best model system,

consideration should be given to the genetic stability and heterogeneity of transplanted cell lines, immunogenicity within the host animal, and the appropriate biological endpoints.

4.3.7 Restrictions on the Use of Animals in Research

There is increasing legislation and pressure on the research community to reduce, or even eliminate the use of animals in research. However, relevant animal model systems provide the appropriate interface between the laboratory bench and a patient's bedside for continued progress in cancer research and drug development. The data generated will also be essential in establishing and testing numerical models as systems approaches are developed and validated. As in many other diseases, ever more sophisticated lung cancer models will be needed in the future as the complexities of this devastating disease are slowly unravelled.

4.4 Patient Biobanks

Biobanks are a collection of biological samples, such as blood, tissues or cell lines, or DNA with associated research data, as well as epidemiological and clinical data of patients and healthy persons. Providing access to human cancer tissue for basic science was the basis on which the National Cancer Institute's Cooperative Human Tissue Network (CHTN) was established in 1987, providing prospective investigator-defined procurement of malignant, benign, diseased and uninvolved (normal adjacent) tissues (http://chtn.nci.nih.gov). The availability of high-quality specimens for research purposes ideally requires standardized methods for their collection, long-term storage, retrieval and distribution. The International Society for Biological and Environmental Repositories (ISBER, http://www.isber.org) is a network of biobanks. Within the scope of their activity to disseminate information and guidance on safe and effective management of specimen collection, the ISBER, in 2008 published guidelines for repositories on how to collect, store, retrieve, and distribute biological material for research. Adherence to these guidelines is however strictly voluntary.

Directive 2004/23/EC of the European Parliament and Council of March 2004 set the standards of quality and safety for the donation, procurement, testing, processing, preservation, storage and distribution of human tissues and cells. These directives have been transposed into national laws in most European Union member states.

Similarly, in the United States, the National Cancer Institute (NCI) initiated the Office of Biorepositories and Biospecimen Research (OBBR) (http://biospecimens.cancer.gov) to address the problem of limited availability of carefully collected biospecimens with adequate epidemiological and clinical data. Objectives include the establishment of biobanks, development of accreditation programmes and distribution of guidelines.

In 2008, the Biobanking and Biomolecular Resources Research Infrastructure (BBMRI) was set up as a pan-European and international broadly accessible research infrastructure, with a network of existing and *de novo* biobanks and biomolecular resources (http://bbmri.eu). It includes samples from patients and healthy persons, representing different European populations.

Clinical scientists commonly have the opportunity of immediate access to patient tumour samples in their hospital environment. Throughout the years, hospitals have acquired large resources of normal and tumour samples of patients undergoing diagnostic or surgical procedures. Advantages of this immediate access to samples are the availability of complete medical and histologic history as well as follow-up data on each patient whose tissue is utilized for research purposes; the opportunity for histological review of the sample for diagnostic accuracy; and finally the potential for collecting viable tissue in a randomized prospective manner.

4.4.1 Paraffin Embedded Tissues

Formalin fixation and paraffin embedding (FFPE) is the clinical standard for processing samples in pathology departments aiming at long-term storage of samples for legal and other purposes. The advantage of this preservation method is the low storage cost. These large archives of material stored in hospitals are an invaluable resource both for testing known biomarkers and discovering new markers. Despite the availability of these large amounts of tissue samples, utilization of this non-viable tissue for large scale genomic oriented research has been limited. Formalin fixation has the disadvantage of generating cross-links between proteins and nucleic acids, resulting in structural denaturation and fragmentation of the nucleic acids. This is particularly problematic with RNA, as it is more susceptible to the degradation process. Factors influencing this process may include the age of the sample, length of storage and the tissue sample size. Many quantitative analytical techniques, such as the gene expression microarray technique, require full length and normal structured RNA. Even though recent development of novel high-throughput technical platforms has enabled the utilization of lower quality RNA and DNA for genomic research, these techniques remain inferior to those using high-quality nucleic acids generated from fresh or snap-frozen tissue.

4.4.2 Snap-frozen Tissues

Translational research to date has traditionally relied on the availability of fresh tissue harvested at the time of surgical resection, since fresh or snap-frozen banked tissue remains the optimal material for genetic research studies, providing high-quality nucleic acids. However, the preservation methods for snap-freezing samples are not yet as standardized in pathology departments as for FFPE samples.

Studies have shown that the interval between harvesting and banking tissue after resection of the specimen from the patient (or even after interruption of blood supply to the resected area, for example ligation of the branches of the pulmonary artery and vein during lung resections) may be crucial for gene expression studies. Intervals of 20–60 min after resection or blood flow interruption are recommended to preserve stability of RNA in the tissue. Furthermore, as tumours are composed of various cell types, pooling from different areas of a tumour should be considered, to minimize artefacts from tumour heterogeneity.

4.4.3 Linking Molecular and Clinical Measurements

Molecular research in cancer has been driven by the observation that patients with equivalent underlying tumour histology, at the same pathological stage, may end up with a markedly different prognosis. These observations suggest that patients may benefit from a more individualized treatment approach to their disease. The ultimate goal, therefore, is to broaden the current classical staging system based on tumour size, lymph node involvement and metastases (TNM) by molecular markers so as to be able to derive individualized treatment decisions.

The prerequisite for this type of approach is thorough annotation (clinical and histologic information) for the utilized tissue samples, as well as long-term follow-up data, including survival and potential disease recurrence or metastasis. Since this information can therefore be used for correlation with findings in molecular studies to identify clinically relevant molecular markers associated with poor prognosis or survival, detailed clinical follow-up information is as essential as high-quality tissue. If possible, medical data should be collected at the patients' follow-up visits with their medical or surgical oncology specialist. However, patients with early-stage cancer may not routinely be followed by medical or surgical oncology specialists, or may choose a health care provider closer to home who is not affiliated with their cancer treating institution. Standardized questionnaires as commonly utilized by certified cancer centres may therefore be sent out to the patient or their primary care physician, forming a helpful tool to obtain information on disease recurrence or death. Publicly available databases such as social security death indices merely provide information on the vital status of the patient, which can be utilized in overall survival analyses but not to determine disease-free survival. As discussed above, the availability of high-quality tissue is crucial to obtaining reliable and reproducible molecular information upon which clinical decisions can be made. Ideally, tissue specimens should be obtained at the time of surgical resection or, if a chemo- or radio-therapeutic approach is indicated, before initiation of treatment. Along with the tumour specimen, matched normal tissue should be obtained either adjacent to or remotely distant from the tumour lesion. Optimal long-term storage, such as in tissue banks with liquid nitrogen facilities, should be a goal. Moreover, obtaining tissue representative of disease recurrence may provide additional useful information. An infrastructure should be established which ensures that data of patients are collected

prospectively in a database with yearly updates, and that fresh frozen tissue with blood samples is obtained and stored in a dedicated tissue bank. This can be achieved either by initiating a coordinated institutional cancer centre effort, or through an individual research unit within a department; it requires significant and reliable coordination between the different specialties. However, building and maintaining such an infrastructure requires considerable funding, and thus may not be feasible in all healthcare systems. Nevertheless, as a critical resource for future research efforts aiming at developing superior treatment approaches, its realization should be striven for in academic centres, and may be feasible as a joint effort of the different oncologic specialties within departments of pathology, epidemiology and statistics.

Furthermore, the establishment of international networks to pool collected clinical data and tissue resources internationally would be a remarkable advantage for studying different ethnicities and populations of diverse environmental exposure as compared with the more limited investigation of the local population. It may well provide useful insights into distinct molecular disease pathologies.

4.5 Role of Interactome Maps and Crucial Pathways

Interactome mapping describes the systematic mapping of protein-protein interactions in organisms. Initially established in model organisms, such as yeast, with a well-defined genome and large network coverage, the concept of biomolecular interaction networks has been adapted to humans and human diseases (Rual et al. 2005; Stelzl et al. 2005). The concept that network modules may supply structure to the interpretation of other genomic data has been demonstrated by now in a number of publications concerning multiple tumour types. An emerging theme is the idea that individual tumour types may be diseases of network perturbation. According to this construct, disease processes may be incompletely understood if entities such as gene expression or copy number variation are investigated alone without a network framework. This is evidenced by the fact that genes associated with poor prognosis in many cancers seem not to be merely randomly positioned in these networks, but rather are clustered together in central network positions with high interconnectivity (Ideker and Sharan 2008). Hence, understanding disease-related networks may facilitate a more reliable adoption of gene expression signatures or candidate genes as biomarkers or prognostic markers, as well as providing novel drug targets.

4.5.1 Links to Specific Types of Cancer

Wachi and colleagues assessed the interactome of differentially expressed genes in squamous-cell carcinoma of the lung, showing that amplified genes are generally well connected, compared to the suppressed genes (Wachi et al. 2005). Similarly, Jonsson and colleagues showed that cancer proteins have a very different network topology from proteins known not to be mutated in cancer (Jonsson and Bates

2006). Interestingly, cancer proteins have an increased frequency of interactions in which they participate, as well as a high ratio of highly promiscuous domains with respect to the number of proteins they interact with.

Chen and colleagues examined 10 published prognostic expression-based gene signatures in breast cancer and their interconnectivity through a protein-protein-interaction network, finding that genetically distinct prognostic signatures from independent data sets have comparable interactions between proteins of differentially expressed genes (Chen et al. 2010). Developing pathway-based cancer roadmaps from established gene expression signatures may have the potential for the identification of novel drug targets. Furthermore, these networks may predict the risk of disease progression or metastatic spread of tumours more accurately than clinical factors or gene signatures alone. Chuang and colleagues showed that genes identified within a network classify metastatic breast cancers versus non-metastatic breast cancers more reliably than individual biomarkers outside a network approach (Chuang et al. 2007).

4.5.2 Synthetic Lethality as a Network-derived Treatment Success-story

Large-scale analysis of genetic interaction networks in model organisms like yeast suggests that disease-specific networks should not only give rise to destructive phenotypes (e.g. uncontrolled cell growth), but will also present unique vulnerabilities that can be exploited for novel therapies. For example, a second perturbation in the context of a perturbed disease network might be lethal, while the same perturbation is harmless in normal cells. Synthetic lethality represents an exciting new avenue, disrupting cancer cells for targeted treatment. Two genes are said to be in a synthetic lethal relationship if a mutation in either gene alone is not lethal, but mutations in both cause the death of the cell. In applying synthetic lethality to the discovery of cancer drugs, a screening program is designed to reveal a target gene that, when mutated or chemically inhibited, kills cells that harbour a specific cancer-related alteration (such as a mutated tumour-suppressor gene or an activated oncogene), but spares otherwise identical cells lacking the cancer-related alteration. This concept was recently exploited in the development of PARP inhibitors as novel chemotherapies for breast cancer. Poly (ADP-ribose) polymerase (PARP) is a protein involved in a number of cellular processes involving mainly DNA repair and programmed cell death. While PARP is not essential in normal cells, BRCA mutant cells are completely dependent on PARP for their survival. This genetic interaction is the basis of a novel treatment strategy for triple-negative breast cancers currently undergoing clinical trials, which provides an exciting proof-of-concept for synthetic lethality as a drug discovery strategy for human disease. Importantly, this would enable targeting of tumour suppressors which are otherwise undruggable. This, in turn, would allow leveraging the expanding number of well-characterized mutations associated with human cancer for drug discovery strategies.

4.6 Integration into Systems and Computational Approaches

The volume of experimental observations from both patient samples and tumour models in the cancer literature is overwhelming. Despite this, the variations in the molecular alterations that can give rise to cancer can be broadly grouped into a handful of traits that cancer cells must acquire for malignant transformation to occur. The original seminal description by Hanahan and Weinberg of the "hallmarks" of cancer (see Chap. 9), still valid today, looks beyond the detailed molecular discoveries governing malignant transformation, and integrates them into a conceptual framework underlying all cancers (Hanahan et al. 2000). This framework simply but insightfully states that molecular alterations can be classified by dysfunction in as many as six different regulatory systems that must be perturbed for a normal cell to become cancerous (Khalil and Hill 2005). These include many diverse and seemingly non-overlapping biological processes, including (1) self-sufficiency in growth signals; (2) insensitivity to anti-growth signals; (3) evasion of apoptosis; (4) limitless replicative potential; (5) sustained angiogenesis, and (6) tissue invasion and metastasis. Genetic instability is defined as an enabling characteristic that facilitates the acquisition of other mutations due to defects in the repair of DNA. Although a limited number of cancer subtypes are defined by a single genetic alteration leading to a primary defect in one of the above listed processes, most solid tumours, responsible for the largest burden of human illness, are heterogeneous lesions characterized by many, if not all, defects observable simultaneously. This includes lung, breast, prostate, colon, and central nervous system tumours among others. In our attempts to understand tumorigenesis by reductionism, much work has gone into the study of the individual biological processes known to as the "hallmarks" of cancer. Increased understanding of many of these biological modules has unfortunately not resulted in parallel understanding of the root cause of the majority of cancers, or how best to treat them.

The concept of cancer as a system failure, and the potential of systems biology and systems medicine for a better understanding of the disease, is generating significant discussion in the literature, as investigators grapple with methods and approaches (Kitano 2002; Alberghina et al. 2004; Spencer et al. 2004; Khalil and Hill 2005; Hornberg et al. 2006). The mere recognition of cancer as a systems biology disease is a key first step. This hypothesis views the individual defects observable in solid tumours cumulatively as system failures either at the cellular or multicellular level. (See Chap. 12 and 17 for detailed discussions). A systematic study and understanding of oncogenic network rewiring (Pawson and Warner 2007) offers the possibility of using systems biology approaches to generate testable models of different tumours, an exciting and as yet unexplored realm of cancer biology.

An integrated system level understanding, the approach advocated in systems biology, requires a change in our notion of what to look for in biology (Kitano 2002). While an understanding of individual genes and proteins continues to be important, the focus needs to shift towards understanding the structure, function

and dynamics of cancer viewed as a system. Why has systems biology received so much recent attention? In short, it is because the key first step of defining system structures has quickly advanced from fantasy to reality in the post-genomic era. The achievement of full genome sequencing projects in many organisms, including *Homo sapiens*, has defined the parts list for growth, development, and normal physiological function. The technological development associated with these achievements has spawned the nascent fields of genomics, proteomics, and multiple "-omic" disciplines defined by their systematic, data-driven approaches to biological experimentation. These approaches are increasingly being applied to the question of understanding cancer.

4.7 The Future: Data Integration to Systems-level Experiments

Integrating across multiple systems is a formidable challenge. Each area alone is extremely complex. However, with system structures now being defined, the key first steps toward system level integration are becoming possible. Much work remains, however, before we can feasibly study multiple system modules as a whole. The role of computational biology and mathematical modelling as an integral part of these advances is becoming increasingly clear. The next round of major advances will arise from the combined efforts of integrated study groups with expertise in both computational and biological experimentation.

References

Adams JM, Cory S (1991) Transgenic models of tumor development. Science 254(5035):1161–1167

Alberghina L, Chiaradonna F, Vanoni M (2004) Systems biology and the molecular circuits of cancer. Chembiochem 5(10):1322–1333

An Z, Wang X, Kubota T et al (1996) A clinical nude mouse metastatic model for highly malignant human pancreatic cancer. Anticancer Res 16(2):627–631

Arguello F, Sterry JA, Zhao YZ et al (1996) Two serologic markers to monitor the engraftment, growth, and treatment response of human leukemias in severe combined immunodeficient mice. Blood 87(10):4325–4332

Belinsky SA, Stefanski SA, Anderson MW (1993) The A/J mouse lung as a model for developing new chemointervention strategies. Cancer Res 53(2):410–416

Bignell GR, Greenman CD, Davies H et al (2010) Signatures of mutation and selection in the cancer genome. Nature 463(7283):893–898

Blackhall FH, Pintilie M, Wigle DA et al (2004) Stability and heterogeneity of expression profiles in lung cancer specimens harvested following surgical resection. Neoplasia 6(6):761–767

Boven E, Winograd B, Berger DP et al (1992) Phase II preclinical drug screening in human tumor xenografts: a first European multicenter collaborative study. Cancer Res 52(21):5940–5947

Brinster RL, Chen HY, Trumbauer ME et al (1985) Factors affecting the efficiency of introducing foreign DNA into mice by microinjecting eggs. Proc Natl Acad Sci U S A 82(13):4438–4442

Bussey KJ, Chin K, Lababidi S et al (2006) Integrating data on DNA copy number with gene expression levels and drug sensitivities in the NCI-60 cell line panel. Mol Cancer Ther 5:853–867
Butler TP, Gullino PM (1975) Quantitation of cell shedding into efferent blood of mammary adenocarcinoma. Cancer Res 35(3):512–516
Cancer Facts (2009) Cancer facts and figures 2009. http://www.cancer.org/docroot/PRO/content/PRO_1_1_Cancer_Statistics_2009_Presentation.asp. Accessed 1 Apr 2010
Chen J, Sam L, Huang Y et al (2010) Protein interaction network underpins concordant prognosis among heterogeneous breast cancer signatures. J Biomed Inform [Epub ahead of print]
Chuang HY, Lee E, Liu YT et al (2007) Network-based classification of breast cancer metastasis. Mol Syst Biol 3:140
Corbett TH, Griswold DP Jr, Roberts BJ et al (1975) Tumor induction relationships in development of transplantable cancers of the colon in mice for chemotherapy assays, with a note on carcinogen structure. Cancer Res 35(9):2434–2439
Corbett TH, Roberts BJ, Leopold WR et al (1984) Induction and chemotherapeutic response of two transplantable ductal adenocarcinomas of the pancreas in C57BL/6 mice. Cancer Res 44(2):717–726
Crawford LV, Pim DC, Gurney EG et al (1981) Detection of a common feature in several human tumor cell lines—a 53,000-dalton protein. Proc Natl Acad Sci U S A 78(1):41–45
Dan S, Tsunoda T, Kitahara O et al (2002) An integrated database of chemosensitivity to 55 anticancer drugs and gene expression profiles of 39 human cancer cell lines. Cancer Res 62(4):1139–1147
Davies B, Brown PD, East N et al (1993) A synthetic matrix metalloproteinase inhibitor decreases tumor burden and prolongs survival of mice bearing human ovarian carcinoma xenografts. Cancer Res 53(9):2087–2091
DeVita VT, Schein PS (1973) The use of drugs in combination for the treatment of cancer: rationale and results. N Engl J Med 288(19):998–1006
Der CJ, Krontiris TG, Cooper GM (1982) Transforming genes of human bladder and lung carcinoma cell lines are homologous to the ras genes of Harvey and Kirsten sarcoma viruses. Proc Natl Acad Sci U S A 79(11):3637–3640
Dinney CP, Fishbeck R, Singh RK et al (1995) Isolation and characterization of metastatic variants from human transitional cell carcinoma passaged by orthotopic implantation in athymic nude mice. J Urol 154(4):1532–1538
Fidler IJ (1986) Rationale and methods for the use of nude mice to study the biology and therapy of human cancer metastasis. Cancer Metastasis Rev 5(1):29–49
Fidler IJ (1991) Orthotopic implantation of human colon carcinomas into nude mice provides a valuable model for the biology and therapy of metastasis. Cancer Metastasis Rev 10(3):229–243
Fidler IJ, Hart IR (1982) Biological diversity in metastatic neoplasms: origins and implications. Science 217(4564):998–1003
Fidler IJ, Kripke ML (1977) Metastasis results from preexisting variant cells within a malignant tumor. Science 197(4306):893–895
Fidler IJ, Naito S, Pathak S (1990) Orthotopic implantation is essential for the selection, growth and metastasis of human renal cell cancer in nude mice [corrected]. Cancer Metastasis Rev 9(2):149–165
Fisher GH, Wellen SL, Klimstra D et al (2001) Induction and apoptotic regression of lung adenocarcinomas by regulation of a K-Ras transgene in the presence and absence of tumor suppressor genes. Genes Dev 15(24):3249–3262
Fowlis DJ, Balmain A (1993) Oncogenes and tumour suppressor genes in transgenic mouse models of neoplasia. Eur J Cancer 29A(4):638–645
Furukawa T, Kubota T, Watanabe M et al (1993) A novel "patient-like" treatment model of human pancreatic cancer constructed using orthotopic transplantation of histologically intact human tumor tissue in nude mice. Cancer Res 53(13):3070–3072
Galski H, Sullivan M, Willingham MC et al (1989) Expression of a human multidrug resistance cDNA (MDR1) in the bone marrow of transgenic mice: resistance to daunomycin-induced leukopenia. Mol Cell Biol 9(10):4357–4363

Gaur A, Jewell DA, Liang Y et al (2007) Characterization of microRNA expression levels and their biological correlates in human cancer cell lines. Cancer Res 67:2456–2468
Gazdar AF, Kurvari V, Virmani A et al (1998) Characterization of paired tumor and non-tumor cell lines established from patients with breast cancer. Int J Cancer 78(6):766–774
Glaves D (1986) Detection of circulating metastatic cells. Prog Clin Biol Res 212:151–167
Gordon JW, Ruddle FH (1983) Gene transfer into mouse embryos: production of transgenic mice by pronuclear injection. Methods Enzymol 101:411–433
Grever MR, Schepartz SA, Chabner BA (1992) The National Cancer Institute: cancer drug discovery and development program. Semin Oncol 19(6):622–638
Hanahan D, Weinberg RA (2008) The hallmarks of cancer. Cell 2000 Jan 7;100(1):57–70
Heppner GH (1984) Tumor heterogeneity. Cancer Res 44(6):2259–2265
Holbeck S, Chang J, Best AM et al (2010) Expression profiling of nuclear receptors in the NCI60 cancer cell panel reveals receptor-drug and receptor-gene interactions. Mol Endocrinol [Epub ahead of print]
Hornberg JJ, Bruggeman FJ, Westerhoff HV et al (2006) Cancer: a systems biology disease. Biosystems 83(2–3):81–90
Howard RB, Chu H, Zeligman BE et al (1991) Irradiated nude rat model for orthotopic human lung cancers. Cancer Res 51(12):3274–3280
Howard RB, Mullen JB, Pagura ME et al (1999) Characterization of a highly metastatic, orthotopic lung cancer model in the nude rat. Clin Exp Metastasis 17(2):157–162
Ideker T, Sharan R (2008) Protein networks in disease. Genome Res 18(4):644–652
Ikediobi ON, Davies H, Bignell G et al (2006) Mutation analysis of 24 known cancer genes in the NCI-60 cell line set. Mol Cancer Ther 5(11):2606–2612
International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. Nature 431(7011):931–945
Jackson EL, Willis N, Mercer K et al (2001) Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. Genes Dev 15(24):3243–3248
Jaenisch R (1980) Retroviruses and embryogenesis: microinjection of Moloney leukemia virus into midgestation mouse embryos. Cell 19(1):181–188
Jaenisch R, Jähner D, Nobis P et al (1981) Chromosomal position and activation of retroviral genomes inserted into the germ line of mice. Cell 24(2):519–529
Johnson L, Mercer K, Greenbaum D et al (2001) Somatic activation of the K-ras oncogene causes early onset lung cancer in mice. Nature 410(6832):1111–1116
Johnston MR, Mullen JB, Pagura ME et al (2001) Validation of an orthotopic model of human lung cancer with regional and systemic metastases. Ann Thorac Surg 71(4):1120–1125
Jonsson PF, Bates PA (2006) Global topological features of cancer proteins in the human interactome. Bioinformatics 22(18):2291–2297
Jähner D, Jaenisch R (1980) Integration of Moloney leukaemia virus into the germ line of mice: correlation between site of integration and virus activation. Nature 287(5781):456–458
Kerbel RS, Cornil I, Theodorescu D (1991) Importance of orthotopic transplantation procedures in assessing the effects of transfected genes on human tumor growth and metastasis. Cancer Metastasis Rev 10(3):201–215
Kerr KM (2001) Pulmonary preinvasive neoplasia. J Clin Pathol 54(4):257–271
Khalil IG, Hill C (2005) Systems biology for cancer. Curr Opin Oncol 17(1):44–48
Khleif SN (1997) Animal models in drug development. In: Holland JF (eds) Cancer medicine, B.R.J.M.D.F.E.K.D.W.R. Williams & Wilkins, Baltimore
Kim SH, Lee CS (1996) Induction of benign and malignant pulmonary tumours in mice with benzo(a)pyrene. Anticancer Res 16(1):465–470
Kitano H (2002) Systems biology: a brief overview. 295(5560):1662–1664
Kozaki K, Miyaishi O, Tsukamoto T et al (2000) Establishment and characterization of a human lung cancer cell line NCI-H460-LNM35 with consistent lymphogenous metastasis via both subcutaneous and orthotopic propagation. Cancer Res 60(9):2535–2540
Kuo TH, Kubota T, Watanabe M et al (1992) Orthotopic reconstitution of human small-cell lung carcinoma after intravenous transplantation in SCID mice. Anticancer Res 12(5):1407–1410

Kuo TH, Kubota T, Watanabe M et al (1993) Site-specific chemosensitivity of human small-cell lung carcinoma growing orthotopically compared to subcutaneously in SCID mice: the importance of orthotopic models to obtain relevant drug evaluation data. Anticancer Res 13(3):627–630

Lin WC, Pretlow TP, Pretlow TG 2nd et al (1990) Bacterial lacZ gene as a highly sensitive marker to detect micrometastasis formation during tumor progression. Cancer Res 50(9):2808–2817

Liotta LA, Kleinerman J, Saidel GM (1974) Quantitative relationships of intravascular tumor cells, tumor vessels, and pulmonary metastases following tumor implantation. Cancer Res 34(5):997–1004

Livingood L (1986) Tumors in the mouse. Johns Hopkins Bull 66(67):177

Malkinson AM (1989) The genetic basis of susceptibility to lung tumors in mice. Toxicology 54(3):241–271

Malkinson AM (1992) Primary lung tumors in mice: an experimentally manipulable model of human adenocarcinoma. Cancer Res 52(9 Suppl):2670s–2676s

Manzotti C, Audisio RA, Pratesi G (1993) Importance of orthotopic implantation for human tumors as model systems: relevance to metastasis and invasion. Clin Exp Metastasis 11(1):5–14

Maronpot RR, Palmiter RD, Brinster RL et al (1991) Pulmonary carcinogenesis in transgenic mice. Exp Lung Res 17(2):305–320

Masters JR (2002) HeLa cells 50 years on: the good, the bad and the ugly. Nat Rev Cancer 2(4):315–319

Mattern J, Bak M, Hahn EW et al (1988) Human tumor xenografts as model for drug testing. Cancer Metastasis Rev 7(3):263–284

Mattison J, Kool J, Uren AG et al (2010) Novel candidate cancer genes identified by a large-scale cross-species comparative oncogenomics approach. Cancer Res 70(3):883–895

McLemore TL, Eggleston JC, Shoemaker RH et al (1988) Comparison of intrapulmonary, percutaneous intrathoracic, and subcutaneous models for the propagation of human pulmonary and nonpulmonary cancer cell lines in athymic nude mice. Cancer Res 48(10):2880–2886

McLemore TL, Liu MC, Blacker PC et al (1987) Novel intrapulmonary model for orthotopic propagation of human lung cancers in athymic nude mice. Cancer Res 47(19):5132–5140

Meuwissen R, Linn SC, Linnoila RI et al (2003) Induction of small cell lung cancer by somatic inactivation of both Trp53 and Rb1 in a conditional mouse model. Cancer Cell 4(3):181–189

Meuwissen R, Linn SC, Valk M Van Der et al (2001) Mouse model for lung tumorigenesis through Cre/lox controlled sporadic activation of the K-Ras oncogene. Oncogene 20(45):6551–6558

Miyoshi T, Kondo K, Ishikura H et al (2000) SCID mouse lymphogenous metastatic model of human lung cancer constructed using orthotopic inoculation of cancer cells. Anticancer Res 20(1A):161–163

Morikawa K, Walker SM, Jessup JM et al (1988) In vivo selection of highly metastatic cells from surgical specimens of different primary human colon carcinomas implanted into nude mice. Cancer Res 48(7):1943–1948

Mulvin DW, Howard RB, Mitchell DH et al (1992) Secondary screening system for preclinical testing of human lung cancer therapies. J Natl Cancer Inst 84(1):31–37

Naito S, Eschenbach AC von, Giavazzi R et al (1986) Growth and metastasis of tumor cells isolated from a human renal cell carcinoma implanted into different organs of nude mice. Cancer Res 46(8):4109–4115

Nicolson GL (1984) Generation of phenotypic diversity and progression in metastatic tumor cells. Cancer Metastasis Rev 3(1):25–42

Nicolson GL (1987) Tumor cell instability, diversification, and progression to the metastatic phenotype: from oncogene to oncofetal expression. Cancer Res 47(6):1473–1487

Nissen KK, Vogel U, Nexø BA (2009) Association of a single nucleotide polymorphic variation in the human chromosome 19q13.3 with drug responses in the NCI60 cell lines. Anticancer Drugs 20(3):174–178

Paget S (1989) The distribution of secondary growths in cancer of the breast. 1889. Cancer Metastasis Rev 8(2):98–101

Park ES, Rabinovsky R, Carey M et al (2010) Integrative analysis of proteomic signatures, mutations, and drug responsiveness in the NCI 60 cancer cell line set. Mol Cancer Ther 9(2):257–267

Pawson T, Warner N (2007) Oncogenic re-wiring of cellular signaling pathways. Oncogene 26(9):1268–1275

Povlsen CO, Rygaard J (1971) Heterotransplantation of human adenocarcinomas of the colon and rectum to the mouse mutant Nude. A study of nine consecutive transplantations. Acta Pathol Microbiol Scand A 79(2):159–169

Price JE (1994) Analyzing the metastatic phenotype. J Cell Biochem 56(1):16–22

Rapp UR, Todaro GJ (1980) Generation of oncogenic mouse type C viruses: in vitro selection of carcinoma-inducing variants. Proc Natl Acad Sci U S A 77(1):624–628

Rashidi B, Yang M, Jiang P et al (2000) A highly metastatic Lewis lung carcinoma orthotopic green fluorescent protein model. Clin Exp Metastasis 18(1):57–60

Ring BZ, Chang S, Ring LW et al (2008) Gene expression patterns within cell lines are predictive of chemosensitivity. BMC Genomics 9:74

Ross DT, Scherf U, Eisen MB et al (2000) Systematic variation in gene expression patterns in human cancer cell lines. Nat Genet 24(3):227–235

Rual JF, Venkatesan K, Hao T et al (2005) Towards a proteome-scale map of the human protein-protein interaction network. Nature 437(7062):1173–1178

Russell PJ, Ho Shon I, Boniface GR et al (1991) Growth and metastasis of human bladder cancer xenografts in the bladder of nude rats. A model for intravesical radioimmunotherapy. Urol Res 19(4):207–213

Sandmöller A, Halter R, Suske G et al (1995) A transgenic mouse model for lung adenocarcinoma. Cell Growth Differ 6(1):97–103

Scherf U, Ross DT, Waltham M et al (2000) A gene expression database for the molecular pharmacology of cancer. Nat Genet 24(3):236–244

Schneider U, Schwenk HU, Bornkamm G (1977) Characterization of EBV-genome negative "null" and "T" cell lines derived from children with acute lymphoblastic leukemia and leukemic transformed non-Hodgkin lymphoma. Int J Cancer 19(5):621–626

Schuster JM, Friedman HS, Archer GE et al (1993) Intraarterial therapy of human glioma xenografts in athymic rats using 4-hydroperoxycyclophosphamide. Cancer Res 53(10 Suppl):2338–2343

Shankavaram UT, Reinhold WC, Nishizuka S et al (2007) Transcript and protein expression profiles of the NCI-60 cancer cell panel: an integromic microarray study. Mol Cancer Ther 6:820–832

Shell S, Park SM, Radjabi AR et al (2007) Let-7 expression defines two differentiation stages of cancer. Proc Natl Acad Sci U S A 104(27):11400–11405

Shimkin MB, Stoner GD (1975) Lung tumors in mice: application to carcinogenesis bioassay. Adv Cancer Res 21:1–58

Shockett PE, Schatz DG (1996) Diverse strategies for tetracycline-regulated inducible gene expression. Proc Natl Acad Sci U S A 93(11):5173–5176

Shoemaker RH (2006) The NCI60 human tumour cell line anticancer drug screen. Nat Rev Cancer 6(10):813–823

Shoemaker RH, Abbott BJ et al (1988) Human tumor xenograft models for use with an in vitro-based, disease-oriented antitumor drug screening program. In: B.W.M.P.a.H.P. (eds) Human tumor xenografts in anticancer drug development. Springer, Berlin

Shoemaker RH, Dykes DJ, Plowman J et al (1991) Practical spontaneous metastasis model for in vivo therapeutic studies using a human melanoma. Cancer Res 51(11):2837–2841

Shoemaker RH, Smythe AM, Wu L et al (1992) Evaluation of metastatic human tumor burden and response to therapy in a nude mouse xenograft model using a molecular probe for repetitive human DNA sequences. Cancer Res 52(10):2791–2796

Soriano P, Jaenisch R (1986) Retroviruses as probes for mammalian development: allocation of cells to the somatic and germ cell lineages. Cell 46(1):19–29

Spencer SL, Berryman MJ, García JA et al (2004) An ordinary differential equation model for the multistep transformation to cancer. J Theor Biol 231(4):515–524

Steel GG, Courtenay VD, Peckham MJ (1983) The response to chemotherapy of a variety of human tumour xenografts. Br J Cancer 47(1):1–13

Stelzl U, Worm U, Lalowski M et al (2005) A human protein-protein interaction network: a resource for annotating the proteome. Cell 122(6):957–968
Stern HM, Zon LI (2003) Cancer genetics and drug discovery in the zebrafish. Nat Rev Cancer 3(7):533–539
Stoner GD (1991) Lung tumors in strain A mice as a bioassay for carcinogenicity of environmental chemicals. Exp Lung Res 17(2):405–423
Suda Y, Aizawa S, Hirai S et al (1987) Driven by the same Ig enhancer and SV40 T promoter ras induced lung adenomatous tumors, myc induced pre-B cell lymphomas and SV40 large T gene a variety of tumors in transgenic mice. EMBO J 6(13):4055–4065
Thomas H, Balkwill F (1995) Assessing new anti-tumour agents and strategies in oncogene transgenic mice. Cancer Metastasis Rev 14(2):91–95
Wachi S, Yoneda K, Wu R (2005) Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. Bioinformatics 21(23):4205–4208
Wallqvist A, Rabow AA, Shoemaker RH et al (2002) Establishing connections between microarray expression data and chemotherapeutic cancer pharmacology. Mol Cancer Ther 1:311–320
Wang H, Huang S, Shou J et al (2006) Comparative analysis and integrative classification of NCI60 cell lines and primary tumors using gene expression profiling data. BMC Genomics 7:166
Wang X, Fu X, Hoffman RM (1992) A new patient-like metastatic model of human lung cancer constructed orthotopically with intact tissue via thoracotomy in immunodeficient mice. Int J Cancer 51(6):992–995
Weerden WM van, Romijn JC (2000) Use of nude mouse xenograft models in prostate cancer research. Prostate 43(4):263–271
Weinstein JN, Myers TG, O'Connor PM et al (1997) An information-intensive approach to the molecular pharmacology of cancer. Science 275(5298):343–349
Wikenheiser KA, Clark JC, Linnoila RI et al (1992) Simian virus 40 large T antigen directed by transcriptional elements of the human surfactant protein C gene produces pulmonary adenocarcinomas in transgenic mice. Cancer Res 52(19):5342–5352
Wistuba II, Bryant D, Behrens C et al (1999) Comparison of features of human lung cancer cell lines and their corresponding tumors. Clin Cancer Res 5(5):991–1000
Wu M, Pastor-Pareja JC, Xu T (2010) Interaction between Ras(V12) and scribbled clones induces tumour growth and invasion. Nature 463(7280):545–548
Zhao B, Magdaleno S, Chua S et al (2000) Transgenic mouse models for lung cancer. Exp Lung Res 26(8):567–579
Zubrod C (1972) Chemical control of cancer. Proc Natl Acad Sci U S A 69(4):1042–1047

Chapter 5
Expression and Genetic Variation Databases for Cancer Research

Johan Rung and Alvis Brazma

Abstract The amount of data generated in cancer research is growing rapidly. High-density array-based technologies, such as genome-wide single nucleotide polymorphism (SNP) genotyping and gene expression microarrays, are producing data that is not only larger in size, but also in complexity, with regard to study design and associated meta-data. This chapter discusses how the flood of genomic and transcriptomic data is managed in databases, often by large collaborative consortia developing new approaches in informatics to maximize the availability and utility of data. Genetic variation databases are most often designed for a particular layer of detail, such as single disease-causing variants associated with specific phenotypes; databases for genome-wide variation, both for SNPs and structural variants; and large repositories for complete genome-wide association studies. Gene-expression microarray data is stored in large repositories, and new services have been developed that take advantage of the increasing number and diversity of stored experiments. By associating data with biological information and integrative analysis, it can be transformed from high dimensionality to a summary level that is directly usable by bench biologists.

5.1 Introduction

Over the years, cancer research has generated large amounts of data and results. This serves not only as a record of the progress of work that has been done, but also as a basis for future research. Thanks to researchers depositing data and recording analysis methods, others in the community are able not only to reproduce results and confirm their validity, but also to re-analyse data in the light of more recent findings and to make new discoveries in old data. It is a core principle in science that work should be conducted with great transparency, and that results should be

J. Rung (✉)
EMBL—European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1SD, Cambridge, UK
e-mail: johan@ebi.ac.uk

A. Brazma
e-mail: brazma@ebi.ac.uk

A. Cesario, F. B. Marcus (eds.), *Cancer Systems Biology, Bioinformatics and Medicine,*
DOI 10.1007/978-94-007-1567-7_5,

held up to the scrutiny of peers. Databases, data management systems and standards that facilitate the exchange of data are crucial tools to achieve that openness.

In this chapter, we will review some of the informatics work and databases that are used for cancer research. Due to the vastness of the resources available, we will focus on human data, but we recognise the huge importance of the research done in animal models (e.g. accessible via Ensembl) and the high quality of resources developed for the different communities.

Our understanding of the effects that gene expression and genetic variation have on human morbidity, and cancer in particular, has grown rapidly in recent years because of technical advances allowing for rapid and parallel assays. In particular, hybridization-based techniques on microarray chips have enabled data to be generated that has revolutionized the fields of genomics and transcriptomics. Handling this explosion of data has been a challenge for bioinformaticians worldwide, and with the current expansion of sequencing-based technologies we predict that this trend will continue for many years to come.

5.2 Genetic Variation

Databases of polymorphisms as genetic markers were in widespread use long before a full human genome reference sequence was first released. Genetic studies, from the earliest linkage studies done in animal systems, to the present boom of genome-wide association studies facilitated by the advances in genotyping and sequencing technologies, depend on maps of variants that can be computationally associated with diseases or other phenotypes.

Online resources for genetic variation of relevance for cancer research can be classified as follows:

i. Single Nucleotide Polymorphisms (SNP) databases
ii. Databases of structural variants
iii. Databases of disease-causing variants
iv. Large-scale repositories for experimental results
v. Reference genomes

With such vast amounts of biomaterial and data being generated, there are many new challenges for informatics (Thorisson et al. 2009). For example, the increased bulk of the data requires new technologies to search and distribute the actual measurements. The privacy of each participating individual must be guaranteed, and ethical considerations have to be taken into account for each case of access and use in a new study, imposing a requirement for technical solutions to ensure proper handling of these issues, as well as solutions that minimize the number of situations where individual privacy is at risk. In addition, the complexity of the data, with so many data types, types of users, and differences in clinical practices, requires new technical solutions for understanding and managing the content without obscuring the scientific value in the complex details.

5.2.1 SNP Databases

The SNP database (dbSNP), hosted at the USA National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/projects/SNP), has emerged as the leading database storing all SNPs identified and submitted, with full genomic information including validation status. The project started in 1998, and the current release, dbSNP build 131, contains 23,652,081 reported human refSNPs. It cross-links to most of the databases storing SNP-based data (Sherry et al. 2001). The main gene and genome databases, such as Ensembl (http://www.ensembl.org) (Hubbard et al. 2009), and Entrez Gene (http://www.ncbi.nlm.nih.gov/gene), also map dbSNP entries. The dbSNP database can be viewed as the central reference for all SNP data and their mapping to the genome.

5.2.2 Databases of Structural Variants

Structural genomic variation, comprising large genetic regions (>1 kb) containing inversions, translocations, deletions or insertions, and in particular copy number variations (CNV), are not as extensively mapped as SNPs. Detection of CNVs was facilitated by the introduction of genome-wide genotyping technology, and recent years have seen an increase in this type of data as well. Since alterations of larger genetic regions are of great importance in most cancer types, databases for structural variation are often cross-linked to cancer specific resources. The Database of Genomic Variants (DGV—http://projects.tcag.ca/variation) is one such database focusing on larger genomic alterations (Iafrate et al. 2004). The release in November 2010 had 101,923 variants, mostly CNVs and insertions/deletions. The database contains data only from healthy individuals and can be thus used as an important reference data source in epidemiological studies. The NCBI database for structural variation, (dbVar—http://www.ncbi.nlm.nih.gov/dbvar) contains data for 433,675 variants in human, collected from 11 studies, as of Mar 2010. This database stores data also for smaller variants, less than 1 kb in size, but not including SNP data. It is dominated by insertion/deletion data from Mills et al. (2006). At the European Bioinformatics Institute (EBI), the Database of Genomic Variants Archive (DGVa—http://www.ebi.ac.uk/dgva/) currently holds 20 studies mapping structural variants in a variety of species. DGVa and dbVar are exchanging data with each other.

DatabasE of Chromosomal Imbalance and Phenotype in Humans (DECIPHER, https://decipher.sanger.ac.uk) (Firth et al. 2009) is a resource for this type of data hosted at the Wellcome Trust Sanger Institute. Its important feature is linking genomic variations to particular cancer types. It makes use of the Ensembl genome browser, and is primarily devoted to analysing at array Comparative Genomic Hybridization (CGH) data, and to mapping variations thus detected. Allowing for restricted access, DECIPHER can also link patient phenotype data to the genomic

variation data. The current version 4.4 contains data collected from over 4400 patients, spanning 59 different syndromes studied in 170 studies.

5.2.3 Databases for Disease-causing Variants

Initial research efforts in genetic epidemiology focused primarily on what was feasible with small sample sizes and panels of genetic markers. Samples from families with rare Mendelian diseases were collected, and using the pedigree structure of the family, linkage studies were designed that correlated a phenotype with strong effects on variation at genetic markers. The data shared from such studies consists commonly of results alone, because there is a reluctance to share expensively assembled sample specific data and phenotypes. Even with sample sizes in biobanks now replete with millions of specimens, we still encounter that reluctance, which is one of the major obstacles in the way of complete openness and reproducibility. Nevertheless, the release of genetic regions linked to Mendelian diseases has proved very helpful in narrowing down the number of candidate genes located in these regions. Thanks to meticulous work by molecular biologists worldwide, it has been possible to find the exact disease-causing genes and genetic variants. For the many cancer types caused by single rare mutations, repositories of these results remain especially relevant and useful.

The Online Mendelian Inheritance in Man (OMIM—http://www.ncbi.nlm.nih.gov/Omim/), is arguably the most complete and well-known resource for knowledge resulting from this type of research. It has grown into a repository not only for Mendelian diseases, but also for the complex polygenic diseases that are now being studied in more detail using genome-wide association studies. Results from the literature have been collected and annotated since the early 1960s.

Examples of databases that contain information linked to specific disease-causing variants include the HGVbase, HGMD, and COSMIC. The Human Genome Variation database of Genotype-to-Phenotype information (HGVbaseG2P, http://www.hgvbaseg2p.org/) has developed from being a catalogue of human genetic variation, to its current state of connecting genetic variation information to phenotypes by matching results from large-scale association studies with the genome and known genetic variants. As of March 2010, it contained data from 366 studies for 631 phenotypes.

The Human Gene Mutation Database, (HGMD—http://www.hgmd.cf.ac.uk), is a database of mutations linked to disease and has been available since 1996. It is now running as a commercial operation with a delayed release to a public version of the database. This delay time is currently 2 1/2 years.

The Catalogue of Somatic Mutations in Cancer (COSMIC—http://www.sanger.ac.uk/genetics/CGP/cosmic) is a gene-centric database run by the Cancer Genome Project at the Sanger Institute. In release r45, it held data from 1,654,274 experiments on 434,364 tumours, specifically the genetic changes detected therein. It focuses on genes that have been published as linked to cancer, and detected on mutations in these genes.

In addition to the above, there are multitudes of locus-specific databases collecting information about mutations and functional evidence to link with disease. These databases are increasingly cross-linked with the larger coordinated databases mentioned earlier, to ensure long-term stability of the collected information. Many smaller databases struggle to achieve stable maintenance and updating, especially where the people or organisations that built and maintained these resources are subjected to changes in employment or funding.

5.2.4 Large-scale Repositories for Experiments

With advances in genotyping technology, the cost per sample and genotype has decreased rapidly and as a result, larger sample sizes can now be scanned in a genome-wide manner using the microarray technologies developed, for example, by Affymetrix and Illumina. This has allowed for studies of complex polygenic diseases with powerful enough tools to detect relatively weak effect sizes compared to the ones observed in many Mendelian diseases. These genome-wide association studies (GWAS) have been highly successful in detecting SNPs associated with diseases. Among the early ground-breaking GWASs for cancer types, Easton and Thomas detected SNPs associated with breast and prostate cancer (Easton et al. 2007; Thomas et al. 2008). The data consists of raw intensity data for the hybridization at each spot on the microarrays, forming a basis for genotypes from each sample and SNP to be inferred by applying classification algorithms to the intensity data. Today, the typical association study investigates effects of millions of SNPs on phenotypes in tens of thousands of samples. For example, the Wellcome Trust Case Control Consortium (WTCCC) published in 2007 studies for seven complex diseases, tested on 14,000 cases and 3000 controls (WTCCC 2007). The data is stored at the European Genome-phenome Archive (EGA—http://www.ebi.ac.uk/ega/). This archive, and the GAP database (dbGAP) at the NCBI, are the two main resources intended for general exchange of large scale genotypes and associated data.

The European Genome-phenome Archive (EGA) is hosted at the European Bioinformatics Institute (EBI) and has been designed to meet the demands on data security and ease of submission and accession placed on this type of data. Accession to data stored in EGA is approved by the data owners themselves through data access committees that can vary in structure. The first data in EGA was deposited by the Wellcome Trust Case Control Consortium as mentioned above. A second-phase data collection by WTCCC, testing approximately 120,000 individuals, will also be deposited in EGA.

These large repositories have encountered a multitude of technical challenges. The sheer bulk of data means that storage and transfer has to be approached differently than previously. Data transfer protocols that are more efficient for large files than TCP-based protocols have been employed to this effect; one such is Aspera (http://www.asperasoft.com). Indeed, sometimes the maximum ‘bandwidth’ is reached by physically sending hard drives with data by courier.

The other challenge we face is to protect the privacy both of the genotyped individuals and of the intellectual property contained in the data. Access within Europe is now mainly handled by data access committees for each cohort; these consider applications from researchers who want access to the data. Encryption and secure transfer protocols are used to protect the data. It is regrettable that very little data is entirely public, but restrictions can be justified where they are intended for the protection of individual privacy. There is an ongoing discussion about the usefulness of conclusions drawn from what can be concluded about a specific individual's participation in studies for which only summary data, such as association significance and effect size, is available (Homer et al. 2008).

5.2.5 Reference Genomes

Since 2001, when the first full draft of the human genome sequence was released, our understanding of the variation in the human genome has been remarkably enhanced. The importance of the reference genome cannot be understated—almost all human genetic studies today rely on it. Nevertheless, long before this, it had become clear that patterns of genetic variation were dependent on population structure and ethnicity. Soon after the completion of the human genome, therefore, the HapMap project (http://www.hapmap.org) set out to map variation in a number of reference populations. In the first phase of the project, 270 individuals in trios of mother-father-child from four different populations were genotyped at approximately 1.1 million SNPs. In phase 2, the same individuals were more densely genotyped, totalling along with phase 1 approximately 3.8 million SNPs. The latest phase expanded the project by testing 1.6 million SNPs in 1115 individuals of 11 different populations (International HapMap Consortium 2005; International HapMap Consortium et al. 2007). Based on this data, we have been able to infer haplotype structure for the different populations, and estimate recombination maps. This information is very valuable in association studies, since it is enough to tag genetic regions containing blocks of SNPs, with highly correlated genotypes involving a small number of tested SNPs. The causal SNP can thus be detected by proxy of a tested highly correlated SNP. The haplotype structure of a reference population also serves to impute genotypes at SNPs that were typed in the reference set, but not in the test set. Genotype imputation provides probabilities for each genotype at a non-typed SNP and can be tested in association scans alongside the actually assayed ones (Howie et al. 2009).

With sequencing experiments similarly becoming less expensive it has been possible to undertake major endeavours like the 1000 genomes project (http://www.1000genomes.org). This was initiated to determine the full genome sequence of at least 1000 individuals from approximately 20 different populations. Currently, data for a number of pilot projects has been collected, for example the re-sequencing of four individuals at approximately 30× depth, and the sequencing of a higher number of individuals at less depth. With quarterly releases, this data is fully open and will become a highly important reference set for years to come.

The possibility of sequencing the whole genome of an individual is opening up new avenues for cancer research. Recent studies have employed sequencing technology to generate complete maps of somatic mutations in cancer cells (Pleasance 2010a, b) compared sequences from malignant melanoma and lymphoblastoid cell lines from the same person, and sequenced DNA from a lung cancer cell line. We anticipate more studies in the near future comparing the full genome of tumour cells from various cancer types with normal surrounding tissue, thereby hopefully detecting somatic mutations associated with the specific cancer type.

5.3 Gene Expression

Research relating various cancers to changes in gene expression is not new. Microarrays were the first technology that allowed the study of such relationships on a genomic scale, and indeed some of the very early microarray experiments were aimed at using gene expression for molecular classification of leukaemias (Golub et al. 1999) and predicting survival rates in lymphomas (Alizadeh et al. 2000). Cancer has been one of the main research subjects for microarray gene expression assays ever since. As the result of the MIAME initiative (Brazma et al. 2001), data related to most published papers is now deposited at public repositories: ArrayExpress (Parkinson et al. 2009) or Gene Expression Omnibus (GEO) (Barrett et al. 2009). In February 2010, the ArrayExpress public repository of functional genomics data became available; it consists of some 1600 experiments comprising over 83,000 assays related to cancer. Data from these repositories is being used to build added-value databases, such as Oncomine (Rhodes et al. 2004) or Gene Expression Atlas (Kapushesky et al. 2010).

In Sect. 5.3.1, we distinguish between public repositories (or archives) and added-value databases containing cancer research-relevant gene expression data. The added-value databases can be separated into commercial and publicly available databases. We also consider specialized cancer databases, and general databases that contain large amounts of cancer-related data. We will concentrate mostly on databases that are freely available.

5.3.1 Archives of Gene Expression Data

There are two major archives (public repositories) for high-throughput gene expression data: ArrayExpress (Parkinson et al. 2009) and GEO (Barrett et al. 2009). Cancer-related gene expression data is the major contributor to both databases. Thanks largely to the activities of the Functional Genomics Data Society (formerly MGED society), most journals now require data submissions to one of these archives as a condition of publication (Ball et al. 2002). In consequence, the size of these databases has been growing rapidly; they now contain data from well over 10,000 ex-

periments (studies) and 300,000 microarray or high-throughput sequencing-based assays.

Since ArrayExpress imports all recent data from GEO, we will concentrate on describing this database. The query to find all cancer-related data in ArrayExpress is extremely simplified (interface page at http://www.ebi.ac.uk/arrayexpress). The database uses a cancer nomenclature based on the National Cancer Institute (NCI) Meta-thesaurus (developed at the National Institute of Health), which allows the database to be searched semantically rather than by pure key-word. Thus, a search for 'cancer' by using ontology expansion would allow the user to find data from these cancer experiments that do not explicitly contain the key-word "cancer".

There were 2921 cancer related experiments in ArrayExpress as of 14 October 2010. Each experiment is presented by an expandable brief description. Experiment E-TABM-909, for instance, describes 'miRNA profiling of human acute myeloid leukaemia Kasumi-1 cell line CD34+38– compartment'. It contains data from eight microarray based assays, where four different phenotypes 'CD34–', 'CD34+CD38+', 'CD34+CD38–', 'CD34–,Kasumi total' are assayed, each in two individuals. A link to publications related to this dataset is given for more information.

The user can download the data and find, for example, the complete experimental description in MAGE-TAB format (Rayner et al. 2006), which can be directly uploaded in the popular microarray data analysis tool Bioconductor (Kauffmann et al. 2009). Users can then analyse this data in combination with their own private data, or incorporate them in their own local data warehouses. It must be noted, however, that these public archives are mostly aimed at gene expression data experts; they do not directly allow one to query for genes associated with particular types of cancer. Therefore secondary, added-value databases aimed at cancer researchers and 'bench biologists' are being built. These are discussed in the following section.

5.3.2 Added-value Databases

The two added-value gene expression databases that are most relevant to cancer research are Oncomine (Rhodes et al. 2004) and Gene Expression Atlas (Kapushesky et al. 2010). (Those fully commercial databases, which are not accessible to the authors of this chapter, will not be examined here). Both Oncomine and Gene Expression Atlas have as their goals the transformation and presentation of high-throughput gene expression data in a form accessible to users who are not experts in microarray or other high-throughput data analysis. The essential functionality of these added-value databases is to facilitate queries for genes and for particular disease states, and to present the findings in user friendly formats.

Oncomine integrates cancer data from various microarray platforms with data mining tools. Differential expression analyses comparing most major types of cancer with respective normal tissues, as well as a variety of cancer subtypes and clinical-based and pathology-based analyses, are available for exploration. Data can be

queried and visualized for a selected gene across all analyses or for multiple genes in a selected analysis. Furthermore, gene sets can be limited to clinically important annotations including secreted, kinase, membrane, and known gene-drug target pairs, to facilitate the discovery of novel biomarkers and therapeutic targets.

Oncomine has two editions, Oncomine Research Edition, which is free to not-for-profit and academic cancer research communities; and Oncomine Research Premium Edition, a subscription version that includes additional features such as multi-gene search, expanded analyses, custom concept upload, quarterly data updates, export capability, and enhanced support. Oncomine can be found on https://www.oncomine.org/resource/login.html.

The other important added-value database, the Gene Expression Atlas at the European Bioinformatics Institute (EBI—http://www.ebi.ac.uk/gxa/), is a semantically enriched database of meta-analytical summary statistics over a curated subset of ArrayExpress Archive. It serves queries for condition-specific gene expression patterns as well as broader exploratory searches for biologically interesting genes or samples. This database is not specific to cancer; it contains gene expression data on other diseases and various biological processes. However, cancer-related experiments are among the major data contributors. In February 2010, Gene Expression Atlas contained data from 59 cancer-related experiments comprising data from 5990 microarray assays. Interestingly, over 16,000 human genes show differential expression in at least one cancer-related study, under at least one specific condition.

The database allows the user to query for a gene of interest; and/or a pathway of interest, and under what specific conditions it (the gene of interest) has been shown to be over- or under-expressed, in at least one study. For instance, the user can find which genes of the 'notch' pathway are differentially expressed in any of various carcinomas. He can then zoom into any of the individual studies, necessarily providing evidence that a particular gene is over- or under-expressed in a particular carcinoma.

5.4 Informatics Coordination by International Consortia

A number of large international consortia have been formed to actively generate data in cancer studies and to coordinate informatics efforts benefitting the whole community. Both DNA variation data and gene expression data are being generated.

Among these, the International Cancer Genome Consortium (ICGC—http://icgc.org) is a global consortium that aims at mapping somatic genetic alterations, changes in gene expression, and epigenetic modifications for 50 cancer types. Its goal is to generate data for somatic changes in cancer as compared to healthy controls, and do it with high resolution and completeness in the genome, down to relatively uncommon variants (allele frequency > 3%). Under the umbrella of the consortium, a number of similar cancer genome projects are running in parallel. In addition to the data and results themselves, there are benefits on the informatics side from having such large consortia coordinating the work. One such benefit is the incentive to

Table 5.1 Cancer related bioinformatics databases

Resource	URL	Comment
ArrayExpress	http://www.ebi.ac.uk/arrayexpress	Central repository for gene expression data (EBI)
EGA	http://www.ebi.ac.uk/ega	Central repository for association studies (EBI)
HapMap	http://www.hapmap.org	Dense genotype mapping of genetic variation in reference populations
1000 genomes	http://www.1000genomes.org	Sequencing and dense genotyping in large reference populations
dbSNP	http://www.ncbi.nlm.nih.gov/projects/SNP	The central SNP database
Ensembl	http://www.ensembl.org	Genome browser and genomic database
HGVbaseG2P	http://www.hgvbaseg2p.org	Mapping genetic variation and results from association studies
HGMD	http://www.hgmd.cf.ac.uk	Mutation database in commercial and delayed-release academic versions
COSMIC	http://www.sanger.ac.uk/genetics/CGP/cosmic	Database of somatic mutations in cancer
DGV	projects.tcag.ca/variation	Database of mostly copy number variants and insertions/deletions
DECIPHER	decipher.sanger.ac.uk	Database of structural variants and their links to diseases
GEN2PHEN	http://www.gen2phen.org	Consortium for development and coordination of informatics for association studies
International Cancer Genome Consortium	icgc.org	Consortium for the complete mapping of cancer types
The Cancer Genome Atlas	cancergenome.nih.gov	Consortium for the complete mapping of cancer types
GEO	http://www.ncbi.nlm.nih.gov/geo/	Central repository for gene expression data (NCBI)
ONCOMINE	http://www.oncomine.org	Service for analysis of cancer data cantered around gene expression experiments
OMIM	http://www.ncbi.nlm.nih.gov/Omim	Literature based database of genes and diseases
Gene Expression Atlas	http://www.ebi.ac.uk/gxa	Database of gene expression associations to experimental factors
dbVar	http://www.ncbi.nlm.nih.gov/dbvar	Database of structural variants, mainly insertions/deletions

develop standards for data and experimental annotation, making sharing easier for the entire research community.

Another initiative related to the ICGC is The Cancer Genome Atlas (TCGA, http://cancergenome.nih.gov). It is led by researchers at the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI). With

goals and methods similar to the ICGC, the TCGA deals with 20 different cancer types and hosts informatics resources coordinating data management in the consortium and making data accessible through public databases.

GEN2PHEN (http://www.gen2phen.org) is a consortium of European partner institutions funded by a European Union 7th Framework Programme for Research (EU FP7) grant. The opening of the research area was the result of a related workshop (ftp://ftp.cordis.europa.eu/pub/lifescihealth/docs/geneticvariationworkshopfinalreport_200604.pdf). One of the project's goals is to develop and integrate informatics resources for genetic variation that target the association of genotype to phenotype, also unifying access to human and model organism databases. This consortium is focused on the informatics rather than generating new experimental data. It unifies data from the biology community, making it accessible via ENSEMBL by developing standard Locus Specific Databases (LSD) packages that are now in widespread use for linking to the clinical community.

The data-generating projects are establishing their own databases. ICGC is working on developing a federated database approach, whereby each of the partners would host their data on their own site, and database federation technology would be applied to enable partners to access each other's data for meta-analysis purposes. It is important to realise that none of these databases provides any kind of 'final truth' about the involvement of the particular genes in particular cancers. Evidence from multiple sources needs to be combined before any conclusions can be drawn, and databases are only one evidential source. Several cancer-related bioinformatics databases are shown in Table 5.1.

References

Alizadeh AA et al (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 403:503–511

Ball CA et al (2002) Standards for microarray data. Science 298:539

Barrett T et al (2009) NCBI GEO: archive for high-throughput functional genomic data. Nucleic Acids Res 37(Database issue):D885–D890

Brazma A et al (2001) Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. Nat Genet 29:365–371

Easton DF et al (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 447:1087–1093

Firth HV et al (2009) DECIPHER: database of chromosomal imbalance and phenotype in humans using ensembl resources. Am J Hum Genet 84:524–533

Golub TR et al (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286:531–537

Homer N et al (2008) Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. PLoS Genet 4:e1000167

Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 5:e1000529

Hubbard TJP et al (2009) Ensembl 2009. Nucleic Acids Res 37(Database issue):D690–D697

Iafrate AJ et al (2004) Detection of large-scale variation in the human genome. Nat Genet 36:949–951

International HapMap Consortium (2005) A haplotype map of the human genome. Nature 437:1299–1320
International HapMap Consortium et al (2007) A second generation human haplotype map of over 3.1 million SNPs. Nature 449:851–861
Kapushesky M et al (2010) Gene expression atlas at the European bioinformatics institute. Nucleic Acids Res 38(Database issue):D690–D698
Kauffmann A et al (2009) Importing arrayexpress datasets into R/Bioconductor. Bioinformatics 25:2092–2094
Mills RE et al (2006) An initial map of insertion and deletion (INDEL) variation in the human genome. Genome Res 16:1182–1190
Parkinson H et al (2009) ArrayExpress update—from an archive of functional genomics experiments to the atlas of gene expression. Nucleic Acids Res 37(Database issue):D868–D872
Pleasance ED et al (2010a) A small-cell lung cancer genome with complex signatures of tobacco exposure. Nature 463:184–190
Pleasance ED et al (2010b) A comprehensive catalogue of somatic mutations from a human cancer genome. Nature 463:191–196
Rayner TF et al (2006) A simple spreadsheet-based, MIAME-supportive format for microarray data: MAGE-TAB. BMC Bioinformatics 7:489
Rhodes DR et al (2004) ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia 6:1–6
Sherry ST et al (2001) dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 29:308–311
Thomas G et al (2008) Multiple loci identified in a genome-wide association study of prostate cancer. Nat Genet 40:310–305
Thorisson GA, Muilu J, Brookes AJ (2009) Genotype-phenotype databases: challenges and solutions for the post-genomic era. Nat Rev Genet 10:9–18
Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447:661–78

Chapter 6
Education and Research Infrastructures

Anna Tramontano and Alfonso Valencia

Abstract Until very recently, the acquisition of data through experimentation was the major bottleneck to research in cancer. The recent explosion in high-throughput methodologies radically changed the situation, so that data analysis and interpretation rather than data acquisition is now the greatest challenge, and bioinformatics tools and services as much as the experimental apparatus, are an integral part of the molecular scientist's workbench.

The earliest field to have experienced this revolution was certainly molecular biology, followed soon after by cell biology, owing to significant advances in systems biology. Another area of science significantly affected by the data deluge is clinical research. Within the clinical framework, cancer research is perhaps the field that has been revolutionized to the greatest extent in recent years, as high-throughput technologies can now be used to identify sets of genes potentially related to different processes in cancer. Finding relationships between the molecular and genomic information and the clinical information available, within the domain of medical informatics, is currently driving the development of translational research in biomedicine.

6.1 The Challenge

It is obvious that the key to success in cancer research is the integration of tools and methodologies from the fields of bioinformatics, computational biology, systems biology, chemical biology and medical informatics into mainstream cancer biology research. It is also clear that this cannot be achieved without an appropriate effort to educate new professionals and to train end users.

Cancer scientists need to gain a grasp of the plethora of available tools and data repositories, and to understand the capabilities and limitations of these resources, in order to effectively plan their experiments, interpret their results and make use of the data being generated in world-wide efforts.

A. Tramontano (✉)
Department of Physics, Sapienza University of Rome, P.le Aldo Moro, 5,
00185 Rome, Italy
e-mail: Anna.tramontano@uniroma1.it

A. Cesario, F. B. Marcus (eds.), *Cancer Systems Biology, Bioinformatics and Medicine,*
DOI 10.1007/978-94-007-1567-7_6,

Within this framework, the main cancer genome projects, i.e. the Sanger Institute Cancer Genome Project (http://www.sanger.ac.uk/genetics/CGP/Census/), the Cancer Genome Atlas (TCGA—http://www.cancergenome.nih.gov/), and the International Cancer Genome Consortium (ICGC—http://www.icgc.org) are using high-throughput sequencing to define somatically acquired sequence variants/mutations and to identify those genes critically implicated in the development of human cancers (see Chap. 5). Indeed, a very large list of cancer-related genes has been published for various cancer types, and a growing number of cancer-oriented databases have being reported in recent years. See Baudot et al. (2010) for an updated list of cancer resources.

The ultimate goal of these projects is to produce a comprehensive list of validated mutations that can be used as risk markers for early disease detection, as targets for the development of anti-cancer drugs and as prognostic indicators of which drugs are more effective for treatments. All of this constitutes the framework for the individualized treatment of cancer in the context of the general personalized medicine perspectives, i.e. a therapeutic strategy that takes into account direct genetic sequence information for each individual.

It is increasingly clear that the development of effective treatments will have to consider systems biology perspective, rather than focusing exclusively on specific genes and proteins, and that it should take pathways and networks into account. For example, analysis of pancreatic cancer (Jones et al. 2008) has clearly shown the different pattern of cancer-associated variations among individuals, and demonstrated how these variations can be better interpreted in terms of signalling pathway and functional associations, rather than at the level of individual mutations and genes (for a review see Baudot et al. 2009). Indeed, in a follow-up study, it was possible to demonstrate the effectiveness of using a specific inhibitor for a target gene different from the mutated one, but belonging to the same pathway (Hidalgo 2010; Villarroel et al. 2011).

This example underlines the need for analysing cancer data in a general context and for using systems biology approaches and concepts (see also Chap. 7 and 8). It also poses interesting new scientific challenges for the integration of disease information, molecular and genomic data with drug targets and pharmacological information. These new approaches and challenges will certainly require a close collaboration across fields. Specific training will be needed in the use of those computational methods and tools required for the representation and interpretation of genomic data at the systems level, i.e. molecular interaction networks and pathways, gene regulation control systems, drug target relationships and disease/symptoms representation, all of which should be linked to the underlying experimental data, databases, publications and other sources of information.

No less relevant are the scientific challenges of integrating the goals and glossary of clinicians, pathologists, genomics and bioinformatics scientists who need to be able to collaborate effectively in this complex and demanding environment.

This not only increases the complexity of the expertise required of scientists, but it also poses a challenge to the developers of tools and resources, because of the

wider and more diversified user base. As one gets closer to the translational aspects of the research, the scope of users not only encompasses molecular and cell biologists, but also reaches out to much larger communities, such as clinical researchers, medical doctors, nurses and palliative care personnel, pharmaceutical chemists and, ultimately and more importantly, the general public, both journalists and people generally interested in the life sciences, and (significantly) cancer patients and their families.

As a side product of this new way of doing science, and given the tremendous impact that the Web is having on all aspects of human health and especially cancer-related communities, the role of each of these groups of stakeholders is changing. In this chapter we will discuss what has to be expected in the near future in terms of their needs, and how the scientific community, in particular the computational life scientists, can effectively respond to the challenge of enabling them to make the most out of the available resources.

6.2 The Actors

6.2.1 Molecular and Cell Biologists

The molecular biology field, and more recently cell biology, have been benefitting from computational biology and bioinformatics efforts. Indeed, it is generally the case that as much as half of a molecular biologist's time is spent sitting at a computer rather than at the workbench, so as to take advantage of the many resources available for retrieving molecular data and analysing them.

Although the symbiosis between molecular and computational biology dates back many years, with most of the efforts of the computational community tailored to the biology community, there is still a crucial issue of training and educating bench scientists. While the tools for analysing molecules individually, for example to attempt the prediction of their functional and structural properties, are now routinely employed and well understood by almost every molecular biologist, workers tackling the problem of analysing the data at a system-wide level are not yet being educated effectively. On the other hand, cancer research is the prototypical application of high-throughput technologies, and is the focus of many molecular and cell biology efforts.

"Now, here, you see, it takes all the running you can do, to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that!" the Red Queen says to Alice; and this is probably the most expressive metaphor describing the current relationship between experimental and computational biology. New technologies enable experimentalists to gather more and more diverse sets of data, while computational biologists try to keep up by developing novel tools which cannot but be more complex, given the complexity of the problems at hand. This

implies that users often lag behind in achieving a sufficient level of understanding of the tools to be able to use them effectively, while at the same time, developers would need to "run at least twice as fast as that" to keep up with the technology and to be able to offer simple user-friendly systems for ever much more complex problems.

6.2.2 Chemical Biologists

Chemistry and biology have traditionally been considered different, although adjacent fields of science, and until recently, the scheme of things was that the molecular scientists would identify the target biomolecule, set up an assay for testing the function of putative interfering molecules, and leave all the rest to pharmaceutical and computational chemists. The whole process was focused on one biomolecule at a time, usually tested *in vitro* against a library of chemical compounds.

The rationale behind the building of this "fence" between the two fields is probably that a number of the properties required for developing an inhibitor into a drug are not well understood, and are often better predicted on the basis of the chemist's experience rather than on the basis of computational or molecular biology tools.

However, it is very interesting to observe to what extent this relationship is changing, mostly because of the possibility of running fast high-throughput experiments at the systems rather than single-molecule level; the increased accessibility of cellular and animal models that facilitate *in vivo* testing; and the availability of suitable and accessible chemical libraries.

From an experimental perspective, automation or semi-automation of protein preparation and of detection tools, such as monoclonal and library-derived antibodies, makes it possible in most cases to obtain the target molecules rapidly, efficiently and in a sufficiently pure state to run automated screening of compounds. Advances in functional genomics and cell biology techniques also offer ways of testing compounds at the systems level, thereby permitting complex diseases such as cancer to be targeted.

As an interesting example, the European Molecular Biology Laboratory (EMBL) includes among its core (i.e. service) facilities, protein and monoclonal antibody preparation as well as chemical biology. This implies that, in principle but also often in practice, a user of these facilities could start from a database accession code of a protein, have it cloned and purified, obtain antibodies for purifying it and submit it to a screening for small molecule inhibitors, almost without ever handling the protein!

From a computational point of view, a major revolution has resulted from the availability of databases of chemicals, which were previously rather expensive and therefore only really accessible to non-academic bodies, as well as from the free usage of docking and small molecule optimization tools. This is attested by the interest of major data resource suppliers, such as the European Bioinformatics Institute

(EBI-EMBL), in including chemical libraries such as ChEBI in the portfolio of their available databases and services.

The convergence between the two disciplines of molecular and chemical biology is bound to have a large impact on translational research, and consequently on cancer research, but once again, appropriate training will be required for using the computational resources.

6.2.3 Clinical Oncology Researchers

As already discussed for biology and chemistry, science often progresses through the merging of disciplines originally considered very different. This applies for clinical research and molecular biology/genomics. An increasing number of biologists work in applications closer to therapy and medical practice, and more and more clinicians are developing research projects strongly based on molecular biology and genomic approaches. Both groups need to develop interdisciplinary approaches taking advantage of the expertise that has been accumulated in the other area. The more complex the problem, the more pressing becomes the need for an interdisciplinary approach. This is indubitably the case for translational cancer research, which is in fact one of the areas where the coming together of biomedicine, genomics and bioinformatics has progressed more swiftly.

The increasing application of high-throughput technologies in oncology (DNA microarrays, protein arrays, high-throughput genotyping, mass spectrometry and, more fundamentally, the enormous impact of Next Generation Sequencing technology) has fundamentally altered the approaches to analysis of the molecular basis of cancer even in clinical settings, and it is obvious that the profile of a clinical researcher needs to change accordingly.

The characterization and diagnosis of cancer was traditionally conducted by analysing a small fraction of the genes and proteins related to cancer, usually guided by the identification of chromosomal abnormalities. It is now possible, thanks to data provided by linkage studies in cancer-prone families, or through known functional characteristics of individual genes or gene families, for a clinical researcher to design studies that interrogate the sequence, structure and function of every gene and protein in a comprehensive approach. This allows the development of a multidimensional view of cancer, a phenomenal task ahead of us that will certainly require new developments in computational and systems biology far beyond what is currently available.

As for treatment modalities, the new technologies offer the prospect of achieving personalized cancer treatment by finding the right drug for the right patient. Here, however, the challenge is to orient oneself among the many putative markers reported in the literature, and the many new markers being discovered by the ongoing cancer genome projects.

Ideally, the personalized cancer treatment pipeline implies an initial phase of testing for markers among a small set of known ones. This can lead to first-line

treatment, followed by a more advanced phase, including putative markers for genes or pathways for drugs in phase I and II clinical development, which are used to direct patients to a target-guided clinical trial. Finally, one can perform a global genomic analysis method such as gene expression profile, assessment of copy number, exonic sequencing, or full genome sequencing. These results do not need to be available in real time and may provide, when linked to clinical data, the opportunity to discover biomarkers or signatures of drug response. Indeed, it is not uncommon to find cancer patients who, relying on the gathered information, are willing to pay for the full sequencing of their tumour, something that is actually possible at the technical level but that, unfortunately, cannot be translated directly into prescriptions and treatments, given our current level of knowledge and understanding.

The application of these concepts to clinical practice is, as yet, far from optimal. Issues such as lack of technological, financial, and human resources to accomplish the multiple tasks needed for such an approach are clearly important, but equally important are the needs for information about the potential of the approach, training in interpreting the results, and easy accessibility of the required resources.

The Bioinfomed study (Martin-Sanchez et al. 2004) provided an overview of the degree of collaboration and interaction between medical informatics and bioinformatics. Based on a literature review, the study concluded that collaboration between the two disciplines was rather infrequent, despite its desirability in view of the fact that the increase in genetic data will have significant consequences for health care in the coming years. The study argued that several issues need to be addressed in order to deal with the impact of genetic data on healthcare, ranging from changes in electronic patient records to include genetic data, development of appropriate clinical decision support systems to aid practitioners, and re-assessment of legal and ethical frameworks relating to genetic data.

In the Symbiomatics study, conducted a few years after the Bioinfomed analysis, the degree of interaction between these two disciplines was reassessed. Although more collaboration between the fields was observed (as witnessed by an increasing number of joint publications), the fields seem to have remained distinct and separate. Symbiomatics also provided a prioritized list of research topics at the interface of the two disciplines (Rebholz-Schuhman et al. 2007).

The need to facilitate interdisciplinary research has also resulted in a number of European projects that focus on multiscale modelling (the most ambitious being the Virtual Physiological Human), (http://www.vph-noe.eu); these emphasize the need to combine the contribution of various disciplines in a single, interrelated and coherent environment.

For interdisciplinary approaches to become a reality in everyday practice, there is a particular need to access large-scale information, as well as to get sequencing results analysed in a consistent, reproducible and systematic way. The data clearly has to be presented in a mode appropriate for use by clinical doctors, which will require a phenomenal effort both to make the technology accessible and, at the computational level, to face the challenges of data analysis, information management (including genomics, medical and drug information), and representation of the data.

Only if these latter issues are satisfactorily resolved will it be possible to transform deeply ingrained methods of oncological practice.

The requirement for cross-training between these disciplines and others, and the acquisition of a common language, is an essential desideratum for the future of cancer research. We badly need a standardized nomenclature for clinical and biomedical terms. It is still very difficult to interpret systems-based information on diseases and symptoms, a fact which clearly hampers our ability to handle clinical information and to relate it to molecular biology/genomic data (see Gómez-López and Valencia 2008). This is true despite various serious efforts, such as that of the US National Library of Medicine, which resulted in the development of the Unified Medical Language System (http://www.nlm.nih.gov/research/umls), a repository of biomedical vocabularies with over 2 million names for some 900,000 concepts from more than 60 families of biomedical vocabularies. These include the controlled vocabulary of the Medical Subject Headings thesaurus (http://www.ncbi.nlm.nih.gov/mesh), the Systematized Nomenclature of Medicine-Clinical Terms (SNOMED—http://www.ihtsdo.org), with more than 357,000 hierarchically organized concepts with their definitions, and the Disease Ontology (DOID—http://diseaseontology.sourceforge.net/).

Needless to say, linguistic, ethical and legal issues constitute additional and very serious restrictions to progress in this area.

6.2.4 General Public

Twelve million people will be diagnosed with cancer this year and will experience tremendous physical and psychological stress. Most of the patients and their relatives will try to gain information through friends, newspapers and the Web. Cancer research reporting as disseminated by main-stream media is not always sufficiently accurate to be helpful. Coverage is very variable in reliability, and identification of trustworthy sources of information for patients is not easy. The situation for cancer professionals and advocacy groups is also fraught with difficulty. For example, about 5% of the Web pages dedicated to breast cancer information contain inaccuracies, and sites relating to complementary and alternative medicine are estimated to be 15 times more likely to contain false or misleading health information (Esquivel et al. 2006). It is obviously very difficult, but at the same time very important, to distinguish between accurate and inaccurate sites. Furthermore, the heterogeneity of outcomes of cancer care throughout Europe is reflected in differences in access to information and services, with detrimental social and economical consequences. Another major concern is the difficulty in understanding the special forms of English that abound on the Web (Mancini et al. 2006).

Scientists and society in general have a responsibility to supply patients with the correct information, and to help them to discard false and inaccurate sources. Overall, this is a situation that widespread access to the Internet has made more pressing, a serious challenge.

6.3 Training and Education of the Stakeholders

As noted above, there is a variety of people who need to be enabled to access information and distil its meaning. In the case of scientists and doctors whether they be molecular and/or chemical biologists, oncologists and/or surgeons, it is possible to devise an educational curriculum supplemented by training resources; different strategies, however are needed for the general public.

What we will discuss applies to all life scientists, not just those studying cancer, but cancer scientists will probably be those facing the biggest challenge in the near future, because of the disease's complexity, and are therefore most in need of an adapted, modern and diversified education.

In the following sections, we will first describe what we believe are the necessary skills to be included in the formal education of biologists, chemists and clinicians. We will next discuss how to tackle the problem of the 'moving target', i.e. how to make sure that sufficient attention is given to training in the usage of the ever-evolving tools and resources available. Finally, we will provide some suggestions about the problem of conveying reliable information to the general public.

6.3.1 The Core Subjects in Modern Scientific Education

Modern science is about information, and a successful scientist is one who is able to transform it into knowledge, understanding and coherent action. But the amount of information is enormous, and equally impressive is the diversity in the type of data that biology and medicine continuously gather.

The obvious question that arises, in this and other fields, is whether education should aim at creating generalists or specialists, and at what stage one should specialize in her/his own field.

Educational systems are more and more going towards producing specialists in science, but, although there is proven value in specialization, the amount of complex information, fragmented in so many ways and developing faster and faster, renders it increasingly important for people to develop transdisciplinary ideas for understanding life and medical sciences. This can only be achieved by providing the new generation of scientists with a broad background in as many disciplines as possible. Furthermore, the more reference points a scientist has, the easier will it be to understand something new by building on pre-existing knowledge, just as a piano player has an advantage over others in learning how to play other instruments.

Molecular biologists, geneticists, chemists, and medical doctors need to be trained in bioinformatics, systems biology and chemical biology during their undergraduate studies. They should be sufficiently competent in data analysis and statistics to make sense of complex data without over- or under-interpreting it. They should also be able to navigate through databases of genes, genomes and proteins and understand the rationale behind basic tools for molecular, structural and func-

tional data analysis. Most importantly, they should be aware of the possibilities that computational and systems biology opens to them. At the end of their undergraduate studies, students should feel at ease with the tools they will undoubtedly need to use in their future career, be that in academia or elsewhere.

One extremely relevant point that we would like to stress is that biology and medicine have been considered for a long time non-exact sciences. This is manifestly changing, and this should be made clear to students. Far too often, young people who want to pursue a career in the biological or medical area are led to believe that they can do so without having a solid mathematical, physical and chemical background. This encourages them to underestimate the importance of these subjects during their course of study. The absence of quantitative approaches was already less than optimal in the past, and is unacceptable now.

At the graduate level, a certain degree of specialization is unavoidable in terms of specific research area, but general education in related fields should not be neglected, but rather promoted by supervisors and actively pursued by students. There are many examples showing that integration of data and ideas from related fields is the key to success in modern science.

To quote Abraham Lincoln: 'The dogmas of the quiet past are inadequate to the stormy present. The occasion is piled high with difficulty, and we must rise to the occasion. As our case is new, so we must think anew and act anew.' And the only way to 'think anew' is to explore different ways of approaching the problems and the challenges. New paradigms have also to be explored in educating the next generation of computational and systems biologists, upon whom the burden now falls, and will fall even more heavily in the future, of integrating the data and the resources and making them available in the context of the different fields.

In the recent past and in some countries even now, education of computational biologists and bioinformaticians has followed one of two routes: training of a qualified biologist in computer science, or tutoring of a computer scientist in the rudiments of biology. This was not very satisfactory, and indeed the two types of scientists usually pursued different routes in science, with the former mostly being an expert in using the computational tools, and the latter in developing them. It is rather obvious that the present-day complexity of computational biology and bioinformatics demands that education of these scientists should no longer consist of the conjoining disciplines, but become a curriculum in its own right. This has in fact already begun to happen. Similarly, in the near future we need to educate system biologists and medical informaticians. In this case, although a background in bioinformatics and computational biology is obviously required, other skills are necessary, such as the ability to optimize storage and analysis of a very large amount of data and to analyse and simulate very complex data structures, and/or obtain a deeper understanding of the ins and outs of the relevant medical systems. Clinical research has the additional problems of maintaining patient confidentiality and of reporting the results of clinical trials on new molecules and devices from the pharmaceutical sector, the latter creating associated conflict of interest and competences that need to be taken into account when developing tools and resources. We believe that this type of education is best delivered at the undergraduate level, with gradu-

ate students being already sufficiently knowledgeable to tackle new challenges, and prepared to deepen their understanding of the new types of data that will undoubtedly soon be available. This can only be achieved effectively if a solid background is already in place.

Finally, it is essential for a translational and clinical scientist to be able to combine existing data resources to generate dedicated data-sets on demand (e.g. integrating different types of data for a given disease model). It is clear that the complexity of cancer-related problems demands the collaboration of scientists with very different backgrounds and research environments—medical doctors, biologists, engineering, chemists, computer scientists, and others.

6.3.2 User Training

While education in biological sciences is generally excellent, and homogeneous enough to allow scientists to communicate and collaborate effectively, the situation with respect to bioinformatics user training is less ideal.

The services commendably supplied by many laboratories add to the confusion by creating a plethora of available tools. To make things worse, sometimes these present technical differences and nuances that are difficult for an inexperienced user to appreciate. Often the training material is far from complete, not sufficiently updated, and deployed in a language not necessarily tailored to the end users.

There are several challenges in addressing user training for computational and systems biology resources, arising out of the rapid evolution of bioinformatics resources, and the lack of a centralized body to which users can resort. Another issue concerns the lack of recognition of the training aspect of bioinformatics and systems biology resources, which discourages developers and trainers to invest in this aspect.

It is obvious that failure to provide a proper user-training infrastructure could prevent researchers from taking advantage of the enormous potential of publicly available biological data, which would have a negative impact not only on academic research, but also on industry. It is also obvious, but perhaps worth stressing here, that the cost of training in core biological data resources and tools is exceedingly small, compared with the cost of data collection, storage and organization. A relatively small investment would make a significant difference to the life sciences by facilitating more effective exploitation of the data.

The usefulness of end-user training is attested by the success of an initiative of the Biosapiens Network of Excellence (http://www.biosapiens.info). Entitled the European School in Bioinformatics, it consisted of a hands-on training course directed at bench scientists. It was held twice a year, each time in a different country, with instruction given by post-doctoral researchers.

The initiative was extremely well received by the community. In total the school trained 349 students, more than 40% of which were women; students and instructors came from many countries and institutions. The level of satisfaction, as judged

by the students' questionnaires filled at the end of each session, was very high. The school also provided a central site (http://www.biosapiens.info/page.php?page=esb) for information on training in Bioinformatics in Europe, complete with material and practicals. Furthermore, it fostered interactions among young European scientists, encouraging mobility.

Not only the users, but also the bioinformatics community benefitted from this initiative, in that it allowed developers to assess tool usability by seeing them in action by a set of novice users. It was also instrumental in strengthening the collaboration between several EU networks (ENFIN—http://www.enfin.org, EMBRACE—http://www.embracegrid.info contributed to the school). A recent large-scale Europe-wide survey of user needs in bioinformatics (see http://www.elixir-europe.org/bcms/elixir/Documents/reports/ELIXIR_UserSurvey_FinalReport.pdf) highlighted that there were substantial differences in access to and knowledge of resources, depending upon country location and research domain. Interestingly, the survey also showed that frequent users of bioinformatics resources were almost equally divided into experimental (58.6% of respondents) and computational biologists (67% of respondents). This makes it even more obvious that training should be directed at both categories, and therefore that the involvement of several communities is indispensable. Biological data providers, trainers and educators in the life sciences, biological data resource and tool providers, funding bodies and end users, whether occasional or experienced, must be able to cooperate effectively.

This has been recognized at the European level. In fact, the European Strategic Forum for Research Infrastructures (ESFRI) in its first roadmap in 2005 (http://cordis.europa.eu/esfri/) included bioinformatics among the 35 large-scale European infrastructure projects that were prioritized for selection. The Elixir (European Infrastructure in Bioinformatics) project includes 32 organizations who all signed-up to the mission statement of constructing and operating a sustainable infrastructure for biological information in Europe, in order to support life science research and its translation to medicine and the environment, the bio-industries and society. One of the Elixir's key goals is to develop a strategy for training in bioinformatics and systems biology. Elixir members are working towards setting up a mechanism whereby the development of data resources is tightly linked to the provision of training materials and the establishment of a training support unit. A centralized training registry will be set up that will improve access to bioinformatics user training throughout Europe. It will also support trainers throughout Europe by providing access to regularly updated training materials; develop benchmarking and evaluation systems; provide mechanisms for developing, piloting and evaluating new training programmes; and act as a single point of contact for national and pan-European training infrastructures and projects. Already, the Elixir teams have completed a comprehensive survey of current user training activities, and made this information openly available. Data and pointers to training resources can be found in the Elixir Web site (http://www.elixir-europe.org/bcms/elixir/Documents/reports/WP11-Training_Strategy_Committee_Report.pdf).

As one might expect, access to training is very variable depending on the country or region. Three topics account for almost half of the training documented in the

survey: transcriptomics, general overviews of bioinformatics, and programmatic access to data resources. This is a clear indication of the interest of the communities. Transcriptomics, as mentioned above, is at present central to cancer research and will probably remain so for a few years, notwithstanding the advent of next-generation sequencing techniques, because of its power, speed and relatively low cost in diagnostic and prognostic analysis. The future requires, however, that scientists be equipped with sufficient information and knowledge in systems biology to progress towards translational research.

In some cases, suitable resources are already available, for example primary knowledge databases (lists of biological parts, their properties, their relationships, phenotypes etc.), secondary knowledge (pathways, mathematical models, literatures etc.), standards (reporting guidelines, formats and ontologies), core libraries and tools. In other cases, the resources need time to mature. There is a lack or scarcity of easy-to-access resources containing data on the topology of supra-macromolecular complexes; amounts, concentration and locations of molecules, at sub-cellular, tissue and organism levels, in both physiological and pathological situations; physiological measurements, and temporal processes.

Some interdisciplinary activities are particularly crucial. As discussed above, translational research activities also require active MD-PhD programs. Large cancer genome projects necessitate the collaboration of molecular biologists, pathologists, epidemiologist, doctors, technologists, statisticians, bioinformaticians and others. In all cases, a basic knowledge of the biological context (pathology, epidemiology, molecular biology), of the details of the experiments (genomic technology), and of the methodologies for data analysis and interpretation (bioinformatics, mathematical analysis), is essential. Because the interpretation of results in terms of molecular pathways and biological networks, including mapping and simulation, requires the participation of modellers, mathematicians and physicists, it has become necessary for biologists and bioinformaticians to acquire some understanding of their methods and approaches.

6.3.3 The General Public

The mass media have historically played an important role in communication between medical science and society, being the obvious channel for the dissemination of science to the public, through the efforts of researchers or journalists. Their function has also been, and still is, to provide a forum for discussing risk, trust, and priorities, and to report on cultural or political concerns that feed back into the production of science. However, internet resources are quickly assuming the role of primary source of information for patients and even for medical professionals. The anarchic nature of the Internet, while desirable for fostering open debates without censorship, raises questions about the quality of the provided information.

This is particularly important because incorrect medical information could be a matter of life or death; there is a serious concern about possible damage to patients

by low-quality medical information. Another concern is that there is no "context-based" selection of information, so that patients reading information intended for health professionals may misinterpret it, leading to false fears or expectations; this can be aggravated by the anonymity of authors. Furthermore, health information that is valid in one specific healthcare context may be wrong in another. Different measures may be proposed to solve this issue, for example the introduction of independent third-party evaluation by either human reviewers or automatic software rating, calling upon the organization of communities of users using social network technology, who are naturally closer to people's interests and behaviour. It is obvious that cancer patients need a better organized system with one or more well-organized information points, validated, updated and supported with various levels of information complexity as well as a more complete and accurate system of accessing essential information, such as clinical trials and feedback from other patients with similar symptoms and treatments.

It is also desirable that future systems should be capable of dynamically connecting the information (publications, electronic data, databases) produced by research teams with medical information such as clinical trials, drugs and treatments, and information on symptoms provided by individual or by communities of patients. The challenge will be to provide this information at the proper level of complexity without losing accuracy and keeping pointers to different information levels, with perhaps a context-based selection of what is shown and how to use it.

We believe that the use of information retrieval and text-mining technologies, in the context of Web technology (semantic Web) can play a key role in this future environment. The initial step must be to map the situation and build inventories of the existing information, tools, and access methods, and to relate these to the technologies accessible to specific types of user. Projects such as the European Eurocancercommons initiative (http://www.eurocancercoms.eu) aim at conducting just such an inventory on cancer in Europe, describing possible faults and flaws and future requirements, and addressing the needs of scientists, clinicians, disease associations and the public in general. Training and information exchange will be areas of particular attention during this survey.

6.4 Organization of Cancer Research Centres and their Cross-disciplinary Activities

A publication by the SCImago Institutions Ranking classified more than 2000 institutions dedicated to scientific research all over the world by their Field Normalized Citation Score. Needless to say, these numbers can be analysed in different ways and should not be overinterpreted, but they can serve to help us explore the organization and needs of integrated cancer research centres.

Among the first 30 institutes listed in Table 6.1, 8 are European-based and 11 are dedicated to cancer research, while a number of others have important cancer-related components within the framework of genomics/molecular biology research.

Table 6.1 List of institutes according to their publication impact (SCImago Institutions Ranking) (SIR, http://www.scimagoir.com). Those fully dedicated to cancer research are listed in bold; European-based ones in italic

Wellcome Trust Sanger Institute
Cold Spring Harbor Laboratory
Institute for Systems Biology
Strang Cancer Prevention Center
The Rockefeller University
Salk Institute for Biological Studies
MRC Laboratory of Molecular Biology
Howard Hughes Medical Institute
Ludwig Institute for Cancer Research
Dana Farber Cancer Institute
European Molecular Biology Laboratory
International Agency for Research on Cancer
Wistar Institute
Scripps Research Institute
Genome Institute of Singapore
Space Telescope Science Institute
Friedrich Miescher Institute for Biomedical Research Novartis Research Foundation
Fred Hutchinson Cancer Research Center
SSM Cardinal Glennon Children's Hospital
Centro de Regulacion Genomica
Centro Nacional de Investigaciones Oncológicas (CNIO)
National Institute of Genetics
Whittier Institute for Diabetes
Burnham Institute for Medical Research
Brigham and Women's Hospital
National Institute on Aging
Hamilton Health Sciences
John Innes Center
Memorial Sloan-Kettering Cancer Center
Garvan Institute of Medical Research

Also included are institutions for which cancer research forms part of a general medical/biomedical hospital activity, such as Brigham and Women's Hospital and Hamilton Health Sciences.

We note the activities of the Sanger Centre and of the European Bioinformatics Institute (EBI) outstation of the European Molecular Biology Laboratory (EMBL) in the development of large-scale genome projects, including some of the main cancer-related ones and their databases, along with software for analysing the contextual molecular and biomedical information organized in the corresponding biological databases.

We can summarize the characteristics of the leading cancer centres as based on:

a. a strong research activity in molecular and cellular biology, including the development of animal models for cancer research;
b. the support of research-based core services, including strong resources in genomics, proteomics and animal facilities;

c. the integration of associated hospitals to permit the participation of oncologist and medical doctors performing research and clinical practice;
d. the capacity to develop new molecules with a continuous flow from basic molecular biology to drug development;
e. a flexible structure to deal with intellectual property (IP) and commercialization issues, including the capacity to generate collaboration with industry and spin-off activities;
f. the participation of computational biologists and bioinformaticians to support and perform research, including basic development of tools, databases and methods.

In terms of resources these centres have made considerable investment in those technologies associated with genomics, imaging, computing, high-throughput chemistry, etc. They have also implemented flexible managerial structures to make translational activities possible.

It is clear that 'omics' sciences are an essential contributor to the area of cancer research, and indeed the core facilities of many cancer institutes include sequencing and genomics, proteomics, bioinformatics, imaging, and transgenic mice, among others. This is a very common set-up for a modern cancer research institute, where information at different levels is collected in high-throughput mode and integrated through bioinformatics.

A key element in the organization of the services that support science is the establishment of a strong bioinformatics service unit, clearly separated from, but strongly collaborating with, those groups working in Computational Biology and Bioinformatics. This type of unit has to be capable not only of support biologists in the organization and analysis of their projects, but also of facilitating collaboration between laboratory-based bioinformaticians and the central bioinformatics service platform, providing workers with the required essential technical support and assistance in developments and implementations.

Ideally, the bioinformatics support units should have expertise in database development and maintenance as well as in the handling of large data sets. They require experience of statistics, and often even the ability to conduct expert automated screenings of image analysis followed by microscopy. In general they need to be able to handle data flows coming from high throughput (HT) platforms, in particular those from Next Generation Sequencing (NGS). Particularly important for the functioning of the bioinformatics units is the transparency of the services they offer, and of the processes needed for accessing them and prioritising projects.

A further challenge facing bioinformatics units is their role as mediators of the complex relationships between bio- and medical informatics. This requires the integration of biobanking data, medical information records, and the corresponding genomic information. Finally, the units have the paramount mission of providing training to biologists at a basic level, as well as advanced instruction and support for bioinformaticians in laboratories.

One important aspect common to all cancer research institutes is the strong connection with hospitals, especially for performing clinical trials, which opens the door to testing of drugs developed by the institutes. The connection also allows the

integration of genome-based services in the context of 'personalized medicine' approaches.

These institutes (see Table 6.1) lend support to what we discussed about the interconnection between what has been traditionally considered basic research and its practical applications by tightly integrating both aspects. In particular, the CNIO Institute (http://www.cnio.es) is actively involved in the development of new drugs, channelled via a specific Medical Chemistry programme that implements the complete drug discovery pipeline. In this respect, CNIO does not represent a typical example in Europe, since the incorporation of such a programme within a European research institution can in fact be considered a pioneering endeavour in Europe. However, it is clear that such efforts will soon be implemented in other similar institutions. As discussed in the introduction, even the EMBL, the European basic research Institute "par excellence," has introduced a chemical biology core facility among its services.

6.5 Conclusions

It is commonly agreed that systems biology will be a major component of cancer research in the next few years, but at present, to the best of our knowledge, there are only a few institutions where such a program is actively integrated within translational research activities. Such integration does exist in many of the research institutions listed above and in particular for certain centres with a long tradition of cancer research, such as the Memorial Sloan Kettering Cancer Center (MSKCC—cbio.mskcc.org/research/sander.html) and the Dana Farber Cancer Institute. For example, the department led by C. Sander in MSKCC has included the analysis of genomic data in the context of new systems biology and network-based approaches. This effort is complemented with a strong mathematical modelling and experimental component and merged within the context of active translational research work. The implementation of this type of ambitious integrated research programme creates needs in terms of resource development and training that constitute a new incentive for research and clinical applications.

Acknowledgements This work was supported by grants from the European EUROCANCER-COMS coordinating action and by the COMBIOMED network of the Spanish Instituto Salud Carlos III to A Valencia and from the Italian Ministry of Health contract no.onc_ord 25/07 to A Tramontano.

References

Baudot A, Gómez-López G, Valencia A (2009) Translational disease interpretation with molecular networks. Genome Biol 10(6):221. Epub 2009 Jun 29. Review. PMID: 19591646

Baudot A, de la Torre V, Valencia A (2010) Mutated genes, pathways and processes in tumours. EMBO Rep 11(10):805. Epub 2010 Sep 17 PMID: 20847737

Esquivel A, Meric-Bernstam F, Bernstam EV (2006) Accuracy and self correction of information received from an internet breast cancer list: content analysis. BMJ 332(7547):939–942. Epub 2006 Mar 2. PMID: 16513686

Gómez-López G, Valencia A (2008) Bioinformatics and cancer research: building bridges for translational research. Clin Transl Oncol 10(2):85–95. PMID: 18258507

Hidalgo M (2010) Pancreatic cancer. N Engl J Med 362(17):1605–1617

Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Kamiyama H, Jimeno A, Hong SM, Fu B, Lin MT, Calhoun ES, Kamiyama M, Walter K, Nikolskaya T, Nikolsky Y, Hartigan J, Smith DR, Hidalgo M, Leach SD, Klein AP, Jaffee EM, Goggins M, Maitra A, Iacobuzio-Donahue C, Eshleman JR, Kern SE, Hruban RH, Karchin R, Papadopoulos N, Parmigiani G, Vogelstein B, Velculescu VE, Kinzler KW (2008) Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science 321(5897):1801–1806. Epub 2008 Sep 4. PMID: 18772397

Mancini J, Noguès C, Adenis C, Berthet P, Bonadona V, Chompret A, Coupier I, Eisinger F, Fricker JP, Gauthier-Villars M, Lasset C, Lortholary A, N'Guyen TD, Vennin P, Sobol H, Stoppa-Lyonnet D, Julian-Reynier C (2006) Patients' characteristics and rate of internet use to obtain cancer information. J Public Health (Oxf) 28(3):235–237. Epub 2006 Jun 29

Martin-Sanchez F, Iakovidis I, Nørager S, Maojo V, Groen P de, Van Der Lei J, Jones T, Abraham-Fuchs K, Apweiler R, Babic A, Baud R, Breton V, Cinquin P, Doupi P, Dugas M, Eils R, Engelbrecht R, Ghazal P, Jehenson P, Kulikowski C, Lampe K, De Moor G, Orphanoudakis S, Rossing N, Sarachan B, Sousa A, Spekowius G, Thireos G, Zahlmann G, Zvárová J, Hermosilla I, Vicente FJ (2004) Synergy between medical informatics and bioinformatics: facilitating genomic medicine for future health care. J Biomed Inform 37(1):30–42. PMID: 15016384

Rebholz-Schuhman D, Cameron G, Clark D, Mulligen E van, Coatrieux JL, Del Hoyo Barbolla E, Martin-Sanchez F, Milanesi L, Porro I, Beltrame F, Tollis I, Van Der Lei J (2007) SYMBIOmatics: synergies in medical informatics and bioinformatics—exploring current scientific literature for emerging topics. BMC Bioinformatics 8(Suppl 1):S18

Villarroel MC, Rajeshkumar NV, Garrido-Laguna I, De Jesus-Acosta A, Jones S, Maitra A, Hruban RH, Eshleman JR, Klein A, Laheru D, Donehower R, Hidalgo M (2011) Personalizing cancer treatment in the age of global genomic analyses: PALB2 gene mutations and the response to DNA damaging agents in pancreatic cancer. Mol Cancer Ther 10(1):3–8. Epub 2010 Dec 6. PMID: 21135251

Part III
Bioinformatics and Systems Biology Analysis

Chapter 7
Mathematical Tools in Cancer Signalling Systems Biology

Julio Vera and Olaf Wolkenhauer

Abstract Rather than a strictly formalized methodological framework, we here describe systems biology as a flexible approach in which the modelling strategy used depends on a trade-off between the nature of the biochemical network investigated, the biomedical question to be elucidated, and the quantity (and quality) of the experimental data available. To further substantiate this idea, we chose a number of recent scientific publications in which systems biology was used in the context of cancer cell signalling. Fundamental aspects of the strategy used to set up the mathematical models, integrate available biomedical knowledge, and specifically to generate quantitative data and analyse the system using theoretical and computational tools, are compared and discussed, but new avenues are also suggested.

7.1 Introduction

By cancer systems biology, we mean an interdisciplinary approach, which aims at understanding the spatio-temporal interactions between components of a cell, between cells, and their interaction with the environment, in the context of cancer. It is an approach whereby biomedical questions are addressed through integrating experiments in iterative cycles with mathematical modelling, simulation and theory. Modelling is not the final goal, but a tool used to increase understanding of the system, to develop more directed experiments, and finally to allow predictions (Wolkenhauer et al. 2010). Although there is no set standard for the application of systems biology to the investigation of a cell signalling system, we can distinguish at least the following steps in the procedure (see Fig. 7.1):

1. The biomedical information, existing in relevant publications and databases, is used to identify the proteins involved and the type of interactions. This information should be based on state-of-the-art biomedical knowledge, but will neces-

J. Vera (✉)
Department of Systems Biology and Bioinformatics, University of Rostock,
18051 Rostock, Germany
e-mail: julio.vera@uni-rostock.de

A. Cesario, F. B. Marcus (eds.), *Cancer Systems Biology, Bioinformatics and Medicine,*
DOI 10.1007/978-94-007-1567-7_7,

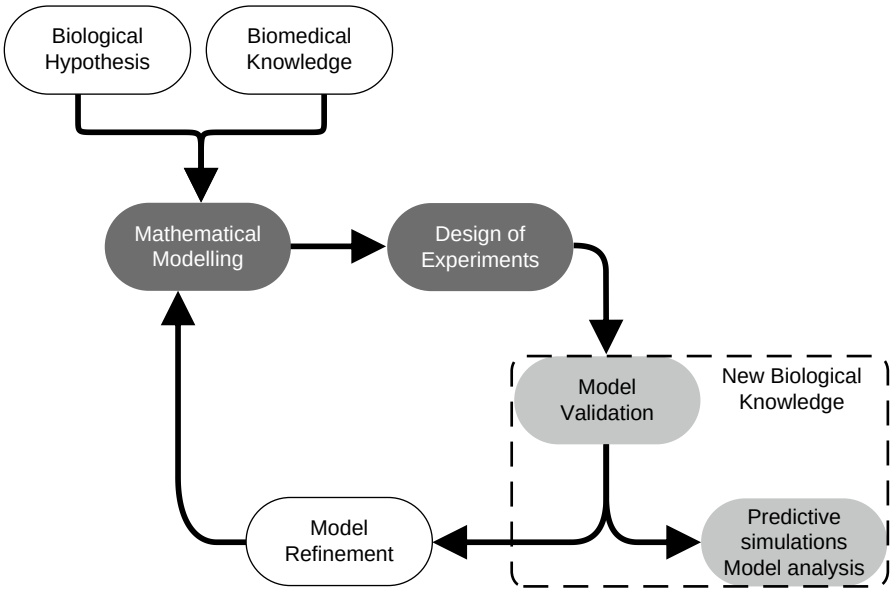


Fig. 7.1 Sketch of the standard iterative approach used in cancer signalling systems biology

sarily also include a number of hypotheses concerning the network structure and dynamics. The goal is then to test these hypotheses by using the systems biology approach.

2. Based on the information gathered from literature and databases, a mathematical model is derived. The mathematical model will consist of equations (e.g. differential or "rate"-equations). Numerical simulations and analytical analyses can then be employed to investigate system properties *in silico*, thereby suggesting experiments.
3. Specific experiments are designed and performed to enrich the characterization of the mathematical model. In some cases, the mathematical model and the quantitative data generated are merged in a computational process called 'model calibration', which assigns values to the parameters characterizing the model equations. In addition, simulations and analysis of this model may also be used to design further experiments aiming to test the biological hypothesis under investigation.
4. The ability of the model to predict and/or match further experiments is used to judge the quality of the model and the validity of the investigated hypothesis. A negative result leads to a reformulation of the hypothesis in terms of refined mathematical model and the design of new experiments.
5. A validated model is already a relevant result because it may confirm or disprove some of the initial biological hypothesis encoded by its equations. Furthermore, a sufficiently validated mathematical model can be used to analyse the system

Table 7.1 Papers on the cancer signalling systems biology analysed for this chapter

Cancer Signalling systems	Selected references
DNA damage response	Ciliberto et al. 2005; Geva–Zatorsky et al. 2006; Ramalingan et al. 2007; Vera et al. 2010
Apoptosis	Rehm et al. 2006; Albeck et al. 2008a, b
Cell cycle control	Qu et al. 2003; Csikász–Nagy et al. 2006; Kim and Ferrell 2007
RTK and MAPK cascades	Bhalla et al. 2002; Schoeberl et al. 2002; Neves et al. 2008; Borisov et al. 2009; Blüthgen et al. 2009; Chen et al. 2009
JAK–STAT signalling	Swameye et al. 2003; Vera et al. 2008
NFkB and inflammatory response	Hoffmann et al. 2002; Nelson et al 2004; Ashall et al. 2009
Cell adhesion and Wnt signalling	Van Leeuwen et al. 2007, Kim et al. 2007, van Leeuwen et al. 2009

(via numerical simulations or analytical approaches) with a variety of purposes, including the faster detection of potential drug targets (Schoeberl et al. 2009) and the identification of new biomarkers for the diagnosis of cancer (Ptitsyn et al. 2008).

This general scheme already delimits a pattern to be followed in any investigation in the field. We analysed a number of recent publications in cancer signalling systems biology to determine ways of performing this strategy. Towards this end, we chose 26 publications in which mathematical modelling has been used to investigate dynamical features of signalling pathways and networks directly related to cancer progression. Our list included papers on DNA damage response and p53 signalling, apoptosis, cell cycle regulation, MAPK cascades and Receptor Tyrosine Kinase (RTK) pathways, JAK-STAT signalling, inflammatory response through NFkB and cell adhesion via Wnt signalling (See Table 7.1). All papers considered were published between 2002 and 2009. Although few older ones could be found and considered (Huang and Ferrell 1996; Kholodenko et al. 1999; Kholodenko 2000), our selection suggests that cancer systems biology is a young discipline.

We further elaborated a questionnaire covering the main aspects that one should address during any research project concerning mathematical modelling based on quantitative experimental data. This questionnaire included general methodological questions, but also questions concerning the technical details of the procedure performed (see Appendix Table 7.1A). More precisely, we wanted to know: (a) how the biomedical context of the investigation, and the set of proteins and interactions considered, were established; (b) the experimental model and technologies used for generating quantitative data and testing the model hypothesis; (c) the details of the modelling strategy pursued including the parameters estimation strategy chosen; (d) the analytical and computational methods used to investigate the dynamical properties of the system; and (e) the strategies used to visualize the data and simulations generated. The analysis of this information allowed us to establish a general proce-

dure for applying systems biology to cancer-related cell signalling systems, which is the topic to be discussed in this chapter.

7.2 The Systems Approach

7.2.1 When to Employ a Systems Biology Approach

In general terms, we may say that systems biology is the appropriate approach to dealing with complexity at the cellular level. This complexity may arise from the structural complexity of the systems investigated (e.g. the number of components and interactions considered), the multiplicity of temporal and spatial scales attained, or the existence of regulatory structures that induce non-linear dynamics at the sub-cellular level. In our analysis we have detected at least four experimental conditions for which systems biology constitutes an adequate (if not a requisite) approach:

1. *Experiment design and validation of hypotheses concerning the dynamics, spatial organisation and structure of the pathway.* Data-based mathematical models are a suitable tool for the formulation and validation of hypotheses about the structure and dynamics of signalling systems (Rehm et al. 2006; Kim and Ferrell 2007; Blüthgen et al. 2009). This idea is supported by the fact that model-based computational simulations make accessible the dynamics of protein states not measurable using state-of-the-art experimental techniques (Swameye et al. 2003); model-based computational simulation also allows the analysis of pathway modifications or features in the signalling spatial organisation, not accessible by purely experimental means (van Leeuwen et al. 2007; Neves et al. 2008). It may also allow the analysis of biochemical properties in receptors and transcription factors whose deregulation may critically modify the performance of the system. Taken together, systems biology can be a convenient approach to investigating the structure of not totally elucidated signalling systems. In addition, it may help towards speedy and reliable design of the appropriate experiments to test a given structural hypothesis.
2. *Investigation of complex and highly interconnected biochemical networks and integration of high-throughput data.* When large signalling networks composed of dozens to hundreds of highly interconnected proteins are investigated, conventional approaches based on direct intuitive analysis and simple knockdown experiments lose efficacy (Schoeberl et al. 2002; Qu et al. 2003; Borisov et al. 2009; Chen et al. 2009). Under these circumstances, mathematical modelling is not only a suitable approach but rather the unique alternative choice for integrating prior knowledge, database information, diverse and tightly connected high-throughput data sources, and biological hypotheses (Csikász-Nagy et al. 2006; Rehm et al. 2006; Albeck et al. 2008a, b). These large-scale models are a perfect

tool for investigating the crosstalk (see Box 7.1) between different signalling systems not previously considered, or for which no experimental set-up is available (Kim et al. 2007).

Box 7.1 Non-linearity in cell signalling systems

Structural motifs inducing non–linear dynamics in cell signalling

Negative feedback loop	Positive feedback loop	Crosstalk	Transcriptional delay

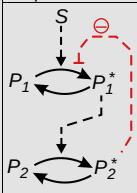
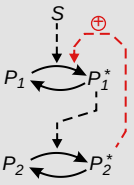
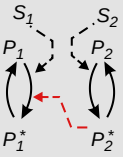
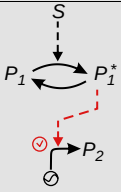

			

Fig. B1.1	Fig. B1.2	Fig. B1.3	Fig. B1.4

A **negative feedback loop** is a regulatory structure in which the activation of a signalling event negative regulates a signalling process upstream of the signalling system considered; it may induce homeostasis in the system (ability to keep stable internal states towards small fluctuations of the input signal) and/or fast signal termination, but also sustained oscillations (see below). A **positive feedback loop** is a regulatory structure in which the activation of a signalling event positive regulates a signalling process upstream; it may provoke signal amplification and ultrasensitivity (see below). Given two essentially isolated signalling systems activated through differentiated input signals, **crosstalk** between them happens when signals from one pathway modulate the activity of the other; crosstalk may be used by the cell to regulate signal integration. The **transcriptional delay in protein expression** accounts for the elapsed time that it takes from the (transcription factor modulated) gene activation to protein synthesis and folding. A sufficiently large transcriptional delay may change the unique equilibrium of the system and induce sustained–oscillations (see below).

Non-linear dynamics behaviour in cell signalling

Self-sustained oscillations	Bi-stability	Ultrasensitivity

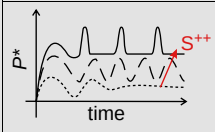
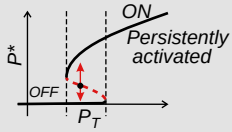
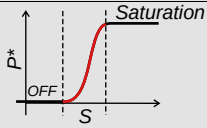

		

Fig. B1.5	Fig. B1.6	Fig. B1.7

In cell signalling systems with **sustained oscillations** the concentration of the involved proteins oscillates in time owing to characteristic internal features of the system,even with constant external stimulation. In systems showing **bi-stability**, for a given experimental regime small perturbations in the set-up may induce totally different fates for the system, inducing for example quick signal termination or persistent activation. **Ultrasensitivity** makes a cell signalling system able to transform graded input signals into discrete all-or-none outputs, creating actual input signal thresholds that determine the activation of the system.

3. *Detection and analysis of non-linear dynamics in cell signalling systems.* In some cases the necessity for mathematical models does not arise from the number of compounds and interactions integrating the system, but from the inherent complexity/non-linearity associated with some of them (Vera et al. 2008, 2010; Albeck et al. 2008a, b). The combination of even a few regula-

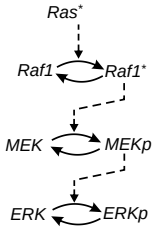
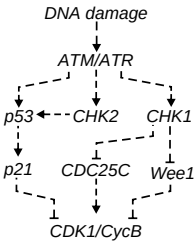
tory motifs in a signalling system such as positive or negative feedback loops already induces complex dynamical behaviour (ultrasensitivity, Kim and Ferrel 2007; multistability, Reynolds et al. 2003; self-sustained oscillations, Nelson et al. 2004; Geva-Zatorsky et al. 2006; Ashall et al. 2009) which escapes any simple intuitive interpretation and requires the use of computational simulations and powerful analytical tools (see Box 7.1). Thus, mathematical modelling may be required to investigate the biological conditions in which feedback loops and time-delay generate non-linear behaviour. Given that medium-size signalling systems include multiple combinations of these regulatory structures, any attempt to investigate the fine tuning and regulation of these systems requires a systems biology perspective.

4. *Model-based detection of drug targets and biomarkers.* Data-based mathematical models may become an excellent tool for detecting the critical signalling processes controlling the essential dynamical properties of a complex system. Furthermore, they may be used to test and prototype strategies for their regulation and manipulation, which may open new/faster ways to develop future therapies. In line with this, systems biology is being already used to identify and prioritize potential cancer biomarkers (Ptitsyn et al. 2008), detect potential drug targets (Schoeberl et al. 2009) and design optimal patterns for cancer chronotherapy (Lévi et al. 2008).

7.2.2 Biological Hypothesis and Set-up of the Signalling System

Inside a cell, signalling systems are highly interconnected, sharing not only the cellular space but even critical intermediates and processes (see definition of crosstalk in Box 7.1). Hence, the exercise of dividing these complex systems into subunits (modularization), for the purpose of making the investigation experimentally and conceptually operative, is a critical initial step in any systems biology investigation. By signalling system, we mean a set of interacting proteins and biomolecules involved in the detection, processing and transduction of a biological signal at the cellular level, which can be considered a system (in terms of structural isolation and reduced number of external inputs/outputs) in a given physiological context. In cell biology, the frontier between different signalling systems is not always clear, and it may shift according to the particular cell type, biological context and the cellular function investigated. Depending on the complexity and structure of the cells, we can distinguish between signalling modules, pathways and networks (See Box 7.2). Our analysis indicates that the selection of the precise set of molecules and interactions that built-up the signalling system under investigation (network, pathway or module) should take account of following criteria at least:

Box 7.2 Cell signalling networks, modules and pathways

Module	Pathway	Network
 Fig. B2.1	 Fig. B2.2	 Fig. B2.3

A **module** is a small signalling system; composed ofa reduced set of interacting proteins, part of a wider system from which it is clearly separable in the investigated biological conditions (left–hand side). A **pathway** is a signalling system, with a unique or a very limited number of input signals, in which protein–protein interactions follow a rather sequential cascade of events (centre). A signalling **network** is a complex, highly interconnected signalling system composed of dozens to hundreds of proteins and several concurrent input signals (right–hand side). Under given biological conditions, networks can be divided into pathways and modules (modularization), a strategy advantageous for investigating small highly regulated systems.

The set of assumptions stated by the researcher to delimit the structure and compounds of the systems has been obtained by bibliographical search. The assumptions are thus clearly supported by the scientific literature in the field and are valid for the considered biological context, *i.e.* the cell type and cellular function investigated (Schoeberl et al. 2002; Ciliberto et al. 2005; Csikász-Nagy et al. 2006). This information may be used to conclude that the system constitutes a closed-type module (with just a few inputs and outputs) of a bigger network, but also to argue that it includes the minimum number of elements that enable the functionality of the studied system.

The model structure is based on one or more previously published mathematical models of the system, which are adapted, expanded or modified for the purpose of the current investigation and the given biological context (Ramalingam et al. 2007; Vera et al. 2008; Kim et al. 2007; van Leeuwen et al. 2009; Vera et al. 2010). This may be combined with the previous criterion, and newer published experimental evidence may be used for the extension or adaptation of the model.

The model components and structure that may arise from the iterative approach suggest, in much the same way that successive cycles of modelling and experimentation are used to sum up or discard information, the need for inclusion of signal inputs, compounds, protein-protein interactions and outputs of the system under the investigated conditions (Rehm et al. 2006; Blüthgen et al. 2009; Albeck et al. 2008a, b). Some of the interactions considered may not be fully confirmed by the literature, a hypothesis that we attempt to verify using our systems biology iterative approach.

In the end, the choice of model components and structure will be a critical combination of all these criteria. Furthermore, their systematic establishment may constitute the actual aim of the investigation rather than its preliminary step.

7.2.3 Mathematical Modelling

The correct choice of modelling framework (the set of precise mathematical premises, rules and structures used to encode the model of the system) is a trade off between several practical and formal issues, that should at least include: (a) details of the existing biomedical knowledge on the system; (b) the quality and quantity of the experimental data available; (c) the nature of the properties of the system under analysis; and (d) the computational requirements of the modelling framework chosen.

The magnitude of the model (e.g. its size in number of compounds and protein interactions) plays an important role in the choice of the modelling strategy pursued. Small models, fully characterized with quantitative data, may be a useful tool for investigating well-defined signalling pathway modules with high non-linear properties (Swameye et al. 2003; Vera et al. 2008, 2010), while large-scale models, with increased computational and experimental data requirements, may be the appropriate tool for the characterization of complex protein-protein interactions networks and systems involving crosstalk (Schoeberl et al. 2002; Chen et al. 2009).

The emergence and relevance of spatial and temporal information in the investigated system is also an important question to consider, because not every mathematical framework accounts for them in an equally effective and detailed manner. Spatial effects may be described using model compartmentalization, in which the model accounts for the dynamical redistribution of the proteins between different compartments in the cell, such as the cytosol, nucleus or mitochondrion, with variables describing the protein level in each compartment (Hoffmann et al. 2002; Albeck et al. 2008a, b). Another possibility is the use of partial differential equations, a special kind of high-level, data-demanding mathematical model specially conceived for the precise description of spatio-temporal processes (Neves et al. 2008). In addition, the emergence of time-delay in the performance of certain cell functions (e.g. gene transcription, subcellular translocation, etc.) may have critical consequences for the dynamics of the signalling system, and must therefore be encoded in the model in some cases (Swameye et al. 2003; Ramalingam et al. 2007; Lai et al. 2009).

In any case, the level of expertise of the initial scientific team should not bias the choice of the modelling framework, and this temptation must be rejected in order to avoid the construction of computationally suitable but biologically meaningless mathematical models. Our analysis suggests that there are three main modelling strategies, based on ordinary differential equations, for setting up mathematical models in cell signalling (see Box 7.3): (a) precise mass-action kinetic models, that describe the complete network of protein-protein interaction considered in a disaggregated manner (Schoeberl et al. 2002; Albeck et al. 2008a, b); (b) simplified mass-action kinetic models used for the construction of purely quantitative data-based small/medium-size models (Swameye et al. 2003; Blüthgen et al. 2009); and (c) models with other non-linear rate equations accounting for biological properties like saturation, cooperativity or repression, such as Michaelis-Menten and Hill

equations or Power-law equations (see Kholodenko 2000; van Leeuwen et al. 2007; Vera et al. 2008, 2010). An alternative approach is the setting up of hybrid models, in which the researcher used his/her experience and intuition to combine several of the previous options in the model (Borisov et al. 2009; Blüthgen et al. 2009). In this way, an experienced researcher may overcome most of the difficulties associated with the choice of a modelling framework alone, but it is difficult to define a list of fixed general rules for this approach.

Box 7.3 Models based on ordinary differential equations

These models describe spatio-temporal changes of protein concentrations and other biological molecules using kinetic equations that describe the variation in time of the populations, or concentration, of the considered biomolecules. Those equations have the following structure:

$$\frac{d}{dt}P = \sum_i F_i(S, k, P)$$

where F_i are the rate equations (the description in mathematical terms of the protein-protein interactions considered), S are variables accounting for the input signals (biological stimulus, external to the system analysed, but affecting it), k the rate parameters (fixed numbers associated with given biochemical properties, that characterize the rate equations) and P are the time-dependent variables accounting for the interacting proteins and molecules. Depending on the interpretation of the biochemical processes and the mathematical structure of the rate equations, we can distinguish the following models in ordinary differential equations.

Mass-Action models. These models account for a precise description of the reaction mechanism, in which every significant intermediate step in the protein-protein interaction is considered. Mono-molecular reactions (like self-degradation, disaggregation or self-activation) are represented with a rate equation which is linear in the variable representing the protein concentration, P: $F(k, P) = k\ P$. Bimolecular reactions (protein-protein interactions) are represented with a bilinear rate equation, proportional to the activator (S) and the activated protein (P): $F(k, S, P) = k\ S \cdot P$. To illustrate the strategy, consider the following activation mechanism of a protein P by its activator S:

$$S + P \leftrightarrow S : P \rightarrow S + P^*$$

Under the given biochemical conditions, the mass-action kinetic model accounting for the dynamics of the different states of the protein P would read as follows:

$$\frac{d}{dt}[S] = -k_b[S] \cdot [P] + k_b^-[S : P]$$

$$\frac{d}{dt}[S:P] = k_b[S]\cdot[P] - k_b^-[S:P] - k_a[S:P]$$

$$\frac{d}{dt}[P^*] = -k_a[S:P]$$

This option is mainly used in case of complex massive networks for which abundant structural information and quantitative data are available. The simplicity in the set-up of the equations and their ease of use is another reason for the success of kinetic models in the representation of cell signalling systems. However: (a) the detailed structure of the interactions is an open question in the majority of the signal transduction pathways; (b) some of the species considered in the model (like protein complexes) are hardly measurable; and (c) the experimental data available is usually not sufficient to estimate parameters in large mass-action models with an acceptable level of confidence.

Reduced mass-action models. These models are a simplified version of the mass-action models, in which the same rate equations are used, but they only consider the dynamics of the protein states that are relevant for the purpose of the investigation, due to limitations in the data available. In our example, under the assumption that the complex [S:P] is not measurable and can be neglected due to biophysical considerations, the previous model would reduce to the following:

$$\frac{d}{dt}[P^*] = k_b[S]\cdot[P]$$

These models surmount most of the difficulties associated with the complete mass-action models, but they are less meaningful from a biological perspective. Although they are good tools for hypothesis validation and dynamical data interpretation in small models, the results are hardly applicable to biological models or contexts different from the one investigated.

Power-law models. In these models the biochemical processes are described with rate equations that are a product of a rate constant γ and the variables of the system involved in the process raised to characteristic kinetic orders g. In our example, this leads to the following equation for the dynamics of activated P:

$$\frac{d}{dt}[P^*] = \gamma[S]^{g1}\cdot[P]^{g2}$$

In contrast with mass-action models, in power-law models the kinetic orders are parameters, estimated from experimental data. Negative values for the kinetic order represent inhibition; zero indicates that the variable does not participate in the described process; and positive values account for satura-

tion or cooperativity, depending on whether the kinetic order is smaller or bigger than one. Power-law models are suitable for the analysis of systems with incomplete structural information or insufficient quantitative data. They become a realistic alternative when simplified models are required. The drawback of power-law models is that the parameter estimation is computationally and data demanding.

Models with non-linear rate equations (e.g. Michaelis-Menten equations, Hill equations…). In this case, non-linear equations are used to describe the dynamics of given processes in the system considered. Some of these non-linear rate equations are obtained by making biochemically motivated approximations, like the quasi steady-state assumptions for some components. This leads to Michaelis–Menten equations and more complex protein-catalysed dynamics (allosteric processes, competitive inhibition, cooperativity…). For instance, under certain assumptions, we may assume that the activation of P follows a Michaelis–Menten equation and our model could be written as:

$$\frac{d}{dt}[P^*] = \nu \frac{[S] \cdot [P]}{K_M + [P]}$$

Interesting enough, under some conditions mass-action models may be reduced to simpler ones where these non-linear equations are used, which gives a biophysical justification for these models. However, the conditions for which these approximations are applicable are not always fulfilled in cell signalling. In addition, there are no fixed rules for choosing between the different complex protein-catalysed dynamics, so that the choice relies mostly upon the intuition and experience of the researcher, making generalizations for this approach difficult.

Although other modelling frameworks have been successfully used for investigating signalling pathways in different contexts (Heiner et al. 2004; Papin et al. 2005; Saez-Rodriguez et al. 2007; Ullah and Wolkenhauer 2009; Aldridge et al. 2009), their use in cancer systems biology remains limited and they are under-represented in our analysis (see Table 7.2). The exceptions are the few recently published models in partial differential equations (Neves et al. 2008) and stochastic equations (Ashall et al. 2009). We note that this trend is based not on fundamental biological motivations but on practical questions. Further developments in experimental techniques, and the understanding of some cellular processes, may render some of these modelling frameworks more relevant or prevalent in the immediate future. It should be observed that no perfect modelling framework exists; and all of the frameworks have strengths and weaknesses as we illustrate in Table 7.3.

Table 7.2 Mathematical frameworks used in cell signalling modelling

Modelling framework	Basic features	References
Models in ordinary differential equations		
Detailed mass-action	*Precise description of protein-protein interactions. Suitable for medium–large signalling modules. Extensive quantitative datasets required. Identifiability problems. Biophysically reliable.*	Aldridge et al. 2006
Reduced mass-action	*Systems description adapted to the data available. Extensive quantitative datasets required. Suitable for small signalling modules.*	Swameye et al. 2003
Models with non-linear equations (e.g. Michaelis-Menten)	*Non–linear rate equations derived from biochemical motivated approximations. Quantitative data required. Difficulties in determiningits validity. Suitable as approximation for high non–linearity.*	Kholodenko 2000
Power-law models	*Simple description of non-linear processes. Feasible with incomplete information on protein–protein interactions. Quantitative data required. Rather phenomenological description.*	Vera, Balsa–Canto et al. 2007
Stoichiometric networks	*Only structural data about protein–protein interactions required. Suitable for structural analysis. Reduced ability to simulate dynamics.*	Papin et al. 2005
Models based in graph-like interpretation of signalling networks		
Petri nets	*Only structural data required. Suitable for structural analysis and model reduction. Reduced ability to simulate dynamics.*	Chaouiya 2007-
Boolean networks	*Only structural data about protein-protein interactions required. Suitable for structural analysis. Reduced ability to simulate dynamics.*	Klamt et al. -2006-
Other modelling frameworks		
Stochastic models	*Precise description of protein–protein interactions under special conditions. Computational requirements scale fast with complexity of model. Biophysically reliable.*	Turner et al.
Fuzzy models	*Structural data and notions of the dynamics required. Suitable for structural analysis and model reduction. Reduced ability to simulate dynamics.*	Aldridge et al. 2009

7.2.4 Experimental Techniques Used for Producing Quantitative Data

As one might expect given the complex nature of cancer research, the experimental models used in cancer systems biology for generating quantitative data and testing the hypothesis are diverse and heterogeneous. In most cases, the ex-

Table 7.3 Pros and cons for different mathematical modelling frameworks[a]

	Realism	Interpretation	Non-linearity	Data	Calibration	Scalability	Computation
Detailed mass-action							
Reduced mass-action							
Non-linear rate equations							
Power-law							
Stoichiometric networks							
Petri Nets							
Boolean Networks							
Stochastic equations							
Fuzzy logic models							

[a] The features analysed account for the capabilities of the model in terms of: a) **Realism:** encoding of a biophysically detailed description of the signalling processes; b) **Interpretation:** simplicity in the interpretation of the simulations and analysis; c) **Non-Linearity**: ability to model and investigate non-linear phenomena; d) **Data**: demand on quantitative data to set up and calibrate the model; e) **Calibration**: Computational and practical difficulties in calibrating the model; f) **Scalability**: Computational and practical requirements for scaling to larger signalling systems in terms of components and interactions; h) **Computation**: computational requirements for running simulations. In this table, black represents a con or drawback (e.g., detailed mass-action models require a lot of data and are very difficult to calibrate because they have a lot of parameters to be estimated), and white accounts for a clear "pro" or advantage of the modelling framework (e.g., detailed mass-action models are supposed to give a very realistic picture of the complexity of signalling events). Grey is neutral.

perimental models used to generate the quantitative data for model calibration or model predictions validation have been standard human cell-lines used in cancer research (HeLa: Rehm et al. 2006; Albeck et al. 2008a, b; HEK 293: Kim et al. 2007; Borisov et al. 2009; BAF3: Swameye et al. 2003; Vera et al. 2008; HTC116: Ramalingam et al. 2007; SK-N-AS: Nelson et al. 2004; Ashall et al. 2009). Only in a few cases were tissue samples from animal models used for experimentation (van Leeuwen et al. 2009; Neves et al. 2008). The emergence of multi-level modelling approaches (integration of intracellular processes and tissue organisation via mathematical modelling) anyway suggests that in the coming decade, data from *in-vivo* tissue samples will be a important resource for building-up meaningful mathematical models with predictive abilities (Ribba et al. 2006; Frieboes et al. 2009).

Our analysis reveals that the experimental techniques used for data generation have depended to a large extent upon the biological conditions under which the system was analysed, as well as upon technological limitations. We have detected abundant use of experimental techniques as such as: (a) live imaging (FRET microscopy and/or fluorescent imaging); (b) quantitative immunoblotting (western/

Table 7.4 Quantitative experimental techniques used in signalling systems biology

Molecule	Technique	Reference
mRNA and microRNA	Quantitative northern blotting	Roth 2002
	Real–Time PCR analysis	Schefe et al. 2006
	Sandwich hybridization	Rautio et al. 2003
	RNA microarrays	Wilhelm and Landry 2009
Proteins and phosphoproteins	Quantitative Western blotting	Schilling et al. 2005
	ELISA kits	Heyman 2006
	Mass spectrometry–based proteomics	Dakna et al. 2009
	Live cell imaging	Mullassery et al. 2008
	Tandem liquid chromatography and mass spectrometry	Gerber et al.2003
	Reverse-phase protein arrays	Sahin et al. 2007
Metabolites	GC-HPLC	Xiayan et al. 2008
	Nuclear magnetic resonance (NMR)	Bollard et al. 2005
	Gas chromatography–mass spectrometry	Schauer et al. 2005
	In vivo labelling	Birkemeyer et al. 2005

northern blots); (c) radioactive protein labelling; (d) flow cytometry; (e) death assays; (f) RT-PCR; and (g) gene knockout experiments. Most of these techniques are standard in cell biology, but in the context of systems biology they need to be setup or modified to produce quantitative or semi-quantitative experimental data, necessary for the characterization of the mathematical model or the validation of the model predictions. Table 7.4 includes a list of relevant experimental techniques specific to cancer signalling systems biology, and also methodological bibliographical references, where the use of the various techniques for generating quantitative data is discussed.

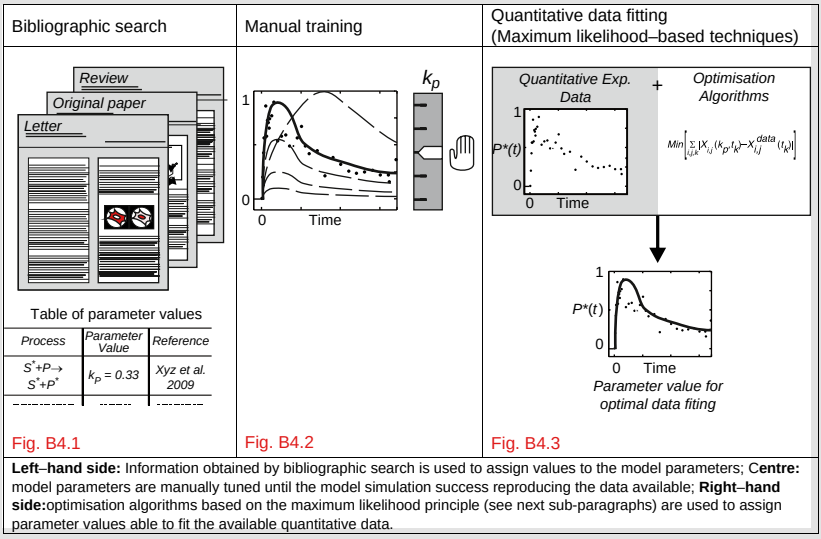
In-depth discussion of experimental techniques and models is not the aim of this chapter, since the book already contains material devoted to these topics (Chap. 3, 4, and 9–12).

7.2.5 *Model Calibration: Parameter Estimation and Model Refinement*

A mathematical model *per se* is a construct encoding in mathematical terms the hypothesis and structure of the signalling system analysed. Very often this is not adequate for the purpose of the investigation, so that the mathematical model must be merged with quantitative data, in a computational process called model calibration. In this process, representative values are assigned to the parameters that characterize the model equations in such a way that the model mimics the behaviour of the system represented by the available experimental data. In addition, iteration of model computational simulations and experiments may be used to further test and validate the structure of the model and the biological hypothesis under investigation.

Regarding the estimation of values for the model parameters, our analysis indicates that there is no unity in the strategies used to assign values. We have distinguished at least four different strategies employed for this purpose (three are illustrated in Box 7.4):

Box 7.4 Strategies for parameter estimation in cell signalling models



Left–hand side: Information obtained by bibliographic search is used to assign values to the model parameters; C**entre:** model parameters are manually tuned until the model simulation success reproducing the data available; **Right–hand side:**optimisation algorithms based on the maximum likelihood principle (see next sub-paragraphs) are used to assign parameter values able to fit the available quantitative data.

1. **Bibliographic search.** Information is sought from publications wherein a similar biological model and similar experimental conditions are considered, and used to assign values to the parameters (Kholodenko 2000; Kim et al. 2007; Albeck et al. 2008a, b). The drawbacks of this strategy are that the reliability and transferability of the data extracted is in general terms dubious, and also that it is difficult, if not impossible, to find enough data in the literature to pursue this strategy in a medium-size mathematical model.

2. **Manual training.** The values of the parameters in the model are manually tuned until the model is capable of properly simulating the set of data available (Ciliberto et al. 2005; Csikász-Nagy et al. 2006), which may include data taken from literature in addition to experimental data (quantitative or qualitative) produced for the investigation. This approach is suitable even for large-scale models. The main problem arising from this strategy relates to the difficulty of merging different kinds of data, quantitative and qualitative, produced using diverse experimental techniques and in unreliably consistent experimental conditions. Furthermore, equilibrium between the complexity of the model and the data available is required to avoid identifiability issues (inability to assign a unique value for one or more parameters in the model).
3. **Quantitative data-fitting.** The quantitative data produced in the investigation is fitted using computational techniques based on the mathematical principle of the "maximum likelihood" (Schoeberl et al. 2002; Swameye et al. 2003; Vera et al. 2008, 2010; Borisov et al. 2009; Chen et al. 2009). Towards this end, the chosen method iteratively assigns values to the model parameters to minimize the differences between the experimental data and the model simulations for identical biological conditions. A variety of optimization algorithms tailor-made for this purpose may be used, e.g. simulated annealing (Gauss-Newton) or evolutionary algorithms (Banga and Balsa-Canto 2008; Chou and Voit 2009). In order to achieve reliability in the results, this approach requires a sufficient amount of quantitative (good quality) data from specifically designed experiments. Until now, our reduced ability to produce enough quantitative data together with some intrinsic mathematical problems of this approach such as identifiability (Raue et al. 2009), or sloppiness (Gutenkunst et al. 2007), make it fully operative only for small-size mathematical models of cell signalling systems.
4. **Hybrid approaches.** These consist of merging parameter values taken from literature, manual training, and finally data-fitting techniques, over a small subset of the parameters in the model (Hoffmann et al. 2002; Rehm et al. 2006; Ramalingam et al. 2007; Albeck et al. 2008a; Neves et al. 2008). This approach is suitable even for large-scale models, but consistency between the different sources of data must be carefully checked before and during the process, and requires supervision by an experienced researcher.

As with the modelling framework, the trend that we indicate here is based not on fundamental biological motivations, but on practical considerations essentially regarding the accessibility of quantitative experimental data. Further developments in quantitative experimental techniques, high-throughput data production, and improved efficiency of optimization algorithms, facilitate the realization of the ideal case of a fully unsupervised quantitative data fitting.

7.2.6 Model Analysis

Validation of the structural and dynamical features encoded by the mathematical modelling may constitute in itself the aim of a systems biology project, if some of these fea-

tures are recent hypotheses previously not verified. In other cases, a properly validated mathematical model may be inspected using analytical tools and simulations, with the purpose of dissecting the sources of its non-linear behaviour or detecting critical interactions regulating the performance of the signalling system such as signalling interactions which may perhaps be susceptible to pharmacological intervention. See Box 7.5.

Box 7.5 Unveiling dynamical biological properties in cancer signalling, using model analysis

$$\frac{d}{dt}EGFR^* = (EGFR_T - EGFR^*)\cdot$$
$$\left[\alpha_1 \cdot \left(EGFR_T - EGFR^*\right) + \alpha_2 \cdot EGFR^* - \gamma \cdot EGFR^* \cdot PTP^*\right]$$
$$\frac{d}{dt}PTP^* = \beta \cdot \left(PTP_T - PTP^*\right) - \delta \cdot EGFR^* \cdot PTP^*$$

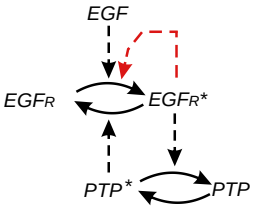


Fig. B5.1

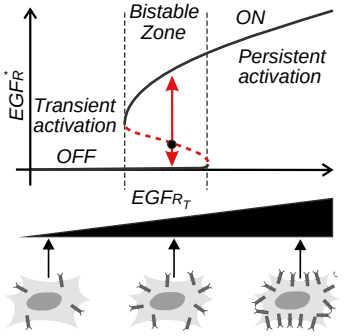


Fig. B5.2

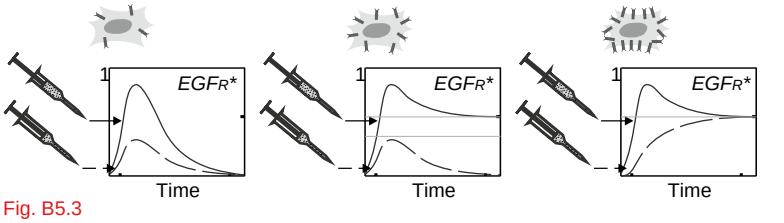


Fig. B5.3

Model investigation analysis using analytical tools may be an excellent strategy for revealing important dynamical features in cancer signalling biology. In Reynolds et al. (2003) the authors used a simple mathematical model and bifurcation analysis to investigate the role of lateral signalling propagation (red arrow in the model scheme) via receptor-receptor membrane interaction in the dynamics of a system integrated by the epidermal growth factor receptor, *EGFR*, and its inhibitor the protein tyrosine phosphatase,*PTP*. The model scheme and equations are shown in the top-left side of the box, and they considered active (*EGFR** and *PTP**) and inactive (*EGFR* and *PTP*) states for both the receptor and the phosphatase, with $EGFR_T = EGFR^* + EGFR$ and $PTP_T = PTP^* + PTP$. The analysis suggests that, as a result of the existence of lateral propagation, the system may present bi-stability for an interval of values in the total amount of membrane EGFR, $EGFR_T$ and other model parameters (Top-right side). For a lower amount of total EGFR, the system shows transient activation, for any level of stimulation (bottom-left figure). For $EGFR_T$ values in the interval of bistability, weak stimulation induces transient activation,while sufficiently intense stimulation may lead to a state of persistent activation (bottom-centre figure). For a large amount of $EGFR_T$, the lateral signal propagation may be sufficient to induce persistent activation even at very small system stimulation (bottom-right figure). This dynamical feature may suggest a mechanism for achieving (via receptor over-expression) independence from the growth signal, one of the so-called hallmarks of cancer (Hanahan and Weinberg 2000).

We need to distinguish between approaches based on a theoretical analysis of the system and requiring a considerable amount of symbolic calculation (mostly derived from systems theory) from approaches based on the available plethora of computational techniques, implemented via computational algorithms or software tools. Our analysis concluded that in most cases, researchers prefer computational techniques based on different kinds of model computational simulations with only a few preferring purely theoretical analysis. We distinguished three main kinds of numerical analysis performed in the models:

1. **Predictive simulations.** Once characterized and refined, the model is used to perform computational simulations focused on the investigation of any of its dynamical properties. We could distinguish predictive simulations used for: (a) model simulation-based derivation and validation of hypotheses about the dynamics and structure of signalling pathways (Swameye et al. 2003; Rehm et al. 2006; Blüthgen et al. 2009; Albeck et al. 2008a); (b) investigation of changes in the system's dynamics under different experimental conditions (van Leeuwen et al. 2007; Vera et al. 2008, 2010; Borisov et al. 2009); and (c) analysis of the dynamics of subcellular compartmentalization for the system's compounds (Swameye et al. 2003; Nelson et al. 2004; Rehm et al. 2006).
2. **Local and global sensitivity analysis.** By sensitivity analysis we mean the study of how the variation in the critical outcomes of a given signalling system can be attributed, qualitatively or quantitatively, to different sources of variation in the model, especially changes in the values of model parameters (Nikolov et al. 2010). Our investigation indicates that this methodology may be used at least for model refinement and model identifiability analysis (Chen et al. 2009; Blüthgen et al. 2009) and detection of critical model parameters, whose regulation controls the dynamical features of a system (Swameye et al. 2003; Rehm et al. 2006).
3. **Computational bifurcation analysis.** This tool, derived from the bifurcation theory, is used to detect intervals in the parameter values defining a distinctive dynamical behaviour of the cell signalling system, such as the emergence of oscillations, bistability or instability (Qu et al. 2003; Ciliberto et al. 2005; Csikász-Nagy et al. 2006).

These are not the sum total of the techniques available. Table 7.5 records the critical information and references of a wide spectrum of analytical techniques used in cell signalling systems biology. The selection of the modelling framework affects the kind of analysis that can be performed: most of the analytical techniques have been derived form and are applicable to, only some of the mathematical models considered. The size of the model may also limit the usability of some methodologies, which may only be feasible up to a given size in the models. On the other hand, these limitations in state-of-the-art analytical tools may be overcome in the coming decade by new methodological and computational developments, thanks to the existence of a numerous community of computational and theoretical researchers interested in this topic.

Table 7.5 Analytical techniques used in cell signalling systems biology

Stability Analysis. It is the study of the dynamical stability in biochemical systems. A system is in a stable state when a small disturbance of that state, at a given time, only alters the state, at all subsequent times, by a correspondingly small amount (Jeffrey 1993). If an arbitrary small disturbance to the state of a system has the effect of producing a large change in the state at all subsequent times, then the system is said to be in an unstable state. In systems biology, stability analysis is used as a tool for estimating boundaries of physiologically feasible parameters for a given system, while keeping the system in a stable state. Stability analysis is an important tool used in bifurcation analysis and robustness analysis.

Bifurcation Analysis. Bifurcation analysis is the study of qualitative changes in the behaviour of nonlinear dynamical systems under varied parameters of the system. By qualitative changes we understand the appearance of new solutions to the differential equations that represent the system, or a change in the stability of the current state. As the term suggests, a 'bifurcation' indicates a branching fork of alternatives. As the system reaches this bifurcation point due to the continuous change of one parameter, the output behaviour may change discontinuously, inducing totally different dynamics behaviour, and the change may be irreversible (Fall et al. 2002).

Robustness Analysis. Robustness is the property that allows a system to maintain its function against internal and external perturbations (Kitano 2007) or under conditions of uncertainty (Stelling et al. 2004). In general, robustness can be classified into 'absolute robustness', representing the average functionality of the system under perturbation, and 'relative robustness' quantifying the impact of perturbations on the nominal behaviour (Rizk et al. 2009). In biological systems, the concept of robustness is closely related to the notions of 'stability' and 'homeostasis'. While robustness is a general concept, homeostasis and stability are its particular instances which stay unmodified if the function to be preserved is one maintaining the system state (Kitano 2007).

Sensitivity Analysis. This technique studies how the variation in the output of a model can be apportioned, qualitatively or quantitatively, to different sources of variation, and how the given model depends upon the information fed into it (Saltelli et al. 2000). In systems biology, sensitivity analysis tells us how a single parameter or a group of parameters affects the dynamics of a given variable, and these two pieces of information (parameter-variable) must accompany any sensitivity value. We can distinguish two different approaches: (i) local sensitivity analysis, in which one parameter at time is varied within a small interval around its nominal value; and (ii) global sensitivity analysis, in which all or several model parameters are varied simultaneously, and sensitivity is measured over the entire range of each input parameter using a variance based methodology.

Model Reduction. The goal of model reduction is the structural simplification of multivariate mathematical models with several parameters (Schmidt et al. 2008). During the procedure, the number of model parameters and/or model variables is reduced. The model thus obtained is a reduced representation of the initial one, in which we attempt to eliminate spurious parameters and variables, while keeping those that maintain the model's capacity to describe the essential properties of the system. Model reduction can be based on: (i) lumping of variables (biological intuition of the process describing the combination of different variables on the original model are combined into a new one using mainly biological criteria); (ii) sensitivity/identifiability-based model reduction, which permits discrimination between important, necessary and redundant variables; and (iii) time-scale separation, in which fast processes are reduced using quasi-steady-state approximations, or rapid equilibrium approximations and slow processes can be absorbed within conservation laws (Krüger and Heinrich 2004).

7.2.7 Data Visualization

The visual representation of data is a fundamental issue when one tries to make the results of a system biology project comprehensible to other research communities potentially interested in them, such as classical cell biologists or clinical researchers. Visualization in a systems biology project covers a variety of topics.

A first question to be considered is the graphical representation of signalling pathways in a coherent figure that takes account of the structured set of biological components and biochemical interactions integrating the system. Our analysis suggests that there is currently a lack of a consistent system for dealing with this critical question, since almost every work we analysed used different styles, software tools and conventions for depicting the model. Well-established tools for this purpose like Cell-Designer (Funahashi et al. 2003) are thus under-represented in our sample of publications. Our belief is that this panorama may change in the near future, thanks to the Systems Biology Graphical Notation initiative (SBGN—Le Novère et al. 2009). For the first time, a wide community of modellers has discussed and proposed a standardised graphical notation for the visual representation of biochemical pathways, including cell signalling systems.

Our analysis did at least reveal the following subtypes of figures commonly used to visualize information, data and results in systems biology papers on cancer signalling (see also Box 7.6):

1. *Figures used to show the experimental data* generated for the investigation, including different kinds of microscopy plots (FRET microscopy, Rehm et al. 2006; or confocal microscopy, Ashall et al. 2009), flow cytometry plots (Albeck et al. 2008a) and immunoblotting (western or northern, Hoffmann et al. 2002; Vera et al. 2008). Such figures are not specific to systems biology publications, but are common in conventional works on cell biology.
2. *Figures for comparison between model simulations and experimental data.* This includes: (a) time series fitting plots (model simulations compared with experimental time series, Swameye et al. 2003; Chen at al. 2009); (b) bar plots to compare the measured and predicted protein activity under different explored experimental conditions (Rehm et al. 2006; Vera et al. 2010); and (c) comparison of life imaging photo series (quantified or not) with simulated time series plots for protein levels in different compartments (Nelson et al. 2004; Ashall et al. 2009).
3. *Figures accounting for model dynamics analysis and predictions,* in which the validated model is used to simulate new experimental conditions or to analyse critical dynamical features. In our investigation we distinguished the following subtypes of figures: (a) time-plots of predictive simulations for the compartment distribution of proteins (Nelson et al. 2004; Ashall et al. 2009); (b) dose response plots accounting for the modulation not only of the system's input signals and regulators (Hoffmann et al. 2002; Vera et al. 2008), but also of protein expression levels related to oncogenic conditions (Vera and Wolkenhauer 2008); (c) bifurca-

tion diagrams elucidating the emergence of oscillations or multistability due to modulation of given model parameters (Ciliberto et al. 2005; Csikász-Nagy et al. 2006); and (d) sensitivity analysis plots accounting for correlations of sensitivities with best fits, dependence of sensitivity on experimental conditions, or effect of parameter variation on activation of target genes (Swameye et al. 2003; Rehm et al. 2006; Nikolov et al. 2010).

Box 7.6 Graphical and visual representations of data and results

Model equations:

$$\frac{d}{dt}P_c = F_{synt}\left[P_n^*(t-\tau)\right] - F_{\deg}\left(P_c\right) - F_{act}\left(\frac{S}{I}, P_c\right) + F_{deac}\left(k_i, P_n^*\right)$$

$$\frac{d}{dt}P_n^* = F_{act}\left(\frac{S}{I}, P_c\right) - F_{deac}\left(k_i, P_n^*\right)$$

Model variables:

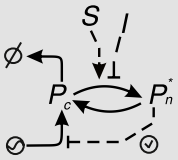
P_c: cytosolic fraction of protein P; P_n^*: nuclear fraction of P

S: Protein Activator level I: Protein inhibitor level

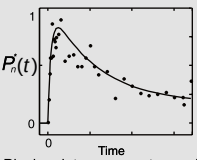
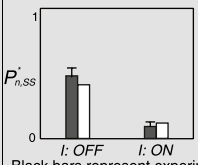
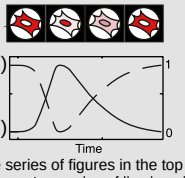
Model parameters:

τ: time-delay in synthesis regulation

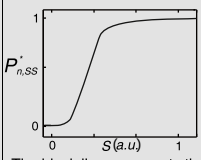
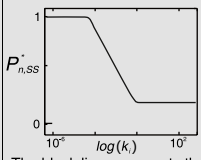
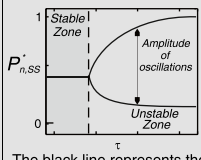
k_i: rate constant of protein deactivation



Experimental data versus model simulations

Time-series fitting	Protein activity bar plot	Compartmentalisation
		
Black points represent experimental time -series data for the activation of P, while the solid black line represents the model simulation for the same biological conditions.	Black bars represent experimental measurements on protein activity under two expression levels of inhibitor I, while the white bars are the model predictions for identical conditions.	The series of figures in the top represents a series of live imaging photos describing the sub-cellular localisation of P during the experiment, while the figure in the bottom is a model simulation for cytosolic and nuclear P during the experiment.

Model analysis

Dose response	Parameter Sensitivity	Bifurcation plot
		
The black line represents the variation in the levels of activated *P* when the intensity of the input signal *S* is modulated.	The black line represents the incidence on the simulated expression levels for *P** of changes in value of the parameter *ki*.	The black line represents the incidence on the simulated *P** of changes in value of the time-delay τ. For a delay higher than a given value (bifurcation point) the value of *P** shows sustained oscillations, and the distance between both black lines indicates the amplitude.

7.3 Discussion

In this chapter we have presented systems biology as a novel approach in cancer sciences, based on integration by mathematical modelling of multiple sources of biomedical information and quantitative experimental data. This new approach relies on a close interaction between (quantitative) experimentation and systems-theoretic methods, in order to formulate and validate hypotheses on the molecular basis of cancer progression. The main point of our contribution has been to highlight the fact that systems biology, far from being a rigid, strictly formalized methodological framework, is a flexible method in which decisions about the precise modelling strategy used depend on the investigated system, the quantity and quality of the experimental data available, and the nature of the biological question to be elucidated.

To illustrate this, we have selected from the recent literature a collection of scientific publications in which systems biology is used to investigate crucial features of several cancer-related cell signalling systems. We further analysed fundamental aspects of these papers to take account of the particular strategies used to set up the mathematical models, to characterize them in terms of available biomedical knowledge and quantitative data, and to evaluate the theoretical and computational tools used in the investigations.

The first conclusion of our analysis is that modern systems biology provides an advantageous approach not only for the formulation and validation of hypotheses concerning non-linear dynamics, spatial organization and structure of signalling pathways (the original motivation posited for the use of mathematical modelling), but also that it is actually the unique possible approach for investigating complex

multi-level highly interconnected biochemical networks and for integrating the massive amount of high throughput data that modern experimental techniques can produce. From a biomedical perspective, mathematical models constructed using a systems biology approach are already in use as an alternative technique to boost the detection of potential drug targets and biomarkers.

Mathematical models are set up by combining the knowledge contained in state-of the-art scientific literature and previously published mathematical models, and also by formulating in mathematical terms the biological hypotheses about the system to be examined. The structure of the mathematical model is ultimately derived through several iterations of the cycle 'mathematical modelling, experimental calibration and validation, model refinement,' whereby hypotheses on the structure and spatial organisation of the system are either confirmed or disproved.

Systems biology makes extensive use of conventional experimental techniques in molecular biology (e.g. immunoblotting, PCR or live imaging), but requires them to be modified so that they can generate time- and space-resolved (semi-)quantitative data about the molecules involved in the signalling systems, and their different states. Along with these modifications. we anticipate that high throughput experimental technologies coming from the '-omics' universe will soon become available as an essential tool for dealing with the modelling of multi-level massive biochemical networks. Moreover, a flourishing community of researchers is developing experimental techniques specifically designed for systems biology, some of which will become standard in the next decade.

An equally active community of computer and theoretical scientists is dedicated to the adaptation, extension and development of mathematical modelling formalisms for systems biology. Thus, although the current 'mainstream' for modelling of signalling systems relies on the use of ordinary differential equations, future developments in experiment techniques and the understanding of some cellular processes may render other modelling frameworks preferable or even prevalent. In the ideal situation, the theoretical framework used for modelling, as for experimental techniques, will be conditioned by the features of the problem under investigation rather than by the preferences or strengths of the modeller.

An abundance of analytical techniques is available for investigating the dynamical properties of cell signalling systems, some of which are essentially theoretical and based on non-linear systems and bifurcation theory, while others have a strong computational background. In both cases, these techniques, together with promising new developments, will soon be confronted with scalability concerns associated with the extreme complexity (in terms of compounds and interactions) that may need to be incorporated into mathematical models in cancer systems biology.

7.4 Appendix

Table 7.1A Items analysed in the literature survey for cancer signalling systems biology

Publication Details
Title of the paper
First author
Journal
Year of publication
Type of paper (original scientific paper, letter, review/survey, meeting report, opinion, hypothesis, perspective, essay)
Biological/Biomedical Context
What system/network/pathway is considered?
How did the authors justify the definition of their system (pathway etc.)?
What is the cell function on which the papers focus? (e.g. cell cycle, apoptosis, proliferation, growth...)
How did the authors justify a systems biology approach?
How did the authors call/define/refer to their systems biology approach? (e.g. 'computational modelling/biology', 'systems biology', 'kinetic modelling'...)
What is the hypothesis the authors pursue?
What elementary process do the authors focus on? (e.g. gene expression, metabolism, cell signalling…)
Experimental Model
Which experimental system did the authors choose? (e.g. model organism, cell lines, tissue samples…)
How did the authors justify their experimental system?
Data Generation
Which technologies were employed?
Did the authors mention any limitations of their approach?
Did the authors mention any advantage or rational for using their choice of technology?
Are the data made publicly available?
Network/Subsystem/Pathway
What is the "size" of the pathway/network model ? (e.g. no of variables, activation states, parameters…)
How did the authors visualise their pathway/network? (e.g. simple line diagram, Kohn map, Cell Designer Map, Powerpoint...)
Parameter Estimation
Which technique (method, algorithm, software) was used to determine the parameters?
How many parameter values were taken from literature?
How did the authors justify the parameter estimation strategy used?
Did the authors refer to any advantage or disadvantage of their approach?
How do the authors phrase open problems, future work?
Model Analyses
What kind of model is used? (e.g. ODE, PDE, stochastic, Petri nets…)
What type of interpretations is used? (e.g. mass action kinetics, Michaelis–Menten kinetics,power–law modelling…)
How did the authors account for spatial information? (e.g. comparmentalisation, PDEs, delays…)
What type of feedback mechanisms are discussed?
What type of plots/visualisations/simulations are shown? (e.g. time plots, bifurcation plots, dose–response plots…)
What types of analyses are used? (e.g. sensitivity analysis, stability analysis, identifiability, numerical/analytical…)
What limitations of their analyses are mentioned?
What do the authors say about the prediction error?
Is the model made available? (e.g. author web–page, BioModels, JWS Online…)
What units, scaling or non–dimensionalisation techniques are used?
Key Findings
What are the key biological findings/conclusions?
To what alternative approaches did the authors compare their results to?

Acknowledgements We thank the collaboration of the following people in the discussions of the papers analysed in this book chapter: A. Bittig, S. Boldt, S. Frey, J. Isaeva, X. Lai, F. Lange, A. Lao, U. Liebal, T. Millat, S. Pauleweit, P. Raasch, K. Rateischak, Y. Schmidt, U. Schmitz, F. Winter. J.V. is funded by the German Federal Ministry of Education and Research (BMBF) as part of the project CALSYS-FORSYS under contract 0315264 (http://www.sbi.uni-rostock.de/calsys). O.W. acknowledges funding through the Helmholtz Association, as part of the Systems Biology Alliance and the Stellenbosch Institute for Advanced Study (STIAS).

References

Albeck JG, Burke JM, Aldridge BB, Zhang M, Lauffenburger DA, Sorger PK (2008a) Quantitative analysis of pathways controlling extrinsic apoptosis in single cells. Mol Cell 30(1):11–27

Albeck JG, Burke JM, Spencer SL, Lauffenburger DA, Sorger PK (2008b) Modeling a snap-action, variable-delay switch controlling extrinsic cell death. PLoS Biol 6(12):2831–2852

Aldridge BB, Burke JM, Lauffenburger DA, Sorger PK (2006) Physicochemical modelling of cell signalling pathways. Nat Cell Biol 8(11):1195–1203

Aldridge BB, Saez-Rodriguez J, Muhlich JL, Sorger PK, Lauffenburger DA (2009) Fuzzy logic analysis of kinase pathway crosstalk in TNF/EGF/insulin-induced signaling. PLoS Comput Biol 5(4):e1000340

Ashall L, Horton CA, Nelson DE, Paszek P, Harper CV, Sillitoe K, Ryan S, Spiller DG, Unitt JF, Broomhead DS, Kell DB, Rand DA, Sée V, White MR (2009) Pulsatile stimulation determines timing and specificity of NF-kappaB-dependent transcription. Science 324(5924):242–246

Banga JR, Balsa-Canto E (2008) Parameter estimation and optimal experimental design. Essays Biochem 45:195–209

Bhalla US, Ram PT, Iyengar R (2002) MAP kinase phosphatase as a locus of flexibility in a mitogen-activated protein kinase signaling network. Science 297(5583):1018–1023

Birkemeyer C, Luedemann A, Wagner C, Erban A, Kopka J (2005) Metabolome analysis: the potential of in vivo labeling with stable isotopes for metabolite profiling. Trends Biotechnol 23(1):28–33

Blüthgen N, Legewie S, Kielbasa SM, Schramme A, Tchernitsa O, Keil J, Solf A, Vingron M, Schäfer R, Herzel H, Sers C (2009) A systems biological approach suggests that transcriptional feedback regulation by dual-specificity phosphatase 6 shapes extracellular signal-related kinase activity in RAS-transformed fibroblasts. FEBS J 276(4):1024–1037

Bollard ME, Stanley EG, Lindon JC, Nicholson JK, Holmes E (2005) NMR-based metabonomic approaches for evaluating physiological influences on biofluid composition. NMR Biomed 18:143–162

Borisov N, Aksamitiene E, Kiyatkin A, Legewie S, Berkhout J, Maiwald T, Kaimachnikov NP, Timmer J, Hoek JB, Kholodenko BN (2009) Systems-level interactions between insulin-EGF networks amplify mitogenic signaling. Mol Syst Biol 5:256

Chaouiya C (2007) Petri net modelling of biological networks. Brief Bioinform 8(4):210–219

Chen WW, Schoeberl B, Jasper PJ, Niepel M, Nielsen UB, Lauffenburger DA, Sorger PK (2009) Input-output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. Mol Syst Biol 5:239

Chou IC, Voit EO (2009) Recent developments in parameter estimation and structure identification of biochemical and genomic systems. Math Biosci 219(2):57–83

Ciliberto A, Novak B, Tyson JJ (2005) Steady states and oscillations in the p53/Mdm2 network. Cell Cycle 4(3):488–493

Csikász-Nagy A, Battogtokh D, Chen KC, Novák B, Tyson JJ (2006) Analysis of a generic model of eukaryotic cell-cycle regulation. Biophys J 90(12):4361–4379

Dakna M, He Z, Yu WC, Mischak H, Kolch W (2009) Technical, bioinformatical and statistical aspects of liquid chromatography-mass spectrometry (LC-MS) and capillary electrophoresis-mass spectrometry (CE-MS) based clinical proteomics: a critical assessment. J Chromatogr B Analyt Technol Biomed Life Sci 877(13):1250–1258

Fall CP, Marland ES, Wagner JM, Tyson JJ (2002) Computational cell biology. Springer Science & Business Media, New York

Frieboes HB, Edgerton ME, Fruehauf JP, Rose FR, Worrall LK, Gatenby RA, Ferrari M, Cristini V (2009) Prediction of drug response in breast cancer using integrative experimental/computational modeling. Cancer Res 69(10):4484–4492

Funahashi A, Tanimura N, Morohashi M, Kitano H (2003) CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. BIOSILICO 1:159–162

Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc Natl Acad Sci U S A 100(12):6940–6947

Geva-Zatorsky N, Rosenfeld N, Itzkovitz S, Milo R, Sigal A, Dekel E, Yarnitzky T, Liron Y, Polak P, Lahav G, Alon U (2006) Oscillations and variability in the p53 system. Mol Syst Biol 2:2006.0033

Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP (2007) Universally sloppy parameter sensitivities in systems biology models. PLoS Comput Biol 3(10):1871–1878

Hanahan D, Weinberg RA (2000) The hallmarks of cancer. Cell 100(1):57–70

Heiner M, Koch I, Will J (2004) Model validation of biological pathways using Petri nets—demonstrated for apoptosis. Biosystems 75(1–3):15–28

Heyman H (2006) Quantification of activated signal transduction proteins using fast activated cell-based ELISAs (FACETM). Nat Appl Notes. doi:10.1038/an1562

Hoffmann A, Levchenko A, Scott ML, Baltimore D (2002) The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. Science 298(5596):1241–1247

Huang CY, Ferrell JE Jr (1996) Ultrasensitivity in the mitogen-activated protein kinase cascade. Proc Natl Acad Sci U S A 93(19):10078–10083

Jeffrey A (1993) Linear algebra and ordinary differential equations. CRC press, Boca Raton

Kholodenko BN (2000) Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. Eur J Biochem 267(6):1583–1588

Kholodenko BN, Demin OV, Moehren G, Hoek JB (1999) Quantification of short term signaling by the epidermal growth factor receptor. J Biol Chem 274(42):30169–30181

Kim D, Rath O, Kolch W, Cho KH (2007) A hidden oncogenic positive feedback loop caused by crosstalk between Wnt and ERK pathways. Oncogene 26(31):4571–4579

Kim SY, Ferrell JE Jr (2007) Substrate competition as a source of ultrasensitivity in the inactivation of Wee1. Cell 128(6):1133–1147

Kitano H (2007) Towards a theory of biological robustness. Mol Syst Biol 3:137

Klamt S, Saez-Rodriguez J, Lindquist JA, Simeoni L, Gilles ED (2006) A methodology for the structural and functional analysis of signaling and regulatory networks. BMC Bioinformatics 7:56

Krüger R, Heinrich R (2004) Model reduction and analysis of robustness for the Wnt/beta-catenin signal transduction pathway. Genome Inform 15(1):138–148

Lai X, Nikolov S, Wolkenhauer O, Vera J (2009) A multi-scale model accounting for the effects of JAK2-STAT5 signal modulation in Erythropoiesis. Comput Biol Chem 30:312–324

Le Novère N, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, Demir E, Wegner K, Aladjem MI, Wimalaratne SM, Bergman FT, Gauges R, Ghazal P, Kawaji H, Li L, Matsuoka Y, Villéger A, Boyd SE, Calzone L, Courtot M, Dogrusoz U, Freeman TC, Funahashi A, Ghosh S, Jouraku A, Kim S, Kolpakov F, Luna A, Sahle S, Schmidt E, Watterson S, Wu G, Goryanin I, Kell DB, Sander C, Sauro H, Snoep JL, Kohn K, Kitano H (2009) The systems biology graphical notation. Nat Biotechnol 27(8):735–741

van Leeuwen IMM, Byrne HM, Jensen OE, King JR (2007) Elucidating the interactions between the adhesive and transcriptional functions of b-catenin in normal and cancerous cells. J Theor Biol 247(1):77–102

van Leeuwen IM, Mirams GR, Walter A, Fletcher A, Murray P, Osborne J, Varma S, Young SJ, Cooper J, Doyle B, Pitt-Francis J, Momtahan L, Pathmanathan P, Whiteley JP, Chapman SJ, Gavaghan DJ, Jensen OE, King JR, Maini PK, Waters SL, Byrne HM (2009) An integrative computational model for intestinal tissue renewal. Cell Prolif 42(5):617–636

Lévi F, Altinok A, Clairambault J, Goldbeter A (2008) Implications of circadian clocks for the rhythmic delivery of cancer therapeutics. Philos Transact A Math Phys Eng Sci 366(1880):3575–3598

Mullassery D, Horton CA, Wood CD, White MR (2008) Single live-cell imaging for systems biology. Essays Biochem 45:121–133

Nelson DE, Ihekwaba AE, Elliott M, Johnson JR, Gibney CA, Foreman BE, Nelson G, See V, Horton CA, Spiller DG, Edwards SW, McDowell HP, Unitt JF, Sullivan E, Grimley R, Benson N,

Broomhead D, Kell DB, White MR (2004) Oscillations in NF-kappaB signaling control the dynamics of gene expression. Science 306(5696):704–708
Neves SR, Tsokas P, Sarkar A, Grace EA, Rangamani P, Taubenfeld SM, Alberini CM, Schaff JC, Blitzer RD, Moraru II, Iyengar R (2008) Cell shape and negative links in regulatory motifs together control spatial information flow in signaling networks. Cell 133(4):666–680
Nikolov S, Lai X, Liebal UW, Wolkenhauer O, Vera J (2010) Integration of sensitivity and bifurcation analysis to detect critical processes in a model combining signalling and cell population dynamics. Int J Syst Sci 41(1):81–105
Papin JA, Hunter T, Palsson BO, Subramaniam S (2005) Reconstruction of cellular signalling networks and analysis of their properties. Nat Rev Mol Cell Biol 6(2):99–111
Ptitsyn AA, Weil MM, Thamm DH (2008) Systems biology approach to identification of biomarkers for metastatic progression in cancer. BMC Bioinformatics 9(Suppl 9):S8
Qu Z, Weiss JN, MacLellan WR (2003) Regulation of the mammalian cell cycle: a model of the G1-to-S transition. Am J Physiol Cell Physiol 284(2):C349–C364
Ramalingam S, Honkanen P, Young L, Shimura T, Austin J, Steeg PS, Nishizuka S (2007) Quantitative assessment of the p53-Mdm2 feedback loop using protein lysate microarrays. Cancer Res 67(13):6247–6252
Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, Timmer J (2009) Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics 25(15):1923–1929
Rautio J, Barken KB, Lahdenperä J, Breitenstein A, Molin S, Neubauer P (2003) Sandwich hybridisation assay for quantitative detection of yeast RNAs in crude cell lysates. Microb Cell Fact 2(1):4
Rehm M, Huber HJ, Dussmann H, Prehn JH (2006) Systems analysis of effector caspase activation and its control by X-linked inhibitor of apoptosis protein. EMBO J 25(18):4338–4349
Reynolds AR, Tischer C, Verveer PJ, Rocks O, Bastiaens PI (2003) EGFR activation coupled to inhibition of tyrosine phosphatases causes lateral signal propagation. Nat Cell Biol 5(5):447–453
Ribba B, Colin T, Schnell S (2006) A multiscale mathematical model of cancer, and its use in analyzing irradiation therapies. Theor Biol Med Model 3:7
Rizk A, Batt G, Fages F, Soliman S (2009) A general computational method for robustness analysis with application to synthetic gene networks. Bioinformatics 25(12):i169–i178
Roth CM (2002) Quantifying gene expression. Curr Issues Mol Biol 4(3):93–100
Saez-Rodriguez J, Simeoni L, Lindquist JA, Hemenway R, Bommhardt U, Arndt B, Haus UU, Weismantel R, Gilles ED, Klamt S, Schraven B (2007) A logical model provides insights into T cell receptor signaling. PLoS Comput Biol 3(8):e163
Sahin O, Löbke C, Korf U, Appelhans H, Sültmann H, Poustka A, Wiemann S, Arlt D (2007) Combinatorial RNAi for quantitative protein network analysis. Proc Natl Acad Sci U S A 104(16):6579–6584
Saltelli A, Chan K, Scott E (200) Sensitivity analysis. Wiley, New York
Schauer N, Steinhauser D, Strelkov S et al (2005) GC-MS libraries for the rapid identification of metabolites in complex biological samples. FEBS Lett 579:1332–1337
Schefe JH, Lehmann KE, Buschmann IR, Unger T, Funke-Kaiser H (2006) Quantitative real-time RT-PCR data analysis: current concepts and the novel "gene expression's CT difference" formula. J Mol Med 84(11):901–910
Schilling M, Maiwald T, Bohl S, Kollmann M, Kreutz C, Timmer J, Klingmüller U (2005) Computational processing and error reduction strategies for standardized quantitative data in biological networks. FEBS J 272:6400–6411
Schmidt H, Madsen M, Dano S, Cedersund G (2008) Complexity reduction of biochemical rate expressions. Bioinformatics 24(6):848–854
Schoeberl B, Eichler-Jonsson C, Gilles E, Müller G (2002) Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. Nat Biotechnol 20:370–377
Schoeberl B, Pace EA, Fitzgerald JB, Harms BD, Xu L, Nie L, Linggi B, Kalra A, Paragas V, Bukhalid R, Grantcharova V, Kohli N, West KA, Leszczyniecka M, Feldhaus MJ, Kudla AJ,

Nielsen UB (2009) Therapeutically targeting ErbB3: a key node in ligand-induced activation of the ErbB receptor-PI3K axis. Sci Signal 2(77):ra31

Stelling J, Sauer U, Szallasi Z, Doyle FJ, Doyle J (2004) Robustness of cellular functions. Cell 118(6):675–687

Swameye I, Muller TG, Timmer J, Sandra O, Klingmuller U (2003) Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling. Proc Natl Acad Sci U S A 100(3):1028–1033

Turner TE, Schnell S, Burrage K (2004) Stochastic approaches for modelling in vivo reactions. Comput Biol Chem 28(3):165–178

Ullah M, Wolkenhauer O (2009) Investigating the two-moment characterisation of subcellular biochemical networks. J Theor Biol 260(3):340–352

Vera J, Bachmann J, Pfeifer AC, Becker V, Hormiga JA, Darias NV, Timmer J, Klingmüller U, Wolkenhauer O (2008) A systems biology approach to analyse amplification in the JAK2-STAT5 signalling pathway. BMC Syst Biol 2:38

Vera J, Balsa-Canto E, Wellstead P, Banga JR, Wolkenhauer O (2007) Power-law models of signal transduction pathways. Cell Signal 19:1531–1541

Vera J, Schultz J, Raatz Y, Ibrahim S, Wolkenhauer O, Kunz M (2010) Dynamical effects of epigenetic silencing of 14-3-3σ expression. Mol Biosyst 6(1):264–273

Vera J, Wolkenhauer O (2008) A system biology approach to understand functional activity of cell communication systems. Methods Cell Biol 90:399–417

Vera J, Rath O, Balsa-Canto E, Banga JR, Kolch W, Wolkenhauer O (2010) Investigating dynamics of inhibitory and feedback loops in ERK signalling using power-law models. Mol Biosyst 6(11):2174–2191

Wilhelm BT, Landry JR (2009) RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. Methods 48(3):249–257

Wolkenhauer et al (2010) Systems biologists seek fuller integration of systems biology approaches in new cancer research programs. Cancer Res 70(1):12–13

Xiayan L, Legido-Quigley C (2008) Advances in separation science applied to metabonomics. Electrophoresis 29(18):3724–3736

Chapter 8
Computational Tools for Systems Biology

Edda Klipp and Falko Krause

Abstract Analysis of health-relevant biochemical networks is being progressively facilitated by computational methods and Web-based resources. We give an overview of the available tools and methods. The reader is familiarised with the development of standards in computational systems biology regarding description and annotation of models, methods, and data standards which enable communication between different groups working in the same field and permit a systematic and precise exchange of information. A brief explanation is also provided of the capabilities of computer tools for model creation, simulation, and storage, as used in systems biology and cancer research. Web resources and strategies to combine different tools are introduced. We present examples of the successful application of tools and strategies in modelling and model-based prediction in disease-related areas of molecular and cellular biology.

8.1 Introduction

Biology is in the process of turning into a quantitative, information-driven science. Recent years have witnessed the development of powerful experimental methods to quantify many different properties of cellular life, such as mRNA levels, protein amounts, or protein-protein interactions. Large-scale data sets produced by genetics, genomics, molecular biology, and imaging need to be analysed with ever-increasing speed. Making effective use of such data sets requires interdisciplinary collaborations among statisticians, theoreticians, computer scientists as well as experimental biologists.

The databases available for many aspects of molecular and cellular biology, as well as of computational biology, are huge repositories for biological data gathered by various techniques. The information in the databases represents raw material for most types of modelling efforts. Modelling tools help to formulate theoretical ideas and hypotheses and to extract information relevant to these hypotheses from the raw material stored in the databases.

E. Klipp (✉)
Institute for Biology, Theoretical Biophysics, Humboldt-Universität zu Berlin,
Invalidenstr. 42, 10115 Berlin, Germany
e-mail: edda.klipp@rz.hu-berlin.de

A. Cesario, F. B. Marcus (eds.), *Cancer Systems Biology, Bioinformatics and Medicine,*
DOI 10.1007/978-94-007-1567-7_8,

Theoretical approaches and modelling tools have been developed to describe cellular processes in a quantitative way using computational models (see Chap. 7). The process of formulating, parameterizing, and evaluating models is often done by individual researchers, but has also become to some extent a community task. For example, data is retrieved in one lab, while modelling is done by collaborators. In addition, the model may be analysed using a tool provided by another group. Moreover, the model may be re-used by a group different from the original author, to be extended or adapted to newly arising data. Yet another example is the case of pathway maps assembled in common efforts (Herrgard et al. 2008).

Cancer systems biology makes increasing use of upcoming technologies. Over recent years, models have been developed and analysed for various cancer-related aspects of cellular life. Systemic understanding of cellular regulation and cancer development requires new ways of producing and analysing data. These models range from small regulatory networks to full-scale genomic models, explaining different aspects of cellular regulation. Clinically relevant studies, as well, need models that provide testable predictions. A few examples are provided below.

Network models comprise interaction information on all compounds of types such as proteins or genes. Model analysis may reveal insight into hidden relations. A large-scale study has systematically linked disorders and genes associated with the diseases (Goh et al. 2007). Another study has linked breast cancer susceptibility with centrosome dysfunction (Pujana et al. 2007).

Cells receive and process external information by signalling pathways which are intensively studied in experimental and computational systems biology. Amongst the most carefully investigated pathways are: the EGFR (Samaga et al. 2009; Schoeberl et al. 2002); the Wnt (Lee et al. 2003); the Jak/STAT (Swameye et al. 2003); the TGFβ (Kerrien et al. 2007; Schmierer et al. 2008); and NFkB pathways (Calzone et al. 2010; Fisher et al. 2006; Visvanathan et al. 2009). The mathematical models are represented as sets of ordinary differential equations (ODEs) describing the temporal dynamics of protein abundance and/or activity (see, as well, Chap. 7). The models serve various objectives, not least of which is the goal of understanding architecture and the observed dynamics. Another important aim is to rationalize the inter-dependence of independently measured data, such as ligand supply, phosphorylation states, protein interactions, and gene expression effects (e.g. Chen et al. 2009). Dynamic features, such as the effect of positive or negative feedback (Bluthgen et al. 2009; Kholodenko 2000), or crosstalk between different pathways (Borisov et al. 2009), have been studied extensively.

Cell cycle progression is highly regulated, and failure to respond to checkpoints or external signals may constitute a first step to cancer. Mathematical modelling of the cell cycle has initially focused on model organisms such as frog eggs or yeast (Chen et al. 2000, 2004), but is now also applied to mammalian cells (Alfieri et al. 2009). As mentioned above, these models primarily serve to test our understanding of the structure or wiring of the cell cycle machinery, but are also used

to investigate detailed questions such as the effect of mutations or of unreplicated DNA on cell cycle progression (Zwolak et al. 2009). The inclusion of more and more details, e.g. of the protein machinery, into models and simulations allows for a better prediction of cell cycle regulation by nutrition, signalling, or checkpoint activation.

Drug target prediction is an important goal of systems biology approaches. Given sufficiently well described networks and properties of the interaction of compound, mathematical models can be useful for predicting targets for treatment and testing the outcome of different target positions, treatment strengths, target combinations or temporal combination scenarios (Fitzgerald et al. 2006; Schulz et al. 2009). Even though this field is in its infancy, because of the continuing lack of sufficiently well understood networks, there are already very promising examples among signalling pathway models. One such is study of the ErbB network using sensitivity analysis (Schoeberl et al. 2009), which identified ErbB3 as a key node in response to ligands. Boolean modelling (assigning discrete activity values such as 'on' and 'off' or 0 and 1 to nodes which are updated in time, based on their inputs from other nodes) of the G1/S transition as regulated by the ErbB receptor, has revealed new potential targets in the case of *de novo* trastuzumab resistance in breast cancer (Sahin et al. 2009).

It has now been established that cancer treatment has different effects if supplied at different times of the day (Levi and Schibler 2007). Chronotherapy (see Chap. 15) takes advantage of the fact that some targets have varying accessibility at different times in the day, because living beings exhibit various rhythms, such as cell cycle, circadian or annual rhythms. Circadian rhythms have been investigated, their components identified (Ueda et al. 2005) and their dynamics described with mathematical models (Brown et al. 2008).

Taken together, different directions in systems biology combining experimental and computational analysis have already made promising first contributions to the understanding of aspects of cellular regulation and dysfunction. Further investigations will need precise and reproducible data and sensible mathematical descriptions in order to produce predictive and helpful models.

Systems biology and cancer research have to face problems inherent to biology: diversity, variability, and temporary inaccessibility. These factors are also reflected in the ambiguity of description. Biology was always characterized by the emergence and formulation of concepts. In our current times of rapid development, the Internet, and high-throughput operation, it has become imperative for concepts to be compatible. In practice, this requirement has led to the development of many standards, some of which we will discussed below.

Standards are being developed in all areas of cell and molecular biology, defining how data should be gathered and how results should be described.

In order to facilitate sharing of information, ideas, and efforts, new tools and data formats have been developed during the last few years. In the following section, we will give a short introduction about standards; present different language formats that are used in cell biology research; explain the concept of on-

tologies; list web resources for pathways, experimental information, and chemical compounds; and conclude with a selection of computational tools (e.g. Copasi, CellDesigner, VirtualCell), and an introduction to workflows (Taverna) and toolboxes. These explanations will provide links to modelling approaches and successful models of diseases, including in particular cancer-related processes, will be discussed.

8.2 Standards in Systems Biology

8.2.1 Standards Support Communication in Biological Research

Systems biology is frequently characterized by the paradigm of the 'iterative cycle'. Experimental investigations lead to scientific questions which require computational methods to answer them; on the basis of initial sets of experimental data, computational models are formulated that allow initial simulations and predictions; the predictions are tested by further experiments that yield additional data; the models are revised accordingly and improved models are created that lead to new experiments; and so it goes on. This iterative cycle is only possible where there is intensive communication between all partners involved in the process. This information exchange between experts in different fields, such as molecular biologists, computer scientists, mathematicians, biophysicists, engineers, and others, requires the development of common principles on how to formulate the results of research.

Moreover, many laboratories in the world work on related subjects. The building up of large data bases, e.g. for protein sequences or for results of specific experiments, requires contributions from many scientists and laboratories. Again, common principles of formulating and reporting research results are essential to facilitate the sharing of knowledge.

Currently and in the recent past, spreadsheets are the most common form of storing and exchanging data for single experiments, as well as for custom-made databases for sets of experiments. Since the structure of each spreadsheet has to be defined by the creator of the spreadsheet, two major problems arise. Firstly, the creator of the spreadsheet is not obliged to add all information necessary to understand the data stored in the spreadsheet (e.g. units of values are not declared). This will limit its reuse to people with inside knowledge about the dataset. Secondly, the structure of the spreadsheet is not fixed, e.g. column positions of data can vary in spreadsheets storing data from repeated experiments. This prevents the automated computational reuse of the data. To overcome this problem in cell biology research in general, and specifically in computational systems biology, a number of guidelines and standards have been developed.

Historically, one of the first standards was MIAME, a guideline to describe the *M*inimum *I*nformation *a*bout *M*icroarray *E*xperiment (Brazma et al. 2001). Since then, more than 100 guidelines have been formulated by the scientific research communities concerned, and many of them are formulated as Minimum Information (MI) about a specific subject.

When it became evident that agreement on minimum information guidelines is crucial for shaping a scientific research field and for enabling precise information exchange, the *M*inimum *I*nformation for *B*iological and *B*iomedical *I*nvestigations initiative (MIBBI, http://www.mibbi.org; (Taylor et al. 2008)) was created. MIBBI provides overviews of existing guidelines, and research initiatives have the options of registering their project in MIBBI. Moreover, MIBBI supports scientific communities in the formulation of new standards by offering a framework covering general requirements and structures.

While minimum information guidelines solve the problem of understanding data, they leave untouched the problem of making data computationally accessible. This is why more and more guidelines are being used as a basis for creating data formats that fit these guidelines. For the MIAMI guidelines, two data formats are officially recommended: MAGE-TAB which is based on spreadsheets and MAGE-ML (based on XML).

While the above is valid for experimental data, the process of storage and exchange of biochemical reaction networks and dynamic models poses some extra problems.

8.2.2 Language Formats

Biochemical models can be expressed in various formats. The most common and universal form of description is the textual description found in millions of research articles. However, as mentioned above, textual descriptions are not computationally accessible. This is why many research articles are accompanied by additional material making the structure and dynamics of the presented model accessible to computation.

For describing systems biology models in a standardized way, the most important formats are the *S*ystems *B*iology *M*arkup *L*anguage (SBML), the Cell Markup Language (CellML), the Biological Pathway Exchange (BioPAX) format, and The Proteomics Standards Initiative—Molecular Interaction exchange format (PSI-MI) (Klipp et al. 2007). Even though each of these formats was developed by different scientific communities to fit the needs of the home community, there are converters that can exchange most of the XML-based formats into each other, albeit with possible loss of information.

Box 8.1 gives a short overview of the properties of the different formats. Figure 8.1 presents an SBML model as an example.

Box 8.1 Different language formats

Format	Short Description	Reference
SBML	The special strength of SBML is for exchanging biochemical reaction networks whose dynamic is described by reaction kinetics. A multitude of tools enables the creation and simulation (by creating ordinary differential equation networks or stochastic representations) of SBML models. The format also allows the annotation of model elements with meta-information such as biological and mathematical terms.	(Hucka et al. 2003)
CellML	Similarly to SBML, CellML can express biochemical reaction networks and their dynamics. Its framework is more generic and not confined to expressing biochemical reaction networks. It supports modularization of models that enable the reuse of models and model components. CellML has a meta-data framework that allows the annotation of its elements.	(Lloyd et al. 2004)
BioPAX	Biological Pathway Exchange (BioPAX) format is based on the Web Ontology Language (OWL)/Resource Description Framework (RDF) formats. This means that BioPAX models are XML based knowledge representations that allow (for example) automated reasoning. While expressing no dynamic information, BioPAX's strength is that it can describe structural aspects of biochemical pathways in a semantically meaningful way.	Review in (Stromback and Lambrix, 2005)
PSI-MI	The Proteomics Standards Initiative-Molecular Interaction exchange format focuses on molecular interactions. Interactions can be annotated, by using specialized domain ontology, and subsequently grouped into pathways.	(Hermjakob et al. 2004)

The currently most widely used format in the Systems Biology community is SBML. It has a large user base that has created over 180 software systems that can create, modify, simulate and analyse information using SBML as a base for exchanging information (a sample will be presented below). There are specialized SBML model repositories, and a variety of databases include the export of SBML instances of their information (see below). SBML has a hierarchical structure with the following basic elements:

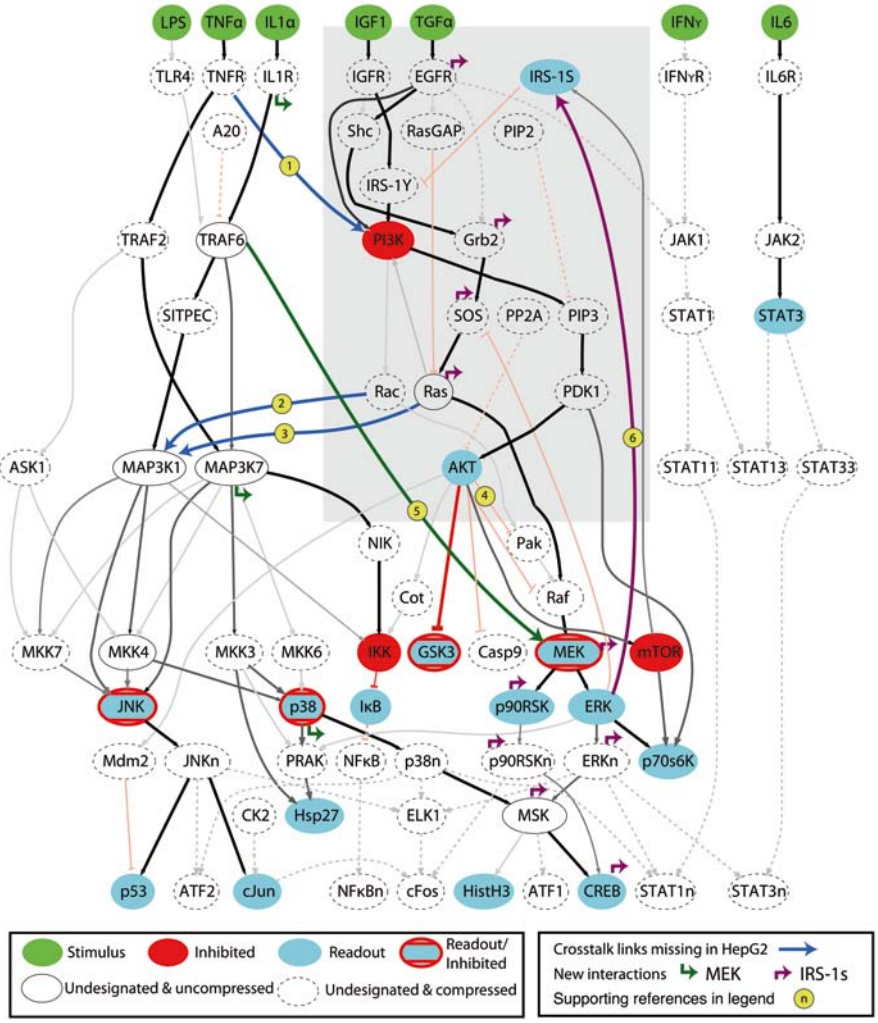


Fig. 8.1 Visualization of a signalling network in HepG2 cells using GraphViz. (taken from (Saez-Rodriguez et al. 2009))

Compartment: A container that has a defined size (Volume or Area in case of membranes such as the cytosol or the cell membrane). Each species in a model must be located in one compartment.

Species: A pool of (biochemical) entities that participate in reactions e.g. ATP or Glucose 6-phosphate.

Compartment Type and Species Type: Groupings of Compartments or Species (disregarding the location of each species)

Parameter: Generic constants and variables (e.g. the temperature).

Reaction: A reaction can describe a transformation, a transport or binding process of one or more species. Reactions can contain a mathematical function that describes the kinetics of the reaction.

Function Definition: A mathematical expression that is used throughout the model e.g. in the reaction kinetic

Initial Assignment: A mathematical expression that calculates the value of a variable at the start of a simulation.

Rule: An additional mathematical expression that defines the values of variables that cannot be defined by the kinetic of a reaction alone or by an initial assignment. In contrast to Initial Assignments, rules apply during the whole simulation time.

Constraint: A mathematical expression constraining the values of a variable of the model

Event: Events are triggered when a certain condition is satisfied (e.g. a species concentration is above a certain threshold). The event can change a set of variables.

Unit Definition: Either a unit of measurement that can overwrite the default unit (e.g. of the initial amount of a species), or a custom unit that is used in an expression of quantities in a model.

Even though SBML is not intended to be viewed by non-programmers, a glance at the code of an SBML document helps us to understand the structure. In Box 8.2, a sample of an SBML document is shown, illustrating one reaction in the cytosol where ERK proteins are phosphorylated by the MEK enzymes.

Box 8.2 Sample SBML file that describes the reaction ERK+ATP⇔ ERK_P+ADP. The model is annotated with one MIRIAM annotation and one SBO term

SMBL code	Explanation
`<?xml version="1.0" encoding="UTF-8"?>`	XML declaration
`<sbml xmlns="http://www.sbml.org/sbml/level2/version3"`	SBML delclaration
`level="2" version="3">`	level and version
`<model metaid="example_meta" id="example" name="example">`	
`<listOfCompartments>`	declaration of the compartment *cell*
`<compartment metaid="cell_meta" id="cell" name="cell" size="1">`	MIRIAM annotaion:
`<annotation>`	
`<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"`	
`xmlns:dc="http://purl.org/dc/elements/1.1/"`	
`xmlns:dcterms="http://purl.org/dc/terms/"`	
`xmlns:vCard="http://www.w3.org/2001/vcard-rdf/3.0#"`	
`xmlns:bqbiol="http://biomodels.net/biology-qualifiers/"`	
`xmlns:bqmodel="http://biomodels.net/model-qualifiers/">`	
`<rdf:Description rdf:about="#cell_meta">`	qualifier: is
`<bqbiol:is>`	
`<rdf:Bag>`	web resource: Gene Ontology term cell
`<rdf:li rdf:resource="urn:miriam:obo.go:GO%3A0005623"/>`	
`</rdf:Bag>`	

```
            </bqbiol:is>
          </rdf:Description>
        </rdf:RDF>
      </annotation>
    </compartment>
  </listOfCompartments>
  <listOfSpecies>
    <species metaid="ATP_meta" id="ATP" name="ATP" compartment="cell"/>
    <species metaid="ERK_meta" id="ERK" name="ERK" compartment="cell"/>
    <species metaid="ADP_meta" id="ADP" name="ADP" compartment="cell"/>
    <species metaid="ERK_P_meta" id="ERK_P" name="Phosphorylated ERK"
        compartment="cell"/>
    <species metaid="MEK_meta" id="MEK" name="MEK"
        compartment="cell" sboTerm="SBO:0000014"/>
  </listOfSpecies>
  <listOfReactions>
    <reaction metaid="reaction_0_meta" id="reaction_0"
        name="ATP_plus_ERK_&lt;=&gt;_ADP_plus_ERK_P">
      <listOfReactants>
        <speciesReference species="ATP"/>
        <speciesReference species="ERK"/>
      </listOfReactants>
      <listOfProducts>
        <speciesReference species="ADP"/>
        <speciesReference species="ERK_P"/>
      </listOfProducts>
      <listOfModifiers>
        <modifierSpeciesReference species="MEK"/>
      </listOfModifiers>
    </reaction>
  </listOfReactions>
</model>
</sbml>
```

declaration of species:

ATP

ERK

ADP

Phosphorylated ERK

MEK (Enzyme)

As mentioned above, SBML permits the adding of annotation to its elements. Annotations for the model itself can be the name and contact data of the model creator, and a text describing the model in words. In addition, the model and any element of the model can be annotated with what is usually referred to as MIRIAM annotations.

The Minimal Information Required in the Annotation of Models (MIRIAM—Le Novere et al. 2005) is based on a collaborative research publication that defines a set of rules for encoding and annotating biochemical models. The rules cover technical aspects, naming conventions, and description of the modelled object in a reference document, while making provision for all parameters of presented simulations. They provide a guideline for authors on how to achieve so careful an annotation of their model that it can be understood, re-implemented, and curated by others.

In SBML a part of these rules was adopted in the form of Resource Description Framework (RDF)-based annotations. The RDF format permits the communication

Table 8.1 Relations in model annotations

Subject	Predicate	Object
D-glucose	Is version of	urn:miriam:obo.chebi:CHEBI%3A17234 http://www.ebi.ac.uk/chebi/entry for glucose
SBML Element	BioModels Qualifier	MIRIAM urn

of relations of model elements to external Web resources, by describing the biological meaning of model elements. This is done by a subject-predicate-object-tuples, similar to natural language, where the subject is the model element, the object is a 'database object' of a web resource, and the predicate is the relationship between the model element and the external resource. See Table 8.1 for an example.

The predicate or relationship can be chosen from a list of so-called qualifiers that can be found at http://www.biomodels.net/qualifiers/, and the objects or Web resources that can be used can be found at http://www.ebi.ac.uk/miriam/. This ensures that all relationships are well defined and that the Web resources can be found through unique names. Both lists are under constant development and open to additions.

SBML elements can be further annotated with Systems Biology Ontology terms (SBO—Le Novere 2006) that describe the meaning of elements in a modelling process, clarifying, for example that a parameter is a Michaelis constant. In the following sub-section there will be a description of what ontologies are, and why they are important in systems biology.

8.2.3 Ontologies

For a long time, biological research has employed multifarious notations for describing objects, processes or observations, which has often led to confusion about the meaning of a statement. A prominent example is the variety of names for proteins, depending on who described them, under what conditions and in which organism. Ambiguity in names and concepts hinders the exchange of information and the advance of understanding. This problem is approached by the development of ontologies and controlled vocabularies (CV). While the terms ontology and CV are often used synonymously, they do not have the same meaning.

A CV is in the simplest case a list of defined words, each representing a concept that is used to describe a certain subject. Agreeing on a CV can ease the communication between two scientific communities that use identical words with different meanings. In more advanced cases, CVs group together terms that describe a similar subject. This grouping might be done in a hierarchical structure, where the root is the most general term, and the child elements define the subject in more detail. This is also referred to as taxonomy, best known for species links.

An ontology (as in the meaning within information science) is also based on a list of defined concepts, but extends it with a list of defined relations between the concepts. There are several language formats that store ontologies. The computer readability of these formats enables the creation of software that can do automatic reasoning (e.g. if A "is a" B and B "is a" C then A "is a" C) and detect logical contradictions (e.g. if A "is not a" B and B "is a" C then it cannot be true that A "is a" C).

The most widely used ontology in systems biology is the Gene Ontology (GO). It is one of the oldest biological ontologies, and its controlled vocabulary of genes and gene product attributes can be used across different species. It is organised into three main branches: biological processes (e.g. GO:0009893: positive regulation of metabolic process); cellular components (e.g. GO:0005829: cytosol); and molecular functions (e.g. GO:0003700: transcription factor activity). Since the creation of Gene Ontology, other communities have developed other domain-specific ontologies. To avoid a duplication of effort, and ensure the quality of newly developed ontologies, the Open Biomedical Ontologies (OBO) Foundry was created. Today it hosts almost 100 ontologies that can be freely downloaded from their website, http://obofoundry.org/. Among these ontologies is the above-mentioned SBO; another is the Human Phenotype Ontology that defines different types of cancer. Given the large number of existing ontologies, it can be hard to find the right one for a specific field; for that reason, the EBI provides an Ontology Lookup Service, located at http://www.ebi.ac.uk/ontology-lookup/.

8.3 Web Resources

Web resources (WR) of various types are a rich source of data for modelling in systems biology. There are comprehensive WRs summarizing experimental results of large scientific communities; these resources often also contain rich annotation of, and inter-linkage between, different types of data.

In contrast, many projects nowadays come with their own project-related databases, furnishing the community with direct access. An important problem is how to encode information in a form convenient for users to find, retrieve, and interpret correctly. Moreover, the conditions under which the data has been measured need to be known.

The computational systems biology community is not only a recipient of data provided by experimental researchers. It also supplies the results of computational analysis in appropriate databases for computational models, such as those of JWS online and BioModels discussed below.

The world of web resources (WS) is steadily growing, and it is not our intention to give a complete overview here. Those WS chosen are resources relevant for systems biology. Although most databases have specific purposes, they often contain

Table 8.2 Classes of web resources in Pathguide with selected examples

Class	Examples
Protein-Protein Interactions	BIND—Biomolecular Interaction Network DB DIP—Database of Interacting Proteins IntAct—Protein-Protein Interactions MINT—Molecular Interaction DB PC—Pathway Commons
Metabolic Pathways	EcoCyc—Encyclopedia of E. coli Genes and Metabolism BRENDA—Braunschweig Enzyme Database Reactome—Reactome Knowledge Base Yeast consensus metabolic network
Signalling Pathways	eMIM—Electronic Molecular Interaction Map SigPath—Signalling Pathway Information System STKE—Signal Transduction Knowledge Environment
Pathway Diagrams	BioCarta—BioCarta Pathway Diagrams KEGG—Kyoto Encyclopaedia of Genes and Genomes
Transcription Factors/Gene Regulatory Networks	aMAZE—Protein Function and Biochemical Pathways Project miRWalk—Predicted and Validated microRNA Targets Targetfinder—Finding transcription factor targets
Protein-Compound Interactions	BindingDB—The Binding Database TTD—Therapeutic Target Database PC—Pathway Commons PDB-Ligand—PDB-Ligand
Genetic Interaction Networks	BIND—Biomolecular Interaction Network Database GeneNet—Genetic Networks

duplicated and overlapping information. Some WRs are redundant and some are cross-referenced.

An overview on available pathway-related databases is provided by the pathway resource list Pathguide, http://www.pathguide.org, (Bader et al. 2006). The authors provide a link, a short description and information about availability and compatibility with various standards, as well as the classification given in Table 8.2.

For their value in computational systems biology, we wish to add the following categories:

- Databases for protein information, UNI-PROT, SGD, Swiss-PROT
- Database for chemical compounds: ChEBI, PubChem-substance
- Database for experimental information: Array Express

Below, we will describe in more detail a few databases, which are used for model creation and storage:

8.3.1 JWS Online and BioModels Database

JWS online and the BioModels Database are both repositories for curated SBML models. Most of their content is available from both resources. While JWS on-

line focuses on providing an online simulation platform, the BioModels Database specializes in providing model annotations, displaying them and interlinking the models with external resources, as well as providing other language formats from automated conversions. Currently, JWSonline contains more than 90 models and BioModels around 240 curated models, among them a number that are directly relevant for cancer, for example:

- A model for Wnt signalling and the roles of APC and Axin was derived from experimental and theoretical analysis, with relevance for colon cancer (Lee et al. 2003).
- A mathematical model describes the regulation of the G1 phase of Rb +/+ and Rb–/– mouse embryonic fibroblasts and an osteosarcoma cell line (Obeyesekere et al. 1997).
- General models are available for eukaryotic cell cycle regulation (e.g. Csikasz-Nagy et al. 2006).

JWSonline—http://jjj.biochem.sun.ac.za, (Olivier and Snoep 2004) is a database that stores and provides access to curated models of biochemical pathways. The models are available in SBML and Mathematica format. It is possible to simulate the model directly via the Internet, as well as to change parameters or calculate steady states and metabolic control coefficients, i.e. a version of local sensitivity coefficients.

BioModels—http://www.biomodels.net, (Le Novere et al. 2006) is a free, centralized database of curated published quantitative kinetic models of biochemical and cellular systems. The BioModels.net project is an international effort to (1) define agreed standards for model curation, (2) define agreed vocabularies for annotating models with connections to biological data resources, and (3) provide a free, centralized publicly-accessible database of annotated computational models in SBML and other structured formats. Before entering the database, each model is carefully curated, i.e. checked manually, to verify that it corresponds to the reference publication and that it gives the proper numerical results. The components of the models are annotated with terms from controlled vocabularies (CVs) and links to other relevant data resources. This allows the users to search accurately for the models they need.

8.3.2 KEGG

The Kyoto Encyclopaedia of Genes and Genomes (KEGG), http://www.genome.jp/kegg/ (Kanehisa and Goto 2000) is a knowledge base for systematic analysis of gene functions, linking genomic information with higher order functional information. KEGG provides, along with much else information on pathway maps and modules (KEGG PATHWAY); on functional hierarchies and ontologies (KEGG BRITE); ortholog annotation (KEGG ORTHOLOGY); genomes, genes, and proteins (KEGG GENES); chemical compounds (KEGG COMPOUND); glycans and

reactions (KEGG LIGAND); and finally drugs (KEGG DRUGS). In KEGG, diseases are viewed as perturbed states of the molecular system, and drugs as agents of perturbation to the molecular system (Kanehisa et al. 2010). Disease information is computerized in two forms: pathway maps and gene/molecule lists. The KEGG PATHWAY database contains pathway maps for the molecular systems in both normal and perturbed states. In the KEGG DISEASE database, each disease is represented by a list of known disease genes; any known environmental factors at the molecular level; diagnostic markers; and therapeutic drugs, which may reflect the underlying molecular system. The KEGG DRUG database contains chemical structures and/or chemical components of all drugs in Japan, including crude drugs and Traditional Chinese Medicine (TCM) formulae, as well as drugs in the USA and Europe. KEGG was used to assign and annotate experimental data to protein networks in cancer research, e.g. for bioinformatics analysis of data on rat bladder cancer (Arum et al. 2010), target identification in cases of gastrointestinal stromal tumours (Hur et al. 2010); or for detection of novel breast cancer metastasis-associated proteins (Ho et al. 2009).

8.3.3 Reactome

Reactome—http://www.reactome.org, (Matthews et al. 2009; Vastrik et al. 2007) is a free online open-source, curated pathway database encompassing many areas of human biology. Information is supplied by expert biological researchers, maintained by the Reactome editorial staff, and cross-referenced to the NCBI Entrez Gene, Ensembl and UniProt databases, the UCSC and HapMap Genome Browsers, the KEGG Compound and ChEBI small molecule databases, PubMed, and GO. Its current release features 2975 human proteins, 2907 reactions, and 4455 literature citations. A new entity-level pathway viewer, and improved search and data mining tools, facilitate searching and visualizing pathway data and the analysis of user-supplied high-throughput data sets. Reactome's data content and software can all be freely used and redistributed under open source terms. Reactome contains many signalling pathways related to cancer, such as Wnt and TGFβ signalling. It has been used for network retrieval, reconstruction, and annotation in disease-relevant studies (Bauer-Mehren et al. 2009; Raman et al. 2009; Renkonen et al. 2010)

8.3.4 BioCyc

BioCyc—http://www.biocyc.org, (Romero and Karp 2004) is a gateway to Pathway/Genome databases based on the software Pathway Tools (free to academic

users). Currently BioCyc contains 505 different organism databases. Each database can be used on-line, downloaded and used in a local version of the Pathway Tools Software. It permits searching for database objects like pathways or genes with a fine-grained control. Pathway Tools also provides bioinformatics applications for data analysis and visualization. The BioCyc databases are classified into three categories by their level of human curation, starting with the two best-curated databases EcoCyc (*Escherichia coli*) and MetaCyc (Multiorganism Metabolic Pathway and Enzyme Database), followed by 23 databases that are moderately well curated, and 482 databases with automated annotations only. BioCyc encourages the creation of new Cyc databases. The creation process usually starts by uploading the genome data of a newly sequenced species, which is then automatically curated (e.g. by prediction of metabolic pathways) using the Pathway tools software. The Pathway Tools Homepage is http://bioinformatics.ai.sri.com/ptools/.

8.3.5 BRENDA

BRENDA—http://www.brenda-enzymes.org/, (Barthelmes et al. 2007), the BRaunschweig Enzyme Database, is the largest publicly available enzyme information system in the world. Its information on function, structure, sequence, localization, disease-relation, isolation, stability, and/or ligand-relation has been extracted either manually or by text-mining (in AMENDA) from primary literature. About 5000 different enzymes are covered. Enzymes are classified according to the Enzyme Commission list of enzymes. Each entry is linked to the enzyme source and literature reference. Programmers can assess BRENDA using the Simple Object Assess Protocol (SOAP) interface. The tables too can be downloaded. For systems biology projects, the BRENDA database is highly valuable for finding kinetic parameter values such as IC50, Km, or Ki values or turnover numbers, either directly or by inferring parameter values from compatible values measured in other species or for related proteins.

8.3.6 SABIO-RK

SABIO-RK—http://sabio.villa-bosch.de/, (Wittig et al. 2006), the *S*ystem for the *A*nalysis of *Bio*chemical Pathways—*R*eaction *K*inetics is a Web-based application based on the SABIO relational database, that contains information about biochemical reactions and related kinetic equations annotated with parameters, and the experimental conditions under which these parameters were measured. It aims to support modellers in the setting-up of models of biochemical networks,

but is also useful for experimentalists or researchers with an interest in biochemical reactions and their kinetics. Information about reactions and their kinetics can be exported in SBML format. The modelling tool CellDesigner (see below) integrates SABIO-RK to support model building (Funahashi et al. 2007).

8.4 Computational Tools

The computational part of systems biology is increasingly being eased by the availability of specialized and general-purpose modelling tools. These tools are usually employed to perform the following tasks:

- *Model coding*, either as set of equations or mathematical expressions, or by a graphical representation of the model structure (as, for example offered by CellDesigner).
- *Model analysis* including, for example, calculation of steady states and sensitivities, stability, eigenvalues of system matrices, consistency checks, parameter scans, etc.
- *Parameter estimation* from data
- *Model simulation*
- *Output* of results in both graphical and textual form.

The need for communication and exchange has led to the development of standardized ways to describe models in machine-readable format. In this connection we mention the Systems Biology Markup Language (SBML) and the Cell Markup Language (CellML). Interestingly, both approaches are used by different scientific communities, but share many common properties and developments. Communication and exchange between communities cannot but facilitate effective use of the best solutions for various problems.

For a more comprehensive overview on computational modelling tools, the reader can consult recent literature and textbooks (Klipp et al. 2007, 2009) and SBML websites (sbml.org).

Most of the tools that will be presented here can be found on the Linux based live DVD **SB.OS** which can be freely downloaded from http://www.sbos.eu. SB.OS can be started on any computer with a DVD drive or a USB port that is able to boot from a USB stick, without modifying the current operating system or installing additional software. While using the pre-installed software it is possible to read and write from local or external devices (e.g. a build in hard drive or a USB stick). Box 8.3 gives an overview on the tools that are introduced below.

Box 8.3 Frequently used tools for analysis and simulation of mathematical models for biochemical processes

Tool	Short description/functionality	Reference
COPASI	Implementation, simulation and analysis of biochemical models in ODE format. Enables also stochastic simulations	(Hoops et al. 2006)
CellDesigner	Graphical representation of biochemical networks; implementation and simulation of ODE models	(Funahashi et al. 2003)
XPP-AUT	Analysis and simulation for mathematical model of various types; bifurcation analysis	(Ermentrout and Chow 2002)
semantic SBML	Creating, annotation, and merging of SBML models using KEGG identifiers	(Krause et al. 2010)
SBML-PET	Parameter estimation and simulation for SBML models	(Zi and Klipp 2006)
Virtual Cell	Deterministic and stochastic simulation of cell processes including diffusion and membrane transport; uses client-server system	(Moraru et al. 2008)
MesoRD	Stochastic simulation of reaction and diffusion systems in 3D	(Hattne et al. 2005)
BioNetGen	Generation and simulation of rule-based models for biochemical systems	(Blinov et al. 2004)
CellNetAnalyzer	Analysis of regulatory networks based on network topology; network visualization	(Klamt et al. 2006)
Squad	Dynamic simulation of signalling networks based on discretization	(Di Cara et al. 2007)
BooleanNet	Simulation of biochemical networks in form of Boolean models; has various updating rules	(Albert et al. 2008)
SBML Toolbox	Collection of functionalities for SBML models based on MATLAB, including reading, writing, graphical presentation, simulation, and many more	(Keating et al. 2006)
SB Toolbox	MATLAB toolbox for simulation of ODE models as well as parameter estimation, steady state and bifurcation analysis, model reduction	(Schmidt and Jirstrand 2006)
SBW	Facilitates communication and combined use of heterogeneous applications for SBML models	(Hucka et al. 2002)

8.4.1 Tools for Model Formulation and Simulation

Here, we introduce tools frequently used in the systems biology community and briefly explain their functionality. For a comprehensive overview the reader is referred to the SBML website (sbml.org).

COPASI—Complex Pathway Simulator, http://www.copasi.org, (Hoops et al. 2006) is an application for the simulation and analysis of biochemical networks. It works with both the chemical reaction network and the mathematical set of differential equations. It features deterministic and stochastic time course simulations, steady state analysis, metabolic control analysis, elementary mode analysis, time scale separation analysis, parameter scans, and optimization of arbitrary target functions, parameter estimation using experimental data, and import and export of SBML. Versions exist for Windows, Linux, MacOS X and Solaris.

COPASI is frequently used for modelling cancer and health-relevant processes. Examples are modelling of TGFβ signalling (Adra et al.), lipid metabolism (Kuhnel et al. 2008), and angiotension receptor occupation (Vauquelin et al. 2006).

CellDesigner—celldesigner.org, (Funahashi et al. 2003; Kitano et al. 2005) is a structured diagram editor for drawing gene-regulatory and biochemical networks. Networks are drawn based on the process diagram, with a graphical notation system proposed by Kitano, and are stored using SBML. The model can be annotated and referenced to databases. An internal simulation tool allows immediate time course simulation. SBML exports and imports and use of analysis packages through Systems Biology Workbench (SBW) are possible. CellDesigner is more and more being used to present networks of health-relevant regulation processes in a standardized manner. Examples of such networks include the regulation of the cell cycle by the retinoblastoma protein (RB/RB1) mutations involved in many human cancers (Calzone et al. 2008); the interaction of *Helicobacter pylori* with epithelial cells which may cause gastric cancer (Dampier and Tozeren 2007); and mapping of the epidermal growth factor receptor (EGFR) signalling pathway (Oda et al. 2005).

XPPAUT—http://www.math.pitt.edu/~bard/xpp/xpp.html, (Ermentrout 2002) is a tool for simulating, analyzing, and animating dynamical systems. It facilitates the solving of differential equations, difference equations, delay equations, integral equations, some partial differential equations, functional equations, stochastic equations and boundary value problems. It enables faster and more flexible numerical integration as many general-purpose tools. XPPAUT contains the code for the popular bifurcation program, AUTO, making it possible to switch back and forth between XPPAUT and AUTO, using the values of one program in the other and vice-versa. XPPAUT is frequently used to explore biochemical systems showing oscillations, and was employed, for example, to analyse a mathematical model of pancreatic duct cell secretion (Whitcomb and Ermentrout 2004), and to study dynamics in neuronal cells after dopamine stimulation (Lindskog et al. 2006).

semanticSBML—http://www.semanticsbml.org, (Krause et al. 2010) can be used to create, annotate and merge SBML. For the creation of an SBML model a list of KEGG identifiers can be inserted from which the model is constructed. The annotation features of semanticSBML allow the user to search, add, and update MIRIAM annotations as well as SBO terms. The search for MIRIAM annotations is conducted on a local database that combines the information of different web resources (KEGG, ChEBI, PubChem, GeneOntolgy, Reactome, and Taxonomy). The merging or combination of different models written in SBML uses MIRIAM annotations to create an initial suggestion for the merged model that can be refined

by the user. Conflicts that may arise in the merging process (e.g. different initial conditions of substances) are highlighted and can be resolved.

SBML-PET—sysbio.molgen.mpg.de/SBML-PET, (Zi and Klipp 2006). SBML-PET is a *P*arameter *E*stimation *T*ool for SBML models. It enables parameter estimation for biological models including signalling pathways, gene regulation networks and metabolic pathways. SBML-PET supports events in models representing such things as changing external conditions, or the onset of an experiment. It runs on Linux and Cygwin on Windows. It has been used to estimate parameters for a model of the cancer-relevant transforming growth factor β (TGFβ)-activated SMAD signalling pathway (Zi and Klipp 2007).

8.4.2 Spatial and Temporal Simulation

Virtual Cell—vcell.org, (Moraru et al. 2008) facilitates analysis, modelling, and simulation of cellular processes, including reaction kinetics, diffusion and flow, membrane transport, and electrophysiology. It is based on a central database and permits the connection of molecular processes with cell geometries derived from microscopic images. The tool presents a Web-based distributed client-server system that generates the mathematical code needed to run simulations from the biological model introduced by the author. Mathematicians may opt to use the math framework, based on the Virtual Cell Math Language, for creating their own mathematical descriptions. Both deterministic and stochastic algorithms are supported for describing and running non-spatial simulations; a full partial differential equation solver using the finite volume numerical algorithm is available for reaction-diffusion-advection simulations in complex cell geometries, including 3D geometries derived from microscope images. Virtual Cell enables model exchange in SBML, CellML, or MatLab format. It adheres to MIRIAM standards (see below). Virtual Cell is open source.

MesoRD—Mesoscopic Reaction Diffusion Simulator, mesord.sourceforge.net (Hattne et al. 2005) is a tool for the stochastic simulation of 3D reaction and diffusion systems. More precisely, it is an implementation of the Next Subvolume Method, which is an exact method for simulating the Markov process corresponding to the reaction-diffusion master equation. MesoRD also supports mean-field simulations. Hitherto, spatio-temporal simulations were primarily used to understand the impact of spatial distribution and stochasticity on the behaviour of cellular systems in general. Specific applications include the characterization of the morphogen gradient in early *Drosophila* embryos, an important model system, (Wu et al. 2007) and clustering of AMPA receptor in Purkinje cell synapses (Launey 2007).

8.4.3 Boolean and Logical Models

BioNetGen—cellsignalling.lanl.gov/bionetgen, (Blinov et al. 2004) is a tool for automatically generating mathematical models of biological systems from user-

specified rules for biomolecular interactions. Rule-based modelling is based on the representation of molecules as structured objects and of molecular interactions as rules for transforming attributes of the objects. Rules are specified in the BioNetGen language, which enables precise, visual, and extensible representation of biomolecular interactions. The language was designed with protein-protein interactions in mind. It allows for systematic incorporation of specific details of protein-protein interactions, thereby handling the combinatorial explosion related to full consideration of all potential interactions. A user can explicitly indicate the parts of proteins involved in an interaction; the conditions upon which an interaction depends; the connectivity of proteins in a complex interaction; and other aspects of protein-protein interactions.

CellNetAnalyzer—http://www.mpi-magdeburg.mpg.de/projects/cna/cna.html, (Klamt et al. 2006) is a package for MATLAB and provides a comprehensive and user-friendly environment for structural and functional analysis of biochemical and cellular networks. CellNetAnalyzer facilitates the analysis of metabolic as well as signalling and regulatory networks solely on their network topology, i.e. independently of kinetic mechanisms and parameters. The core concept of visualization and interactivity is realized by interactive network maps where the abstract network model is linked with network graphics. CellNetAnalyzer provides a powerful collection of tools and algorithms for structural network analysis which can be started in a menu-controlled manner within the interactive network maps. It is the successor and further development of FluxAnalyzer 5.3, and has been applied to detect targets and effects of drugs relevant to the metabolic disorder Smith-Lemli syndrome (Eapen 2007).

Squad—enfin.org/squad, (Di Cara et al. 2007) allows for dynamic simulation of signalling networks with a standardized dynamical systems approach (Mendoza and Xenarios 2006). It is intended for situations where kinetic parameters or other experimental data elements are scarce or unavailable. The software converts a gene-regulatory network into a discrete dynamical system to identify steady states, and then implements a continuous dynamical system applying defined rules. Squad has been applied in the study of T-helper cell differentiation (Di Cara et al. 2007)

BooleanNet—atlas.bx.psu.edu/booleannet/booleannet.html, (Albert et al. 2008) is a Python module for simulating biological regulatory networks. It is able to perform simulations in several modes using synchronous updates, asynchronous updates, ranked asynchronous updates, time synchronous updates, and piecewise differential updates. Models can be created by using its intuitive mini-language.

8.4.4 General Purpose Tools

Modelling and programming problems can nowadays be solved with many different tools or packages, among them Mathematica®, MatLab®, R, C/C++, Perl, or Python. Mathematica (Wolfram Research, http://www.wolfram.com) is an algebraic and symbolic maths package. Matlab (MathWorks, http://www.mathworks.com) is

predominantly a numerical computation tool. Both tools share many similarities; both are commercial. There are so-called toolboxes for Matlab providing additional functionalities, among them toolboxes for systems biology purposes, as presented below. R is a free software environment for statistical computing and graphics. C/C++ is a standardized static middle-level general-purpose programming language. Perl is a high-level dynamic programming language. The programming language Python is object-oriented, imperative and very powerful. All these tools are frequently used to solve systems biology mathematical or statistical problems. Mathematica for example, is employed for simulating cell cycle model (Alfieri et al. 2009); for non-linear compartmental modelling of breast cancer (Cheung et al. 2008); and for statistical analysis of cancer associated gene expression matrices (Wahde et al. 2002). Matlab is also serviceable as a compartmental model of tumour targeting, to predict the magnitude, specificity, time dependence, and affinity dependence of a tumour (Schmidt and Wittrup 2009); for dosimetry (Gossio et al. 2009); and for image analysis (Rexhepaj et al. 2008). R packages have been developed for various statistical analyses of cancer-related data (Frohlich et al. 2008; Holleczek et al. 2009; Marot et al. 2009; Mattfeldt et al. 2007).

SBML Toolbox—sbml.org/software/sbmltoolbox, (Keating et al. 2006) linking to libSBML, provides a set of functions that permit an SBML model to be imported into MATLAB and stored as a structure within the MATLAB environment. SBMLToolbox provides functions for reading, writing, and validating SBML models; viewing model structures in a simple Graphical User Interface (GUI); converting models into a symbolic form suitable for use with MATLAB's Symbolic Math Toolbox; and simulating models using MATLAB's ordinary differential equation solvers. It includes functions for translating an SBML document into a MATLAB_SBML structure; saving and loading these structures to/from a MATLAB data file; validating each structure (e.g. reaction structure); viewing the structures using a set of GUIs; and converting elements of the MATLAB_SBML structure into symbolic form, thereby allowing access to MATLAB's Symbolic Toolbox. Functions exist for facilitating simulation using MATLAB's ODE solvers, as well as one that will output an SBML document from the MATLAB_SBML structure definition of a model.

SB Toolbox—http://www.sbtoolbox.org, (Schmidt and Jirstrand 2006) is an open-source Systems Biology Toolbox for MATLAB. Its functionality includes import of SBML models, deterministic and stochastic simulation, steady-state and stability analysis, parameter estimation and sensitivity analysis, network identification, bifurcation analysis, optimization, determination of stoichiometric matrix and simple model reduction. It is built in modular fashion with the object classes SBmodel and SBdata, which represent models and experimental data, and employs ODEs.

SBW—http://www.sys-bio.org, (Hucka et al. 2002), the Systems Biology Workbench (SBW), is a software framework that allows heterogeneous application components, written in diverse programming languages and running on different platforms, to communicate and use each other's capabilities via a fast binary encoded-message system. The interfaces to the system are encapsulated in client-side libraries provided for different programming languages. The list of SBW-enabled programs contains programmes specializing in the graphical creation of reaction

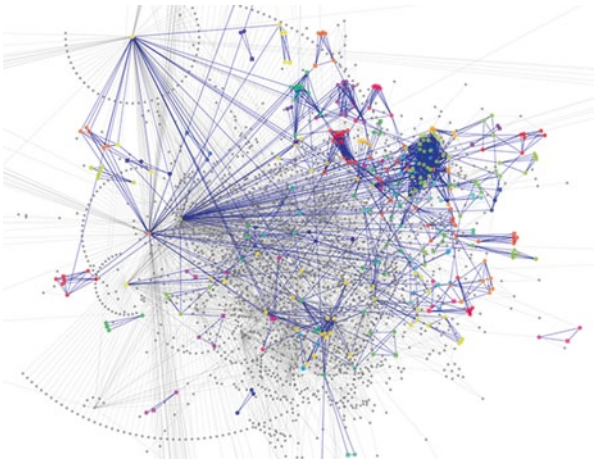


Fig. 8.2 Visualization of a large graph created with Cytoscape

networks (JDesigner and CellDesigner); simulation tools (Jarnac and TauLeapService); analysis and optimization tools (Metatool, Bifurcation and Optimization); and utilities like the Inspector module, which provides information about other modules.

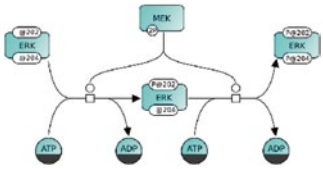
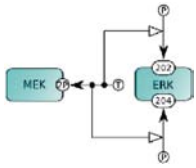
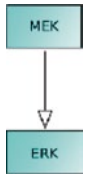
8.5 Visualizing Networks

Graph visualizations are often used for data representation and for the display of biochemical networks described by systems biology models.

There are two common methods of creating a graph-type representation. The first method is to draw the graph manually; this generally means that a vector-graphic editor software (OpenOffice Draw, Inkscape, Adobe Illustrator), or graphic editor components of software tools (like PowerPoint), are used to draw nodes, edges and other components of the graph image. This, however, is only feasible for very small networks, and for networks that do not have to be redrawn frequently, in which case it is common for graph visualizations to be generated automatically. Among the most popular packages for graph visualization is the open-source software Graphviz. It has a simple graph notation language called dot, and is easily customizable. Its most important features are its graph layout algorithm that automatically places nodes and edges of a graph in a professional-looking format.

While the graph layout algorithms for Graphviz work well with small networks (about 50 nodes), the exponential growth of biological data has created the challenge of visualizing networks with thousands to millions of nodes. For this purpose the open-source bioinformatics platform Cytoscape was created. Whereas its main goal was originally to visualize molecular interaction networks, in its current state it has transcended this initial aim and now offers a wide variety of plug-in-based extensions. Its strength, however, lies in its capacity to visualize networks containing millions of nodes in a good-looking format on an average desktop computer within a reasonable timeframe. See Fig. 8.2.

Table 8.3 SBGN languages

The SBGN *Process Description* (Saevels et al. 1996) language shows the temporal courses of biochemical interactions in a network. It can be used to show all the molecular interactions taking place in a network of biochemical entities, with the same entity appearing multiple times in the same diagram.	
The SBGN *Entity Relationship* (ER) language allows users to see all the relationships in which a given entity participates, regardless of the temporal aspects. Relationships can be seen as rules describing the influences of entity nodes on other relationships. ER can be used to describe the molecular biology behind a model.	
The SBGN *Activity Flow* (AF) language depicts the flow of information between biochemical entities in a network. It omits information about the state transitions of entities, and is particularly convenient for representing the effects of perturbations, whether genetic or environmental in nature. It can be used to describe coarse-grained interaction networks.	

Graph visualizations face the problem that they should be clearly comprehensible to their observers. This problem was recently addressed for graph visualizations of biochemical networks. The Systems Biology Graphical Notation (SBGN) project created three sets of symbol languages respectively describing process descriptions, entity relationships and activity flows in a standardized way (see Table 8.3). The idea behind the SBGN languages stems from the notation of circuit schematics that is universally used, and can be understood by every electronics engineer familiar with it.

It is important to remember that the SBGN languages are sets of rules on how to draw graph visualizations independently of the software that creates SBGN-compliant images.

Standards for graphical representation of biochemical and cellular networks such as Systems Biology Graphical Notation (SBGN), (Le Novere et al. 2009) and Minimum Interaction Maps (MIM), (Kohn 2001; Kohn and Aladjem 2006; Kohn et al. 2006a, b) aim at unambiguous representation of the stoichiometry and type of biochemical interactions.

8.6 Workflows

Scientific projects in computational systems-biology and/or computer-aided medical research are composed of a number of steps that re-occur in different settings and make use of various tools, often publicly available. A so-called workflow com-

prises all the activities in the project from start to finish, organised in an ordered way. In a project involving data retrieval, computational tasks, comparisons, and so on, the various tasks and services must be combined, sorted and executed in a sensible fashion, for example in the form of data analysis pipeline.

Workflows can be roughly divided into three categories: (1) manually constructed workflows; (2) custom workflow software; and (3) workflow engines using standardized remote procedure calls.

Manually constructed workflows consist usually of a set of scripts (e.g. in Perl) that combine a small number of services. They are highly customized and are thus in most cases not easily reusable or expandable. To overcome this, custom workflow software defines one or more interfaces to its core functionalities, to make it capable of being used by similar services. Such software facilitates extension, for example by writing small programs that adapt the output of similar services to the common interface (a wrapper), as well as permitting the creation of an interface that does not require users to have programming knowledge. The customized interfaces have the advantage that all data that is passed through the workflow has a common format that does not need to be converted. The disadvantage is that a wrapper has to be created for each new service. An example of a customised interface is EnVISION (http://www.ebi.ac.uk/enfin-srv/envision/index.html).

The third category, that of workflow engines employing standardized remote procedure calls, implies a process whereby a standardized interface of a software application is accessed over a network. The most popular uses of remote procedure calls are based on two XML language formats. The first language is called Web Service Description Language (WSDL). WSDL documents describe the interface of software and the protocol whereby the input and output information is transferred. This protocol is the second language, called Simple Object Access Protocol (SOAP). The SOAP protocol defines how data objects such as lists of strings and messages, such as the error messages produced by programming errors, are encoded. All major programming languages have libraries providing methods for generating SOAP interfaces for existing software. Most such interfaces come equipped with tools for the automatic creation of WSDL documents, and are referred to as a Web service. SOAP and WSDL were originally created for automated business-to-business communication (e.g. shops ordering goods from wholesalers), but they are becoming increasingly popular in bioinformatics. The basic concept of SOAP Web services is depicted cartoon-fashion in Fig. 8.3.

While Web services can be used to create manual and custom workflow systems, standardization also theoretically allows the creation of workflow management systems, i.e. software tools for the creation of workflows with a graphical user interface that requires minimal or no programming knowledge. In practice, this only holds true if the inputs and outputs of the different Web services match. Where incompatibilities arise, there is a need for translation. To solve this problem, most workflow management systems provide a large number of modules that enable simple translations. However, because these management systems were created for use in business-to-business communication, they lack the proper translation tools, thereby creating problems with managing workflows for bioinformatics applica-

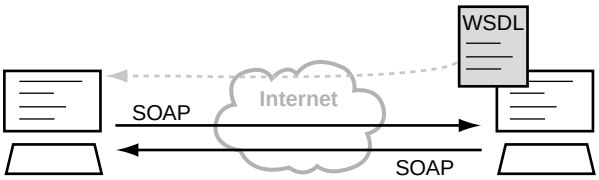


Fig. 8.3 The server on the right provides a web service. It exposes its web service interfaces through a WSDL document. This document is read by the client on the left who can now call the service on the right with a SOAP message and receive returning information, also encoded in SOAP

tions. The most popular workbench is Taverna (see below). The community around Taverna also developed a portal for exchanging workflows (http://www.myexperiment.org/), and another for registering and identifying Web services (http://www.biocatalogue.org).

8.6.1 Taverna Workbench

The Taverna workbench is a free software tool for designing and executing workflows. It was created by the *myGrid* project—taverna.org.uk, (Hull et al. 2006; Oinn et al. 2006), with its basis the integration of different software tools, including Web services, from different domains. Among the Bioinformatics services are those provided by NCBI (National Center for Biotechnology Information); EBI (European Bioinformatics Institute); DDJB (DNA Databank of Japan); SoapLab; BioMOBY; and EMBOSS. The Taverna workbench allows the construction of highly complex analyses over public and private data and computational resources, all from a standard PC, UNIX box or Apple computer. This is achieved by providing a desktop authoring environment and enactment engine for scientific workflows, simplifying labour-intensive aspects of data analyses by dispensing with physical steps like cut-and-paste, pressing buttons, and so on. Taverna provides access to remote services, thereby permitting the exploitation of resources provided by other institutes and applications developed by the community.

8.7 Discussion

The requisites of understanding cellular regulation, representing complex biochemical networks, and exploring their dynamics are today increasingly being facilitated by dedicated computational tools offering a broad spectrum of functionality. Interestingly, the development of tools, approaches and standards is frequently being undertaken as a community effort, thus strengthening exchange and accelerating the availability of information and tools.

Science relies both on the effort, enthusiasm, and brilliance of individual researchers, and on the accumulation of knowledge and research infrastructure provided by the scientific community. The development of new standards for molecular and systems biology is an impressive example showing that individual and collective efforts must be combined to form the foundation for future research and to enable effective exchange and communication of information.

Although there exist a number of tools in systems biology that appear to serve the needs of model development, data analysis, model analysis, and prediction generation, it is frequently observable that research groups in the field of systems and cancer biology are having to develop their own tools, pipelines and approaches, in order to have tailor-made functionalities at hand (e.g. Saez-Rodriguez et al. 2009). A major reason for this is that data suited for parameter estimation is still very heterogeneous and sparse, in spite of multiple high-throughput measuring efforts. Specific types of data can often be generated for one part of a network, but not for another. On the other hand, researchers in cellular and systems biology or bioinformatics have developed many tools and approaches that are devoted to the analysis of cancer-relevant data and to the modelling of cell processes during tumour genesis and tumour treatment. Despite a number of very promising results, these tools and approaches are still not extensively used and we are some considerable distance away from having computational methods, as distinct from pure statistics, form a natural part of the methods repertoire.

Upcoming technologies producing more extensive, more precise and more dedicated data are bound to generate further development and application of computational analysis and simulation. Only then will it certainly become possible to predict with any certainty drug targets and the effects of drug administration, depending on the genetic make-up and the current state of a patient.

Acknowledgements This work was supported by the Network of Excellence of the European Commission, Project ENFIN, contract number LSHG-CT-2005-518254 and by the German Ministry of Education and Research (BMBF), SysMO project Translucent, contract number 0313982A, to E.K.

References

Adra S, Sun T, MacNeil S, Holcombe M, Smallwood R (n d) Development of a three dimensional multiscale computational model of the human epidermis. PLoS One 5:e8511

Albert I, Thakar J, Li S, Zhang R, Albert R (2008) Boolean network simulations for life scientists. Source Code Biol Med 3:16%U. http://www.scfbm.org/content/3/1/16

Alfieri R, Barberis M, Chiaradonna F, Gaglio D, Milanesi L, Vanoni M, Klipp E, Alberghina L (2009) Towards a systems biology approach to mammalian cell cycle: modeling the entrance into S phase of quiescent fibroblasts after serum stimulation. BMC Bioinformatics 10(Suppl 12):S16

Arum CJ, Anderssen E, Tommeras K, Lundgren S, Chen D, Zhao CM (2010) Gene expression profiling and pathway analysis of superficial bladder cancer in rats. Urology 75:742–749

Bader GD, Cary MP, Sander C (2006) Pathguide: a pathway resource list. Nucleic Acids Res 34:D504–D506

Barthelmes J, Ebeling C, Chang A, Schomburg I, Schomburg D (2007) BRENDA, AMENDA and FRENDA: the enzyme information system in 2007. Nucleic Acids Res 35:D511–D514

Bauer-Mehren A, Furlong LI, Rautschka M, Sanz F (2009) From SNPs to pathways: integration of functional effect of sequence variations on models of cell signalling pathways. BMC Bioinform 10(Suppl 8):S6

Blinov ML, Faeder JR, Goldstein B, Hlavacek WS (2004) BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. Bioinformatics 20:3289–3291

Bluthgen N, Legewie S, Kielbasa SM, Schramme A, Tchernitsa O, Keil J, Solf A, Vingron M, Schafer R, Herzel H, Sers C (2009) A systems biological approach suggests that transcriptional feedback regulation by dual-specificity phosphatase 6 shapes extracellular signal-related kinase activity in RAS-transformed fibroblasts. Febs J 276:1024–1035

Borisov N, Aksamitiene E, Kiyatkin A, Legewie S, Berkhout J, Maiwald T, Kaimachnikov NP, Timmer J, Hoek JB, Kholodenko BN (2009) Systems-level interactions between insulin-EGF networks amplify mitogenic signaling. Mol Syst Biol 5:256

Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet 29:365–371

Brown SA, Kunz D, Dumas A, Westermark PO, Vanselow K, Tilmann-Wahnschaffe A, Herzel H, Kramer A (2008) Molecular insights into human daily behavior. Proc Natl Acad Sci U S A 105:1602–1607

Calzone L, Gelay A, Zinovyev A, Radvanyi F, Barillot E (2008) A comprehensive modular map of molecular interactions in RB/E2F pathway. Mol Syst Biol 4:173

Calzone L, Tournier L, Fourquet S, Thieffry D, Zhivotovsky B, Barillot E, Zinovyev A (2010) Mathematical modelling of cell-fate decision in response to death receptor engagement. PLoS Comput Biol 6:e1000702

Chen KC, Csikasz-Nagy A, Gyorffy B, Val J, Novak B, Tyson JJ (2000) Kinetic analysis of a molecular model of the budding yeast cell cycle. Mol Biol Cell 11:369–391

Chen KC, Calzone L, Csikasz-Nagy A, Cross FR, Novak B, Tyson JJ (2004) Integrative analysis of cell cycle control in budding yeast. Mol Biol Cell 15:3841–3862

Chen WW, Schoeberl B, Jasper PJ, Niepel M, Nielsen UB, Lauffenburger DA, Sorger PK (2009) Input-output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. Mol Syst Biol 5:239

Cheung SY, Evans ND, Chappell MJ, Godfrey KR, Smith PJ, Errington RJ (2008) Exploration of the intercellular heterogeneity of topotecan uptake into human breast cancer cells through compartmental modelling. Math Biosci 213:119–134

Csikasz-Nagy A, Battogtokh D, Chen KC, Novak B, Tyson JJ (2006) Analysis of a generic model of eukaryotic cell-cycle regulation. Biophys J 90:4361–4379

Dampier W, Tozeren A (2007) Signaling perturbations induced by invading H. pylori proteins in the host epithelial cells: a mathematical modeling approach. J Theor Biol 248:130–144

Di Cara A, Garg A, De Micheli G, Xenarios I, Mendoza L (2007) Dynamic simulation of regulatory networks using SQUAD. BMC Bioinform 8:462

Eapen BR (2007) Photosensitivity in Smith-Lemli-Opitz syndrome: a flux balance analysis of altered metabolism. Bioinformation 2:78–82

Ermentrout B (2002) Simulating, analyzing, and animating dynamical systems: a guide to XPPAUT for researchers and students. Society for Industrial Mathematics, Philadelphia

Ermentrout GB, Chow CC (2002) Modeling neural oscillations. Physiol Behav 77:629–633

Fisher WG, Yang PC, Medikonduri RK, Jafri MS (2006) NFAT and NFkappaB activation in T lymphocytes: a model of differential activation of gene expression. Ann Biomed Eng 34:1712–1728

Fitzgerald JB, Schoeberl B, Nielsen UB, Sorger PK (2006) Systems biology and combination therapy in the quest for clinical efficacy. Nat Chem Biol 2:458–466

Frohlich H, Beissbarth T, Tresch A, Kostka D, Jacob J, Spang R, Markowetz F (2008) Analyzing gene perturbation screens with nested effects models in R and bioconductor. Bioinformatics 24:2549–2550

Funahashi A, Jouraku A, Matsuoka Y, Kitano H (2007) Integration of CellDesigner and SABIO-RK. In Silico Biol 7(2 Suppl):S81–90

Funahashi A, Morohashi M, Kitano H (2003) CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. BIOSILICO 1:159–162

Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL (2007) The human disease network. Proc Natl Acad Sci U S A 104:8685–8690

Gossio S, Carando DG, Gonzalez SJ (2009) A computational dosimetry tool for the study of tumor doses and skin toxicities in BNCT. Appl Radiat Isot 67:S145–S148

Hattne J, Fange D, Elf J (2005) Stochastic reaction-diffusion simulation with MesoRD. Bioinformatics 21:2923–2924

Hermjakob H, Montecchi-Palazzi L, Bader G, Wojcik J, Salwinski L, Ceol A, Moore S, Orchard S, Sarkans U, Mering C von, Roechert B, Poux S, Jung E, Mersch H, Kersey P, Lappe M, Li Y, Zeng R, Rana D, Nikolski M, Husi H, Brun C, Shanker K, Grant SG, Sander C, Bork P, Zhu W, Pandey A, Brazma A, Jacq B, Vidal M, Sherman D, Legrain P, Cesareni G, Xenarios I, Eisenberg D, Steipe B, Hogue C, Apweiler R (2004) The HUPO PSI's molecular interaction format—a community standard for the representation of protein interaction data. Nat Biotechnol 22:177–183

Herrgard MJ, Swainston N, Dobson P, Dunn WB, Arga KY, Arvas M, Bluthgen N, Borger S, Costenoble R, Heinemann M, Hucka M, Le Novere N, Li P, Liebermeister W, Mo ML, Oliveira AP, Petranovic D, Pettifer S, Simeonidis E, Smallbone K, Spasic I, Weichart D, Brent R, Broomhead DS, Westerhoff HV, Kirdar B, Penttila M, Klipp E, Palsson BO, Sauer U, Oliver SG, Mendes P, Nielsen J, Kell DB (2008) A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. Nat Biotechnol 26:1155–1160

Ho J, Kong JW, Choong LY, Loh MC, Toy W, Chong PK, Wong CH, Wong CY, Shah N, Lim YP (2009) Novel breast cancer metastasis-associated proteins. J Proteome Res 8:583–594

Holleczek B, Gondos A, Brenner H (2009) periodR—an R package to calculate long-term cancer survival estimates using period analysis. Methods Inf Med 48:123–128

Hoops S, Sahle S, Gauges R, Lee C, Pahle J, Simus N, Singhal M, Xu L, Mendes P, Kummer U (2006) COPASI—a COmplex PAthway SImulator. Bioinformatics 22(24):3067–3074

Hucka M, Finney A, Sauro HM, Bolouri H, Doyle J, Kitano H (2002) The ERATO Systems biology workbench: enabling interaction and exchange between software tools for computational biology. Pac Symp Biocomput 450–461

Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, Cuellar AA, Dronov S, Gilles ED, Ginkel M, Gor V, Goryanin II, Hedley WJ, Hodgman TC, Hofmeyr JH, Hunter PJ, Juty NS, Kasberger JL, Kremling A, Kummer U, Le Novere N, Loew LM, Lucio D, Mendes P, Minch E, Mjolsness ED, Nakayama Y, Nelson MR, Nielsen PF, Sakurada T, Schaff JC, Shapiro BE, Shimizu TS, Spence HD, Stelling J, Takahashi K, Tomita M, Wagner J, Wang J (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics 19:524–531

Hull D, Wolstencroft K, Stevens R, Goble C, Pocock MR, Li P, Oinn T (2006) Taverna: a tool for building and running workflows of services. Nucleic Acids Res 34:W729–W732

Hur K, Lee HJ, Woo JH, Kim JH, Yang HK (2010) Gene expression profiling of human gastrointestinal stromal tumors according to its malignant potential. Dig Dis Sci 55(9):2561–2567, DOI: 10.1007/s10620-009-1061-4 Springer

Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27–30

Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res 38:D355–D360

Keating SM, Bornstein BJ, Finney A, Hucka M (2006) SBMLToolbox: an SBML toolbox for MATLAB users. Bioinformatics 22:1275–1277

Kerrien S, Orchard S, Montecchi-Palazzi L, Aranda B, Quinn AF, Vinod N, Bader GD, Xenarios I, Wojcik J, Sherman D, Tyers M, Salama JJ, Moore S, Ceol A, Chatr-Aryamontri A, Oesterheld M, Stumpflen V, Salwinski L, Nerothin J, Cerami E, Cusick ME, Vidal M, Gilson M, Armstrong J, Woollard P, Hogue C, Eisenberg D, Cesareni G, Apweiler R, Hermjakob H (2007) Broadening the horizon—level 2.5 of the HUPO-PSI format for molecular interactions. BMC Biol 5:44

Kholodenko BN (2000) Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. Eur J Biochem 267:1583–1588

Kitano H, Funahashi A, Matsuoka Y, Oda K (2005) Using process diagrams for the graphical representation of biological networks. Nat Biotechnol 23:961–966

Klamt S, Saez-Rodriguez J, Lindquist JA, Simeoni L, Gilles ED (2006) A methodology for the structural and functional analysis of signaling and regulatory networks. BMC Bioinformatics 7:56

Klipp E, Liebermeister W, Helbig A, Kowald A, Schaber J (2007) Systems biology standards—the community speaks. Nat Biotechnol 25:390–391

Klipp E, Liebermeister W, Wierling C, Kowald A, Lehrach H, Herwig R (2009) Systems biology. Wiley-VCH, Weinheim

Kohn KW (2001) Molecular interaction maps as information organizers and simulation guides. Chaos 11:84–97

Kohn KW, Aladjem MI (2006) Circuit diagrams for biological networks. Mol Syst Biol 2:2

Kohn KW, Aladjem MI, Weinstein JN, Pommier Y (2006a) Molecular interaction maps of bioregulatory networks: a general rubric for systems biology. Mol Biol Cell 17:1–13

Kohn KW, Aladjem MI, Kim S, Weinstein JN, Pommier Y (2006b) Depicting combinatorial complexity with the molecular interaction map notation. Mol Syst Biol 2:51

Krause F, Uhlendorf J, Lubitz T, Schulz M, Klipp E, Liebermeister W (2010) Annotation and merging of SBML models with semanticSBML. Bioinformatics 26:421–422

Kuhnel M, Mayorga LS, Dandekar T, Thakar J, Schwarz R, Anes E, Griffiths G, Reich J (2008) Modelling phagosomal lipid networks that regulate actin assembly. BMC Syst Biol 2:107

Launey T (2007) A computational approach to the study of AMPA receptor clustering at Purkinje cell synapses. Arch Ital Biol 145:299–310

Le Novere N, Finney A, Hucka M, Bhalla US, Campagne F, Collado-Vides J, Crampin EJ, Halstead M, Klipp E, Mendes P, Nielsen P, Sauro H, Shapiro B, Snoep JL, Spence HD, Wanner BL (2005) Minimum information requested in the annotation of biochemical models (MIRIAM). Nat Biotechnol 23:1509–1515

Le Novere N (2006) Model storage, exchange and integration. BMC Neurosci 7 Suppl 1, S11

Le Novere N, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, Demir E, Wegner K, Aladjem MI, Wimalaratne SM, Bergman FT, Gauges R, Ghazal P, Kawaji H, Li L, Matsuoka Y, Villeger A, Boyd SE, Calzone L, Courtot M, Dogrusoz U, Freeman TC, Funahashi A, Ghosh S, Jouraku A, Kim S, Kolpakov F, Luna A, Sahle S, Schmidt E, Watterson S, Wu G, Goryanin I, Kell DB, Sander C, Sauro H, Snoep JL, Kohn K, Kitano H (2009) The systems biology graphical notation. Nat Biotechnol 27:735–741

Lee E, Salic A, Kruger R, Heinrich R, Kirschner MW (2003) The roles of APC and axin derived from experimental and theoretical analysis of the Wnt pathway. PLoS Biol 1:116–132

Levi F, Schibler U (2007) Circadian rhythms: mechanisms and therapeutic implications. Annu Rev Pharmacol Toxicol 47:593–628

Lindskog M, Kim M, Wikstrom MA, Blackwell KT, Kotaleski JH (2006) Transient calcium and dopamine increase PKA activity and DARPP-32 phosphorylation. PLoS Comput Biol 2:e119

Lloyd CM, Halstead MD, Nielsen PF (2004) CellML: its future, present and past. Prog Biophys Mol Biol 85:433–450

Marot G, Foulley JL, Mayer CD, Jaffrezic F (2009) Moderated effect size and P-value combinations for microarray meta-analyses. Bioinformatics 25:2692–2699

Mattfeldt T, Eckel S, Fleischer F, Schmidt V (2007) Statistical modelling of the geometry of planar sections of prostatic capillaries on the basis of stationary strauss hard-core processes. J Microsc 228:272–281

Matthews L, Gopinath G, Gillespie M, Caudy M, Croft D, Bono B de, Garapati P, Hemish J, Hermjakob H, Jassal B, Kanapin A, Lewis S, Mahajan S, May B, Schmidt E, Vastrik I, Wu G, Birney E, Stein L, D'Eustachio P (2009) Reactome knowledgebase of human biological pathways and processes. Nucleic Acids Res 37:D619–D622

Mendoza L, Xenarios I (2006) A method for the generation of standardized qualitative dynamical systems of regulatory networks. Theor Biol Med Model 3:13

Moraru II Schaff JC, Slepchenko BM, Blinov ML, Morgan F, Lakshminarayana A, Gao F, Li Y, Loew LM (2008) Virtual cell modelling and simulation software environment. IET Syst Biol 2:352–362

Obeyesekere MN, Knudsen ES, Wang JY, Zimmerman SO (1997) A mathematical model of the regulation of the G1 phase of Rb+/+ and Rb-/- mouse embryonic fibroblasts and an osteosarcoma cell line. Cell Prolif 30171–30194

Oda K, Matsuoka Y, Funahashi A, Kitano H (2005) A comprehensive pathway map of epidermal growth factor receptor signaling. Mol Syst Biol 1:10

Oinn T, Addis M, Ferris J, Marvin D, Senger M, Greenwood M, Carver T, Glover K, Pocock MR, Wipat A, Li P (2004) Taverna: a tool for the composition and enactment of bioinformatics workflows. Bioinformatics 20:3045–3054

Olivier BG, Snoep JL (2004) Web-based kinetic modelling using JWS online. Bioinformatics 20:2143–2144

Pujana MA, Han JD, Starita LM, Stevens KN, Tewari M, Ahn JS, Rennert G, Moreno V, Kirchhoff T, Gold B, Assmann V, Elshamy WM, Rual JF, Levine D, Rozek LS, Gelman RS, Gunsalus KC, Greenberg RA, Sobhian B, Bertin N, Venkatesan K, Ayivi-Guedehoussou N, Sole X, Hernandez P, Lazaro C, Nathanson KL, Weber BL, Cusick ME, Hill DE, Offit K, Livingston DM, Gruber SB, Parvin JD, Vidal M (2007) Network modeling links breast cancer susceptibility and centrosome dysfunction. Nat Genet 39:1338–1349

Raman K, Vashisht R, Chandra N (2009) Strategies for efficient disruption of metabolism in mycobacterium tuberculosis from network analysis. Mol Biosyst 5:1740–1751

Renkonen J, Mattila P, Parviainen V, Joenvaara S, Toppila-Salmi S, Renkonen R (2010) A network analysis of the single nucleotide polymorphisms in acute allergic diseases. Allergy 65:40–47

Rexhepaj E, Brennan DJ, Holloway P, Kay EW, McCann AH, Landberg G, Duffy MJ, Jirstrom K, Gallagher WM (2008) Novel image analysis approach for quantifying expression of nuclear proteins assessed by immunohistochemistry: application to measurement of oestrogen and progesterone receptor levels in breast cancer. Breast Cancer Res 10:R89

Romero PR, Karp PD (2004) Using functional and organizational information to improve genome-wide computational prediction of transcription units on pathway-genome databases. Bioinformatics 20:709–717

Saevels J, Schepdael A van, Hoogmartens J (1996) Determination of the kinetic parameters of adenosine deaminase by electrophoretically mediated microanalysis. Electrophoresis 17:1222–1227

Saez-Rodriguez J, Alexopoulos LG, Epperlein J, Samaga R, Lauffenburger DA, Klamt S, Sorger PK (2009) Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. Mol Syst Biol 5:331

Sahin O, Frohlich H, Lobke C, Korf U, Burmester S, Majety M, Mattern J, Schupp I, Chaouiya C, Thieffry D, Poustka A, Wiemann S, Beissbarth T, Arlt D (2009) Modeling ERBB receptor-regulated G1/S transition to find novel targets for de novo trastuzumab resistance. BMC Syst Biol 3:1

Samaga R, Saez-Rodriguez J, Alexopoulos LG, Sorger PK, Klamt S (2009) The logic of EGFR/ErbB signaling: theoretical properties and analysis of high-throughput data. PLoS Comput Biol 5:e1000438

Schmidt H, Jirstrand M (2006) Systems biology toolbox for MATLAB: a computational platform for research in systems biology. Bioinformatics 22:514–515

Schmidt MM, Wittrup KD (2009) A modeling analysis of the effects of molecular size and binding affinity on tumor targeting. Mol Cancer Ther 8:2861–2871

Schmierer B, Tournier AL, Bates PA, Hill CS (2008) Mathematical modeling identifies Smad nucleocytoplasmic shuttling as a dynamic signal-interpreting system. Proc Natl Acad Sci U S A 105:6608–6613

Schoeberl B, Eichler-Jonsson C, Gilles ED, Muller G (2002) Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. Nat Biotechnol 20:370–375

Schoeberl B, Pace EA, Fitzgerald JB, Harms BD, Xu L, Nie L, Linggi B, Kalra A, Paragas V, Bukhalid R, Grantcharova V, Kohli N, West KA, Leszczyniecka M, Feldhaus MJ, Kudla AJ, Nielsen UB (2009) Therapeutically targeting ErbB3: a key node in ligand-induced activation of the ErbB receptor-PI3K axis. Sci Signal 2:ra31

Schulz M, Bakker BM, Klipp E (2009) TIde: a software for the systematic scanning of drug targets in kinetic network models. BMC Bioinformatics 10:344

Stromback L, Lambrix P (2005) Representations of molecular pathways: an evaluation of SBML, PSI MI and BioPAX. Bioinformatics 21:4401–4407

Swameye I, Muller TG, Timmer J, Sandra O, Klingmuller U (2003) Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling. Proc Natl Acad Sci U S A 100:1028–1033

Taylor CF, Field D, Sansone SA, Aerts J, Apweiler R, Ashburner M, Ball CA, Binz PA, Bogue M, Booth T, Brazma A, Brinkman RR, Michael Clark A, Deutsch EW, Fiehn O, Fostel J, Ghazal P, Gibson F, Gray T, Grimes G, Hancock JM, Hardy NW, Hermjakob H, Julian RK Jr, Kane M, Kettner C, Kinsinger C, Kolker E, Kuiper M, Le Novere N, Leebens-Mack J, Lewis SE, Lord P, Mallon AM, Marthandan N, Masuya H, McNally R, Mehrle A, Morrison N, Orchard S, Quackenbush J, Reecy JM, Robertson DG, Rocca-Serra P, Rodriguez H, Rosenfelder H, Santoyo-Lopez J, Scheuermann RH, Schober D, Smith B, Snape J, Stoeckert CJ Jr, Tipton K, Sterk P, Untergasser A, Vandesompele J, Wiemann S (2008) Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. Nat Biotechnol 26:889–896

Ueda HR, Hayashi S, Chen W, Sano M, Machida M, Shigeyoshi Y, Iino M, Hashimoto S (2005) System-level identification of transcriptional circuits underlying mammalian circadian clocks. Nat Genet 37:187–192

Vastrik I, D'Eustachio P, Schmidt E, Gopinath G, Croft D, Bono B de, Gillespie M, Jassal B, Lewis S, Matthews L, Wu G, Birney E, Stein L (2007) Reactome: a knowledge base of biologic pathways and processes. Genome Biol 8:R39

Vauquelin G, Fierens F, Van Liefde I (2006) Long-lasting angiotensin type 1 receptor binding and protection by candesartan: comparison with other biphenyl-tetrazole sartans. J Hypertens (Suppl 24):S23–S30

Visvanathan M, Pfeifer B, Baumgartner C, Tilg B, Lushington GH (2009) Integrative approach for combining TNFalpha-NFkappaB mathematical model to a protein interaction connectivity map. Lect Notes Comput Sci 5542:63–74

Wahde M, Klus GT, Bittner ML, Chen Y, Szallasi Z (2002) Assessing the significance of consistently mis-regulated genes in cancer associated gene expression matrices. Bioinformatics 18:389–394

Whitcomb DC, Ermentrout GB (2004) A mathematical model of the pancreatic duct cell generating high bicarbonate concentrations in pancreatic juice. Pancreas 29:e30–e40

Wittig U, Golebiewski M, Kania R, Krebs O, Mir S, Weidemann A, Anstein S, Saric J, Rojas I (2006) SABIO-RK: Integration and curation of reaction kinetics data 3rd international workshop on data integration in the life sciences 2006 (DILS'06), vol 4075. Lecture Notes in Bioinformatics, Hinxton, pp 94–103

Wu YF, Myasnikova E, Reinitz J (2007) Master equation simulation analysis of immunostained Bicoid morphogen gradient. BMC Syst Biol 1:52

Zi Z, Klipp E (2006) SBML-PET: a systems biology markup language-based parameter estimation tool. Bioinformatics 22:2704–2705

Zi Z, Klipp E (2007) Constraint-based modeling and kinetic analysis of the Smad dependent TGF-beta signaling pathway. PLoS One 2:e936

Zwolak J, Adjerid N, Bagci EZ, Tyson JJ, Sible JC (2009) A quantitative model of the effect of unreplicated DNA on cell cycle progression in frog egg extracts. J Theor Biol 260:110–120

Chapter 9
The Hallmarks of Cancer Revisited Through Systems Biology and Network Modelling

Charles Auffray, Trey Ideker, David J. Galas and Leroy Hood

Abstract Since 10 years ago, when the seven hallmarks of cancer were first defined by Hanahan and Weinberg, after decades of molecular, cellular and clinical investigations, new systems-based approaches have provided a wide range of improved investigative methods. These approaches integrate various global data types into mathematical and computational models of molecular and cellular pathways and networks that become dysregulated in cancer, since the models are now able to take into account the large-scale properties of complex biological networks. Genome variation and instability have been revisited through study of genetic and genomic networks; while transcription and protein interaction networks are revealing cancer biomarkers of modular change. Growth, proliferation and apoptosis are being more fully described by signalling network modelling. Sustained angiogenesis and metastasis are being addressed via multiscale modelling. Enhanced understanding of the initial hallmarks of cancer, extended to the control of metabolism and stress, is opening novel avenues for cancer diagnosis and treatment. It is fully expected that further progress will take place through iterative cycles of experimentation and modelling, typical of systems biology. All of this will require advances in molecular data acquisition, multiscale integration of data scales and types, new approaches to data analysis and improved modelling. Success in all these endeavours cannot be achieved without better cross-disciplinary interactions among researchers and technologists.

9.1 Introduction

9.1.1 Hallmarks of Cancer

Ten years ago, to recapitulate half a century of molecular and cellular biology investigations, Douglas Hanahan and Robert Weinberg listed six acquired capabilities and one enabling characteristic of cancer cells, together forming the seven 'hallmarks of cancer' (Hanahan and Weinberg 2000).

C. Auffray (✉)
Functional Genomics and Systems Biology for Health, CNRS Institute of Biological Sciences -7, rue Guy Moquet, BP8, 94801 Villejuif, France
e-mail: charles.auffray@vjf.cnrs.fr

A. Cesario, F. B. Marcus (eds.), *Cancer Systems Biology, Bioinformatics and Medicine,*
DOI 10.1007/978-94-007-1567-7_9,

Their intention was to address the following questions: 'How many distinct regulatory circuits within each type of target cell must be disrupted for such a cell to become cancerous? Does the same set of cellular regulatory circuits suffer disruption in the cells of the disparate neoplasms arising in the human body? Which of these circuits operate on a cell-autonomous basis, and which are coupled to the signals that cells receive from the surrounding microenvironment within a tissue? Can the large and diverse collection of cancer-associated genes be tied to the operations of a small group of regulatory circuits?'

They suggested that 'self-sufficiency in growth signals, insensitivity to growth-inhibitory (antigrowth) signals, evasion from programmed cell death (apoptosis), limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis' were the 'six essential alterations in cell physiology that collectively dictate malignant growth', enabled by the characteristic genomic instability of tumours.

In their synthesis, they foresaw that: 'Two decades from now, having fully charted the wiring diagram of every cellular signalling pathway, it will be possible to lay out the complete integrated circuit of the cell upon its current outline. We will then be able to apply the tools of mathematical modelling to explain how specific genetic lesions serve to reprogramme this integrated circuit in each of the constituent cell types so as to manifest cancer'.

9.1.2 Network Properties

Concurrently, building on another large body of work in the engineering and social sciences, Albert-László Barabási and Réka Albert highlighted commonalities in network properties in natural and artificial systems (Barabasi and Albert 1999).

They reported 'the existence of a high degree of self-organization characterizing the large-scale properties of complex networks', such as social and business networks, the World Wide Web, and transportation or electrical power line networks, which 'self-organize into a scale-free state'. They further established that 'growth and preferential attachment, two key features of real networks, are responsible for the power-law scaling observed', implying that the 'results are relevant to a large class of networks observed in nature'.

They went on to propose that 'possible scale-free features of genetic and signalling networks could reflect the networks' evolutionary history, dominated by the growth and aggregation of different constituents, leading from simple molecules to complex organisms', thus laying the foundations for network-based biology and medicine (Kreeger and Lauffenburger 2010; Barabasi and Oltvai 2004; Barabasi 2007; Pawson and Linding 2008).

9.1.3 Advances in Hallmark Analysis and Networks

In this chapter, we will review, through selected examples, some of the advances that have been made over the past decade in addressing the questions raised by

Hanahan & Weinberg, and discuss how and to what extent network properties, highlighted initially for biologists by Barabási & Albert, have contributed to the mathematical modelling they called for of molecular and cellular pathways in cancer.

Our discussion of these two converging and complementary threads will develop in the framework of the emerging field of systems biology and its application to important biomedical problems. We will draw attention to several other important conceptual, mathematical, computational, and experimental advances that are providing essential contributions to the implementation of systems approaches in biology and medicine, as discussed in extensive reviews of the field by ourselves and others (Ideker et al. 2001; Wolkenhauer 2001; Kitano 2002; Noble 2002; Auffray et al. 2003; Hood et al. 2004; Nicholson et al. 2004; Westerhoff and Palsson 2004; Kirschner 2005; Zerhouni 2005; Ahn et al. 2006a, 2006b; Butte 2008; Auffray et al. 2009).

We will point to novel avenues of investigation worthy of further exploration, and indicate challenges needing to be addressed for a full development of the systems biology roadmap in cancer research (Wolkenhauer et al. 2010; Auffray and Nottale 2008; Nottale and Auffray 2008; Aebersold et al. 2009; Clermont et al. 2009; Tegner et al. 2009), as also discussed in other chapters of this book.

9.2 The Potential of Systems Approaches to Disease

9.2.1 Principles of Systems Biology

Living organisms are systems processing two fundamental types of information: the digital information stored in the genome, and the environmental signals interacting with the organism to generate its phenotype. The information arising from the interaction of digital genetic and environmental signals is captured, transmitted, integrated and mediated by biological networks. These complex networks process their information and regulate molecular machines that execute the functions of life. For that reason, an important challenge facing systems biology is to understand how genotype and environmental signals are translated into phenotype through these two information-handling structures; biological networks, and molecular machines. Finally, biological information is hierarchical and multiscalar, ranging from the purely molecular (DNA, RNA, protein, metabolites, and their interactions), to molecular networks and to cells, organs, individuals, populations, and ecologies. Because environmental signals modify each of these levels of information, in order to understand the functioning of biological systems, one must capture relevant information at each level and carry out an integration of different data types into models that explain the multiple environmental contributions to the functioning of the system.

A systems view of disease is based on the simple idea that disease arises as a consequence of disease-perturbed networks in the relevant organ, i.e. that a dis-

ease state of a cell differs in some module or sub-module of the cellular network from a 'normal' state, so that questions of stability and/or instability of these altered states must also be considered. Disease perturbations that instigate these state changes may arise from genetic and/or environmental influences. The initial disease perturbation has two consequences. Firstly, the envelope of information which the disease-perturbed network expresses is altered compared to that of its normal counterpart. Because the architecture and nodal points (mRNAs, miRNAs, proteins, metabolites, etc.) of the network change during the course of the disease, there is an altered envelope of information expressed throughout disease progression. Since this altered, dynamically changing, envelope of information underlies the pathophysiology of the disease, exploration of these changes may suggest new insights into the diagnosis and therapy of the disease. Secondly, the initial disease-perturbed network can cause other networks also to become disease-perturbed as the disease progresses, resulting in effects that can be greatly amplified and extended well beyond the initial perturbation.

Systems approaches are hypothesis-driven: models must be created from the available information. Hypotheses must be formulated; perturbations must be designed to test the predictions of the model; and new data must be collected and compared with the original model predictions. The model must then be modified as appropriate after which the cycle begins again and iterates until model and experimental data are in agreement. A systems approach to disease requires that data acquisition be as global or comprehensive as possible, ideally being collected on appropriate temporal and spatial scales.

9.2.2 Challenges in Modelling Networks in Cancer

There are several key questions that arise in connection with modelling networks that are related to cancer initiation and progression. First is the question of which known pathways or sub-networks are relevant to the model. A corollary to this question is whether there are some universal and some tissue-type specific pathways and networks. The answer to this latter question appears to be that there are both (Vogelstein and Kinzler 2004); however, the completeness of the list and the connections is as yet far from certain. The next question is whether there are stable or quasi-stable states of these key networks and whether they may progress from state to state naturally, without further stimulation or perturbation. Since the models remain largely qualitative, it is impossible to answer these questions directly. Certain sub-networks, as for some signal transduction pathways, have begun to yield to quantitative analysis, but they remain short of significant predictive power (Yuan and Hu 2006). Finally, one of the more important questions related to these future models is what kinds of quantitative data on the significant molecular interactions are available. Much such data has been accumulated (Beckman and Loeb 2005; Weinberg 2006), but it remains to be seen how complete and accurate the data will be found to be when put into quantitative models.

There are many challenges to quantitative network models of cancer. Cancer has a wide range of known genetic and environmental triggers. Its heterogeneity also reveals itself at the molecular level. Within a tumour, different cells may actually be in different states, and have different gene expression and protein profiles. It is not yet possible to create clear molecular categories for this type of heterogeneous disease. The current difficulty encountered in cancer diagnosis using molecular markers is partly due to such heterogeneity. Mathematically, molecular expression levels should probably be treated as stochastic variables in any model. Thus, one method employed to study the protein expression levels in cancer would be stochastic differential equations. Even so, it is clear that the initial models will involve simplifications of the known or hypothesized networks (Ao et al. 2008).

Finally, the questions of the stability or instability of network states, an important issue in studying cancer, must be considered as a characteristic of the topology and structure of networks. It is clear, as pointed out by Barabasi and Albert (1999), that biological networks are characteristically power-law networks, which means that the nodes in the network (e.g. genes, proteins or metabolites) are connected to other nodes to a degree proportional to the inverse power of the degree of connectedness. This is also sometimes called a scale-free network (which is not rigorously true). The occurrence of power-law networks is by no means limited to biology, and we now know that there are many ways of generating these networks (the Internet is a power-law network, for example). While the significance of this particular characteristic (amongst others) of biological networks is unknown probably the most notable characteristic of power-law networks is that they have a small number of very highly connected nodes, sometimes called hubs. These may be very important in cancer, because modifications of, or damage to, hubs can have large effects on the network. A major hub in a network of central importance to cancer is the p53 protein (Vogelstein and Kinzler 2004). There are important and subtle properties of networks near criticality, i.e. the boundary between highly stable and highly chaotic behaviour, which may be very significant for cancer cells and their susceptibility to uncontrolled progression. We are now in a position to begin to examine some of these network structure issues as related to function and to the global systems properties (Carter et al. 2009).

9.2.3 Network Inference Through Machine Learning

A number of new ideas are currently being developed that may serve to allow integration of a truly large range of data types in the future. For example, a major challenge involves the integration of genetic data with the global molecular data types previously discussed, and we can add to that the challenge of using imaging data of various kinds, e.g. other medical diagnostic, as well as metabolomic data. The application to biology and medicine of various machine learning methods, including Bayesian network methods, has recently been enhanced by the application of Markov Logic Networks (Sakhanenko and Galas 2010). This is one of the most

general approaches to statistical relational learning, a sub-field of machine learning that combines two kinds of modelling: probabilistic graphical models, namely Markov random fields; and first-order logic (Bayesian networks employ Boolean logic). Probabilistic graphical models offer a way to represent joint probability distributions of sets of random variables in a compact fashion, and permit the development of numerous algorithms for learning and inference, making these models a good choice for handling uncertainty and noise in data. On the other hand, first-order logic allows the representation of inferences over complex, relational domains. Propositional (Boolean) logic, with which biologists are most familiar, describes the truth state on the level of specific instances, while first-order logic allows us to make assertions about the truth state of complex relations between subsets (classes) of instances. Using first-order logic we can represent recursive and potentially infinite structures such as Markov chains, where a temporal dependency of the current state on the state at the previous time step can be instantiated to an infinite time series. Thus, first-order logic is a very flexible method for representing general knowledge, as we encounter in the biology of cancer.

9.2.4 An Illustrative Example: Systems Biology of Prion Disease

Systems biology and systems genetics approaches are now able to take advantage of the power of whole-transcriptome and whole-genome sequencing for the identification of the genetic basis of various diseases (Roach et al. 2010). An illustrative example of what can be expected from a systems approach to disease is that of the neurodegenerative prion disease studied in mice (Hwang et al. 2009). Integration of brain transcriptomes, relevant biological networks, histopathological examination of the infected brains, clinical signs, and the distribution of infectious prion proteins across the brains were studied at 10 time points across the course of the disease in eight different inbred mouse strain and prion strain combinations. This enabled identification of four fundamental and dynamically changing disease processes: prion accumulation and replication, glial cell activation, destruction of the dendrites and axons of nerve cells and nerve cell apoptosis, which together explained virtually every known aspect of the pathophysiology of prion disease. It also identified at least six networks not previously known to be associated with prion disease, opening avenues for exploring new dimensions of prion disease.

This ability to deconvolute the prion disease process into sequentially activated networks has interesting implications for future approaches to the identification of drug targets. One approach to drug target identification is based on the possibility that drugs (probably two or more) may be selected to force disease-perturbed networks to revert to their normal state or to behave in a more normal manner. Obviously, in the case of prion disease, one would focus on the initially perturbed networks to identify these potential drug targets, because the ability to re-engineer the behaviour of these networks would check the disease at its initial stages. Another interesting question is whether any of the detected differentially expressed genes

encode proteins that are secreted into the blood, and whether their altered levels of transcription are reflected in the blood by altered protein levels. The answer to both of these questions is yes. Hence it is possible to carry out presymptomatic diagnosis of prion disease. Obviously this systems approach to prion disease has many implications for the study of cancer, both in model animal systems and in humans.

9.3 Transcription and Protein Interaction Networks Revealed by Modular Cancer Biomarkers

9.3.1 Networks and Biomarkers

Transcription is the natural channel for transmission of the information stored in the genome sequence and structure, on the one hand, and the expression of functional proteins and RNA on the other. Transcriptome analyses using microarrays have therefore played a prominent role in the identification of biomarkers for classification of cancer types, disease progression, and assessment of responses to drug treatments, some of which have already demonstrated clinical utility (Carro et al. 2010; Ransohoff and Gourlay 2010; Ludwig and Weinstein 2005; Rhodes and Chinnaiyan 2005; Varambally et al. 2005; Bild et al. 2006a; Bild et al. 2006b; Carrivick et al. 2006; Tomlins et al. 2007; Chang et al. 2008; Itadani et al. 2008; Ornish et al. 2008; Sawyers 2008; Schlabach et al. 2008; Ransohoff 2009). Systematic surveys of transcriptomes through microarrays and RNA sequencing have revealed more extensive transcription of the genome than had been predicted (Velculescu and Kinzler 2007; Mortazavi et al. 2008; Wu et al. 2008). This new discovery includes the existence of a large set of non-coding RNAs, among which a family of micro-RNAs involved in a previously unsuspected layer of genetic regulation after transcription, which can now be exploited experimentally in functional screens for the identification of essential network components in cancer (Ngo et al. 2006; Shimoni et al. 2007; Silva et al. 2008). Such studies have been supported by a growing variety of tools designed for inference of transcriptional and regulatory networks at work in normal and cancer cells (Yeger-Lotem et al. 2004; Schlitt and Brazma 2005; Barrett and Palsson 2006; Bonneau et al. 2006; Gianchandani et al. 2006; Alon 2007; Bonneau 2008; Madar and Bonneau 2009).

9.3.2 Proteomics and Pathways

The very significant experimental and computational advances of the past decade in the field of proteomics have enabled much more detailed and comprehensive analysis and data-driven modelling of protein phosphorylation networks and sig-

nalling cascades than was possible before (Aldridge et al. 2006; Janes and Yaffe 2006; Linding et al. 2007; Huang et al. 2008; Tan et al. 2009). The use of explicit or fuzzy-logic representations has facilitated the analysis of the dynamics of signalling pathways, and the development of predictive models relevant to disease states (Bosl 2007; Aldridge et al. 2009; Samaga et al. 2009). Similar advances have been made in the development of protein-protein and protein-DNA interaction databases of verified binding sites of increasing coverage and quality, resulting in the discovery of relevant links with disease phenotypes and drug responses (Cusick et al. 2005; Rhodes et al. 2005; Rachlin et al. 2006; Goh et al. 2007; Pujana et al. 2007; Yildirim et al. 2007; Ideker and Sharan 2008; Taylor et al. 2009). The power of these methods, and the value of these extensive datasets are best revealed when they are integrated and analysed together with other data types through computational and visualization tools. This makes it possible to identify modular structures as important pathway components of the global cellular network of interacting components that are relevant to cancer development, progression and response to drug treatment (Hwang et al. 2005a; Hwang et al. 2005b; Zheng et al. 2005; Du et al. 2006; Graudens et al. 2006; Chuang et al. 2007; Cline et al. 2007; Cui et al. 2007; Legewie et al. 2008; Mani et al. 2008; Chang et al. 2009; Debily et al. 2009; Huang and Fraenkel 2009).

9.4 Growth, Proliferation and Apoptosis Revisited Through Signalling Network Modelling

9.4.1 Signalling Pathways

Among the signalling pathways that play a central role in transmitting and processing intracellular signals triggered by binding of growth factors to their cell surface receptors, an especially important one is the SOS-Ras-Raf-MAPK cascade. This is often altered in complex ways in various types of cancers, and acts to sustain autonomous growth. Hence, there is a strong interest for the sake of devising countermeasures of therapeutic value, in deciphering the precise mechanisms of this deregulation, particularly how it is related to other components of the cellular machinery. In this context, the analysis of extensive genomic, transcriptomic and proteomic data from a collection of breast cancer cell lines has enabled their integration into a model composed of several hundred molecular states and signalling rules, with results indicating the existence of sub-networks which may play an important role in the growth characteristics of the cancer cell lines (Heiser et al. 2009). This led the authors to suggest Pak1 as an important regulator of the EGFR-MAPK pathway, which could trigger sensitivity to Mek inhibitors when overexpressed; and indeed this hypothesis could be verified experimentally, a promising avenue to explore in the clinical setting for treatment of the specific subset of luminal breast cancers overexpressing Pak1.

A similar strategy was used to develop a mathematical model of Ras signalling and to investigate the mechanism whereby only certain point mutations result in growth autonomy. The predicted behaviour of the cell lines was validated through experiments, again suggesting a possible point of therapeutic intervention (Stites et al. 2007). Later analyses of the MAPK signalling pathway in Ras-transformed fibroblasts identified one dual-specificity phosphatase as a negative-feedback transcriptional regulator at the origin of the ultrasensitive activation kinetics of MAPK. This was achieved through integration of targeted measurements of kinase activities along with global gene expression and regulation data into predictive mathematical models, followed by experimental verification (Bluthgen et al. 2009). A mass action model (see Chap. 7) was used to disentangle the interplay of multiple growth factors with the MAPK and PI3K/Atk signalling pathways, revealing that interactions between the pathway components are not only highly non-linear, but also heavily dependent upon the context of different oncogenic mutations (Chen et al. 2009).

9.4.2 Growth Factors and Apoptosis

The intracellular signalling pathways that revolve around the Ras hub are also intricately linked to the effect of extrinsic growth factors on programmed cell death, to the extent that various Ras mutations in colon cancer are associated with different levels of apoptosis through differential induction signalling cascades modulated by ERK (Kreeger et al. 2009). These observations have been elucidated by computational modelling of the interplay between an autocrine positive-feedback loop involving a growth factor and a chemokine with opposite effects, and a negative-feedback loop mediated by another dual-specificity phosphatase, with the apoptotic outcome in the different cell lines being influenced by additional interlinked pathways in a predictable manner. Computational modelling was also used to investigate the crosstalk mechanisms that amplify ERK signalling upstream of Ras, identifying the critical network nodes that influence amplification of the mitogenic potential induced by insulin (Borisov et al. 2009).

It is thus apparent that systems approaches combining various types of data-driven mathematical modelling for hypothesis generation, followed by experimental validation in appropriate cancer cellular models, have already provided fresh insights into the mechanisms that underpin the peculiar properties of cancer cells with regard to growth and apoptosis. Although these advances open novel avenues for identification of subsets of patients in which targeted intervention could be effective, they are certainly not the end of the story, but rather the beginning. The iterative nature of the systems approach means that further refinement of the initial models to make them more quantitative, explanatory and predictive, may lead to even more profound and significant insights into the molecular and cellular mechanisms leading to the cancerous phenotype, as already demonstrated in several studies (Citri and Yarden 2006; Kumar et al. 2008; Lazzara and Lauffenburger 2009; Schoeberl et al. 2009).

9.5 Sustained Angiogenesis and Metastasis Revisited Through Multiscale Modelling

9.5.1 Mathematical Modelling

For more than half a century, mathematicians have taken a theoretical approach to modelling cancer, with particular emphasis on events occurring at the cellular, tissue, organ, and organism levels in human and animal models. Although this research thread originally developed independently of experimental and clinical biology, it has provided a number of insights on the fundamental principles and stages of cancer development, and is now merging with the data-driven modelling efforts highlighted in the previous section (Byrne 2010; Wang et al. 2007, 2009; Anderson and Quaranta 2008). These formal modelling approaches have proved to be particularly well suited to the study of vascular and avascular growth (Ferreira et al. 2002; Jiang et al. 2005; Macklin et al. 2009), and the complex processes underlying tissue invasion and metastasis formation and dissemination from a cell population dynamics perspective (Johnston et al. 2007).

9.5.2 Angiogenesis

Once a tumour is established, its maintenance becomes highly dependent upon the provision of oxygen through the development of blood vasculature. Thus, control of angiogenesis is an important target of cancer treatment through radiotherapy or chemotherapy (Folkman 2007; Dewhirst et al. 2008). In this context, it is important to understand the complex sequence of events starting with migration and proliferation of endothelial cells sustained by changes in the extracellular matrix, that lead to the formation of capillaries supplying a dynamically controlled blood flow adapted to the needs of the tumour. In fact, mathematical and theoretical models have already provided insights for the development of anti-angiogenic treatments of solid tumours (Chaplain et al. 2006). Integration of the pharmacological properties of pro- and anti-angiogenic compounds together with the cell cycle states of cancer and blood vessel endothelial cells into a multiscale mathematical model makes it possible to simulate and predict the efficacy of therapies targeting angiogenesis, thus supporting the design and monitoring of prospective clinical trials (Billy et al. 2009).

9.5.3 Metastasis

Tumours become clinically malignant and life-threatening when their growth becomes invasive and they metastasize to distant sites; hence the great attention paid to understanding this complex process, which involves interaction of tumour cells

with stromal cells and mesenchymal stem cells of bone marrow origin (Karnoub et al. 2007), and to preventing it through targeted therapies (Mazzone and Comoglio 2006; Hu and Polyak 2008; Nguyen et al. 2009; Polyak et al. 2009). As a first step, it is important to be able to assess the potential of metastatic formation through appropriate biomarkers. In this context, the integration of gene expression and protein interaction data has facilitated identification of functional sub-networks that provide increased accuracy for the classification of metastatic versus non-metastatic tumours, as well as suggesting novel hypotheses concerning the molecular mechanisms involved in tumour progression to metastasis in colon and breast cancer (Varambally et al. 2005; Auffray 2007; Chuang et al. 2007). This compares favourably with previous biomarkers based solely on transcriptome signatures (van't Veer and Bernards 2008). Integration of genetic and expression data into functional and regulatory networks, and the study of their topology and dynamics, are also starting to reveal the role of specific microRNAs, acting as tumour suppressors, in the control of the expression of target genes associated with tumour progression, suggesting that restoration of microRNA expression may be a valid anti-metastatic therapeutic strategy (Lee et al. 2010).

Multiscale mathematical models of the process by which tumours cells become invasive, and attach to vascular endothelial cells, as a first step of their migration to distant sites, have been developed. These models take into account various features such as the biochemical properties of cell adhesion molecules, the hypoxic condition of the microenvironment, and the physical properties of the cells and the vessels, leading to realistic simulation of actual properties observed in real experiments, and suggestions on how to counteract tumour invasion in the clinical context (Anderson et al. 2006; Ramis-Conde et al. 2009).

9.6 The Hallmarks of Cancer Extended to the Control of Metabolism and Stress

9.6.1 Cancer as a Metabolic Disease

Ever since Otto Warburg reported, more than half a century ago, the pronounced dependence of tumours on aerobic glycolysis rather than oxidative phosphorylation (the Warburg effect), cancer has increasingly been seen as a metabolic disease linked to impaired mitochondrial function that impacts upon the classical hallmarks of cancer as well as genome instability, to the extent that mitochondrial dysfunction is now considered as an additional essential hallmark, taken together with adaptation to stress conditions such as hypoxia and escape from immune surveillance (Seyfried and Shelton 2010; Zou 2005; Pouyssegur et al. 2006; Solimini et al. 2007; Kroemer and Pouyssegur 2008; Colotta et al. 2009; Luo et al. 2009; Ruan et al. 2009; Sheng et al. 2009; Tennant et al. 2009).

The wealth of very detailed biochemical information collected on metabolism over several decades, and published in the biological and biomedical literature, has

recently been incorporated into global computational representations amenable to mathematical modelling and simulation in the context of global -omics data sets in a variety of species including human (Duarte et al. 2007; Ma et al. 2007; Mo et al. 2007; Mo and Palsson 2009; Oberhardt et al. 2009). The value of this integration for cancer systems biology is rooted in the rich tradition of mathematical modelling in biochemistry and enzymology, which led to advanced tools for analysis of metabolic flux and control, with applications in metabolic engineering being already particularly well developed in microbiology and plant biology (Papin et al. 2003; Fell 2005; Lee et al. 2006; Hornberg et al. 2007).

9.6.2 Beyond Oncogene Addiction

It is clear, from the foregoing, that research communities in the mathematical, computational, engineering, biological and clinical sciences, who hitherto have worked independently on various aspects of metabolism, are now converging to integrate their expertise into systems approaches to disease, including cancer. Similar advances can now be expected in the understanding of cancer cell oncogene and non-oncogene addiction; tumour cells whose behaviour is 'driven' by particular genomic aberrations may be more dependent on aberrant pathways than normal cells (Weinstein 2002; Weinstein and Joe 2008; Luo et al. 2009). Cancer cells appear to rely on an ability to manage the consequences of oncogene-induced DNA damage and DNA replication stress (Gorgoulis et al. 2005; Bartkova et al. 2006; Harper and Elledge 2007; Halazonetis et al. 2008); mitotic stress resulting in aneuploidy (Whitesell and Lindquist 2005; Ganem et al. 2007; Torres et al. 2008; Williams et al. 2008); and metabolic stress related to the imbalance in production of reactive oxygen species, mitochondrion function and autophagy (Mathew et al. 2007; DeBerardinis et al. 2008; Dewhirst et al. 2008; DiPaola et al. 2008; Gogvadze et al. 2008; Diehn et al. 2009; Mathew et al. 2009). The challenge will be to unravel how these different basic processes interact with each other to produce cancer initiation and progression.

9.7 Conclusions and Perspectives

9.7.1 Genome Variation and Instability Revisited Through Genetic and Genomic Networks

The work of Theodor Boveri, over a century ago, showed that chromosomal instability and aneuploidy are common characteristics of tumours, and are associated with their progression to malignancy (Holland and Cleveland 2009). With the rapid advances in array-based and high-throughput sequencing methods of the past few

years, it has become possible to characterize extensively the landscape of genome and expression variation of different tumour types (Greenman et al. 2007), including breast and colon cancer (Sjoblom et al. 2006; Wood et al. 2007), lung carcinoma (Ding et al. 2008), pancreatic cancer (Jones et al. 2008), and glioblastoma (Cancer Genome Atlas Research Network 2008; Parsons et al. 2008). This extensive information is integrated in curated databases to provide informative genotype-phenotype relationships (Thorisson et al. 2009). The sequencing of the entire genome of normal and cancer cells of an individual leukaemia patient has enabled the discovery of far more cancer-initiating mutations than previously known from decades of molecular and cellular investigations (Ley et al. 2008). A similar study of a breast cancer has also revealed a number of mutagenic events that may be cancer-initiating (Ding et al. 2010).

9.7.2 Novel Avenues for Diagnosis, Therapy and Disease Network Modelling

Cancer is arguably a complex systems disease in 2010 which disease-perturbed networks and regulatory circuits are associated with non-linear stochastic dynamics, leading to robust cellular states that behave as stable attractors, trajectories in phase space towards which the biological system tends to converge over time (Lin et al. 2005; Hornberg et al. 2007; Ao et al. 2008; Bizzarri et al. 2008; Lee et al. 2008; Huang et al. 2009). It is also widely recognized that metabolic transformation is associated with the acquired characteristics and genome instability that together are the classical and newly defined hallmarks of cancer. This suggests a potential for metabolomics studies that take into account intracellular micro-compartmentation and metabolic channelling to contribute to increased understanding of the cancerous phenotypes, and to open alternative avenues for therapeutic interventions (Tennant et al. 2009; Boros et al. 2002; Cascante et al. 2002; Ovadi and Saks 2004; Cascante and Marin 2008).

With systems approaches being increasingly used to unravel the complexity of cancer states, we are already witnessing a shift of emphasis away from oncogenes and tumour suppressors studied and targeted in isolation, based on the century-old concept of the ‘magic bullet’ introduced by Paul Ehrlich that once provided significant therapeutic advances, but is now facing important limitations (Strebhardt and Ullrich 2008). The deciphering of the vast array of genetic, epigenetic and biochemical changes, modulated together in an orchestrated manner, is revealing network modules and protein interactions that have become interesting novel diagnostic and combined therapeutic leads (Cho et al. 2006; Pawson and Warner 2007; Wells and McClendon 2007; Tonon 2008; Wong et al. 2008; Brynildsen and Collins 2009; Chang et al. 2009). This is the case, for example, for synthetic lethal interactions (see Chap. 4) revealed through RNA interference functional screens (Lord and Ashworth 2009; Farmer et al. 2005; Kaelin 2005; Luo et al. 2009), or network components underlying DNA damage or oxidative stress (Benz and Yau 2008; Garinis et al. 2008).

Progress in experimental data collection, network topology, and dynamics analysis methods will be needed to enable assessment of those stochastic fluctuations that occur at the molecular and cellular levels, with important consequences on the behaviour of cancer cell fate (Tyson and Novak 2008; Tyson et al. 2001; Klemm and Bornholdt 2005; Novak and Tyson 2008; Spencer et al. 2009). It can thus be expected that network theory, as introduced by Barabasi and Albert, will contribute to bring understanding of cancer one step closer to the expectations of Hanahan and Weinberg.

Network analysis methods are being applied to very extensive clinical datasets to help in discovering the primary site of cancer development when metastases have already formed, or to predict the spread of metastatic disease to additional sites (Chen et al. 2009). While most of these networks remain at the qualitative level, this is a major step towards the predictive quantitative models that are a critical goal in the systems biology of cancer. Networks are also used to reveal the intricate relationships between metabolites and multiple diseases, and the influence of the topology of the metabolic network on the co-occurrence of diseases. New avenues are being opened for understanding the underlying mechanisms, as well as possible targets for diagnosis, prevention and therapy through metabolic control and nutrition (Braun et al. 2008; Ornish et al. 2008)

9.7.3 Frontier Challenges: Multiscale Integration and Cross-disciplinarity

We are still a long way, however, in the cancer field, from the level of detail and depth of coverage already achieved in the deciphering the regulatory logic of the early development of the sea urchin embryo (Davidson et al. 2002; Oliveri et al. 2008), or the transcriptional control of physiology in *halobacteria* (Bonneau et al. 2007). In order for such progress to occur with cancer, there are significant hurdles that will have to be overcome.

One that is becoming increasingly apparent is the use of different mathematical and computational formalisms and tools to model biological systems at their different levels of structural organisation and temporal scales. In cancer, as in systems biology in general, integration across these multiple scales is one of the next frontier challenges that needs to be addressed, for example, by using the conceptual and mathematical tools of scale relativity theory, and those developed in the context of the Physiome Project (Auffray and Nottale 2008; Noble 2002; Nottale and Auffray 2008; Bassingthwaighte et al. 2009; Kohl and Noble 2009).

Another major hurdle to be overcome is related to the cross-disciplinary nature of systems biology. It can be viewed as physiology and medicine revisited by formal quantitative and engineering sciences, as occurred at their foundation by William Harvey and Claude Bernard (Noble 2008; Auffray and Noble 2009). This vision of the future will require a shift from the conventional approaches of cellular, molecular and clinical biologists, following a linear predefined succession of experimental

investigations focusing on increasing levels of detail, to a more flexible framework of thinking in which iterative cycles of experimentation and rigorous modelling are designed and performed to address biological questions, generate understanding, and derive novel hypotheses and predictions that can be tested through refined modelling, simulation and experimental validation. We expect that over the next decade this will become the hallmark '*par excellence*' of cancer systems biology.

Acknowledgements We thank Odile Brasier for secretarial assistance. Eruption of the Eyjafjöll provided the opportunity to complete this chapter on time.

References

Aebersold R, Auffray C et al (2009) Report on EU-USA workshop: how systems biology can advance cancer research (27 October 2008). Mol Oncol 3(1):9–17

Ahn AC, Tewari M et al (2006a) The clinical applications of a systems approach. PLoS Med 3(7):e209

Ahn AC, Tewari M et al (2006b) The limits of reductionism in medicine: could systems biology offer an alternative? PLoS Med 3(6):e208

Aldridge BB, Burke JM et al (2006) Physicochemical modeling of cell signaling pathways. Nat Cell Biol 8(11):1195–1203

Aldridge BB, Saez-Rodriguez J et al (2009) Fuzzy logic analysis of kinase pathway crosstalk in TNF/EGF/insulin-induced signaling. PLoS Comput Biol 5(4):e1000340

Alon U (2007) Network motifs: theory and experimental approaches. Nat Rev Genet 8(6):450–461

Anderson AR, Quaranta V (2008) Integrative mathematical oncology. Nat Rev Cancer 8(3):227–234

Anderson AR, Weaver AM et al (2006) Tumor morphology and phenotypic evolution driven by selective pressure from the microenvironment. Cell 127(5):905–915

Ao P, Galas D et al (2008) Cancer as robust intrinsic state of endogenous molecular-cellular network shaped by evolution. Med Hypotheses 70(3):678–684

Auffray C (2007) Protein subnetwork markers improve prediction of cancer outcome. Mol Syst Biol 3:141

Auffray C, Chen Z et al (2009) Systems medicine: the future of medical genomics and healthcare. Genome Med 1(1):2

Auffray C, Imbeaud S et al (2003) From functional genomics to systems biology: concepts and practices. C R Biol 326(10–11):879–892

Auffray C, Noble D (2009) Origins of systems biology in William Harvey's masterpiece on the movement of the heart and the blood in animals. Int J Mol Sci 10(4):1658–1669

Auffray C, Nottale L (2008) Scale relativity theory and integrative systems biology: 1. Founding principles and scale laws. Prog Biophys Mol Biol 97(1):79–114

Barabasi AL (2007) Network medicine—from obesity to the "diseasome". N Engl J Med 357(4):404–407

Barabasi AL, Albert R (1999) Emergence of scaling in random networks. Science 286(5439):509–512

Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet 5(2):101–113

Barrett CL, Palsson BO (2006) Iterative reconstruction of transcriptional regulatory networks: an algorithmic approach. PLoS Comput Biol 2(5):e52

Bartkova J, Rezaei N et al (2006) Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. Nature 444(7119):633–637

Bassingthwaighte J, Hunter P et al (2009) The Cardiac Physiome:perspectives for the future. Exp Physiol 94(5): 597–605
Beckman RA, Loeb LA (2005) Genetic instability in cancer: theory and experiment. Semin Cancer Biol 15(6):423–435
Benz CC, Yau C (2008) Ageing, oxidative stress and cancer: paradigms in parallax. Nat Rev Cancer 8(11):875–879
Bild AH, Potti A et al (2006a) Linking oncogenic pathways with therapeutic opportunities. Nat Rev Cancer 6(9):735–741
Bild AH, Yao G et al (2006b) Oncogenic pathway signatures in human cancers as a guide to targeted therapies. Nature 439(7074):353–357
Billy F, Ribba B et al (2009) A pharmacologically based multiscale mathematical model of angiogenesis and its use in investigating the efficacy of a new cancer treatment strategy. J Theor Biol 260(4):545–562
Bizzarri M, Cucina A et al (2008) Beyond the oncogene paradigm: understanding complexity in cancerogenesis. Acta Biotheor 56(3):173–196
Bluthgen N, Legewie S et al (2009) A systems biological approach suggests that transcriptional feedback regulation by dual-specificity phosphatase 6 shapes extracellular signal-related kinase activity in RAS-transformed fibroblasts. FEBS J 276(4):1024–1035
Bonneau R (2008) Learning biological networks: from modules to dynamics. Nat Chem Biol 4(11):658–664
Bonneau R, Reiss DJ et al (2006) The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. Genome Biol 7(5):R36
Bonneau R, Facciotti MT et al (2007) A predictive model for transcriptional control of physiology in a free living cell. Cell 131(7):1354–1365
Borisov N, Aksamitiene E et al (2009) Systems-level interactions between insulin-EGF networks amplify mitogenic signaling. Mol Syst Biol 5:256
Boros LG, Cascante M et al (2002) Metabolic profiling of cell growth and death in cancer: applications in drug discovery. Drug Discov Today 7(6):364–372
Bosl WJ (2007) Systems biology by the rules: hybrid intelligent systems for pathway modeling and discovery. BMC Syst Biol 1:13
Braun A, Samann A et al (2008) Effects of metabolic control, patient education and initiation of insulin therapy on the quality of life of patients with type 2 diabetes mellitus. Patient Educ Couns 73(1):50–59
Butte AJ (2008) Medicine. The ultimate model organism. Science 320(5874):325–327
Brynildsen MP, Collins JJ (2009) Systems biology makes it personal. Mol Cell 34(2):137–138
Byrne HM (2010) Dissecting cancer through mathematics: from the cell to the animal model. Nat Rev Cancer 10(3):221–230
Cancer Genome Atlas Research Network (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 455(7216):1061–1068
Carrivick L, Rogers S et al (2006) Identification of prognostic signatures in breast cancer microarray data using Bayesian techniques. J R Soc Interface 3(8):367–381
Carro MS, Lim WK et al (2010) The transcriptional network for mesenchymal transformation of brain tumours. Nature 463(7279):318–325
Carter GW, Galas DJ et al (2009) Maximal extraction of biological information from genetic interaction data. PLoS Comput Biol 5(4):e1000347
Cascante M, Boros LG et al (2002) Metabolic control analysis in drug discovery and disease. Nat Biotechnol 20(3):243–249
Cascante M, Marin S (2008) Metabolomics and fluxomics approaches. Essays Biochem 45:67–81
Chang JT, Carvalho C et al (2009) A genomic strategy to elucidate modules of oncogenic pathway signaling networks. Mol Cell 34(1):104–114
Chang LW, Payton JE et al (2008) Computational identification of the normal and perturbed genetic networks involved in myeloid differentiation and acute promyelocytic leukemia. Genome Biol 9(2):R38

Chaplain MA, McDougall SR et al (2006) Mathematical modeling of tumor-induced angiogenesis. Annu Rev Biomed Eng 8:233–257
Chen LL, Blumm N et al (2009) Cancer metastasis networks and the prediction of progression patterns. Br J Cancer 101(5):749–758
Cho KH, Kim JR et al (2006) Inferring biomolecular regulatory networks from phase portraits of time-series expression profiles. FEBS Lett 580(14):3511–3518
Chuang HY, Lee E et al (2007) Network-based classification of breast cancer metastasis. Mol Syst Biol 3:140
Citri A, Yarden Y (2006) EGF-ERBB signaling: towards the systems level. Nat Rev Mol Cell Biol 7(7):505–516
Clermont G, Auffray C et al (2009) Bridging the gap between systems biology and medicine. Genome Med 1(9):88
Cline MS, Smoot M et al (2007) Integration of biological networks and gene expression data using Cytoscape. Nat Protoc 2(10):2366–2382
Colotta F, Allavena P et al (2009) Cancer-related inflammation, the seventh hallmark of cancer: links to genetic instability. Carcinogenesis 30(7):1073–1081
Cui Q, Ma Y et al (2007) A map of human cancer signaling. Mol Syst Biol 3:152
Cusick ME, Klitgord N et al (2005) Interactome: gateway into systems biology. Hum Mol Genet 14(Spec No. 2):R171–R181
Davidson EH, Rast JP et al (2002) A genomic regulatory network for development. Science 295(5560):1669–1678
Deberardinis RJ, Sayed N et al (2008) Brick by brick: metabolism and tumor cell growth. Curr Opin Genet Dev 18(1):54–61
Debily MA, Marhomy SE et al (2009) A functional and regulatory network associated with PIP expression in human breast cancer. PLoS One 4(3):e4696
Dewhirst MW, Cao Y et al (2008) Cycling hypoxia and free radicals regulate angiogenesis and radiotherapy response. Nat Rev Cancer 8(6):425–437
Diehn M, Cho RW et al (2009) Association of reactive oxygen species levels and radioresistance in cancer stem cells. Nature 458(7239):780–783
Ding L, Ellis MJ et al (2010) Genome remodeling in a basal-like breast cancer metastasis and xenograft. Nature 464(7291):999–1005
Ding L, Getz G et al (2008) Somatic mutations affect key pathways in lung adenocarcinoma. Nature 455(7216):1069–1075
DiPaola RS, Dvorzhinski D et al (2008) Therapeutic starvation and autophagy in prostate cancer: a new paradigm for targeting metabolism in cancer therapy. Prostate 68(16):1743–1752
Du Y, Wang K et al (2006) Coordination of intrinsic, extrinsic, and endoplasmic reticulum-mediated apoptosis by imatinib mesylate combined with arsenic trioxide in chronic myeloid leukemia. Blood 107(4):1582–1590
Duarte NC, Becker SA et al (2007) Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proc Natl Acad Sci U S A 104(6):1777–1782
Farmer H, McCabe N et al (2005) Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature 434(7035):917–921
Fell DA (2005) Enzymes, metabolites and fluxes. J Exp Bot 56(410):267–272
Ferreira SC Jr, Martins ML et al (2002) Reaction-diffusion model for the growth of avascular tumor. Phys Rev E Stat Nonlin Soft Matter Phys 65(2 Pt 1):021907
Folkman J (2007) Angiogenesis: an organizing principle for drug discovery? Nat Rev Drug Discov 6(4):273–286
Ganem NJ, Storchova Z et al (2007) Tetraploidy, aneuploidy and cancer.Curr Opin Genet Dev 17(2):157–162
Garinis GA, van der Horst GT et al (2008) DNA damage and ageing: newage ideas for an age-old problem. Nat cell Biol 10(11):1241–1247
Gianchandani EP, Brautigan DL et al (2006) Systems analyses characterize integrated functions of biochemical networks. Trends Biochem Sci 31(5):284–291

Gogvadze V, Orrenius S et al (2008) Mitochondria in cancer cells: what is so special about them? Trends Cell Biol 18(4):165–173
Goh KI, Cusick ME et al (2007) The human disease network. Proc Natl Acad Sci U S A 104(21):8685–8690
Gorgoulis VG, Vassiliou LV et al (2005) Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. Nature 434(7035):907–913
Graudens E, Boulanger V et al (2006) Deciphering cellular states of innate tumor drug responses. Genome Biol 7(3):R19
Greenman C, Stephens P et al (2007) Patterns of somatic mutation in human cancer genomes. Nature 446(7132):153–158
Halazonetis TD, Gorgoulis VG et al (2008) An oncogene-induced DNA damage model for cancer development. Science 319(5868):1352–1355
Harper JW, Elledge SJ (2007) The DNA damage response: ten years after. Mol Cell 28(5):739–745
Hanahan D, Weinberg RA (2000) The hallmarks of cancer. Cell 100(1):57–70
Heiser LM, Wang NJ et al (2009) Integrated analysis of breast cancer cell lines reveals unique signaling pathways. Genome Biol 10(3):R31
Holland AJ, Cleveland DW (2009) Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis. Nat Rev Mol Cell Biol 10(7):478–487
Hood L, Heath JR et al (2004) Systems biology and new technologies enable predictive and preventative medicine. Science 306(5696):640–643
Hornberg JJ, Bruggeman FJ et al (2007) Metabolic control analysis to identify optimal drug targets. Prog Drug Res 64:171, 173–189
Hu M, Polyak K (2008) Microenvironmental regulation of cancer development. Curr Opin Genet Dev 18(1):27–34
Huang S, Ernberg I et al (2009) Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. Semin Cell Dev Biol 20(7):869–876
Huang SS, Fraenkel E (2009) Integrating proteomic, transcriptional, and interactome data reveals hidden components of signaling and regulatory networks. Sci Signal 2(81):ra40
Huang YJ, Hang D et al (2008) Targeting the human cancer pathway protein interaction network by structural genomics. Mol Cell Proteomics 7(10):2048–2060
Hwang D, Lee IY et al (2009) A systems approach to prion disease. Mol Syst Biol 5:252
Hwang D, Rust AG et al (2005a) A data integration methodology for systems biology. Proc Natl Acad Sci U S A 102(48):17296–17301
Hwang D, Smith JJ et al (2005b) A data integration methodology for systems biology: experimental verification. Proc Natl Acad Sci U S A 102(48):17302–17307
Ideker T, Galitski T et al (2001) A new approach to decoding life: systems biology. Annu Rev Genomics Hum Genet 2:343–372
Ideker T, Sharan R (2008) Protein networks in disease. Genome Res 18(4):644–652
Itadani H, Mizuarai S et al (2008) Can systems biology understand pathway activation? Gene expression signatures as surrogate markers for understanding the complexity of pathway activation. Curr Genomics 9(5):349–360
Janes KA, Yaffe MB (2006) Data-driven modeling of signal-transduction networks. Nat Rev Mol Cell Biol 7(11):820–828
Jiang Y, Pjesivac-Grbovic J et al (2005) A multiscale model for avascular tumor growth. Biophys J 89(6):3884–3894
Johnston MD, Edwards CM et al (2007) Mathematical modeling of cell population dynamics in the colonic crypt and in colorectal cancer. Proc Natl Acad Sci U S A 104(10):4008–4013
Jones S, Zhang X et al (2008) Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science 321(5897):1801–1806
Kaelin WG, Jr. (2005) The concept of synthetic lethality in the context of anticancer therapy. Nat Rev Cancer 5(9):689–698
Karnoub AE, Dash AB et al (2007) Mesenchymal stem cells within tumour stroma promote breast cancer metastasis. Nature 449(7162):557–563

Kirschner MW (2005) The meaning of systems biology. Cell 121(4):503–504
Kitano H (2002) Systems biology: a brief overview. Science 295(5560):1662–1664
Klemm K, Bornholdt S (2005) Topology of biological networks and reliability of information processing. Proc Natl Acad Sci U S A 102(51):18414–18419
Kohl P, Noble D (2009) Systems biology and the virtual physiological human. Mol Syst Biol 5:292
Kreeger PK, Lauffenburger DA (2010) Cancer systems biology: a network modeling perspective. Carcinogenesis 31(1):2–8
Kreeger PK, Mandhana R et al (2009) RAS mutations affect tumor necrosis factor-induced apoptosis in colon carcinoma cells via ERK-modulatory negative and positive feedback circuits along with non-ERK pathway effects. Cancer Res 69(20):8191–8199
Kroemer G, Pouyssegur J (2008) Tumor cell metabolism: cancer's Achilles' heel. Cancer Cell 13(6):472–482
Kumar N, Afeyan R et al (2008) Multipathway model enables prediction of kinase inhibitor cross-talk effects on migration of Her2-overexpressing mammary epithelial cells. Mol Pharmacol 73(6):1668–1678
Lazzara MJ, Lauffenburger DA (2009) Quantitative modeling perspectives on the ErbB system of cell regulatory processes. Exp Cell Res 315(4):717–725
Lee JM, Gianchandani EP et al (2006) Flux balance analysis in the era of metabolomics. Brief Bioinform 7(2):140–150
Lee DS, Park J et al (2008) The implications of human metabolic network topology for disease comorbidity. Proc Natl Acad Sci U S A 105(29):9880–9885
Lee Y, Yang X et al (2010) Network modeling identifies molecular functions targeted by miR-204 to suppress head and neck tumor metastasis. PLoS Comput Biol 6(4):e1000730
Legewie S, Herzel H et al (2008) Recurrent design patterns in the feedback regulation of the mammalian signaling network. Mol Syst Biol 4:190
Ley TJ, Mardis ER et al (2008) DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature 456(7218):66–72
Lin B, White JT et al (2005) Evidence for the presence of disease-perturbed networks in prostate cancer cells by genomic and proteomic analyses: a systems approach to disease. Cancer Res 65(8):3081–3091
Linding RL, Jensen J et al (2007) Systematic discovery of in vivo phosphorylation networks. Cell 129(7):1415–1426
Lord CJ, Ashworth A (2009) Bringing DNA repair in tumors into focus. Clin Cancer Res 15(10):3241–3243
Ludwig JA, Weinstein JN (2005) Biomarkers in cancer staging, prognosis and treatment selection. Nat Rev Cancer 5(11):845–856
Luo J, Solimini NL et al (2009) Principles of cancer therapy: oncogene and non-oncogene addiction. Cell 136(5):823–837
Ma H, Sorokin A et al (2007) The Edinburgh human metabolic network reconstruction and its functional analysis. Mol Syst Biol 3:135
Macklin P, McDougall S et al (2009) Multiscale modeling and nonlinear simulation of vascular tumour growth. J Math Biol 58(4–5):765–798
Madar A, Bonneau R (2009) Learning global models of transcriptional regulatory networks from data. Methods Mol Biol 541:181
Mani KM, Lefebvre C et al (2008) A systems biology approach to prediction of oncogenes and molecular perturbation targets in B-cell lymphomas. Mol Syst Biol 4:169
Mathew R, Karantza-Wadsworth V et al (2007) Role of autophagy in cancer. Nat Rev Cancer 7(12):961–967
Mathew R, Karp CM et al (2009) Autophagy suppresses tumorigenesis through elimination of p62. Cell 137(6):1062–1075
Mazzone M, Comoglio PM (2006) The Met pathway: master switch and drug target in cancer progression. FASEB J 20(10):1611–1621

Mo ML, Jamshidi N et al (2007) A genome-scale, constraint-based approach to systems biology of human metabolism. Mol Biosyst 3(9):598–603
Mo ML, Palsson BO (2009) Understanding human metabolic physiology: a genome-to-systems approach. Trends Biotechnol 27(1):37–44
Mortazavi A, Williams BA et al (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5(7):621–628
Ngo VN, Davis RE et al (2006) A loss-of-function RNA interference screen for molecular targets in cancer. Nature 441(7089):106–110
Nguyen DX, Bos PD et al (2009) Metastasis: from dissemination to organ-specific colonization. Nat Rev Cancer 9(4):274–284
Nicholson JK, Holmes E et al (2004) The challenges of modeling mammalian biocomplexity. Nat Biotechnol 22(10):1268–1274
Noble D (2002) Modeling the heart—from genes to cells to the whole organ. Science 295(5560):1678–1682
Noble D (2008) Claude Bernard, the first systems biologist, and the future of physiology. Exp Physiol 93(1):16–26
Novak B, Tyson JJ (2008) Design principles of biochemical oscillators. Nat Rev Mol Cell Biol 9(12):981–991
Nottale L, Auffray C (2008) Scale relativity theory and integrative systems biology: 2. Macroscopic quantum-type mechanics. Prog Biophys Mol Biol 97(1):115–157
Oberhardt MA, Palsson BO et al (2009) Applications of genome-scale metabolic reconstructions. Mol Syst Biol 5:320
Oliveri P, Tu Q et al (2008) Global regulatory logic for specification of an embryonic cell lineage. Proc Natl Acad Sci U S A 105(16):5955–5962
Ornish D, Magbanua MJ et al (2008) Changes in prostate gene expression in men undergoing an intensive nutrition and lifestyle intervention. Proc Natl Acad Sci U S A 105(24):8369–8374
Ornish D, Magbanua MJ et al (2008) Changes in prostate gene expression in men undergoing an intensive nutrition and lifestyle intervention. Proc Natl Acad Sci U S A 105(24):8369–8374
Ovadi J, Saks V (2004) On the origin of intracellular compartmentation and organized metabolic systems. Mol Cell Biochem 256–257(1–2):5–12
Papin JA, Price ND et al (2003) Metabolic pathways in the post-genome era. Trends Biochem Sci 28(5):250–258
Parsons DW, Jones S et al (2008) An integrated genomic analysis of human glioblastoma multiforme. Science 321(5897):1807–1812
Pawson T, Warner N (2007) Oncogenic re-wiring of cellular signaling pathways. Oncogene 26(9):1268–1275
Pawson T, Linding R (2008) Network medicine. FEBS Lett 582(8):1266–1270
Polyak K, Haviv I et al (2009) Co-evolution of tumor cells and their microenvironment. Trends Genet 25(1):30–38
Pouyssegur J, Dayan F et al (2006) Hypoxia signaling in cancer and approaches to enforce tumour regression. Nature 441(7092):437–443
Pujana MA, Han JD et al (2007) Network modeling links breast cancer susceptibility and centrosome dysfunction. Nat Genet 39(11):1338–1349
Rachlin J, Cohen DD et al (2006) Biological context networks: a mosaic view of the interactome. Mol Syst Biol 2:66
Ramis-Conde I, Chaplain MA et al (2009) Multi-scale modeling of cancer cell intravasation: the role of cadherins in metastasis. Phys Biol 6(1):016008
Ransohoff DF (2009) Promises and limitations of biomarkers. Recent Results Cancer Res 181:55–59
Ransohoff DF, Gourlay ML (2010) Sources of bias in specimens for research about molecular markers for cancer. J Clin Oncol 28(4):698–704
Rhodes DR, Chinnaiyan AM (2005) Integrative analysis of the cancer transcriptome. Nat Genet (37 Suppl):S31–S37

Rhodes DR, Tomlins SA et al (2005) Probabilistic model of the human protein-protein interaction network. Nat Biotechnol 23(8):951–959
Roach JC, Glusman G et al (2010) Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science 328(5978):636–639
Ruan K, Song G et al (2009) Role of hypoxia in the hallmarks of human cancer. J Cell Biochem 107(6):1053–1062
Sakhanenko N, Galas DJ (2010) Markov logic networks in the analysis of genetic data. J Comp Biol 17(11):1491–1508
Samaga R, Saez-Rodriguez J et al (2009) The logic of EGFR/ErbB signaling: theoretical properties and analysis of high-throughput data. PLoS Comput Biol 5(8):e1000438
Sawyers CL (2008) The cancer biomarker problem. Nature 452(7187):548–552
Schlabach MR, Luo J et al (2008) Cancer proliferation gene discovery through functional genomics. Science 319(5863):620–624
Schlitt T, Brazma A (2005) Modeling gene networks at different organisational levels. FEBS Lett 579(8):1859–1866
Schoeberl B, Pace EA et al (2009) Therapeutically targeting ErbB3: a key node in ligand-induced activation of the ErbB receptor-PI3K axis. Sci Signal 2(77):ra31
Seyfried TN, Shelton LM (2010) Cancer as a metabolic disease. Nutr Metab (Lond) 7:7
Sheng H, Niu B et al (2009) Metabolic targeting of cancers: from molecular mechanisms to therapeutic strategies. Curr Med Chem 16(13):1561–1587
Shimoni Y, Friedlander G et al (2007) Regulation of gene expression by small non-coding RNAs: a quantitative view. Mol Syst Biol 3:138
Silva JM, Marran K et al (2008) Profiling essential genes in human mammary cells by multiplex RNAi screening. Science 319(5863):617–620
Sjoblom T, Jones S et al (2006) The consensus coding sequences of human breast and colorectal cancers. Science 314(5797):268–274
Solimini NL, Luo J et al (2007) Non-oncogene addiction and the stress phenotype of cancer cells. Cell 130(6):986–988
Spencer SL, Gaudet S et al (2009) Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. Nature 459(7245):428–432
Stites EC, Trampont PC et al (2007) Network analysis of oncogenic Ras activation in cancer. Science 318(5849):463–467
Strebhardt K, Ullrich A (2008) Paul Ehrlich's magic bullet concept: 100 years of progress. Nat Rev Cancer 8(6):473–480
Tan CS, Bodenmiller B et al (2009) Comparative analysis reveals conserved protein phosphorylation networks implicated in multiple diseases. Sci Signal 2(81):ra39
Taylor IW, Linding R et al (2009) Dynamic modularity in protein interaction networks predicts breast cancer outcome. Nat Biotechnol 27(2):199–204
Tegner JN, Compte A et al (2009) Computational disease modeling—fact or fiction? BMC Syst Biol 3:56
Tennant DA, Duran RV et al (2009) Metabolic transformation in cancer. Carcinogenesis 30(8):1269–1280
Thorisson GA, Muilu J et al (2009) Genotype-phenotype databases: challenges and solutions for the post-genomic era. Nat Rev Genet 10(1):9–18
Tomlins SA, Mehra R et al (2007) Integrative molecular concept modeling of prostate cancer progression. Nat Genet 39(1):41–51
Tonon G (2008) From oncogene to network addiction: the new frontier of cancer genomics and therapeutics. Future Oncol 4(4):569–577
Torres EM, Williams BR et al (2008) Aneuploidy: cells losing their balance. Genetics 179(2):737–746
Tyson JJ, Chen K et al (2001) Network dynamics and cell physiology. NatRev Mol Cell Biol 2(12):908–916
Tyson JJ, Novak B (2008). Temporal organization of the cell cycle. Curr Biol 18(17):R759–R768

van't Veer' LJ, Bernards R (2008) Enabling personalized cancer medicine through analysis of gene-expression patterns. Nature 452(7187):564–570
Varambally S, Yu J et al (2005) Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression. Cancer Cell 8(5):393–406
Velculescu VE, Kinzler KW (2007) Gene expression analysis goes digital. Nat Biotechnol 25(8):878–880
Vogelstein B, Kinzler KW (2004) Cancer genes and the pathways they control. Nat Med 10(8):789–799
Wang Z, Zhang L et al (2007) Simulating non-small cell lung cancer with a multiscale agent-based model. Theor Biol Med Model 4:50
Wang Z, Birch CM et al (2009) Cross-scale, cross-pathway evaluation using an agent-based non-small cell lung cancer model. Bioinformatics 25(18):2389–2396
Weinberg RA (2006) The biology of cancer. New York, Taylor and Francis
Weinstein IB (2002) Cancer. Addiction to oncogenes—the Achilles heal of cancer. Science 297(5578):63–64
Weinstein IB, Joe A (2008) Oncogene addiction. Cancer Res 68(9):3077–3080; discussion 3080
Wells JA, McClendon CL (2007) Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. Nature 450(7172):1001–1009
Westerhoff HV, Palsson BO (2004) The evolution of molecular biology into systems biology. Nat Biotechnol 22(10):1249–1252
Whitesell L, Lindquist SL (2005) HSP90 and the chaperoning of cancer. Nat Rev Cancer 5(10):761–772
Williams BR, Prabhu VR et al (2008) Aneuploidy affects proliferation and spontaneous immortalization in mammalian cells. Science 322(5902):703–709
Wolkenhauer O (2001) Systems biology: the reincarnation of systems theory applied in biology? Brief Bioinform 2(3):258–270
Wolkenhauer O, Auffray C et al (2010) Systems biologists seek fuller integration of systems biology approaches in new cancer research programs. Cancer Res 70(1):12–13
Wong PK, Yu F et al (2008) Closed-loop control of cellular functions using combinatory drugs guided by a stochastic search algorithm. Proc Natl Acad Sci U S A 105(13):5105–5110
Wood LD, Parsons DW et al (2007) The genomic landscapes of human breast and colorectal cancers. Science 318(5853):1108–1113
Wu JQ, Du J et al (2008) Systematic analysis of transcribed loci in ENCODE regions using RACE sequencing reveals extensive transcription in the human genome. Genome Biol 9(1):R3
Yeger-Lotem E, Sattath S et al (2004) Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction. Proc Natl Acad Sci U S A 101(16):5934–5939
Yildirim MA, Goh KI et al (2007) Drug-target network. Nat Biotechnol 25(10):1119–1126
Yuan JM, Hu DW (2006) Time-dependent sensitivity analysis of biological networks: coupled MAPK andPI3K signal transduction pathways. J Phys Chem A110:5361–5370
Zerhouni EA (2005) Translational and clinical science – time for a new vision. N Engl J Med 353(15):1621–1623
Zheng PZ, Wang KK et al (2005) Systems analysis of transcriptome and proteome in retinoic acid/arsenic trioxide-induced cell differentiation/apoptosis of promyelocytic leukemia. Proc Natl Acad Sci U S A 102(21):7653–7658
Zou W (2005) Immunosuppressive networks in the tumour environment and their therapeutic relevance. Nat Rev Cancer 5(4):263–274

Chapter 10
Systems Biology Analysis of Cell Death Pathways in Cancer: How Collaborative and Interdisciplinary Research Helps

Boris Zhivotovsky and Emmanuel Barillot

Abstract Programmed cell death, together with proliferation and differentiation, is responsible for maintaining the population of cells in all tissues. Dysregulation in this genetically-regulated process plays an important role in the pathophysiology of different disorders. Excessive cell death is pathogenic when it concerns post-mitotic cells and results in various diseases, such as infarction, stroke or neurodegenerative disorders. Insufficient cell death also results in a series of diseases, among which is cancer. Since cancer cells are so different, these cells can be used as an example (model) of how one can understand the whole system qualitatively and quantitatively. In this chapter the various cell death pathways are described. These pathways are central to cancer progression and to the understanding and optimal treatment of individual cancers. The complexity and interweaving of these pathways and the various feedback mechanisms indicate that a systems biology approach with mathematical modelling is needed to fully describe and understand the processes involved. The models based on these studies are crucial to effective drug development, drug testing, and optimal choice of combination therapies, since extensive molecular data for each type of cancer, personalized to each patient's history and genetic makeup, can be analysed. In addition, we focus on the question of why highly interdisciplinary approaches are required, and why common data, languages, and approaches are needed for optimum data analysis and clinical applications. This can often be effectively accomplished with multiple laboratory collaboration. An example of these collaborative efforts is given by the European Commission funded programs. Collaborative research helps to achieve the main goal of developing new approaches to rational strategies for cancer treatment.

B. Zhivotovsky (✉)
Institute of Environmental Medicine, Division of Toxicology, Karolinska Institutet, 17177, Stockholm, Sweden
e-mail: Boris.Zhivotovsky@ki.se

A. Cesario, F. B. Marcus (eds.), *Cancer Systems Biology, Bioinformatics and Medicine,*
DOI 10.1007/978-94-007-1567-7_10,

10.1 Introduction

For many years it was well known that the population of cells in any tissue is controlled by three genetically-regulated processes, namely proliferation, differentiation and death, suggesting that programmed cell death is one of the most important mechanisms in cell biology. In the early 1970s apoptotic cell death was described as a basic biological process with wide-ranging implications in tissue kinetics (Kerr et al. 1972). In the normal healthy human adult, apoptosis affects several millions of cells per second. During the four decades since Kerr's seminal paper was published, various modes of cell death have been described, in ways which often caused misunderstanding or misinterpretation. Recently, the Nomenclature Committee on Cell Death (NCCD) proposed unified criteria for the definition of cell death, based on biochemistry and morphology, while formulating several caveats against the misuse of words and concepts that slow down progress in the area of cell death research (Kroemer et al. 2009). Among the various cell death modes, apoptosis or 'caspase-dependent programmed cell death' is the most investigated. Apoptosis, autophagy and 'programmed' necrosis play an active role in physiological cell turnover during adult life, as in embryonic development and normal functioning of the immune system. On the other hand, dysregulation of cell death plays an important role in the pathophysiology of different disorders (Thompson 1995). Thus, excessive cell death is pathogenic when it concerns post-mitotic cells such as cardiomyocytes or neurons (for instance in infarction, stroke or neurodegenerative diseases) or when it affects renewable cells at a pace that cannot be compensated by proliferation and differentiation, as in the course of AIDS. Insufficient cell death also results in a series of diseases, including cancer.

Carcinogenesis is described as a multistep process, each step reflecting genetic alterations which drive the progressive transformation of normal cells into highly malignant derivatives (Vogelstein et al. 1988). Hanahan and Weinberg (2000; see Chap. 9) suggested that a manifestation of six essential alterations (plus one enabling characteristic) in cell physiology collectively dictates malignant growth (the so-called seven hallmarks of cancer). Among these alterations are: self-sufficiency in growth signals; insensitivity to anti-growth signals; tissue invasion and metastatic potential; limitless proliferation; and sustained angiogenesis, as well as evasion of cell death. Luo, Solimini, and Elledge proposed a set of additional cancer hallmarks that depict the stress phenotypes, such as DNA damage, oxidative, mitotic, proteotoxic and metabolic stresses, as well as evading immune surveillance (Luo et al. 2009). Conceptual progress during last several years pushed recently Hanahan and Weinberg publish "the new generation" of hallmarks of cancer (Hanahan D, Weinberg RA (2011) Hallmarks of Cancer: The Next Generation. Cell, 144:646–674). To the list of previously-described physiological alterations the authors suggested adding genomic instability, which generates the genetic diversity, and inflammation, which brings up numerous hallmarks functions, reprogramming of energy metabolism and escaping immune destruction. In addition, tumours create the specific microenvironments that add another level of complexity. Altogether it suggests that cancer is a complex systemic illness, different in many ways from other diseases.

Inactivation of the cell death program contributes to cancer at two levels: firstly during initial oncogenesis (one of the defining features of cancer is suppressed cell

death); and secondly during the acquisition of therapy resistance (cell death induction is the goal of chemotherapy and radiotherapy).

Within the last few years, a substantial step forward to understanding complex cell death pathways has been achieved through utilization of systems biology approaches. These approaches involve building mathematical models of biological networks, embodying our knowledge of the system, and using established facts as well as observations from high-throughput technologies (microarrays, deep sequencing, proteomics, imaging, etc) at several levels of description (DNA, RNA, protein, complexes, organelles, cells, organs and individuals). International funders, notably the European Commission, recently undertook major initiatives to support collaborative research in applying systems biology methods to cancer research. Both national and international funding agencies are supporting programmes of multilateral cooperation with focused collaborations using interdisciplinary approaches to understand pathways, leading ultimately to a more successful fight against cancer. In this chapter we describe the current systems biology studies devoted to cell-death signalling pathways for combating cancer. In addition, we focus on how collaborative and interdisciplinary research forwards the development of new approaches to optimised cancer treatment.

10.2 Cell Death Pathways

Autophagy, necrosis and mitotic catastrophe, in addition to apoptosis, may contribute to cell loss. All of them are genetically controlled and evolutionarily conserved modes of cell death, and may be triggered by a variety of biological, chemical or physical stimuli. In an evolutionary context, they have evolved as adaptive responses to damage in multicellular organisms, in order to prevent the emergence of aberrant and potentially malignant cells. In what follows, we describe these pathways in extensive detail, to illustrate the necessity for systems approaches fully and quantitatively to integrate and understand the functioning of these pathways.

Apoptosis is the most prominent mechanism of cell death. The most evident morphological alterations that characterize apoptotic cell death are cell shrinkage, nuclear chromatin condensation, nuclear fragmentation, and blebbing of the plasma membrane. The most important biochemical features relate to the activation of a specialized class of proteases, namely caspases, and a ladder-patterned DNA fragmentation. Although the basic molecular machinery that controls and unleashes the apoptotic program has been well characterized in model organisms, such as *Caenorhabditis elegans* (2002 Nobel Prize in Medicine or Physiology for Robert H. Horvitz (Horvitz 2003)), it is clear that mammalian cell death is subjected to an extremely complex regulation in which stimulus- and cell type-specific signal transduction pathways determine the apoptotic outcome in a context-dependent fashion.

Several protein families are involved in the regulation of the multistep apoptotic process. Some of these proteins (a family of "adaptor" proteins) are required for activation of different complexes, such as DISC (death-inducible signalling complex), apoptosome, or PIDDosome complexes. The second set of proteins (the Bcl-2 family of proteins) is involved in the activation of, or protection from, cell death. The third

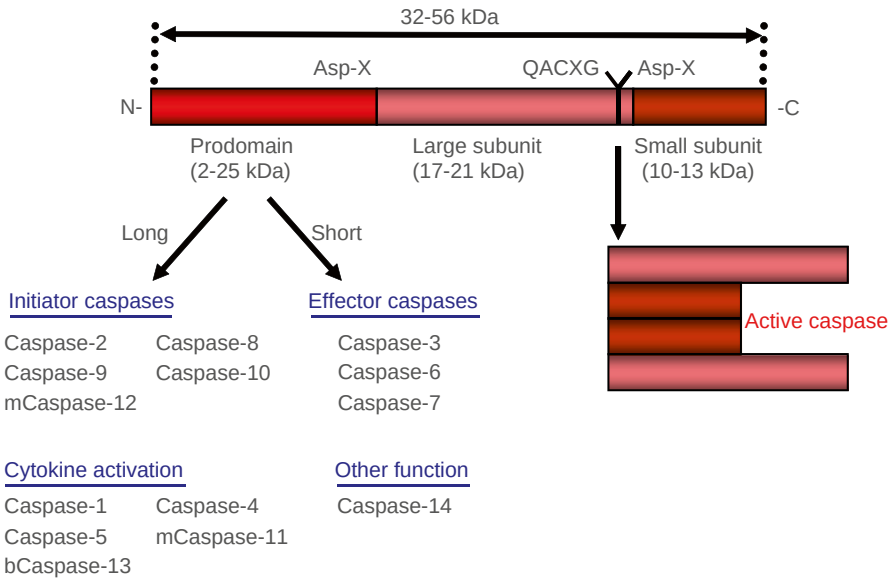


Fig. 10.1 The structural and functional organization of caspases

family of proteins, the so-called caspases, regulates the activation or execution of the apoptotic process. Currently, 14 different mammalian caspases have been described (11 in humans). Caspases are synthesized as inactive zymogens with differently sized N-terminal pro-domains, which are usually proteolytically removed to become active enzymes after oligomerization into tetramer complexes. Caspases are classified into "initiator" (caspase-1, -2, -4, -8, -9, -10 and -12) and "effector" caspases (caspase-3, -6, -7, -11 and -13) according to their function and the size of the N-terminal pro-domain, which is long in initiator caspases and short in effector caspases (Fig. 10.1). The N-terminal pro-domain of initiator caspases contains interaction epitopes for adaptor proteins that regulate their activation (Degterev et al. 2003).

On the other hand, effector caspases are unable to interact with adaptor proteins and hence are activated by other proteases, generally by upstream active caspases. Active effector caspases participate in cell disassembly by either cleaving and inactivating inhibitors of apoptosis, dismantling cellular structures, or regulating the activity of other proteins, resulting in the morphological and functional changes associated with apoptosis.

The two principal pathways of apoptosis are the death receptor-mediated pathway, initiated by the engagement of death receptors, and the mitochondria-mediated pathway, induced by various forms of stress-like intracellular damage, developmental cues, and external stimuli. Accumulating evidence shows that other intracellular compartments and/or organelles, such as the nucleus, the endoplasmatic reticulum, and the lysosomes, all have a role in apoptotic signalling (Orrenius et al. 2003). How the signals emerging from these organelles bifurcate into extrinsic and/or intrinsic signalling pathways is currently under intense investigation for the potential improvement of cancer therapy.

10.2.1 The Death Receptor-mediated Pathway

This pathway, also known as the caspase-8/-10-dependent or extrinsic pathway, is a plasma membrane receptor-mediated mechanism that results in the activation of caspase-8 and/or -10. The receptor-mediated apoptosis involves participation of members of the TNF superfamily of surface receptors and ligands. Among the more than 20 members of this family, the best known in apoptosis are Fas/Apo-1/CD-95, tumour necrosis factor-alpha (TNF-α) receptor 1 (TNF-R1), and the TRAIL receptors DR4 and DR5/Killer. The receptor-mediated pathway is initiated by the binding of the associated ligand. These ligands are released in response to different stimuli, such as cytotoxic response to microbial or viral infections, as part of the cellular or humoral immune responses, during activation of dendritic cells, upon stimulation or survival of T-, B-, and natural killer (NK) cells or in oncogenic transformation. Upon receptor-ligand association, the receptor undergoes oligomerization, which allows the recruitment of intracellular adaptor proteins, such as FADD and TRADD, for Fas and TNF receptors respectively. The adaptor proteins are required for further efficient recruitment and autoproteolytical activation of the initiator caspase-8 and -10 in the DISC. The downstream signalling is cell type-dependent, and accordingly, cells can be classified as type I or type II (Scaffidi et al. 1999). In the former, the initiator caspases directly activate the effector caspase-3 and -7. In contrast, for type II cells, activation of caspase-8 results in the cleavage of Bid (a member of the Bcl-2 family of proteins), whose truncated form, tBid, participates in the oligomerization of Bax-like proteins on the outer mitochondrial membrane, promoting the release of cytochrome *c* and other factors from the mitochondria, the converging point within the intrinsic signalling pathway.

10.2.2 The Mitochondria-mediated Pathway

This pathway is also known as the Bcl-2 inhibitable or intrinsic apoptotic pathway. It requires properly functioning mitochondria, which are the core organelles for detection of "death signals" and generation of apoptotic response. In certain cases mitochondria might amplify the apoptotic response by release of proteins that in normal conditions are confined to the mitochondrial intermembrane space and fulfil other specific functions (Cory et al. 2003). The key players in the intrinsic pathway are the Bcl-2 family of proteins, which together with several related proteins represent a large group of more than 30 pro- (Bax, Bad, Bak, Bcl-Xs, Noxa, Puma, etc) and anti-apoptotic (Bcl-2, Bcl-X_L, Bag-1, etc) members. The majority of these proteins operate on the mitochondrial level, participating in the stabilization or destabilization of these organelles, but also fulfilling some functions on the levels of the ER and the nuclei. The members of the family contain one or more conserved Bcl-2 homology domains (BH1 to BH4) and a C-terminus transmembrane domain that anchors them to membranes. The pro-apoptotic Bcl-2 family of proteins can be divided into two subgroups. The first group (Bax, Bok, Bcl-Xs, Bak and Bcl-Gl) contain two or three BH regions and share extensive structural similarity with their

pro-survival relatives. In contrast, other pro-apoptotic proteins (Bid, Bik, Noxa, Puma, Bmf) share with each other and with other Bcl-2 family members, only the short BH3 domain, and they are therefore often called BH3-only proteins. Both subgroups of pro-apoptotic Bcl-2 family members are essential for transmitting apoptotic signals. Following the reception of stress signals, the pro-apoptotic Bcl-2 family of proteins is activated and subsequently interacts with and inactivates the anti-apoptotic Bcl-2 proteins. Some of the pro-apoptotic Bcl-2 proteins are translocated from the cytosol to the mitochondrial outer membrane to form pore-like structures and thereby promote the release of apoptogenic factors, such as cytochrome *c*, AIF (apoptosis inducing factor), Smac/DIABLO (second mitochondria-derived activator of caspases/direct IAP-binding protein with low pI), and endonuclease G and Omi/HtrA2 (Wang 2001). In the cytosol, cytochrome *c*, together with the adaptor protein Apaf-1 (apoptotic protease-activating factor-1) and pro-caspase-9, builds up the apoptosome complex in the presence of dATP, activating pro-caspase-9. Processed caspase-9 directly activates the effector caspase-3 and -7, resulting in the cleavage of downstream apoptotic substrates and subsequent morphological changes, such as chromatin condensation, oligonucleosome-sized DNA fragmentation, and eventually cell death.

The possible mechanisms underlying mitochondria permeabilization are still under intense investigation. Earlier in time it was shown that in response to DNA damage, activation of caspase-2 takes place, followed by mitochondrial membrane permeabilization and release of cytochrome *c* and Smac/DIABLO. Other studies using *in vitro* model systems indicate that release of cytochrome *c* might occur independently of the outer membrane permeabilization (Gogvadze et al. 2004). In this case, the release of cytochrome *c* requires the modulation of the mitochondrial volume in circumstances under which the mitochondria remain intact and fully active.

10.2.3 Modulators of Caspase Activity: The IAP Family of Proteins and Their Regulators

Both caspase processing and activation are under the tight control of various regulatory proteins. Among them is a conserved family of inhibitors of apoptosis proteins (IAPs), containing baculoviral IAP repeat (BIRs) domains, which mediate the binding to and inhibition of caspases. Different BIRs participate in the inhibition of distinct caspases. Thus, BIR3 mainly interacts with caspase-9, whereas BIR2 inhibits activity of caspase-3 and caspase-7. In addition to the BIR domain, some of the IAPs members also contain a zinc-binding motif, the RING domain, which can catalyse the transfer of ubiquitin onto targeted proteins and label them for downstream degradation in the proteasome (Vaux and Silke 2005). Eight human IAPs have been identified so far, based on the presence or absence of BIR and RING domains. Although almost all of these proteins are expressed in most human tissues, the highest level of expression of a particular protein is tissue specific. Upon binding to a broad

range of caspases (caspase-3, -7, and -9), IAPs cause a steric block, which impedes the interaction and cleavage by caspases of the downstream substrates. Survivin is the most versatile member within this family in terms of its bifunctional role. It can participate in both cell cycle regulation and inhibition of caspase-3 and -7 in the apoptosis cascade. Survivin is one of the components of the mitotic spindle, which regulates the timing of the sister chromatides segregation during mitosis (Altieri 2003). While highly expressed in proliferating tumour cells, it is almost undetectable in quiescent cells, a feature that has attracted renewed interest on targeting survivin using new anticancer therapies with a double effect, blocking its inhibitory function on caspases and suppressing proliferation of tumour cells.

The effect of IAPs is antagonized by the mitochondria-localized proteins Smac/DIABLO and Omi/HtrA2, both released after mitochondrial permeabilization. They contain IAP-binding motifs which participate in the neutralization of the anti-apoptotic activity of IAPs. Thus, a balance between expression of IAPs and their inhibitors Smac/DIABLO and/or Omi/HtrA2 may determine whether a certain apoptotic stimulus will eventually trigger apoptosis.

10.2.4 Cross-talk Between Various Modes of Cell Death

Cells often succumb to non-apoptotic cell death that is manifested as necrosis or autophagic cell death. In contrast to apoptosis, necrosis for many years was considered as a passive process occurring as a result of acute damage that disables the cell from maintaining the basic energetic functions. It has been shown that a severe change in the energy balance after the injury can make the cell die by necrosis instead of apoptosis (Nicotera et al. 1998). However, recent evidence suggests that a cell can die by necrosis in a manner that initiates both inflammatory and/or reparative responses in the host. By initiating these adaptive responses, programmed cell necrosis may serve to maintain tissue and organism integrity. In this case necrosis represents a well-orchestrated pathway of biochemical reactions, which culminate in plasma membrane rupture.

Similarly to apoptosis, cell death with a necrotic appearance can contribute to embryonic development and adult tissue homeostasis. Some gene products, such as TNFR, CD95, TRAIL-R and RIP1, might trigger apoptosis and/or necrosis, depending on their interaction with other proteins. Moreover, there is crosstalk between these two cell death modalities. For example, the inactivation of caspases might cause a shift from apoptosis to necrosis, or to a mixture of the two. The term necroptosis has been introduced to designate a special type of programmed necrosis that depends on serine/threonine kinase RIP1 activity (Degterev et al. 2008). In a genome-wide siRNA screen for regulators of necroptosis, a set of 432 genes was identified. Among these genes was a subset of 32 genes that act downstream of, and/or as regulators of RIP1 kinase, and 7 genes involved in both necroptosis and apoptosis. Interestingly, Bmf, a BH3-only, Bcl-2 family member, was shown to be required for death receptor-induced necroptosis.

Recent data demonstrated that autophagy, which is referred to as type II programmed cell death, can also be involved in the regulation of cell death or survival (Eisenberg-Lerner and Kimchi 2009). Similar to apoptosis, autophagy is an evolutionarily conserved pathway, which is characterized by the sequestration and delivery of cytoplasmic material to the lysosomes, where it is degraded and recycled. The formation of autophagosomes is initiated by class III phosphoinositide 3-kinase and an autophagy-related gene (Atg) 6 (also known as Beclin-1). In addition, two further systems are involved, composed of the ubiquitin-like protein Atg8 (known as LC3 in mammalian cells) and the Atg4 protease in the one, and the Atg12-Atg5-Atg16 complex in the other.

It was shown that inhibition of autophagy may be cytoprotective and that its inhibition triggers apoptotic cell death under stress conditions. However, some of the proteins may control both autophagy and apoptosis, and some of them may contribute to apoptosis by inducing autophagy. Importantly, mitochondria might act as regulators of both apoptosis and autophagy. There is evidence for the involvement of mitochondria-generated ROS in the induction of autophagy, although the precise mechanism is unknown. Both apoptosis and autophagy can be controlled by the Bcl-2 family of proteins. Mitochondrial Bcl-2 blocks apoptosis via sequestration of the pro-apoptotic Bcl-2 protein family, while Bcl-2 localized in the ER can interact with Beclin-1 and inhibit autophagy. In addition to the pro-apoptotic functions of the BH3-only protein Bad, its overexpression can stimulate the formation of autophagy-associated complexes. This effect is lost when the BH3 domain of Bad is disrupted. In addition to its role in apoptosis, the activation of p53 can inhibit mTOR activity and regulate autophagy. It was also reported that, independent of the Atg5 function in autophagy, this protein might be pro-apoptotic and target mitochondria after being cleaved by calpain. It seems that the Bcl-2 protein family is essential for the regulation of the majority of programmed cell death modalities.

Depending on the type of lethal stimulus, the cell death process can be initiated in different intracellular compartments, and crosstalk between these compartments is essential for all cell death modalities. Moreover, it seems that various organelles might trigger cell death by specific stress sensors and transmit cell-death-modulating signals throughout the cell. This inter-organelle crosstalk apparently involves several molecular ‘switches’ within the signalling network. Thus, p53 can be activated in response to DNA damage or to changes in redox balance in the mitochondria. The Bcl-2 family of proteins might act at the level of the mitochondria, ER or nucleus. Importantly, depending on the nature of the stimulus, its severity, and on the cell type, the hierarchy of inter-organelle crosstalk might result in different cell death modalities. Moreover, in some cases suppression of the function of a particular compartment might switch one mode of cell death to another. For example, inhibition of mitochondrial function (lowering of ATP) can change the mode of cell death from apoptosis to necrosis. Similarly, inhibition of caspase activity might result in necrosis or autophagic cell death, whereas, as mentioned above, activation of calpain-mediated cleavage of autophagy-regulated protein, Atg-5, switches cell death from autophagy to apoptosis. However, in spite of these observations, the crosstalk between various modes of cell death remains not entirely explained.

10.3 Dysregulation of Cell Death Pathways in Cancer

It is accepted that tumour cells possess many different mechanisms for becoming insensitive to external insults or internal damages and thereby thwarting the activation of the cell death machinery. However, many current therapeutic strategies, including chemotherapy and radiotherapy, trigger damage which is sometimes not sufficient to activate the cell death process in tumour cells, suggesting that defects in cell death machineries might be responsible for unsuccessful treatment. This feature of cancer cells constitutes a growth advantage in that it selects resistant cells in a tumour among those that are sensitive; the resulting increase in rate of invasiveness and potential appearance of metastatic cells makes for a poor prognosis. Therefore, one of the main goals of new anticancer therapies is to reactivate defective cell death and implement a combinatorial strategy to kill tumour cells, in addition to existing radio- or chemotherapy treatments.

10.3.1 Defects in the Apoptotic Machinery of Tumour Cells

Cancer cells can harbour defects that disable either detectors of damage/activators of apoptosis (p53, Bax, Apaf-1, etc), or upregulate inhibitors of apoptosis (Bcl-2, Bcl-X_L, FLIPs, IAPs, etc). Such defects might influence the functioning of both receptor- and mitochondria-mediated pathways. Recent advances in cancer research are focused on the development of new drugs that try to overcome the ability of cancer cells to bypass the activation of apoptosis.

a. Defects in the death-receptor pathway and relevance to apoptosis resistance

The death-receptor pathway has been the subject of intense study since its significance for the initiation and signalling of apoptosis in lymphocytes, in response to gamma-radiation and chemotherapy, was observed (Fulda et al. 1998). Present knowledge reveals that many anticancer drugs induce apoptosis via activation of the death-receptor pathway. Alterations in this pathway can lead to resistance to anticancer drugs. It was reported that small-cell lung carcinomas (SCLCs) were resistant to FasL and TRAIL-induced apoptosis. These tumours, compared to non-small-cell lung carcinomas (NSCLCs), were characterized by reduced expression of Fas and DR4 as well as pro-caspase-8, due to methylation of the respective genes. Treatment of these tumour cells with demethylating agent 5'-aza-2-deoxycytidine increased the sensitivity to FasL and TRAIL-induced death (Hopkins-Donaldson et al. 2003).

As mentioned above, caspase-8 plays a crucial role in the signalling of the receptor-mediated pathway. The relevance of the expression of a mutated form of caspase-8 to the development of hepatocellular carcinomas (HCC) has been demonstrated. A frame-shift deletion in caspase-8 causes a premature termination of translation in the p10 subunit of the protease, implying its role in the early stages of

HCC, the loss of its cell death function, and its contribution to the pathogenesis of HCC (Soung et al. 2005). Absence or silence of caspase-8 was also observed and associated with development of neuroblastomas, suggesting the function for this apical caspase as a potential tumour suppressor. However, a more recent detailed investigation of the status of caspase-8 in neuroblastoma cells showed that suppression of caspase-8 expression occurs during the establishment of neuroblastoma metastases *in vivo*, and that reconstitution of caspase-8 expression in deficient neuroblastoma cells suppressed their metastasis. These findings define caspase-8 as a metastasis suppressor gene that regulates the survival and invasive capacity of neuroblastoma cells.

The reduced expression of the caspase-5 gene was shown in highly metastatic subpopulations of lung cancer cells and tissues, as demonstrated by cDNA array-based expression profiling (Hosomi et al. 2003). The combination of this method with polymerase chain reaction-single strand conformation polymorphism (PCR-SSCP) analysis suggests that caspase-5 might be a suppressor gene of high metastatic potential in lung cancer. Comparing differential gene expression profiles established by cDNA microarrays between normal cells, primary carcinoma cells and metastatic carcinoma cells, it was shown that down-regulation of caspase-5 as well as specific methylation of this gene occurs only in metastatic cells, suggesting that caspases might indeed act as tumour suppressor genes. How specific this phenomenon is for caspase-5 and -8, as well as how it might influence sensitivity of tumour cells to treatment, remains unclear and requires additional investigation.

Mutations of one of the CD95 alleles were shown to be sufficient for acquired resistance to apoptosis. The vast majority of these mutations are located in the death-domain (DD) (Muschen et al. 2002), and consequently the resistance provoked by these mutations is particularly efficient in type I cancer cells, where the recruitment of DD-associated proteins to constitute the DISC is more efficient as compared to type II cells. This resistance was shown to be overcome when protein synthesis or NFκB was inhibited, which was a first example of inducing apoptosis in CD95 defective cells through non-canonical activation of caspase-8 pro-enzyme (Barnhart et al. 2005). Overexpression of inhibitors of the death receptor-mediated pathway might also play an important role in the resistance of cancer cells. The overexpression of c-FLIP mediates resistance of Hodgkin/Reed-Sternberg cells to death-receptor-induced apoptosis (Mathas et al. 2004). In this case, overexpression of the caspase-8 inhibitor, c-FLIP, was observed in 55 out of 59 Hodgkin's lymphoma cell lines and was mediated by NFκB transactivation. The mere down-regulation of c-FLIP by siRNA was sufficient to sensitize cells to CD95-induced apoptosis. In some cases, defects in the adaptor proteins FADD/TRADD have been also reported.

The fact that some tumours with an intact death-receptor pathway are still resistant to treatment suggests that defects in other apoptotic pathways may also account for the lack of their response to chemotherapeutic interventions. In support of this assumption, it has been documented that Bax-deficient human colon carcinomas remain resistant to death-receptor activation of apoptosis, whereas Bax-expressing counterparts were sensitive to undergoing apoptosis through the Apo2L/TRAIL receptor, DR5 (LeBlanc et al. 2002). This clearly shows the connection between

the death-receptor pathway and the mitochondrial pathway, and how defects in the latter can affect the former, at least in type II cells. Furthermore, overexpression of inhibitors of apoptosis specific for the mitochondrial pathway, such as Bcl-2, Bcl-X_L and survivin, can desensitize tumour cells to TRAIL-induced apoptosis, and this overexpression/inhibition can be recovered by silencing of the anti-apoptotic regulators.

b. Failures of the mitochondrial apoptotic pathway

The signalling, execution and amplification of apoptosis through the mitochondria is of major importance for induction of the intrinsic as well as the extrinsic pathways in type II cells. Therefore, alterations in the components of this pathway are critical, and may strongly influence the sensitivity of the cancer cells to apoptosis upon chemotherapeutic treatment. This pathway is complex, and in cancer cells can be altered at different levels. Thus, such factors as the wrong balance between the anti- or pro-apoptotic regulators of the Bcl-2 family of proteins, alterations in stability of the outer mitochondrial membrane, efficiency of cytochrome *c* release, defects in the constitution of a functional apoptosome complex, and inhibition of the executioner stage, can all influence the successful functioning of this pathway and efficient execution of cell death.

By default, the mitochondrial pathway is activated in response to different stimuli, such as cytotoxic drugs, DNA damaging agents, etc. In some tumour cells the activation of the mitochondria-mediated apoptotic route requires, upon detection of genotoxic damage, transcription of genes via p53-dependent or -independent mechanisms. In fact, p53 might either trans-activate pro-apoptotic proteins of the Bcl-2 family and trans-repress anti-apoptotic proteins, or interact directly with anti-apoptotic proteins of the Bcl-2 family, shifting the global balance towards apoptosis (Slee et al. 2004). Hence, for some tumour cells the Bcl-2:Bax ratio has been shown to be an important parameter for predicting tumour resistance to apoptosis. However, in other cancer cells the elevated level of Bcl-2 expression might not always explain the resistance to chemo- or radiotherapy. Moreover, several studies have shown that Bcl-2 protein expression is higher in SCLC than in NSCLC, which leads to an apparent paradox, since SCLC is usually much more sensitive to chemotherapy than NSCLC. Accordingly, measurement of the level of Bcl-2 was not able to explain the resistance to cisplatin of the SCLC H69 cells, and in several NSCLCs, a low Bcl-2:Bax ratio was correlated with a positive prognosis. Taking all these results into consideration, it seems that there is/are a missing factor(s) that could more accurately predict the prognosis of lung cancer from treatment. In fact, the highest levels of Bcl-2 expression and Bcl-2:Bax ratio have been correlated with mutated p53 phenotype expression, suggesting that aggressiveness, response to therapy, and prognosis in lung cancer might be linked to the Bcl-2 family of apoptosis regulators. However, the prognosis cannot be completely determined by the relative level of protein expression of particular regulatory proteins.

As mentioned above, there are different points where the mitochondrial pathway can be disabled in tumours. For instance, 50% of colorectal cancer samples harbour a biallelic bax(G)8 frameshift that inactivates mutations of the pro-apoptotic Bax.

These mutations confer growth advantage on the tumour by decreasing apoptosis, compared to Bax-positive tumours (Miquel et al. 2005). Similarly, a significant frequency of mutations in the bak gene has been found to explain radiotherapy resistance in carcinomas of the uterine cervix, compared to healthy tissues. The absence of Puma and the loss of Bim, both belonging to so-called BH3-only pro-apoptotic proteins of the Bcl-2 family, efficiently protected cells from treatment with gamma-radiation or glucocorticoid hormones, establishing these proteins as key regulators of anticancer drug-induced apoptosis in lymphoid cells *in vivo* (Erlacher et al. 2005). It seems that not only alterations in the regulatory proteins belonging to the Bcl-2 family, but also chemical and physical alterations of the mitochondria, such as enhanced cytosolic Ca^{2+}-buffering capacity or abnormally elevated level of the basal mitochondrial membrane potential (Ψm), can be involved in acquisition of resistance to apoptosis. For instance, the level of intracellular Ca^{2+}, an important second messenger in apoptosis signalling (McConkey and Orrenius 1997) and responsible for the mitochondria permeability transition, can affect the responsiveness of cancer cells to certain treatments and their ability to commit apoptosis. In leukemic cells resistant to the purine nucleoside 2-chlorodeoxyadenosine (CdA), the higher mitochondrion Ca^{2+}-buffering capacity has been correlated with a lower level of cytochrome *c* release due to minimized mitochondrial membrane permeability transition, and correspondingly diminished apoptosis. Furthermore, some tumour cells are also characterized by a higher mitochondrial membrane potential (Ψm) than normal cells, which can influence their response to drug-induced mitochondrial depolarization and apoptosis (Gogvadze et al. 2008).

Importantly, even when all the parameters discussed earlier, including the release of cytochrome *c*, are properly functioning upon induction, the last step essential for activation of pro-caspase-9, the most upstream caspase linked to the mitochondrial pathway, can be disturbed by a defective assembly of the apoptosome complex. As mentioned above, in the cytosol, released cytochrome *c* binds to Apaf-1 in the presence of dATP, triggering a structural change in Apaf-1 that allows the arrangement of a wheel-shaped multimeric complex of seven spokes, called apoptosome. This structure catalyses the recruitment of pro-caspase-9 monomers, dimerization and production of active caspase-9, which in turn activates caspase-3, and the death machinery is ready to start the cell dismantlement. It was shown that in malignant melanoma, a very refractive type of cancer where most of the therapies have encountered severe challenges, even though it retains active p53, impairment in apoptosome complex formation reflects the importance of proper apoptosome functioning for successful anticancer therapy. Loss of heterozygosity of the Apaf-1 gene has been found in 42% of analysed samples of human malignant melanomas (Soengas et al. 2001). In general, a diminished level of Apaf-1 expression appears as a late event in malignant transformation, and the expression is completely abrogated in metastatic melanomas. Supporting these observations, *in vitro* overexpression of Apaf-1 significantly increased the sensitivity of melanoma cells to anticancer treatment, suggesting the possibility that Apaf-1 restoration can be a potential strategy for skin cancer treatment. Surprisingly, two groups found that only very few metastatic melanomas were negative for Apaf-1, rendering questionable the role of this protein in tumour

formation and sensitivity (Allen et al. 2005; Peltenburg et al. 2005). It seems that more work is required to clarify the role of this protein in both processes.

Recent observations indicate that defective activation of pro-caspase-9 associated with the apoptosome can also explain resistance to cisplatin-induced apoptosis in the ovarian cancer cell line SKOV3 (Liu et al. 2002). Caspase-9 activation and/or activity might also be regulated by phosphorylation. Hence, the cyclic AMP (cAMP) signal transduction pathway is implicated, and its activation is associated with inhibition of apoptosis. Moreover, using an *in vitro* system, it was found that protein kinase-A might inhibit apoptosome-associated caspase-9 activation in phosphorylation-dependent and -independent processes, downstream of cytochrome *c* release. However, it is unclear how specific this inhibition is for cancer cells.

In addition to the basic regulatory factors involved in the mitochondria-mediated pathway, there are many extra regulatory proteins whose levels and status might directly affect the activation of caspase-9 associated with the apoptosome complex formation. Among these factors it is important to mention the regulatory role of Hsp27, Hsp70, Hsp90, and Akt, as well as the interplay between IAPs and their inhibitors Smac/DIABLO and Omi/HtrA2. There is some evidence that almost all of these factors or their regulation are impaired in many forms of cancer and might thereby influence the sensitivity of tumour cells to radio- and chemotherapy.

10.3.2 Defects in Autophagy-regulated Machinery

The expression of some autophagy-related genes is changed in different types of cancer cells. For instance, it was found that there is a significant loss of Beclin-1 protein in breast cancer cells compared to normal cells, while expression of Beclin-1 increased in the malignant colorectal and gastric epithelial cells, compared to the normal mucosal cells (Qu et al. 2003). Upregulation of another autophagy-related gene LC3 was demonstrated in various gastrointestinal cancers, and it was suggested that LC3 expression is an advantage for cancer development in early phases of carcinogenesis (Takahashi et al. 2007). We found that the expression of LC3 protein correlates with sensitivity of NSCLC cell to radiotherapy and the surviving fraction after 2 Gy exposure (Kaminsky and Zhivotovsky, manuscript in preparation).

Suppression of autophagy might accelerate tumour development as a result of DNA damage. This is especially true if the mechanisms facilitating cell cycle arrest and/or apoptosis are also defective. Consistent with this, gene amplification, DNA damage and aneuploidy were prominent in Beclin 1 heterozygote cells that overexpressed Bcl-2, which cannot undergo apoptosis.

Tumour suppressors that activate autophagy, including inhibitors of the mTOR pathway, include TSC1/TSC2, PTEN (a PIP3 phosphatase that antagonizes PI(3)K/Akt signalling), and p53 (through AMPK). Other tumour suppressors play a more direct role in regulating autophagy, such as DAPk, Beclin 1, p19ARF (smARF) and UVRAG. Improved understanding of the role of resistance-related genes may have

an impact on establishment of new methods to predict therapy response, as well as to define new therapeutic targets. Many of these proteins also influence apoptosis, again stressing the prominence of the crosstalk between the two pathways. With the exception of Beclin 1, none of the Atg genes have been shown so far to be tumour suppressors or oncogenes. However, this does not mean that autophagy machinery is not essential for tumour progression. It is clear that the crosstalk between various modes of cell death manifests itself in several layers. This crosstalk is complex, and a full understanding of this multifaceted relationship will be critical for the assessment of anticancer strategies.

10.4 Mathematical Modelling of Cell Death Pathways

10.4.1 Different Models of Cell Death

Even though some of the players intervening in the apoptotic pathway are correctly determined, with some of their interactions apparently well established, others, along with the role of some species in different cell types or different cell conditions, may not always be well understood.

Mathematical models provide a systematic tool for investigating all possible scenarios in the cell, and proposing networks or mechanisms recapitulating experimental observations. They are a translation of biological knowledge into a mathematical language. Their purpose is to shed light on some contradictions and paradoxes exposed in the literature; summarize all that is known about a specific process; and give some insight into phenomena that remain obscure. Indeed, any kind of hypothesis can be tested *in silico* before performing wet experiments, once one is confident enough with the model. For a model to be used as a tool for biologists, it must reproduce existing data, from wild-type conditions to mutant phenotypes and drug treatment outputs. This requires close interactions between theoreticians and experimentalists, and short feedback loops from model results to experimental assessment for flexible model revision. Only then can the modeller or the biologist trust the model and formulate predictions.

Mathematical models become useful when the network—or the amount of information it represents—is too complex to understand with intuition only. They answer a specific question that needs to be stated clearly. Thus, some models will be simple with very few components, while others will be complex and very detailed, but all of them can bring something valuable to our general understanding of the apoptotic functioning and dysfunction. Models can also produce important quantitative results. The interest in apoptotic mathematical modelling includes both the activation of mitochondrial-dependent apoptosis in response to DNA damage via p53, and the receptor-triggered apoptosis via death receptors such as Fas or TNFR.

Both intrinsic and extrinsic apoptosis are described in variable detail according to the data and the problem the model is tackling. In 2000, Fussenegger was the first

to present a model that contained both pathways in a reasonable amount of detail (Fussenegger et al. 2000). Very often, these models give rise to emerging properties of the system that are not apparent in very detailed models (Bhalla and Iyengar 1999; Tyson et al. 2003). For instance, the all-or-nothing response linked to bistability in caspase activation (Eissing et al. 2004) and in Bax/Bcl2 interaction (Cui et al. 2008) is brought into relief in models of no more than four and five variables respectively. The bistability and irreversibility concepts (see Chap. 7) are shown at several levels, and appear to be a recurrent feature in apoptotic models. Their necessary presence is treated in almost all models.

In type I cells, the positive feedback involving caspase-8 and caspase-3 has been thoroughly studied (Eissing et al. 2004), as well as the perturbation of this feedback by specific caspase inhibitors (Stucki and Simon 2005). Indeed, it has been proved that the fight between caspases and their inhibitors can either destabilize or enhance the bistable behaviour (Choi et al. 2007; Legewie et al. 2006). Other bistable switches are observed in the intrinsic pathway. Indeed, some groups have concentrated on the events regulating mitochondria permeabilization (Bagci et al. 2006; Chen et al. 2007), and consequential release of essential components leading to effector caspases' activation (Eissing et al. 2004; Nakabayashi and Sasaki 2006; Rehm et al. 2009). Many models demonstrated that the amount of inhibitors could suppress the apoptotic phenotype. For instance, Bcl-2 was shown to have the capability to block type-II apoptosis (Hua et al. 2005; O'Connor et al. 2006). In a very detailed model of extrinsic apoptosis, the switch-like behaviour brought about by the positive feedback implicating caspase-8 and caspase-3, and the role of the permeabilization of the mitochondria membrane, were studied in depth (Albeck et al. 2008). Similarly, the role of c-FLIP in the regulation of type-I cell death has been demonstrated (Bentele et al. 2004).

The intrinsic apoptotic pathway is initiated by p53 in response to diverse internal stimuli such as DNA damage. The dynamics of p53 activation have raised many questions over the years, as evidenced by over 55,000 publications concerning this gene. How does the cell know when DNA cannot be repaired and apoptosis turned on, or how are p53 dynamics linked to initiation of apoptosis? Some periodic pulses of p53 with varying amplitudes have been observed by several experimental groups (Lahav et al. 2004; Lev Bar-Or et al. 2000) and studied theoretically (Ciliberto et al. 2005; Geva-Zatorsky et al. 2006; Ma et al. 2005; Wagner et al. 2005). These p53 oscillations may arise from the negative feedback involving Mdm2 and p53, but how this behaviour really triggers apoptosis remains obscure. This part of the apoptotic pathway needs more investigation both from an experimental and a theoretical point of view, and requires above all unifying models to explain the extensive observations.

At the level of the receptors, some extensive studies have been performed on the formation of the DISC and its implication in cell fate decision between death and survival (Bentele et al. 2004; Lavrik et al. 2007). The receptors not only trigger death, but also lead to cell growth and differentiation, depending on the concentration of crucial players such as cFLIP (Han et al. 2008). Aguda proposed, early

in the history of cell fate modelling, a very qualitative model, which exposed the types of feedbacks and non-linearity needed to enter the cell cycle or provoke apoptotic death (Aguda and Algar 2003). Since then, more models of cell fate decision have been proposed (Gaudet et al. 2005; Lavrik et al. 2007; Philippi et al. 2009), but they mainly concentrate on early decisions.

Most of the models of the apoptosis pathway use a continuous framework based on ordinary differential equations in order to follow the rate of change of protein concentration. Rhem and colleagues proposed the first spatio-temporal model of mitochondria outer membrane permeabilization, using partial differential equations (Rehm et al. 2009). Some other discrete formalisms have also proved their validity and offered a useful contribution to the analysis of apoptotic mechanisms, such as Boolean (Tournier and Chaves 2009; Zhang et al. 2008) and Petri Nets modelling (Heiner et al. 2004; Li et al. 2010). This review of models describing apoptosis and cell-fate decision is not exhaustive, but illustrates the growing interest, over the past 10 years, in understanding the subjacent mechanisms of cell death. Most of the models referenced here can be freely accessed in the BioModels database (Li et al. 2010).

10.4.2 Modelling Cell-fate Decision Between Survival, Apoptosis and Necrosis

To this long list of models, we can append the work developed in our group on cell-fate decision based on a Boolean formalism. In this study (Calzone et al. 2010), we investigated the connectivity and crosstalk between three pathways: survival, apoptosis, and necrosis, in response to death-receptor engagement. As of today, no mathematical model of necrosis has been proposed. We chose to concentrate on the pathway of programmed necrosis mainly because recent experiments have shown that it is tightly connected to the other two pathways and triggered by similar mechanisms (Declercq et al. 2009; Galluzzi and Kroemer 2008). By constructing a network of these three pathways, we tried to determine when, how, and under which conditions the cell fate decision is made between the three phenotypes.

More specifically, we modelled the engagement of TNF and FasL cytokines. Both cytokines can trigger survival or cell death according to the cell type or cellular conditions. On one hand, the cell can survive by activating the immune system via the NFkB signalling pathway. On the other hand, it can die by apoptosis or necrosis. Apoptosis is an orchestrated way of disrupting cellular components and packing them into specialized vesicles that are removed by phagocytes, whereas necrosis occurs with plasma membrane disruption and release of intracellular components into the surrounding tissues, possibly causing an inflammatory response and severe injury.

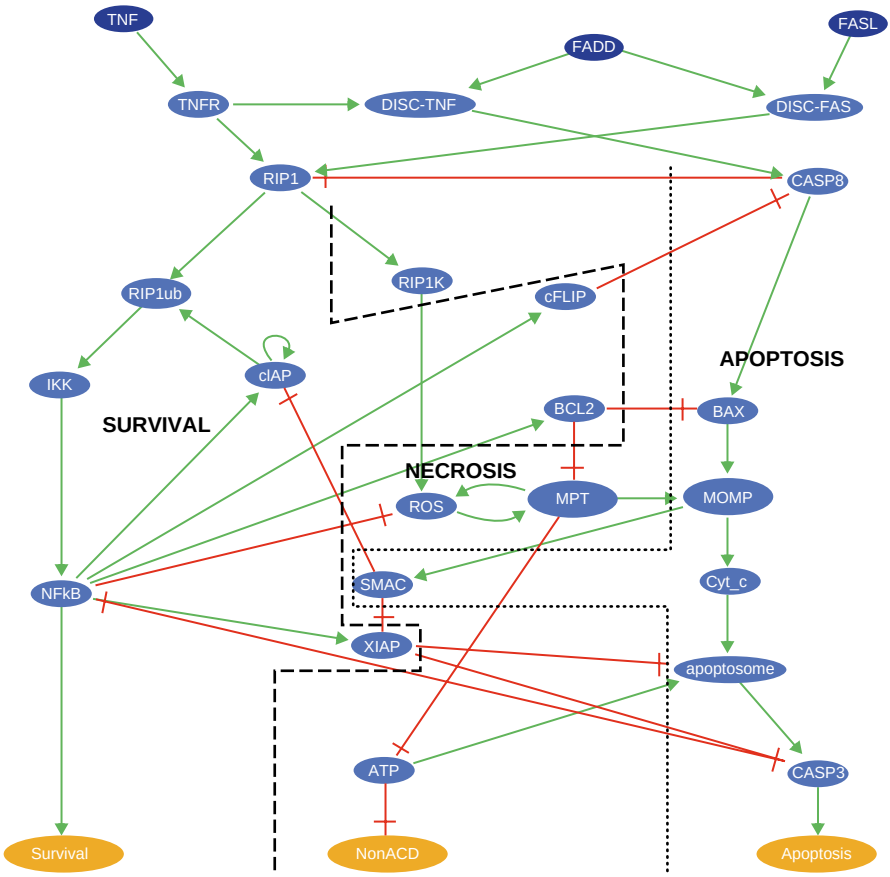


Fig. 10.2 Influence network of three pathways activated downstream of the death receptors Fas and TNFR. The nodes are biological species. They can be: ligands, receptors, proteins, metabolites, complexes of proteins, specific modified forms of a protein, cellular processes and phenotypes. Green arrows represent an activating influence of one node onto another node, and red arrows an inhibitory influence. Even though the pathways are highly intertwined, the network is separated into three pathways: the survival, non-apoptotic cell death (necrosis) and apoptotic pathways

We conducted a thorough search of the literature, recapitulated, integrated, annotated, and summarized known and published biological data into an influence diagram describing the molecular events leading to one or the other phenotype. The resulting diagram (Fig. 10.2) is composed of 28 nodes of which 3 are inputs (TNF, FasL, and FADD), and 3 are outputs (survival, non-apoptotic cell death and apoptosis). The nodes can represent proteins, protein modifications or domains of the protein (*e.g.* ubiquitinated form or kinase activity of RIP1), molecular processes (*e.g.* mitochondria outer membrane permeabilization—MOMP) or group of proteins (*e.g.* apoptosome). The choice of the species to include in the network was

governed by our reading and consequential intuition of the mechanisms involved in the decision. The connection of the three pathways revealed a highly intertwined network.

The network was then translated into a Boolean model. In such a framework, each node can have two values (1 or 0) corresponding to the presence or absence of the entity it represents, or the state of its activity. The Boolean rules setting the value of the nodes are chosen according to what is known about the regulation of each species. For instance, the initiator caspase (CASP8) can be active in the presence of the DISC of Fas or that of TNF, but always in the absence of the inhibitor cFLIP.

Simulations of wild-type, mutant cells and drug treatments qualitatively matched the existing data. Both steady-state and dynamical analyses confirmed that the network was coherent with published information, which constitutes a first validation of the model. Some novel mutant phenotypes could then be predicted along with general characteristics of the cell fate decision mechanism. First, we speculated that the activity of some key protein, RIP1, common to the necrotic and the survival pathways, was maintained in the first pathway but transient in the latter. Secondly, we made qualitative predictions on the increase or decrease of the probability of different cell fates in mutants, compared to wild-type cell fate probabilities. Thirdly, we showed that when reduced to the simplest network, the three pathways exhibit mutual inhibition. This is a requirement for reproducing all mutant phenotypes.

It was mentioned earlier that the mathematical model could be used as a tool for *in silico* experiments. With this idea, we tested the effect of the length of TNF induction in the cell fate decision process. In all the wild-type, mutant and drug treatment simulations, we assumed that the signal from the cytokines was maintained during the whole experiment. Here, we investigated the consequences it would have on the cell fate, if the signal were somehow removed at different time points. In this way we identified where the decision was made for the wild type and each mutant case. It then becomes possible to propose potential points of intervention in order to rescue a dysfunctional phenotype, or to suggest the increase or decrease in protein amounts to redirect the decision process.

Although the Boolean modelling does not allow a very refined description, our model still unravels contradictory facts about some interactions, constitutes a useful tool to test *in silico* experiments, and provides a framework for reasoning on specific biological questions.

In an article in which three mathematical models on cell fate were reviewed, Hardy and Stark (2002) concluded with this quotation: '*Who knows, in another ten years, mathematical models may become a standard tool in the armoury of the biologist trying to understand the interaction of apoptosis and proliferation*'. If this quote is generalized to cell-fate decision and not only to apoptosis vs. proliferation, in view of the increasing number of articles that combine experiments and models, this possibility would seem to have been confirmed five years earlier than expected.

10.5 Elements for Interdisciplinary Approaches to Cancer Research

10.5.1 Cancers Susceptible to Integrated Systems Approaches

In the majority of cases cancer is an evolving genetic disease acquired over time. A series of somatic mutations obtained as a result of endogenous and exogenous danger factors are accumulated in cells, eventually leading to uncontrolled cell growth and accumulation. Importantly, accumulation of these mutations is not sufficient to lead to a fully fledged cancer. As mentioned above, the manifestation of several essential alterations in cell physiology collectively dictate malignant growth, supporting the notion that cancer is a complex systemic illness that distinguishes it from many other diseases. The old concept that one gene is related to one cancer paradigm is defunct.

Many of the genes associated with cancer interact with several signalling pathways. For example, p53 promotes apoptosis and autophagy, activates DNA repair, regulates the level of ROS, and when mutated in cooperation with integrins, promotes invasion and metastasis. Many oncogenes encode various kinases to propagate molecular signalling. All these activities can operate in several signalling pathways. Therefore, the detailed investigation of individual components (genes and their products), as well as their interactions, is essential for systemic analysis.

It is important to note that even one cancer localization in the body is not a unified form of this disease. For example, lung carcinomas are classified into four histological types: small-cell lung carcinoma (SCLC), squamous-cell lung carcinoma (SCC), adenocarcinoma (AC), and the more undifferentiated large-cell lung carcinoma (LC). The histological features and clinical evolution makes SCLC a separate entity. The other three types are referred to as non-small-cell lung cancer (NSCLC) (see Chap. 2). Although all of these types of cancer are found in the lung, their pathophysiology, biochemistry and insensitivity to treatment are different. To achieve the main goal of successful treatment of the tumour, an interdisciplinary approach to analyse the reasons for these variations in local cancer physiology needs to be undertaken. It is important to identify and verify key milestones to evaluate our understanding of the system. An additional focus should be on the structure and dynamics of the system. Since cancer cells possess such variety, they can be used as examples (models) of how to qualitatively and quantitatively understand the whole system.

Biological processes are highly reproducible, a feature referred to as ‘robustness’, despite variability in genetic makeup as well as harsh and fluctuating environmental conditions. This robustness (see Chap. 17) also applies to cancer cells, where it is detrimental to drug development: in acute cases cancer cells have acquired insensitivity to many drugs (multidrug resistance phenomenon), which fail to deflect the cell from survival and proliferation. Nevertheless it is known that robust systems also have fragility points which, if identified, can be targeted for triggering a collapse (Kitano 2004). This paradigm applies directly to cancer and

to more precise definitions of cancer's defects, which should improve the odds of finding a successful drug. Improved molecular definition of tumour properties leads to a better understanding of the sensitivity or responsiveness to treatment. Even an advanced cancer has an Achilles' heel amenable to treatment. The role of systems biology is to find the most rational route to determine new treatment procedures. How can this be accomplished, and what is needed? The approach in cancer research is mostly the same as in other areas of systems biology, which is based on viewing biology as an informational science. An important difference is that each type of cancer should be analysed separately, as follows:

1. Functioning of different types of biological networks and their relevance to each type of cancer should be investigated.
2. Detailed information should be obtained concerning the disturbances in each network that might lead to tumour progression.
3. Knowledge of disturbances in each pathway that might lead to tumour resistance is of great importance and needs to be determined.

A cycle of hypothesis-driven investigation includes defining a biological or clinical question to investigate; preparing the model; collection and validation of experimental data or known facts; and integration into the model. If the model does not recapitulate experimental observations, iterations with model modification are necessary. This cycle should be repeated until the model and data are in agreement.

Systems approaches put together a unique combination of biological, mathematical, biomedical and biostatistical know-how that should lead to the generation of large data banks, as well as new methods of data analysis. The aims of this approach include refining the molecular diagnosis of various types of cancer, optimizing the calculation of prognostic and predictive parameters, and guiding new strategies for the amelioration of existing treatments and the identification of novel targets for therapeutic modulation of cell death pathways. Since one of the main goals of systems biology is to find the most rational route, a unified 'language' of modelling should be used by bioinformaticians, which should include the maximum possible information essential for understanding the proper functioning of the biological system.

In a Millennium review paper, Hanahan and Weinberg argued that 'cancer biology and treatment…will become a science with a conceptual structure and logical coherence that rivals that of chemistry or physics' (Hanahan and Weinberg 2000). They declared that within two decades, cell biologists will have derived a complete integrated circuit of the cell's signalling pathways, allowing us to model how specific genetic perturbations cause cancer, and to predict how to correct the problem by using drugs acting on key points in the circuit.

10.5.2 Laboratory and Clinical Measurements and Resources

Technology development is critical for the accumulation of data essential for further mathematical modelling. Recent successes in the improvement of imaging meth-

ods in living organisms allow space- and time-resolved quantitative measurements. Thus, combinations of advanced technologies in genomics, proteomics, life imaging and metabolomics will provide data for biostatistic, bioinformatic and systems biology analyses crucial for mathematical modelling of pathways, both at the stage of model definition, and for model challenging and validation. However, proper biological and clinical annotations of materials and their handling are also required, so as to lead to the discovery of new biomarkers and novel therapeutic targets. Validated information will provide new knowledge for patient management. Therefore, the next step for the proper use of system approaches is to integrate data and pathway definitions based on a common repository of well-annotated biobank resources, including surgical specimens and blood derivatives, selected cell lines and innovative preclinical models, with a high level of recorded histological quality control. The use of these shared biospecimens and cancer models, especially within an integrated project involving diversified groups which employ complementary methodologies, optimizes data integration for the application and pre-clinical/clinical validation of mathematical models.

The main conceptual novelties in this paradigm are unique methods of integrating data and generating models, based on the correlation of traditional measurements with certain tumour phenotypes, disease stages or response to treatment. Use of clinical specimens is essential for correctly weighting the reciprocal dynamic cooperation between critical driver and secondary effector pathways. Weighted ranking is based on the frequency assessed in the cell population. *In vitro* experimentation applies sequential or combined knockout or therapeutic intervention strategies for each lethal driving pathway and measures the relative impact on associated effector pathways. Such ranking contributes to a clear vision of the dynamics of the pathway that act in cooperative networks. This knowledge is absolutely essential for the integration of data and the design of robust models that will in all likelihood have important medical implications for diagnostic and therapeutic approaches. Moreover, such modelling will enhance knowledge about the critical points of intervention for biomedical healthcare strategies.

The main goal of this conceptual vision is to go beyond traditional medical paradigms, and to open the gateway for future mechanism-based biomedicine. Along with today's high-throughput tools for molecular investigation, this approach may enable the dream of personalised medicine to be realized. This methodology is a crucial step in translating measured data into clinically useful knowledge for patient management. The concept can be applied to all types of tumours and requires a complex processing of biological samples: biopsies from both normal and tumour tissues, along with blood samples and tumour samples from innovative preclinical *in vivo models* of cancer. Breakthroughs deriving from this conceptual approach rely upon the use of common repositories of biobank resources, so that all collaborating researchers and partners within research consortia can work on the same biospecimens, making use of their interdisciplinary abilities and complementary technological resources. This ensures a uniquely high level of data integration and opens new perspectives for modelling the complexity of cancer, thus generating truly holistic models of the disease. Since this systems biology strategy requires as

its basis large-scale experimentally generated and quantitative data sets, it should lead to highly reliable, validated and relevant models of human cancer. The optimal way, and often the only possible way for using this complex approach, is to put together resources from various research groups located in different universities and hospitals, which will permit the exchange of experimental and clinical materials. Legal and ethical controls are usually required, involving careful attention to legal guidelines, within collaborative and interdisciplinary projects.

As a consequence of such collaborations, the generation of useful knowledge, combined with innovative tools in the fields of early diagnosis, prognostic and predictive biomarkers, and with rational choices of treatments, will provide the basis for eventual improvements in patient management. Improvements should also be obtained in the ratio of output efficacy in research activities to economic investment in the areas of basic to translational and clinical research, with respect to societal needs.

10.6 How to Share Knowledge About Systems Biology Approaches to Cancers (See Also Chap. 6)

10.6.1 A Common Language

Systems approaches bring together specialists from diverse fields (biology, clinics, mathematics, computer science) whose interactions should enable questions to be solved that one specialist cannot tackle alone. As in any interdisciplinary project, the first and main impediment to a smooth interaction between persons from different cultures is language mismatch. Additionally, in a rapidly growing field like systems biology, many software programs and databases flourish in the community. Making them fully compatible, so that the network descriptions and models can be readily reused and perpetuated, is also an issue. Not surprisingly, establishing a common language has been one of the early endeavours of the systems biology community. Today several standards have emerged. (See also Chap. 6–8).

The problem of agreeing on common semantics for all terms and concepts used in describing biological networks has been addressed by using so-called ontologies. Ontology is a formal and explicit definition of concepts, their properties, and the relationships they share. This provides a common vocabulary and facilitates reasoning on the concepts. BioPAX (Biological Pathway Exchange—http://www.biopax.org) is such an ontology-based language, developed by a community of systems biology researchers and coordinated by the Computational Biology Center at Memorial Sloan-Kettering Cancer Center in New York. It aims at developing a common exchange format for biological pathways data. Its ontology covers several types of biological networks: metabolic pathways, molecular interactions, signalling pathways, gene regulatory networks and genetic interactions. The BioPAX format implements this ontology and offers a common syntax for describing these networks.

SBML (Systems Biology Markup Language—http://sbml.org) is another popular standard in systems biology, albeit with a different purpose. It aims at describing mathematical models of biochemical networks, such as signalling pathways, metabolic pathways, biochemical reactions, or gene regulation. SBML includes biochemical details of the models, not only reactant and products, but also stoichiometry, kinetic laws, and rules and constraints about all the reaction variables. The SBML effort was launched by a community led by the ERATO Kitano Systems Biology Project (Hucka et al. 2003). It is based on a XML definition of a shared format, which means that it is intended as a machine readable standard, making it possible to share and reuse models in any software which is compliant with this format, i.e. almost 200 software packages today.

10.6.2 Visualizing Networks as a Stimulus to Reasoning and Exchanges

BioPAX and SBML offer standard semantics and formats for describing biological networks and their models. These provide a basis for enabling exchanges, but remain at an abstract level. Since a key issue is how to facilitate discussions between experimentalists and theoreticians, in particular for debating network charting, a visual tool is necessary. Codification of graphical representation is required as a common, well specified and unambiguous language. This is the goal of the SBGN (Systems Biology Graphical Notation—http://sbgn.org) language, introduced by the Kitano Systems Biology group (Kitano et al. 2005; Le Novere et al. 2006). This visual language is human-readable, but it also enables automatic translation into SBML, making it readily available for algorithmic processing. SBGN consists of three complementary languages, each representing one aspect of biochemical networks:

- The process diagram depicts the molecular interactions between biochemical species and their results, and therefore focuses on state transition.
- The entity relationship diagram represents influences that species can have on other transformations: e.g. a Cyclin Kinase Inhibitor will inhibit the Retinoblastoma protein phosphorylation.
- The activity flow diagram ignores the details of biochemical processes, retaining only the influences between chemical species.

The CellDesigner software (http://www.celldesigner.org) is a graphical editor of process diagrams, whose usage can be mastered by any biologist or mathematician with one hour of training. It is compatible with the SBGN language, and has proven to be very useful in designing network descriptions, which united theoreticians and experimentalists in charting the RB/E2F pathway (Calzone et al. 2008), and the mitochondrial pathway of apoptosis. Its graphical interface being appropriate for both communities, it has dramatically facilitated knowledge transfer and promoted lively exchanges at the frontier of biology and mathematical modelling. Its interface to the

SBML language enables direct conversion of the diagram into a mathematical model. CellDesigner is typically used for charting pathways by a bioanalyst who reads the literature of interest and integrates a selection of interactions into a network map through the editor, keeping track of all evidence for further discussion with specialists. CellDesigner was developed and is maintained by the Kitano's group in Tokyo.

10.6.3 Sharing Network Description and Models

The above-presented tools allow scientists to describe, model, and visualize networks in common languages. Many databases have endorsed these languages and offer access to their content in BioPAX or SBML. This is particularly the case for the major pathway databases of importance for cancer research such as KEGG (http://www.genome.jp/kegg), Reactome (http://www.reactome.org), or Biomodels (http://www.ebi.ac.uk/biomodels). Biomodels was introduced at the European Bioinformatics Institute (Le Novere et al. 2006) for preserving and reusing published mathematical models of biological interest. Currently including more than 450 models of networks, it constitutes the largest public resource for such material, and is a central component of systems biology research.

To achieve integration of the BioPAX standard and the SBML/SBGN world, public release was made of a dedicated software, BiNoM—http://bioinfo.curie.fr/projects/binom, (Zinovyev et al. 2008). BiNoM also includes many features for structural analysis of biological networks, and comes as a Cytoscape plug-in (http://www.cytoscape.org).

10.7 Major Collaborative Efforts

Systems biology approaches have a real potential for advancing cancer research, having already proved successful in a very wide variety of cancer-related areas. Several project activities involving bioinformatics and systems biology approaches are currently supported by international funding agencies, such as the European Commission (EC) Framework Programs (http://cordis.europa.eu/fp7/home_en.html). These projects are either highly relevant to cancer research or directly dedicated to it. Here we discuss two projects, and refer the reader to (Marcus 2008) for further information.

10.7.1 Apo-sys: Large Scale Collaborative Research on Apoptosis

The Apo-sys (http://www.apo-sys.eu) 4-year project was initiated in 2008. It aims at achieving major progress in the understanding of apoptosis (and more generally cell death) in human diseases, by combining a series of systems biology approaches: *in silico*; *in vitro* (*in organello* and *in cellula*); and *in vivo*. The project integrates

experimental results with large data sets acquired on tissue samples from patients suffering from diseases that are caused by deregulated apoptosis, in particular cancer and AIDS. A Europe-wide consortium of experimental biologists, biomathematicians, biostatisticians, computer scientists and clinical scientists has been formed to address the striking complexity of human cell death pathways in health and disease using integrated methods involving high-throughput screening and 'omics' approaches applied to biological systems and computational modelling. This should lead to accurate and disease-relevant in silico models of apoptotic signalling triggered along the principal pathways, the extrinsic pathways (stimulated by ligation of death receptors), and the intrinsic pathways (stimulated by intracellular stress causing mitochondrial membrane permeabilization). Furthermore, the consortium is comparatively assessing the systems biology of apoptotic and non-apoptotic cell death (necrosis, autophagy and mitotic catastrophe), in order to understand the extent of overlap in the mechanisms leading to different phenotypic manifestation of cell death, as well as the molecular 'switches' that decide whether cells remain alive, or die through one or the other cell-death pathway.

10.7.2 Cancersys: Medium Scale Collaborative Research on Hepatocellular Carcinoma

The CancerSys (see also Chap. 11) project (http://www.ifado.de/cancersys) is a European Commission (FP7) project which endeavours to build dynamical mathematical models of two major signalling pathways, beta-catenin and Ras, involved in the formation of hepatocellular carcinoma. Integrative studies linking measurements in primary hepatocytes with effects at the organ level address the impact of these signalling networks on proliferation, tissue organization, and formation of hepatocellular carcinoma. In a close collaboration of scientists from theoretical fields and life sciences, the main approach of Cancersys is to combine dynamic mathematical modelling of signalling networks with spatial-temporal modelling of the liver microarchitecture. In an iterative process, the models are validated and adjusted to the *in vivo* situation. Modelling evaluates the consequences within the three-dimensional structure of the liver lobule (cell proliferation, tissue organization and of course carcinogenesis), with the final aim of guiding the design of better therapeutic strategies.

10.8 Supporting Collaborative Research Projects

10.8.1 Enfin: Systems Biology Tool Development and Application to Cancer

The European Commission funded ENFIN Project (http://www.enfin.org) is a Network of Excellence which aims at building a European-wide infrastructure for

systems biology research (see Chap. 8). It started in 2005 and brought together 21 partners each committed to integrate its databases and tools into a common distributed computational platform, in order to enable systems-level interpretation of experimental results.

10.8.2 Gen2phen: Bioinformatics Analysis of Genetic Variation and Application to Colorectal Cancer

The Gen2Phen project (http://www.gen2phen.org) is also funded by the European Commission and started in 2008 as a partnership of 19 institutions, sharing the same vision of what should be a biomedical knowledge environment for genome variation. It aims at unifying access to human and model organism genetic variation databases in a GRID-based mode and through genome browsers.

10.9 Conclusion

Deregulation of cell death plays a major role in carcinogenesis, and with the advent of high-throughput biotechnologies such as microarrays, deep sequencing and cellular imaging, we can now characterize a tumour in detail at the molecular, cellular and organ levels, and in a close to exhaustive manner in many domains (e.g. genome modifications, RNA transcription). This situation has opened the way to new approaches to cancer understanding and therapeutic development. Systems biology is based on the mathematical modelling of the different biological networks that control the mechanism of tumorigenesis and tumour progression, and in particular cell death. As described above, the scientific community has started organizing the building blocks of a shared infrastructure that should enable the continuation of this systems biology effort. National and international funding programs, including that of the European Commission, have significantly supported research in this field. A combination of systems biology approaches and resulting drugs obtained holds great promise of real progress in the ongoing war on cancer over the next decade or two. History shows that therapy comes when you understand the system of the disease (Abbott 2002).

Elements needed for successful interdisciplinary approaches to cancer research include the choice of cancers susceptible to integrated systems approaches: appropriate laboratory and clinical measurements and resources; bioinformatics and systems biology resources; and communication and management structures. Within Europe, as in other countries, many programmes have been launched that are based on the same principle of fostering research strategies based on systems approaches and bring together interdisciplinary consortia. Several major existing efforts involve large- and medium-scale collaborative projects in addiction to supporting those pro-

grammes which were discussed in this chapter; they demonstrate how a systems biology approach might substantially contribute to combating cancer.

Acknowledgements We thank Laurence Calzone for her substantial contribution to this chapter and Andrei Zinovyev for stimulating discussions on the subject. This work was performed within the FP7 (APO-SYS) program. E Barillot is a member of the U900 team "Computational Systems Biology of Cancer," Equipe labellisée par la Ligue Nationale Contre le Cancer. B Zhivotovsky's group is supported by grants from the Swedish Research Council, the Swedish and the Stockholm Cancer Societies, the Swedish Childhood Cancer Foundation and the EC FP-6 (Chemores) programs.

References

Abbott A (2002) On the offensive. Nature 416:470–474

Aguda BD, Algar CK (2003) A structural analysis of the qualitative networks regulating the cell cycle and apoptosis. Cell Cycle 2:538–544

Albeck JG, Burke JM, Spencer SL, Lauffenburger DA, Sorger PK (2008) Modeling a snap-action, variable-delay switch controlling extrinsic cell death. PLoS Biol 6:2831–2852

Allen JD, Zhang XD, Scott CL, Boyle GM, Hersey P, Strasser A (2005) Is Apaf-1 expression frequently abrogated in melanoma? Cell Death Differ 12:680–681

Altieri DC (2003) Survivin in apoptosis control and cell cycle regulation in cancer. Prog Cell Cycle Res 5:447–452

Bagci EZ, Vodovotz Y, Billiar TR, Ermentrout GB, Bahar I (2006) Bistability in apoptosis: roles of bax, bcl-2, and mitochondrial permeability transition pores. Biophys J 90:1546–1559

Barnhart BC, Pietras EM, Algeciras-Schimnich A, Salmena L, Sayama K, Hakem R, Peter ME (2005) CD95 apoptosis resistance in certain cells can be overcome by noncanonical activation of caspase-8. Cell Death Differ 12:25–37

Bentele M, Lavrik I, Ulrich M, Stosser S, Heermann DW, Kalthoff H, Krammer PH, Eils R (2004) Mathematical modeling reveals threshold mechanism in CD95-induced apoptosis. J Cell Biol 166:839–851

Bhalla US, Iyengar R (1999) Emergent properties of networks of biological signaling pathways. Science 283:381–387

Calzone L, Gelay A, Zinovyev A, Radvanyi F, Barillot E (2008) A comprehensive modular map of molecular interactions in RB/E2F pathway. Mol Syst Biol 4:173

Calzone L, Tournier L, Fourquet S, Thieffry D, Zhivotovsky B, Barillot E, Zinovyev A (2010) Mathematical modelling of cell-fate decision in response to death receptor engagement. PLoS Comput Biol 6:e1000702

Chen C, Cui J, Lu H, Wang R, Zhang S, Shen P (2007) Modeling of the role of a Bax-activation switch in the mitochondrial apoptosis decision. Biophys J 92:4304–4315

Choi HS, Han S, Yokota H, Cho KH (2007) Coupled positive feedbacks provoke slow induction plus fast switching in apoptosis. FEBS Lett 581:2684–2690

Ciliberto A, Novak B, Tyson JJ (2005) Steady states and oscillations in the p53/Mdm2 network. Cell Cycle 4:488–493

Cory S, Huang DC, Adams JM (2003) The Bcl-2 family: roles in cell survival and oncogenesis. Oncogene 22:8590–8607

Cui J, Chen C, Lu H, Sun T, Shen P (2008) Two independent positive feedbacks and bistability in the Bcl-2 apoptotic switch. PLoS One 3:e1469

Declercq W, Vanden Berghe T, Vandenabeele P (2009) RIP kinases at the crossroads of cell death and survival. Cell 138:229–232

Degterev A, Boyce M, Yuan J (2003) A decade of caspases. Oncogene 22:8543–8567

Degterev A, Hitomi J, Germscheid M, Ch'en IL, Korkina O, Teng X, Abbott D, Cuny GD, Yuan C, Wagner G et al (2008) Identification of RIP1 kinase as a specific cellular target of necrostatins. Nat Chem Biol 4:313–321

Eisenberg-Lerner A, Kimchi A (2009) The paradox of autophagy and its implication in cancer etiology and therapy. Apoptosis 14:376–391

Eissing T, Conzelmann H, Gilles ED, Allgower F, Bullinger E, Scheurich P (2004) Bistability analyses of a caspase activation model for receptor-induced apoptosis. J Biol Chem 279:36892–36897

Erlacher M, Michalak EM, Kelly PN, Labi V, Niederegger H, Coultas L, Adams JM, Strasser A, Villunger A (2005) BH3-only proteins Puma and Bim are rate-limiting for gamma-radiation- and glucocorticoid-induced apoptosis of lymphoid cells in vivo. Blood 106:4131–4138

Fulda S, Scaffidi C, Pietsch T, Krammer PH, Peter ME, Debatin KM (1998) Activation of the CD95 (APO-1/Fas) pathway in drug- and gamma-irradiation-induced apoptosis of brain tumor cells. Cell Death Differ 5:884–893

Fussenegger M, Bailey JE, Varner J (2000) A mathematical model of caspase function in apoptosis. Nat Biotechnol 18:768–774

Galluzzi L, Kroemer G (2008) Necroptosis: a specialized pathway of programmed necrosis. Cell 135:1161–1163

Gaudet S, Janes KA, Albeck JG, Pace EA, Lauffenburger DA, Sorger PK (2005) A compendium of signals and responses triggered by prodeath and prosurvival cytokines. Mol Cell Proteomics 4:1569–1590

Geva-Zatorsky N, Rosenfeld N, Itzkovitz S, Milo R, Sigal A, Dekel E, Yarnitzky T, Liron Y, Polak P, Lahav G, Alon U (2006) Oscillations and variability in the p53 system. Mol Syst Biol 2:33

Gogvadze V, Robertson JD, Enoksson M, Zhivotovsky B, Orrenius S (2004) Mitochondrial cytochrome c release may occur by volume-dependent mechanisms not involving permeability transition. Biochem J 378:213–217

Gogvadze V, Orrenius S, Zhivotovsky B (2008) Mitochondria in cancer cells: what is so special about them? Trends Cell Biol 18:165–173

Han L, Zhao Y, Jia X (2008) Mathematical modeling identified c-FLIP as an apoptotic switch in death receptor induced apoptosis. Apoptosis 13:1198–1204

Hanahan D, Weinberg RA (2000) The hallmarks of cancer. Cell 100:57–70

Hardy K, Stark J (2002) Mathematical models of the balance between apoptosis and proliferation. Apoptosis 7:373–381

Heiner M, Koch I, Will J (2004) Model validation of biological pathways using Petri nets—demonstrated for apoptosis. Biosystems 75:15–28

Hopkins-Donaldson S, Ziegler A, Kurtz S, Bigosch C, Kandioler D, Ludwig C, Zangemeister-Wittke U, Stahel R (2003) Silencing of death receptor and caspase-8 expression in small cell lung carcinoma cell lines and tumors by DNA methylation. Cell Death Differ 10:356–364

Horvitz HR (2003) Nobel lecture. Worms, life and death. Biosci Rep 23:239–303

Hosomi Y, Gemma A, Hosoya Y, Nara M, Okano T, Takenaka K, Yoshimura A, Koizumi K, Shimizu K, Kudoh S (2003) Somatic mutation of the Caspase-5 gene in human lung cancer. Int J Mol Med 12:443–446

Hua F, Cornejo MG, Cardone MH, Stokes CL, Lauffenburger DA (2005) Effects of Bcl-2 levels on Fas signaling-induced caspase-3 activation: molecular genetic tests of computational model predictions. J Immunol 175:985–995

Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A et al (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics 19:524–531

Kerr JF, Wyllie AH, Currie AR (1972) Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. Br J Cancer 26:239–257

Kitano H (2004) Cancer as a robust system: implications for anticancer therapy. Nat Rev Cancer 4:227–235

Kitano H, Funahashi A, Matsuoka Y, Oda K (2005) Using process diagrams for the graphical representation of biological networks. Nat Biotechnol 23:961–966

Kroemer G, Galluzzi L, Vandenabeele P, Abrams J, Alnemri ES, Baehrecke EH, Blagosklonny MV, El-Deiry WS, Golstein P, Green DR et al (2009) Classification of cell death: recommendations of the nomenclature committee on cell death 2009. Cell Death Differ 16:3–11

Lahav G, Rosenfeld N, Sigal A, Geva-Zatorsky N, Levine AJ, Elowitz MB, Alon U (2004) Dynamics of the p53-Mdm2 feedback loop in individual cells. Nat Genet 36:147–150

Lavrik IN, Golks A, Riess D, Bentele M, Eils R, Krammer PH (2007) Analysis of CD95 threshold signaling: triggering of CD95 (FAS/APO-1) at low concentrations primarily results in survival signaling. J Biol Chem 282:13664–13671

Le Novere N, Bornstein B, Broicher A, Courtot M, Donizelli M, Dharuri H, Li L, Sauro H, Schilstra M, Shapiro B et al (2006) BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. Nucleic Acids Res 34:D689–D691

LeBlanc H, Lawrence D, Varfolomeev E, Totpal K, Morlan J, Schow P, Fong S, Schwall R, Sinicropi D, Ashkenazi A (2002) Tumor-cell resistance to death receptor—induced apoptosis through mutational inactivation of the proapoptotic Bcl-2 homolog Bax. Nat Med 8:274–281

Legewie S, Bluthgen N, Herzel H (2006) Mathematical modeling identifies inhibitors of apoptosis as mediators of positive feedback and bistability. PLoS Comput Biol 2:e120

Lev Bar-Or R, Maya R, Segel LA, Alon U, Levine AJ, Oren M (2000) Generation of oscillations by the p53-Mdm2 feedback loop: a theoretical and experimental study. Proc Natl Acad Sci U S A 97:11250–11255

Li C, Courtot M, Le Novère N, Laibe C (2010) BioModels.net Web Services, a free and integrated toolkit for computational modelling software. Brief Bioinform 11:270–277

Liu JR, Opipari AW, Tan L, Jiang Y, Zhang Y, Tang H, Nunez G (2002) Dysfunctional apoptosome activation in ovarian cancer: implications for chemoresistance. Cancer Res 62:924–931

Luo J, Solimini NL, Elledge SJ (2009) Principles of cancer therapy: oncogene and non-oncogene addiction. Cell 136:823–837

Ma L, Wagner J, Rice JJ, Hu W, Levine AJ, Stolovitzky GA (2005) A plausible model for the digital response of p53 to DNA damage. Proc Natl Acad Sci U S A 102:14266–14271

Marcus FB (2008) Bioinformatics and systems biology: collaborative research and resources. Springer, Berlin

Mathas S, Lietz A, Anagnostopoulos I, Hummel F, Wiesner B, Janz M, Jundt F, Hirsch B, Johrens-Leder K, Vornlocher HP et al (2004) c-FLIP mediates resistance of Hodgkin/Reed-Sternberg cells to death receptor-induced apoptosis. J Exp Med 199:1041–1052

McConkey DJ, Orrenius S (1997) The role of calcium in the regulation of apoptosis. Biochem Biophys Res Commun 239:357–366

Miquel C, Borrini F, Grandjouan S, Auperin A, Viguier J, Velasco V, Duvillard P, Praz F, Sabourin JC (2005) Role of bax mutations in apoptosis in colorectal cancers with microsatellite instability. Am J Clin Pathol 123:562–570

Muschen M, Rajewsky K, Kronke M, Kuppers R (2002) The origin of CD95-gene mutations in B-cell lymphoma. Trends Immunol 23:75–80

Nakabayashi J, Sasaki A (2006) A mathematical model for apoptosome assembly: the optimal cytochrome c/Apaf-1 ratio. J Theor Biol 242:280–287

Nicotera P, Leist M, Ferrando-May E (1998) Intracellular ATP, a switch in the decision between apoptosis and necrosis. Toxicol Lett 102–103:139–142

Orrenius S, Zhivotovsky B, Nicotera P (2003) Regulation of cell death: the calcium-apoptosis link. Nat Rev Mol Cell Biol 4:552–565

O'Connor KC, Muhitch JW, Lacks DJ, Al-Rubeai M (2006) Modeling suppression of cell death by Bcl-2 over-expression in myeloma NS0 6A1 cells. Biotechnol Lett 28:1919–1924

Peltenburg LT, Bruin EC de, Meersma D, Smit NP, Schrier PI, Medema JP (2005) Expression and function of the apoptosis effector Apaf-1 in melanoma. Cell Death Differ 12:678–679

Philippi N, Walter D, Schlatter R, Ferreira K, Ederer M, Sawodny O, Timmer J, Borner C, Dandekar T (2009) Modeling system states in liver cells: survival, apoptosis and their modifications in response to viral infection. BMC Syst Biol 3:97

Qu X, Yu J, Bhagat G, Furuya N, Hibshoosh H, Troxel A, Rosen J, Eskelinen EL, Mizushima N, Ohsumi Y et al (2003) Promotion of tumorigenesis by heterozygous disruption of the beclin 1 autophagy gene. J Clin Invest 112:1809–1820

Rehm M, Huber HJ, Hellwig CT, Anguissola S, Dussmann H, Prehn JH (2009) Dynamics of outer mitochondrial membrane permeabilization during apoptosis. Cell Death Differ 16:613–623

Scaffidi C, Schmitz I, Zha J, Korsmeyer SJ, Krammer PH, Peter ME (1999) Differential modulation of apoptosis sensitivity in CD95 type I and type II cells. J Biol Chem 274:22532–22538

Slee EA, O'Connor DJ, Lu X (2004) To die or not to die: how does p53 decide? Oncogene 23:2809–2818

Soengas MS, Capodieci P, Polsky D, Mora J, Esteller M, Opitz-Araya X, McCombie R, Herman JG, Gerald WL, Lazebnik YA et al (2001) Inactivation of the apoptosis effector Apaf-1 in malignant melanoma. Nature 409:207–211

Soung YH, Lee JW, Kim SY, Sung YJ, Park WS, Nam SW, Kim SH, Lee JY, Yoo NJ, Lee SH (2005) Caspase-8 gene is frequently inactivated by the frameshift somatic mutation 1225_1226delTG in hepatocellular carcinomas. Oncogene 24:141–147

Stucki JW, Simon HU (2005) Mathematical modeling of the regulation of caspase-3 activation and degradation. J Theor Biol 234:123–131

Takahashi Y, Coppola D, Matsushita N, Cualing HD, Sun M, Sato Y, Liang C, Jung JU, Cheng JQ, Mule JJ et al (2007) Bif-1 interacts with Beclin 1 through UVRAG and regulates autophagy and tumorigenesis. Nat Cell Biol 9:1142–1151

Thompson CB (1995) Apoptosis in the pathogenesis and treatment of disease. Science 267:1456–1462

Tournier L, Chaves M (2009) Uncovering operational interactions in genetic networks using asynchronous Boolean dynamics. J Theor Biol 260:196–209

Tyson JJ, Chen KC, Novak B (2003) Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. Curr Opin Cell Biol 15:221–231

Vaux DL, Silke J (2005) IAPs, RINGs and ubiquitylation. Nat Rev Mol Cell Biol 6:287–297

Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, Nakamura Y, White R, Smits AM, Bos JL (1988) Genetic alterations during colorectal-tumor development. N Engl J Med 319:525–532

Wagner J, Ma L, Rice JJ, Hu W, Levine AJ, Stolovitzky GA (2005) p53-Mdm2 loop controlled by a balance of its feedback strength and effective dampening using ATM and delayed feedback. Syst Biol (Stevenage) 152:109–118

Wang X (2001) The expanding role of mitochondria in apoptosis. Genes Dev 15:2922–2933

Zhang R, Shah MV, Yang J, Nyland SB, Liu X, Yun JK, Albert R, Loughran TP Jr (2008) Network model of survival signaling in large granular lymphocyte leukemia. Proc Natl Acad Sci U S A 105:16308–16313

Zinovyev A, Viara E, Calzone L, Barillot E (2008) BiNoM: a cytoscape plugin for manipulating and analyzing biological networks. Bioinformatics 24:876–877

Chapter 11
Systems Biology, Bioinformatics and Medicine Approaches to Cancer Progression Outcomes

Jan G. Hengstler, Mathias Gehrmann, Stefan Höhme, Dirk Drasdo, Joanna D. Stewart and Marcus Schmidt

Abstract Because of the complexity of carcinogenesis and tumour development, it is critical to understand the underlying organizing principles. In this chapter a possible approach is illustrated, starting with a description of breast cancer prognosis as a function of three powerful biological motifs derived from gene expression profiling. A proliferation metagene describing the transition from slow to fast proliferation leads to the most dramatic aggravation of prognosis. A second immune cell metagene represents an opponent of tumour evolution, whereby only fast-proliferating tumours that are not recognized and eliminated by immune cells can progress. In the absence of endocrine treatment, a third motif, the oestrogen receptor metagene, is of limited prognostic importance, although it is required to model the interactions between the proliferation and immune cell axes. The subsequent and critical step will be to model the gene array-derived biological motifs as a function of a manageable number of signalling pathways or signalling network constellations. The advantages of translating biological motifs into pathway signatures are the possibility of identification of pathologic pathways in individual tumours, and the perspective of selecting appropriate drugs. In recent years much progress has been achieved in the field of spatial-temporal tissue modelling. Spatial-temporal models simulating tissue regeneration or 3D tumour infiltration are available. However, the ambitious goal of fully linking spatial-temporal tissue development to single-cell decisions created by molecular models of signalling network constellations is yet to be achieved.

11.1 Introduction: The Concept of Pathway Signatures

It is well accepted that cancer is caused by a sequence of genetic and epigenetic alterations. The complexity of the disease is due not only to the relatively high number of genes that may be involved, but also to the variability of involved genes even between carcinomas of the same type, and the variability of the sequence of events during carcinogenesis and tumour progression. Because of the complexity of the various types and stages of carcinomas, cancer research would benefit from a systems biology approach to understand the critical organizing principles. Although the number and di-

J. G. Hengstler (✉)
Department of Toxicology, IfADo-Leibniz Research Centre for Working Environment and Human Factors, Technical University of Dortmund, Ardeystraβe 67, 44139 Dortmund, Germany
e-mail: hengstler@ifado.de

A. Cesario, F. B. Marcus (eds.), *Cancer Systems Biology, Bioinformatics and Medicine,*
DOI 10.1007/978-94-007-1567-7_11, © Springer Science+Business Media B.V. 2011

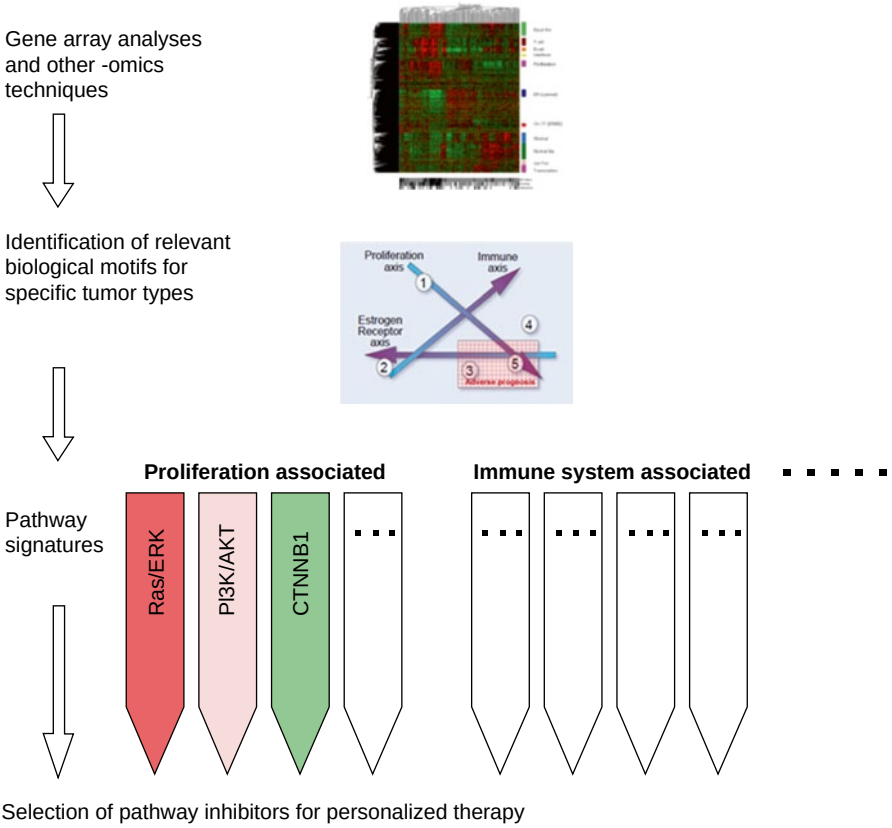


Fig. 11.1 Concept of biological motifs and pathway signatures for personalized cancer therapy

versity of involved proteins is large, a promising concept is emerging (Fig. 11.1). The phenotypic behaviour of tumour cells may be controlled by a manageable number of key pathways (Itadani et al. 2008; Kreeger and Laufenburger 2010). Pathologic states of key pathways can be recognised by '-omics' techniques, of which the most advanced is gene array analysis, where certain expression patterns ('pathway signatures') may indicate the activity of aberrant pathways (review: Itadani et al. 2008). A pathway signature describes a battery of genes whose expression correlates with activity of a specific signalling pathway, or with certain critical signalling network activity constellations. The concept of pathway signatures reduces complexity, because aberrant activity of a pathway may be caused by a large number of mutations or epigenetic events. For example, an aberrant pathway signature for ras-ERK may be caused by many alterations of membrane receptors RAS, RAF, MEK or ERK as well as numerous feedback loops. However, the specific molecular nature of these events is of minor relevance for the pathway signature concept. Nevertheless, a good pathway signature should identify the most upstream pathway activator. This is critical for therapeutic consequences. For example, the use of erbB2 antagonists would make sense if erbB2 amplification is the most upstream event, but not if Ras muta-

tions are the cause of over-activity of the Ras-ERK pathway. Therefore, the concept of pathway signatures aims at identifying the most upstream pathway activators in order to select therapeutic inhibitors of downstream factors (Fig. 11.1).

Furthermore, a pathway signature concept should consider the high flexibility of pathways that fuel tumour progression. This has become obvious from using transgenic mice that allow the expression of the specific oncogene that initiated carcinogenesis to be switched off (Hengstler et al. 2006). Initial remission was often followed by recurrent tumour growth, where, instead, progression of recurrent tumours was triggered by different oncogenes or pathways from those responsible for the onset of the primary tumours. For that reason, it would be an important step to be able to predict the most frequent escape mechanisms to specific pathway inhibitors.

At present we remain some way off from introducing a comprehensive therapeutic concept based on pathway inhibitors into clinical practice. In this chapter we will discuss the state of the art and major obstacles. As well we will focus on current possibilities of describing and understanding carcinomas by genome-wide gene expression patterns.

11.2 Identification of Biological Motifs from Gene Array Data

11.2.1 Gene Expression Profiling

It is well known that microarrays allow measurement of thousands of mRNA species in a single analysis. New versions of microarrays, exon chips, have been designed in such a way that each exon of a gene is represented by several oligos, thereby also allowing analysis of differential splicing. Since breast cancer represents a particularly advanced field of research involving gene array analyses, we will present this tumour type as an example.

Gene expression profiling has demonstrated that breast cancer is not a single disease, but can be categorized into at least four main molecular classes (review: Sotiriou and Pusztai 2009): (i) ‘Basal-like’ carcinomas, which mostly correspond to Her2, oestrogen negative as well progesterone receptor negative tumours; (ii) ‘luminal A’ carcinomas expressing cytokeratins of luminal epithelial cells of normal breast tissue, mostly well differentiated oestrogen receptor positive tumours; (iii) ‘luminal B’ carcinomas, which differ from the luminal A type mostly by higher histological grading; and (iv) ‘Her-2 positive’ carcinomas, which show high expression of her-2 and her-2 dependent genes. The establishment of these subgroups from gene array data represented substantial progress; they correspond well to different cells of origin, and are also relevant for the choice of therapy.

Although the four molecular classes described above, also named the ‘intrinsic classification’ (Perou et al. 2000), represent an important development, categorization alone does not explain breast cancer as a process of dysregulated biological functions. In order to develop an understanding of the interrelated biological processes of this disease, we performed genome-wide gene expression profiling of 200 tumours

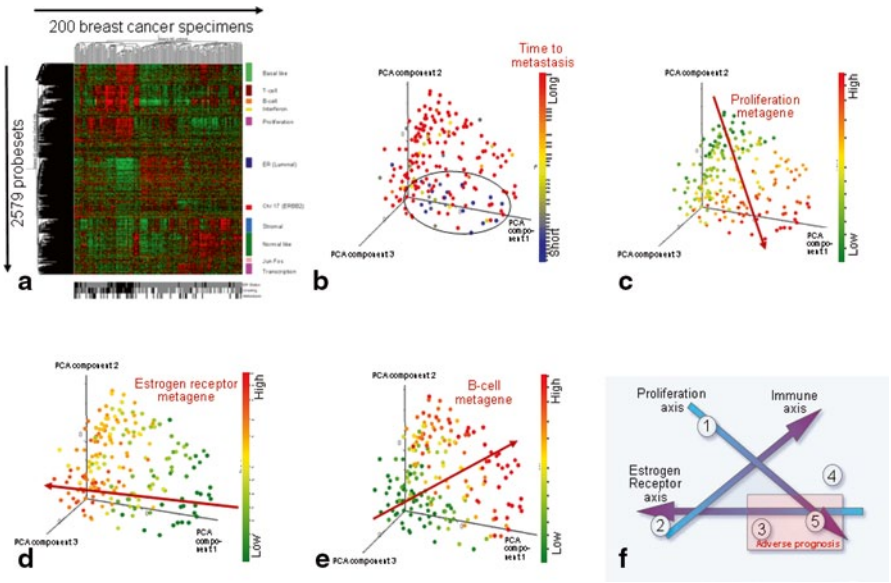


Fig. 11.2 Dominant biological motifs in relation to prognosis correlation in node-negative breast cancer. **a** Hierarchical clustering of gene array data helps to identify biological motifs representing either processes (such as proliferation) or cell types (such as B- or T-cell infiltration). **b** Principal component analysis, based on gene expression patterns, visualizes the relationship between prognosis and gene expression. Each spot represents an individual patient. Patients with bad prognosis (short time until metastasis) cluster to the lower right. **c** Proliferation. **d** Estrogen. **e** B-cell metagenes in relation to prognosis. **f** Overview over the dominant biological motifs abstracted from principal component analysis of whole genome gene expression data of 788 patients with node-negative breast cancer (Schmidt et al. 2008, 2009). High (*purple*) and low (*blue*) expression of genes belonging to their respective axes. The regions indicated by numbers correspond to specific histological types or intrinsic subtypes by Perou's definition (2000): *1* Small tumours, tubular and lobular histology, grade 1 and 2, expression of basal like marker, normal-like subtype; *2* Ductal and lobular histology, highest expression of ER, absence of immune-related transcripts, elderly patients, luminal A subtype; *3* Ductal histology, ER positive and negative, presence and absence of ERBB2 amplification, variable expression of immune-related transcripts, luminal B and ER negative but non-basal-like and ERBB2 subtype; *4* Ductal and medullary histology, grade 2 and 3, ERBB2 negative, high expression of basal-like markers, younger patients, high expression of immune-related transcripts, basal-like A (advantageous prognosis); *5* Ductal histology, grade 2 and 3, ERBB2 negative, high expression of basal-like markers, younger patients, low expression of immune-related transcripts, basal-like B (bad prognosis) (Schmidt et al. 2008, 2009)

derived from women with node-negative breast cancer (Schmidt et al. 2008). To be able to establish a picture of the natural history of breast cancer; we selected patients who had not received systemic chemotherapy. To identify co-regulated genes representing specific biological processes or cell types, we performed an unsupervised hierarchical cluster analysis (Fig. 11.2a). This discovery-driven approach groups the genes and tumours according to overall similarity in relative gene expression, thereby visualizing dominant clusters of co-regulated genes. Analysis of the functions of the clustered genes indicated either a common biological process, or a cell type-specific

origin. These clusters, some of which have already been described by other groups (Sotiriou and Pusztai 2009; van't Veer et al. 2002) can be summarized as follows:

- Biological processes: (i) proliferation, (ii) oestrogen receptor dependent, (iii) ERBB2 dependent, (iv) interferon dependent, (v) jun-fos dependent, (vi) further transcription related genes
- Cell type-specific: (i) T-cell, (ii) B-cell, (iii) stromal, (iv) normal-like, (v) basal-like.

11.2.2 Metagenes for Clusters of Co-regulated Genes

In order to obtain an overview over the interrelations of these biological motifs, we calculated 'metagenes' for each cluster of co-regulated genes. A 'metagene' represents the normalized median expression of genes belonging to a specific cluster. Principal component analysis visualization enables the complex relation to be grasped between metagenes as representatives for biological motifs, and prognosis (Fig. 11.2b). Each spot in Fig. 11.2b represents an individual patient. We used PC1 to PC3 which retains the largest possible variation that can be displayed in three dimensions (28% of total variance). The colour code in Fig. 11.2b visualizes time to metastasis, where dark blue represents patients with extremely short time to metastasis. Clearly, patients with worse prognosis cluster on the bottom right (Fig. 11.2b). In Fig. 11.2c, the same patients are represented, with the only difference that the colour code represents the expression level of the proliferation metagene, with red representing high expression. Importantly, the proliferation metagene seems to be oriented exactly in the direction of the cluster of patients with worse prognosis. Almost no tumours from patients who developed a metastasis were observed in the region of low proliferation metagene expression. In a similar way to the proliferation metagene (Fig. 11.2c) we included two further biological motifs, which in systematic analysis have been shown to be most influential, namely the oestrogen receptor (Fig. 11.2d) and an immune cell, the so-called B-Cell metagene (Fig. 11.2e). This led to a global picture of the natural history of node-negative breast cancer (Fig. 11.2f—Schmidt et al. 2009). The transition from slow to fast proliferation leads to the most dramatic aggravation of prognosis. The orientation of the B-cell metagene (and similarly of T-cell metagene, not shown in Fig. 11.2) is almost perpendicular to the proliferation axis, indicating mutually independent information. In a region where the immune cell metagenes are highly expressed, metastasis occurred rarely, despite high expression of the proliferation metagene. This hypothesis was confirmed by conventional survival analysis in our cohort (Fig. 11.3a) and was validated in two independent cohorts of node-negative breast cancer (Fig. 11.3b, c), as well as in multivariate analyses adjusted for the oestrogen receptor status, tumour stage and grade, and the T-cell receptor. When we classified tumours as fast-proliferating or slow-proliferating, the B-cell metagene was associated with good prognosis independent of the oestrogen receptor status (Schmidt et al. 2009). This observation suggests that the well-known

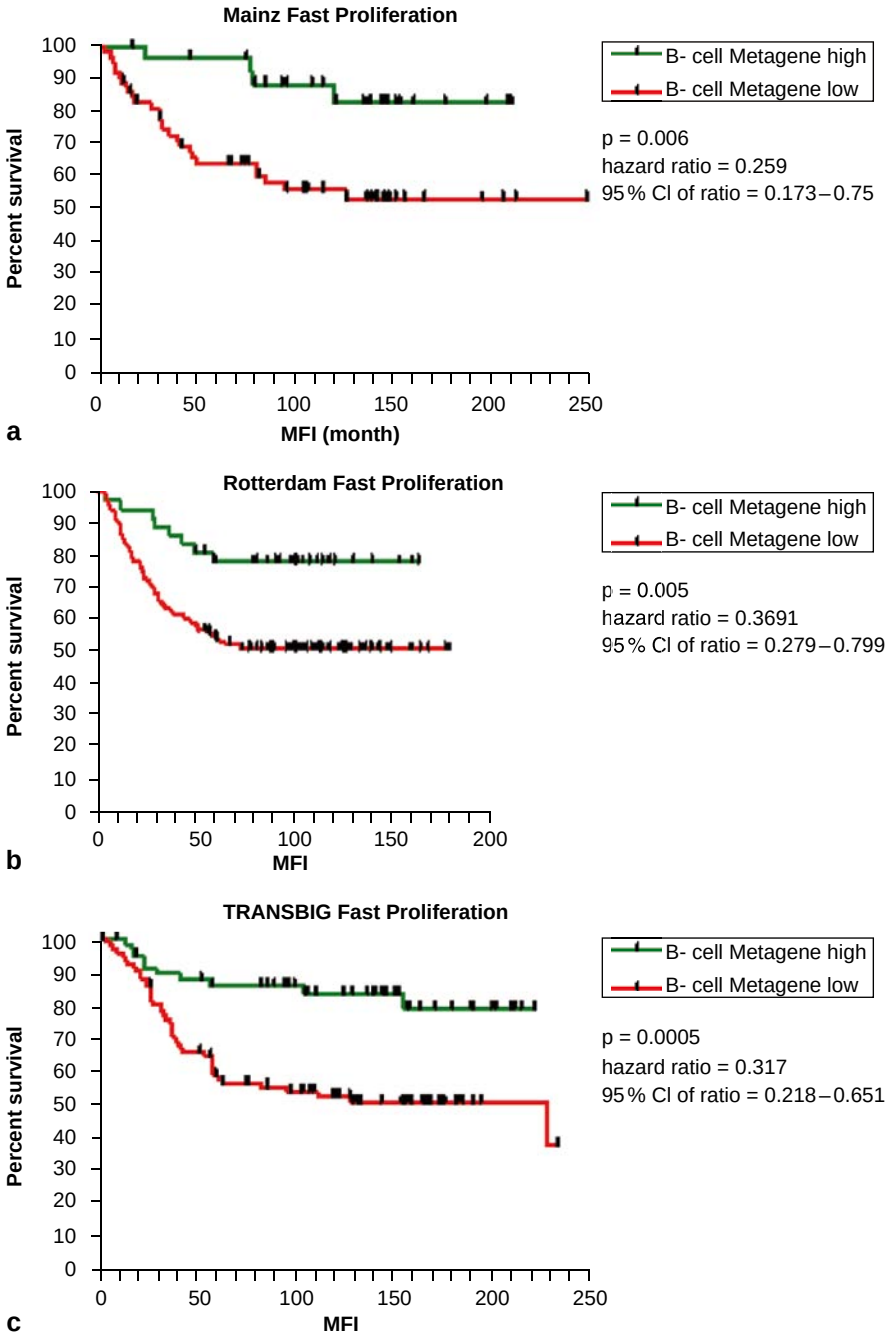


Fig. 11.3 The B-cell metagene is associated with better prognosis in fast proliferating node-negative breast carcinomas. The result could be validated in three independent cohorts of node-negative breast cancer patients, namely the Mainz (**a**), the Rotterdam (**b**) and the Transbig (**c**) cohorts (Schmidt et al. 2008)

adverse prognostic effect is attenuated by the immune system, regardless of oestrogen receptor expression. It is plausible that the immune system represents an opponent of tumour evolution fuelled by the module of proliferation-associated genes. Only fast-proliferating tumours that are not recognized and eliminated by immune cells can progress, form metastases and finally lead to patient death. Therefore, the immune system, represented by the B-cell metagene in Fig. 11.2f, is one of the major players responsible for the selection aspect in tumour evolution.

In the absence of endocrine treatment, steroid hormone receptor expression visualized as the oestrogen receptor metagene in Fig. 11.2f is of limited prognostic significance. Nevertheless, consideration of this third axis is important for orientation, and may avoid misunderstanding. For instance, almost all oestrogen receptor-negative breast carcinomas are characterized by high proliferation (Fig. 11.2e, f). Some researchers have suggested that some prognostic effects in breast cancer, such as the effect of the immune system, depend on the oestrogen receptor status. This should carefully be re-evaluated, as this relation may be the result of the correlation between oestrogen receptor status and proliferation, with proliferation representing the real or independent factor of influence.

In conclusion, the three 'coordinates in the universe of breast cancer', proliferation, oestrogen receptor, and the immune system, facilitate orientation and help to correctly interpret breast cancer biology. Using this system it was also possible to include the four molecular classes of breast cancer according to the 'intrinsic subtype classification' of Perou et al. (2000) (Fig. 11.2f). Considering just three biological motifs can create a large amount of confusion, and in view of the fact that chemotherapy will add an additional layer of complexity, the hunt for a deeper understanding of more complex interrelations has only just begun!

11.3 From Biological Motifs to Pathway Activation

Identification of biological motifs unravelling the interrelations and associations with tumour development already represents a complex matter. However, this is just the starting point, the point at which systems biology comes in. In the previous section we described the proliferation metagene as one of the most relevant biological motifs for tumour progression. However, expression of the numerous genes of the proliferation metagene is controlled by a number of activating or inactivating signalling pathways, such as ras-/ERK, PI3K/Akt, Smad, Stat, p38, JNK, P53 and RB. Ideally, gene expression patterns should be translated into signalling pathway activities. In a primitive way this could be expressed, for instance, as ras/ERK-high, PI3K/AKT-high, p38-low, etc. In a more advanced system, mathematical terms could be used to express signalling network activity constellations. One of the advantages of translating gene expression patterns into signalling statuses is that causative pathways may be identified allowing selection of adequate drugs (Fig. 11.1). For many reasons, translating biological motifs derived from gene ex-

pression patterns into signalling statuses is difficult. The expression of most genes is affected by more than one pathway. In addition, pathways are influenced by both positive and negative feedback loops, further complicating the situation. The associations between specific pathway activities and the respective gene expression signatures are usually identified by comparing samples in which a specific pathway is activated or inactivated (review: Itadani et al. 2008). Frequently applied strategies are comparison of wild-type to knockout or transgenic mice, transgene expression or siRNA silencing in cell lines and chemical perturbations. Although the field of pathway signatures is just emerging, the results obtained so far are encouraging. Some studies suggest that pathway signatures may predict the activation status of signalling pathways more accurately than single gene mutations, and show a better correlation with patient prognosis (review: Itadani et al. 2008). One example is an EGFR mutation status signature (Choi et al. 2007). Through comparison of non-small-cell lung carcinoma (NSCLC) cell lines with known mutation status, gene sets were identified that are associated with constitutive activation of the receptor. Using this EGFR mutation status signature derived from cell lines, the EGFR mutation status of NSCLC patients could be correctly classified using the patient's expression profiles. A second example is a P53 mutation status signature for breast cancer (Miller et al. 2005). A 32 gene signature established by comparing P53 mutant and wild-type carcinomas was superior to DNA sequence-based mutation status in predicting prognosis.

Although these and many more examples represent an important development, they nevertheless consider only a single pathway. To the best of our knowledge, a comprehensive analysis of carcinomas considering signatures of all, or at least the most relevant, pathways has not yet been achieved, and remains one of the major challenges. One of many obstacles in this field of research is the commonly large discrepancy in gene expression patterns in cell lines needed for specific manipulations of pathways and human carcinomas.

11.4 How Realistic is Modelling of Carcinogenesis and Tumour Development in Virtual Tissues and Organs?

11.4.1 Spatial-temporal Models of Tumours

One of the major challenges in systems biology is modelling of virtual tissues. The most ambitious version of such artificial tissues is a multiscale model in which models of intracellular pathways create single-cell fate decisions, such as proliferation, cell death, migration, etc. Single-cell fate decisions further influence cell-cell interactions, tissue architecture, and organ functions, which can all be simulated by an integrated multiscale model. Although we are still far away from this ambitious goal, some important initial steps have been achieved. A precondition for realistic spatial-temporal modelling of carcinogenesis and tumour development is a model

of the healthy tissue of origin. Relatively little is known about how cells coordinately behave to establish functional tissue structure and restore micro-architecture during regeneration; these are processes that are often disturbed during malignant transformation. Research in this field is hampered by a lack of techniques that allow quantification of tissue architecture.

To bridge this gap, we have established a procedure based on confocal laser scans, image processing and three-dimensional tissue reconstruction, as well as quantitative mathematical modelling (Hoehme et al. 2010). The concrete liver lobule reconstructed from confocal laser scans of mouse livers (Fig. 11.4) is still abstract, as hepatocytes are represented by spheres and only the two most frequent cell types of the liver (hepatocytes and sinusoidal endothelial cells) have been considered. However, this reconstructed tissue based on static architectural parameters can be used as a starting point for mathematical modelling. The smallest unit of the model is the individual cell whose behaviour is controlled by so-called process parameters. Experimentally determined process parameters are, for example, the probability of cell death or cell proliferation at a given time and position of the liver lobule. Mathematical modelling parameters are, for instance, ordering principles such as cell alignment along micro-vessels, or cell migration along an oxygen gradient. Using this technique, we can demonstrate that alignment of hepatocyte daughter cells along microvessels (the liver sinusoids) represents a previously unrecognized mechanism essential for maintenance of liver micro-architecture (videos of simulations; websites: CancerSys and SysTox; Hoehme et al. 2010). Based on this spatial temporal model we can simulate tumour micro-architecture during early stages of hepatocellular carcinogenesis (Fig. 11.4g).

11.4.2 Tumour Modelling Perspectives

Spatial-temporal models of tumours so far have described tumour morphology and distribution of necrotic cells (Ferreira et al. 2002), tumour growth in relation to oxygen profiles and angiogenesis (Macklin et al. 2009), local gradients of angiogenic factors (Qutub and Popel 2009) and responses to micro-environmental changes, such as cell density and oxygen tension (Anderson et al. 2006). Within the next 5–10 years, spatial-temporal models are expected to become available that correctly simulate tumour tissue architecture and infiltration of tumour cells into normal tissue. However, a current limitation of these models is that they require experimentally determined process parameters as an input. This means that knowledge of the detailed behaviour of cells, such as proliferation, cell death, migration, destruction of normal cells by proteases, etc., is a precondition for modelling. It has not yet been possible to convincingly explain single-cell behaviour and decisions in tissue by molecular-level networks. This is due to the complexity of mechanisms responsible for cell fate decisions and also to problems of *in vitro* to *in vivo* extrapolation when mechanisms are studied in cell lines. Another critical problem is that relatively little

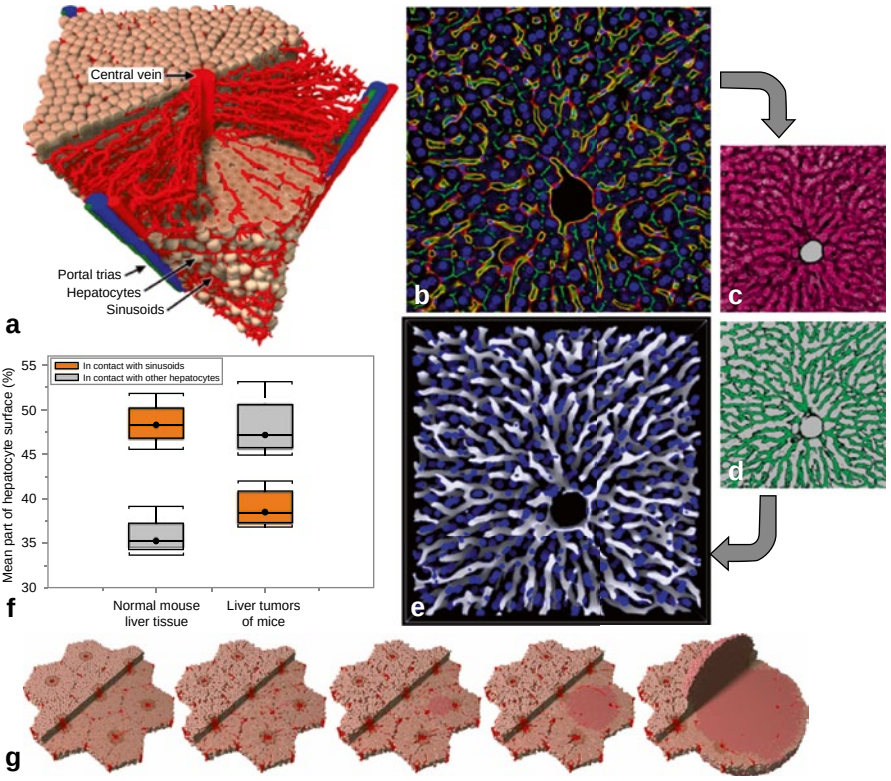


Fig. 11.4 **a** Liver lobule reconstructed from experimental data by the image processing chain B-E and successive image analysis. Such reconstructed abstracted tissues serve as initial state for spatial-temporal mathematical modelling. **b** Typical image obtained by confocal microscopy after adaptive histogram equalization filtering. *Blue* DAPI (hepatocyte nuclei), *yellow* ICAM+DPPIV (sinusoids), *red* ICAM, *green* DPPIV. **c** Effect of generalized erosion filtering (all red pixels are removed). **d** Effect of generalized dilatation filtering (all green pixels are added). **e** Result of image processing chain in 3D. *Blue* Hepatocyte nuclei, *white* sinusoids. Note the complex architecture that links the periportal zone to the central vein in the middle of the lobule. **f** Fraction of surface area of hepatocytes in contact with sinusoids (*orange*) and other hepatocytes (*grey*) in normal liver tissue and hepatocellular carcinomas of mice (from: Hoehme et al. 2010). **g** Spatial-temporal model visualizing the early stages of a liver carcinogenesis. The model recapitulates growth of a poorly differentiated hepatocellular carcinoma that occurs in mice within approximately 6 months after Cre recombinase induced APC knockout

is known about the sequence and time course of accumulation of somatic mutations and epigenetic alterations during carcinogenesis and tumour development. It can probably be expected that these questions will be solved by next-generation sequencing and genome wide determination of methylation patterns within the next 5–10 years. These are important preconditions for the generation of virtual models capable of allowing intracellular events to induce single-cell decisions which then translate to tissue and organ level.

References

Anderson AR, Weaver AM, Cummings PT et al (2006) Tumor morphology and phenotypic evolution driven by selective pressure from the microenvironment. Cell 127:905–915

Choi K, Creighton CJ, Stivers D et al (2007) Transcriptional profiling of non-small cell lung cancer cells with activating EGFR somatic mutations. PLoS One 2(11):e1226

Ferreira SC Jr, Martins ML, Vilela MJ (2002) Reaction-diffusion model for the growth of avascular tumor. Phys Rev E Stat Nonlin Soft Matter Phys 65:021907

Hengstler JG, Bockamp EO, Hermes M et al (2006) Oncogene-blocking therapies: new insights from conditional mouse tumor models. Curr Cancer Drug Targets 6:603–612

Hoehme S, Brulport M, Bauer A et al (2010) Cell alignment along micro-vessels as order principle to restore tissue architecture during liver regeneration: from experiment to virtual tissues and back. Proc Natl Aad Sci U S A 107(23):10371–10376

Itadani H, Mizuarai S, Kotani H (2008) Can systems biology understand pathway activation? Gene expression signatures as surrogate markers for understanding the complexity of pathway activation. Curr Genomics 9:349–360

Kreeger PK, Lauffenburger DA (2010) Cancer systems biology: a network modeling perspective. Carcinogenesis 31:2–8

Macklin P, McDougall S, Anderson AR et al (2009) Multiscale modelling and nonlinear simulation of vascular tumour growth. J Math Biol 58:765–798

Miller LD, Smeds J, George J et al (2005) An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. Proc Natl Acad Sci U S A 102:13550–13555

Perou CM, Sørlie T, Eisen MB et al (2000) Molecular portraits of human breast tumours. Nature 406(6797):747–752

Qutub AA, Popel AS (2009) Elongation, proliferation & migration differentiate endothelial cell phenotypes and determine capillary sprouting. BMC Syst Biol 3:13

Schmidt M, Böhm D, von Törne C et al (2008) The humoral immune system has a key prognostic impact in node-negative breast cancer. Cancer Res 68:5405–5413

Schmidt M, Hengstler JG, von Törne C et al (2009) Coordinates in the universe of node-negative breast cancer revisited. Cancer Res 69:2695–2698

Sotiriou C, Pusztai L (2009) Gene-expression signatures in breast cancer. N Engl J Med 360:790–800

van't Veer LJ, Dai H, van de Vijver MJ et al (2002) Gene expression profiling predicts clinical outcome of breast cancer. Nature 415(6871):530–536

Websites

CancerSys: http://www.ifado.de/cancersys/index.html

SysTox: http://www.ifado.de/forschung_praxis/projektgruppen/susceptibility/forschung/link2/index.html

Chapter 12
System Dynamics at the Physiological and Tumour Level

Robert A. Gatenby

Abstract Cancers are complex dynamical systems dominated by non-linear processes. As a result, most critical system parameters exhibit significant temporal and spatial heterogeneity. This variability, while critical to the ability of cancers to adapt to a wide range of environmental perturbations including therapy, tends to be lost in molecular-level data which is typically an 'average' value for large numbers of heterogeneous tumour cells obtained at a single time point. The role of mathematical modelling in cancers at a tumour level is to identify the first principles that govern tumour growth, invasion, metastases, and response to therapy. Tumour biologists and oncologists often dismiss quantitative methods with the statement that cancer is 'too complex' for mathematical modelling. In fact, lessons from the history of physical sciences demonstrate that the opposite is true. While complex systems may be difficult to model, they are impossible to understand intuitively. Biologically realistic mathematical models are necessary to transform the reductionist approach of modern cancer biology into comprehensive models of the host-cancer interactions that govern the dynamics of tumour growth and therapy.

12.1 Introduction to Mathematical Modelling in Cancer

12.1.1 Lessons from History

In the earlier chapters of this book, justifications have been made concerning the need for systems approaches and modelling at the cellular and subcellular level. However, when moving to the physiological and tumour level as discussed in this chapter, it might be argued that the systems are too complex for such approaches. In fact, the contrary is true, as supported by both a historical analysis of modelling and a detailed consideration of the research questions to be resolved.

R. A. Gatenby (✉)
Departments of Radiology and Integrative Mathematical Oncology, Moffitt Cancer Center, Tampa, USA
e-mail: robert.gatenby@moffitt.org

A. Cesario, F. B. Marcus (eds.), *Cancer Systems Biology, Bioinformatics and Medicine,*
DOI 10.1007/978-94-007-1567-7_12,

In the latter half of the sixteenth century, Tycho Brahe laboriously documented the movement of the known planets in the solar system. His observations were remarkably accurate, particularly since the telescope had not yet been invented, and the performance of earlier instruments by Girolamo Rucellai and Gianbattista Della Porta were far from being optimal. The movements of the planets, however, were difficult to understand, as they often stopped in their normal orbital motion, and even seemed to briefly move backwards as they traversed the night sky. Astronomers of the era had two theoretical views of the universe: the ancient Ptolemaic model in which the earth stood at the centre of the universe and planets revolved around it, and the new Copernican theory in which the sun was at the centre and the planets revolved around it in perfectly spherical orbits (Thoren 1990).

Brahe recognized that the stationary earth model did not fit his data. But, since theological considerations 'required' a stationary earth, he proposed a compromise solution in which the planets orbited the sun and, in turn, revolved around the earth at a central point. However, Brahe also recognized he was in need of a mathematical assistant and, in 1599, hired Johannes Kepler. For 17 years, Kepler studied this vast data set to find patterns and connections. His arithmetical calculations involving just the orbit of Mars filled nearly 1000 sheets of paper (Tiner 1977). The remarkable intensity of the effort is apparent in his description of the moment of insight:

> …and if you want the exact moment in time, it was conceived mentally on 8th March in this year 1618, but submitted to calculation in an unlucky way, and therefore rejected as false, and finally returning on the 15th of May and adopting a new line of attack, stormed the darkness of my mind. So strong was the support from the combination of my labour of 17 years on the observations of Brahe and the present study, which conspired together, that at first I believed I was dreaming, and assuming my conclusion among my basic premises. But it is absolutely certain and exact that the proportion between the periodic times of any two planets is precisely the sesquialternate proportion of their mean distances …. (Kepler 1619)

About 70 years later, Isaac Newton recognized the existence of gravity: '… all matter attracts all other matter with a force proportional to the product of their masses and inversely proportional to the square of the distance between them' (Newton 1846). Through the new mathematics of calculus, Newton derived Kepler's laws from interaction of gravitational and centripetal forces. The model of planetary motion derived from these fundamental principles both reproduced extant data and provided a deep understanding of the forces that govern the system. Furthermore, the modelling results led to predictions, as perturbations in the orbit of Uranus led to the conclusion that another massive object must be nearby. This, in turn, motivated new directed observations that led to the discovery of Neptune.

This experience is not unique. The Balmer lines in the early twentieth century and the unruly 'zoo' of subatomic particles in the 1950s, represented complex experimental observations that defied explanation. In each case, new conceptual frameworks such as quantum mechanics and the 'standard model' brought order to complicated data sets.

12.1.2 Extension to Bioinformatics and Systems Biology

In many ways, Kepler and Newton demonstrate the differences between bioinformatics and mathematical modelling. The former analysed extant data using the most sophisticated computing tools available (his brain plus pen and paper) to find patterns and connections. The latter approached the problem of planetary motion by identifying first principles, modelling the system dynamics, and then demonstrating that these results matched experimental methods. Since the time of Newton, the physical sciences have thrived on a research paradigm that deeply integrates mathematical modelling and empirical data. Richard Feynman, a Nobel Prize winner in physics, expressed the relationship as follows: 'Mathematics is a deep way of describing nature, and any attempt to express nature in philosophical principles, or in seat-of-the-pants mechanical feelings, is not an efficient way' (Feynman 1965).

Despite centuries of successful experience in use of mathematical models to understand complex systems in the physical sciences, this paradigm has been extended into cancer research only recently.

12.2 Mathematical Models in Cancer

12.2.1 The Role of Modelling in Cancer Research

The societal burden of cancer has stimulated decades of intense scientific effort that has resulted in many important new insights and therapies. Yet, despite these advances, the improvement in mortality rates for cancer patients still lags behind that of the other major causes of death such as cardiovascular and cerebrovascular diseases (1950 Mortality Data 2001). Research in cancer biology has been greatly accelerated by new experimental technologies and the revolution in genomics and bioinformatics that have generated overwhelming amounts of bio-molecular data. Lacking, however, are the conceptual frameworks necessary to organize these bodies of data in ways that promote more significant advances in understanding of the disease (Gatenby and Maini 2002, 2003). This state of affairs clearly suggests the need for interdisciplinary research that synthesizes experimental results with mathematical analysis and modelling, to provide new insights into the underlying dynamics governing the disease, and to help organize new experimental and treatment strategies.

Such an idea is not new. As noted above, since the days of Newton, the natural philosopher has used the tools of mathematics to quantify, with consistent success, the physical world around us—fully justifying the statement by Galileo Galilei that 'Philosophy [i.e. physics] is written in this grand book—I mean the universe—which stands continually open to our gaze, but it cannot be understood unless one first learns to comprehend the language and interpret the characters in which it is

written. It is written in the language of mathematics, and its characters are triangles, circles, and other geometrical figures, without which it is humanly impossible to understand a single word of it; without these, one is wandering around in a dark labyrinth' (Galilei 1623). In contrast, the biological world, perhaps by virtue of its remarkable diversity, has been dominated by a tradition of observation, description and classification. The potential role of mathematics in biological research has long been acknowledged. D'Arcy Thompson's monumental treatise 'On Growth and Form' (Thompson 1942) opens with a number of quotations that includes one from Karl Pearson: 'I believe the day must come when the biologist will—without being a mathematician—not hesitate to use mathematical analysis when he requires it'. Over 100 years later an article in 'The Economist' stated 'If cancer is ever to be understood properly, mathematical models such as these will surely play a prominent role' (The Economist 2004).

However, for a scientific community steeped in the Aristotelian culture of empiricism, the introduction of mathematical methods is immensely challenging. Indeed, despite occasional bursts of enthusiasm, the role of mathematical and physical reasoning in the life sciences remains relatively limited.

The explosion of data generated by molecular biology has necessitated a widespread interest in the set of complex data mining tools and techniques generally described as bioinformatics. However, there remains little utilization of mathematical modelling in tumour biology and oncology to frame hypotheses, provide contextual frameworks for organizing data, and generate testable predictions. Modelling of this type in biological systems is very challenging, and the genuine successes of mathematics in biology, such as the Hodgkin-Huxley model in neurobiology and knot theory in DNA conformations, are relatively rare. In part, this reflects the daunting intellectual demands of working, either collaboratively or individually, in such profoundly different disciplines. Biological and clinical investigators typically have little or no training in the applied mathematics necessary to write and analyse mathematical models. Similarly, applied mathematicians usually have little background in the complex, multiscalar (molecular, cellular, tissue, and populations) dynamics in the life sciences. As a result their models are often of little relevance or interest to real biological or clinical problems.

Nevertheless, a Nature article by Gatenby and Maini (2002) suggested that future advances in cancer biology and treatment would require development of a field of study they termed 'mathematical oncology'. It has been pointed out that cancer is a complex multiscalar disease dominated by non-linear dynamics. Such systems, while difficult to model using a wide range of mathematical methods, are impossible to understand through the intuitive 'seat-of-the-pants' approach currently employed by virtually the entire field of tumour biology and oncology. The above comment accordingly concludes: 'In the absence of consistent application of rigorous mathematical models, theoretical medicine will largely remain empirical, phenomenological and anecdotal, successful only in linear systems that can be defined by a single experiment or a few experiments.'

Achievement of an integrative cancer research paradigm combining modelling and empirical research, such as that of the physical sciences, will need to overcome

many historical, philosophical and methodological barriers. The core component of cancer research must always remain bio-molecular research including *in-vitro* and *in-vivo* laboratory and clinical observations. Clearly, mathematical models without data are useless. At the same time, tumour biologists must recognize that data is not science. Hypothesis-driven, biologically-informed mathematical models are necessary to provide theoretical frameworks to organize and understand data and to guide new experiments.

12.2.2 Aspects of Cancer Modelling

Cancer is a complex disease, but it cannot be hopelessly so. Fundamentally, cancer must be governed by natural laws, and like any system in nature, can be understood through first principles. Critical to development of realistic mathematical models is integration with the statistical methods used to analyse the torrent of data generated by modern molecular methods—an approach described as 'integrative mathematical oncology' (Anderson and Quaranta 2008). Although, the dialogue between bioinformatics and bio-molecular experimentation is often systematic, it is not always strategic. For one thing, these methodologies often provide only limited insight into tumour dynamics. Molecular biology typically requires tissue removal and homogenization, with the result that it generates large amounts of 'average data', but limited information on spatial and temporal heterogeneity—critical properties in understanding the dynamics of cancer progression and treatment. Because of the continuous interactions of evolving phenotypes, and the chaotic micro-environment of the biological processes at the molecular, cellular, and tissue levels in cancer, there is a need for appropriate, dynamical mathematical models capable of transforming extensive, and occasionally haphazard, experimental programmes into an integrated conceptual approach (Maini and Gatenby 2006).

Thus, it seems reasonable to propose that a crucial missing methodological component in cancer research is mathematical modelling (Gatenby and Maini 2002, 2003), which can provide the informatics data with a conceptual framework and predictive value. The process of formulating a model invites the development and incorporation of first principles; clarifies assumptions; demands rigorous statement of hypotheses; and identifies key variables and parameters. Analysing the failure of a model can often be as valuable as developing a successful one. By virtue of its predictive power, a good model can help plan experiments by identifying parameter regimes of interesting behaviour—regimes that might otherwise be time-consuming and costly to discover by systematic experimentation. Furthermore, a model can also be used to estimate important parameters by data fitting. In these ways mathematical modelling completes the circle of discovery: experiments provide data that, in turn, informs the construction of new experimental designs.

12.3 Model Development

12.3.1 Historical Perspective—Understanding Tumour as a Complex System

A number of investigators in the past century have advocated the application of quantitative methods to biological investigation. Robert Fisher, for example, used statistics theory to demonstrate that the pioneering genetic studies by Mendell represented a 'selective' use of the data. Schrodinger in his lecture, which he subsequently published as a short book entitled 'What is life?' (Schrödinger 1944), applied physical science methods and (most importantly) thought processes to understanding cell biology. Armitage and Doll (1954) were probably the first to apply mathematical methods to data in cancer, and their results estimating the number of cellular events or stages necessary for the transition from normal to cancer remain influential to this day.

While appreciation of cancer as a genetic disease has a long history, the role of microenvironmental factors on the formation of a tumour 'system' has often been poorly understood. Krogh (1924) constructed a diffusion-reaction model of substrate and metabolite exchange between blood vessels and surrounding cells. A key observation from this work was that the diffusion and consumption of oxygen will result in insufficient concentrations to support life at a distance of 100–200 μm from a blood vessel. Observational studies on pathologic specimens by Tomlinson and Gray (1955) confirmed that Krogh cylinders (which they called tumour cords) can be commonly found. Thus, it became clear that regional hypoxia due to inadequate vascularization is common in cancers. Interestingly, observations over two decades later also demonstrated the importance of temporal variations in blood flow in producing 'acute' hypoxia. This reflects the chaotic nature of blood transit in tumours, so that even intratumoral blood vessels in patients can evince transient episodes of stagnant and even reversed flow.

The emerging picture is one of complexity. Cancer is a disease associated with genetic changes, but a tumour is far more than a ball of mutated cells. In fact, tumours consist of a heterogeneous mix of tumour cells with varying genotypes and phenotypes, which both drive and are driven by regional and temporal variations in microenvironmental conditions. Ultimately, any clinically relevant model of cancer must take account of communication between different tumour populations, between tumour cells and normal mesenchyma (such as blood vessels (Owen et al. 2009)), and the microenvironmental conditions (such as hypoxia) which are both causes and consequences of these interactions.

12.3.2 Building the Tumour System—Starting with Spheroids

Quantitative investigation of the tumour microenvironment became possible with development of the tumour spheroid experimental model. Under certain conditions,

tumour cells coalesce into a spheroid which can achieve a size of several hundred cell diameters. Since these typically float in culture media, they can replicate the diffusion reaction kinetics found *in vivo*, producing regional and even temporal variations in substrate and metabolites within the spheroid. Cellular response to environmental cues allowed observation of tumour proliferation, apoptosis, and necrosis within the spheroids.

Tumour spheroids are justifiably criticized as an overly simplistic model of tumour growth. However, their simplicity is also an advantage, as substrate and metabolite dynamics can be readily linked to tumour phenotype by this straightforward equation (Byrne et al. 2008):

$$\partial c/\partial t = D/r^2 \partial/\partial r(r^2 \partial c/\partial r) - T(c, r) \tag{1}$$

where c is the concentration of a substrate or metabolite, D is the diffusion coefficient, r is the position along the radius and T(c,r) is the local rate of consumption or production by the tumour cells (typically using Michelis-Menten kinetics).

By solving this equation and comparing it to experimental measures, many key metabolic parameters of tumour cells can be reasonably estimated. This becomes very useful in subsequent modelling of tumour systems.

12.4 Iterative Modelling of Tumour Systems

Ultimately, the value of mathematical models of tumour systems is to understand *in-vivo* growth. A number of methods have been developed to capture the key dynamics including ordinary and partial differential equations, agent-based models and modified cellular automata approaches. Each approach has advantages and disadvantages. To be successful, the approach utilized needs both to match the biological question being asked, and also to be capable of integration into the available experimental methods. A common approach uses models adapted from population biology. For example, if N_1 and N_2 denote the cell densities (in cells/cm^3) of the normal and tumour populations, assuming that only these populations compete for available space, then their temporal evolution is governed by the following equations:

$$\begin{aligned} \frac{\partial N_1}{\partial t} &= r_1 N_1 \left(1 - \frac{N_1}{K_1} - \frac{N_2}{K_2}\right) \\ \frac{\partial N_2}{\partial t} &= r_2 N_2 \left(1 - \frac{N_1}{K_1} - \frac{N_2}{K_2}\right), \end{aligned} \tag{2}$$

where $r_{1,2}$ and $K_{1,2}$ are the growth rates (in s^{-1}) and spatial carrying capacities (in cells/cm^3) of the respective populations. If it is also assumed that cells can migrate through space via a process akin to Fickian diffusion, where the diffusion parameters are themselves density-dependent (having a maximum value in empty space and going to zero when cells are close-packed), then equation (2) becomes

$$\frac{\partial N_1}{\partial t} = r_1 N_1 \left(1 - \frac{N_1}{K_1} - \frac{N_2}{K_2}\right) + \nabla \cdot \left[D_{N_1}\left(1 - \frac{N_1}{K_1} - \frac{N_2}{K_2}\right)\nabla N_1\right]$$
$$\frac{\partial N_2}{\partial t} = r_2 N_2 \left(1 - \frac{N_1}{K_1} - \frac{N_2}{K_2}\right) + \nabla \cdot \left[D_{N_2}\left(1 - \frac{N_1}{K_1} - \frac{N_2}{K_2}\right)\nabla N_2\right], \quad (3)$$

where D_{N_1} and D_{N_2} (in cm^2/s) are the 'empty-space' diffusion constant of the normal and tumour cells respectively. For simplicity, we assume that these are approximately equal: $D_{N_1} \approx D_{N_2} = D_N$. The Lotka-Volterra terms ensure that the density-dependent diffusion parameters are always positive-definite $\in [0, D_N]$.

Next, assume that each cell type has an optimal *pH* for survival and that if the local *pH* is perturbed from that optimal value, in either an acidic or an alkaline direction, the cells begin to die. We also assume that the death rate saturates at some maximum value when the environment is extremely acidic or alkaline. The simplest *ad hoc* functional form meeting these criteria is an "inverted Gaussian":

$$f_{1,2}(H) = d_{1,2}\left[1 - \exp\left\{-\left(\frac{H - H_{1,2}^{opt}}{2H_{1,2}^{width}}\right)^2\right\}\right], \quad (4)$$

where H is the local concentration of H^+ ions (in mol/L), $d_{1,2}$ are the saturated death rates (in s^{-1}), $H_{1,2}^{opt}$ are the local H^+ ion concentrations (in mol/L) corresponding to the optimal *pH*'s, and $H_{1,2}^{width}$ are the half-widths of the "inverted Gaussians" (in mol/L). Including the death rates, (4) into (2) gives

$$\frac{\partial N_1}{\partial t} = r_1 N_1 \left(1 - \frac{N_1}{K_1} - \frac{N_2}{K_2}\right) - f_1(H) N_1 + D_N \nabla \cdot \left[\left(1 - \frac{N_1}{K_1} - \frac{N_2}{K_2}\right)\nabla N_1\right]$$
$$\frac{\partial N_2}{\partial t} = r_2 N_2 \left(1 - \frac{N_1}{K_1} - \frac{N_2}{K_2}\right) - f_2(H) N_2 + D_N \nabla \cdot \left[\left(1 - \frac{N_1}{K_1} - \frac{N_2}{K_2}\right)\nabla N_2\right]. \quad (5)$$

We assume that H^+ ions are produced at a rate proportional to the local concentration of tumour and removed by the combined effects of buffering and vascular evacuation, both of which are proportional to microvessel area density. Thus

$$\frac{\partial H}{\partial t} = r_3 N_2 - d_3(H - H_0) + D_3 \nabla^2 H \quad (6)$$

where H is the H^+ ion concentration (in mol/cm^3), r_3 is the H^+ ion production rate (in mol/(cell·s)), d_3 is the H^+ ion uptake rate (in s^{-1}), H_0 is the H^+ ion concentration in serum, and D_3 is the H^+ ion diffusion constant (cm^2/s).

The model results obtained by solving this system of equations numerically provide extensive insights into the tumour-host interactions and into potential tumour treatment strategies such as systemic acidosis. Here we will focus on numerical simulations using parameter estimates based on experimentally determined proliferation rates, acid production, and acid-induced toxicity in the cell lines used in

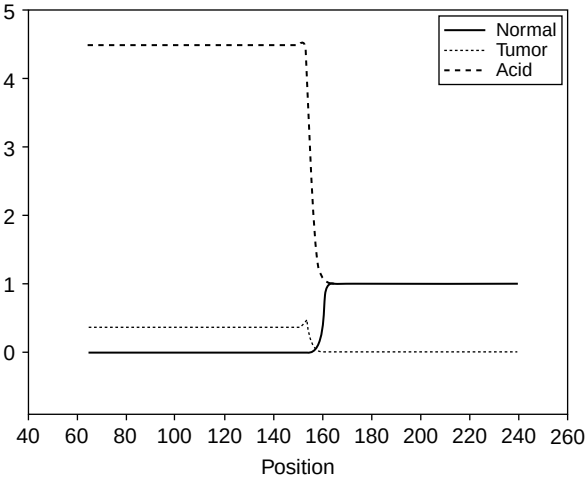


Fig. 12.1 Predicted cellular and microenvironmental dynamics of the tumor-host interface

subsequent experiments. Stability analysis of the steady state solutions of the state equations was also carried out.

As shown in Fig. 12.1, numerical solutions demonstrate that the interface at any given time represents a snapshot of a travelling wave as tumour cells advance and normal cells recede. The tumour wave is preceded by a gradient of excess H^+ extending into adjacent normal tissue. Within the region of peritumoral acidosis, the models predict a loss of normal tissue due to acid-induced cellular toxicity and ECM breakdown. These results supported the feasibility of the acid-mediated invasion model and provide guidance for the experiments to test the model. For example, the models demonstrated that observation of the gradient and associated toxicity would require a spatial resolution in the range of 50 μm or less.

12.5 Experimental Studies of Tumour Invasion

As in the physical sciences, it is essential to validate mathematical models against experimental observations. The modelling results above can be compared to *in-vivo* experiments performed using tumours growing *in-vivo* in a dorsal wound chamber (Fig. 12.2). Typical values of pH_e are shown in Fig. 12.3 and are plotted as a function of normalized distance from the tumour centre along eight angular segments. The experiments confirmed a peritumoral pHe gradient as predicted by the models. As demonstrated in Fig. 12.3, the gradient in the initial (upper) images was typically quite uniform, but less so on later (lower) imaging.

A key component of the model is that the pHe gradient would have substantial consequences, including degradation of the extracellular matrix allowing tumour

Fig. 12.2 Dorsal window chamber in a SCID mouse

invasion. In Fig. 12.4 PAS stains, after measurement of the pHe gradients, demonstrated evidence of considerable degradation in the extracellular matrix immediately adjacent to the tumour edge corresponding to the peritumoral acid gradient. However, the more distant ECM remained normal.

12.6 Tumour Modelling Collaborations

Increasingly, the complexity of cancer dynamics requires multidisciplinary partnerships. The Center for the Development of a Virtual Tumour (CViT 2009) is an example of such a multidisciplinary collaborative organization and has developed and unified a wide range of models, including tumour growth, vascular growth, and multiscale modelling of specific cancers via a cancer virtual tumour platform. CViT's long term goal is to develop a generic module-based toolkit for modelling and simulating selected cancer types of interest, such as breast, brain, and melanoma, following a paradigm-shifting cross-disciplinary complex systems approach. Combined with cutting-edge biomedical data, this modelling toolkit has significant value for experimental cancer research, as it allows researchers to properly study cancer initiation and such critically linked progression features as invasion, angiogenesis, and metastasis.

Many of CViT's activities, along with those of other worldwide projects, were summarized at an EU-USA workshop on Multiscale cancer modelling (2009), with the papers published as a book by Deisboeck and Stamatakos (2010). They emphasize the need for multiscale approaches to modelling cancer, which is a complex disease process that spans both space and time. Driven by cutting-edge mathematical and computational techniques, *in silico* biology provides powerful tools to investigate the mechanistic relationships of genes, cells, and tissues. It enables the creation of experimentally testable hypotheses, the integration of data

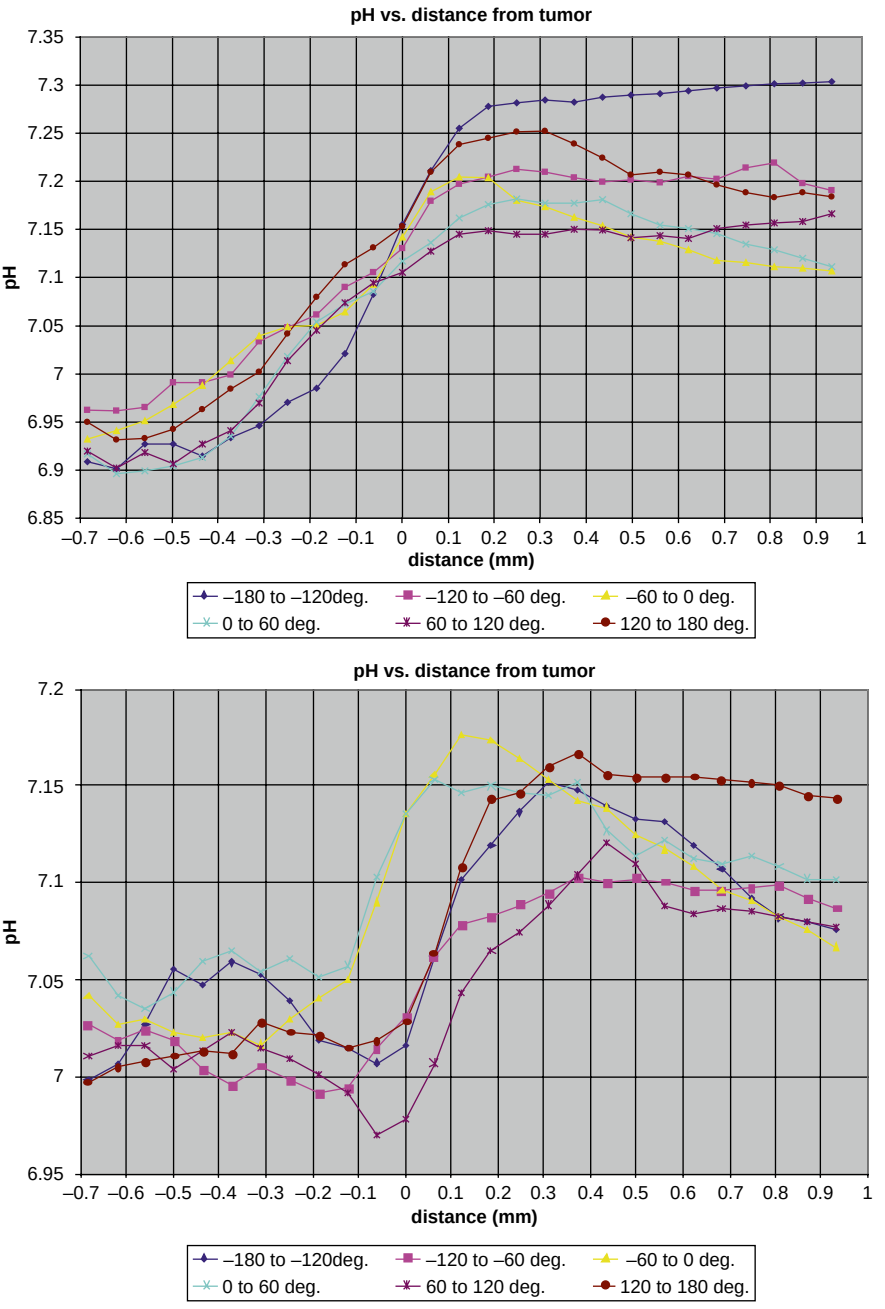


Fig. 12.3 pH_e gradients at the PC3N/eGFP tumour-host interface along radians drawn from the tumour center. The tumor-host interface is designated as the 0 point on the x-axis. *Top* 2 d following placement of the tumor slurry. The relatively avascular tumour demonstrates a fairly uniform pH_e distribution and gradient. *Bottom* 4 d later. During that time significant tumour growth was observed and pH_e distribution is less uniform. These results are in qualitative agreement with the mathematical model predictions although the regional and temporal variation demonstrate the need to modify the model treatment of vascular density and blood flow

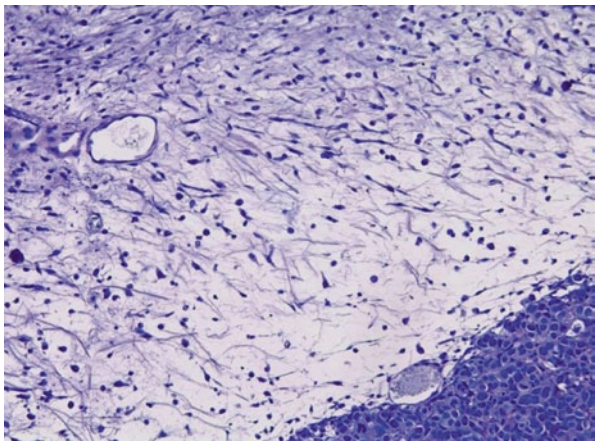


Fig. 12.4 PAS stain of tumor following pHe measurements (see Fig. 12.3) showing ECM breakdown in the peritumoral normal tissues in the region shown to be acidic on FRIM studies, while the more distant ECM appears normal

across scales, and the prediction of tumour progression and treatment outcome (*in silico* oncology). The papers presented at the workshop, and the resulting book, discuss the scientific and technical expertise necessary to conduct innovative cancer modelling research across multiple scales. The ultimate goal of multiscale modelling and simulation approaches is their use in clinical practice, such as supporting patient-specific treatment optimization. Areas discussed in detail include:

- Evolution, regulation and disruption of homeostasis
- Oncogenic signalling and molecular structure
- Linkage and tetraploidy in neoplastic progression
- Genomically unstable tumorigenesis
- Intestinal stem-cell homeostasis
- Colonic crypts and early colorectal cancer
- The microenvironment in somatic evolution
- Cell motion in 3-D environments
- Cancer growth and agent-based models
- Diffusional instability and tumour invasion
- Mesenchymal cell migration and sprouting angiogenesis
- Tumour invasion strategies
- Vascular tumour growth
- Solid tumour growth and vessel networks
- Cancer growth and treatment
- Clinical applications and gliomas
- Reaction-diffusion tumour growth in MR images
- Clinically oriented cancer modelling.

The range of areas discussed demonstrates the high level of activity in multiscale modelling that will provide a strong basis for further developments.

12.7 Detailed Modelling Example

12.7.1 Carcinogenesis Transitions

Detailed analytical modelling of colorectal cancer in the crypt is described by Johnston et al. (2007). Colorectal cancer is initiated in colonic crypts. A succession of genetic mutations or epigenetic changes can lead to homeostasis in the crypt being overcome, and subsequent unbounded growth. The model considers the dynamics of a single colorectal crypt by using a compartmental approach developed by Tomlinson and Bodmer (1995), which accounts for populations of stem cells, differentiated cells, and transit cells. That original model made the simplifying assumptions that each cell population divides synchronously, but these assumptions are relaxed by adopting an age-structured approach that models asynchronous cell division, and by using a continuum model (Vincent and Gatenby 2008). Two mechanisms might regulate the growth of cell numbers and maintain the equilibrium that is normally observed in the crypt. The first maintains an equilibrium for all parameter values, whereas the second would allow unbounded proliferation if the net per capita growth rates were large enough. Results show that an increase in cell renewal, which is equivalent to a failure of programmed cell death or of differentiation, can lead to the growth of cancers. The second model can be used to explain the long lag phases in tumour growth, during which new, higher equilibria are reached, before unlimited growth in cell numbers ensues. The second form of feedback could help explain the observed lag phases after mutations occur, and thus the existence of benign tumours or adenomas before carcinogenesis takes over. Early mutations in the adenoma-carcinoma sequence could raise the net per-capita growth rates but keep them below their critical values, which would create new, higher steady states. Later stage mutations could push the net per-capita growth rates above their critical values, resulting in unregulated cell population growth. However, if no genetic changes occur, then it remains benign. Although the evidence supports the view that almost all colorectal cancers go through an adenoma, or benign phase, it is by no means true that all adenomas develop into carcinomas.

The potential value of modelling is apparent in a sequence of studies investigating carcinogenesis. The stepwise transition of normal cells to the cancer phenotype through a number of pre-malignant intermediates is often described as ‘somatic evolution’. The classic conceptual model is depicted in this evolutionary process as a series of genetic mutations typically found in oncogenes and tumour suppressor genes (Fearon and Vogelstein 1990). Gatenby and Vincent (2003) subjected this concept to rigorous analysis by applying evolutionary game theory. The results of the models demonstrated that mutations in oncogenes and tumour suppressor genes alone did not result in formation of a malignant cancer. In fact, these changes led only to self-limited growth, because as the tumour population increased in size, proliferation was constrained by substrate limitation (Fig. 12.5).

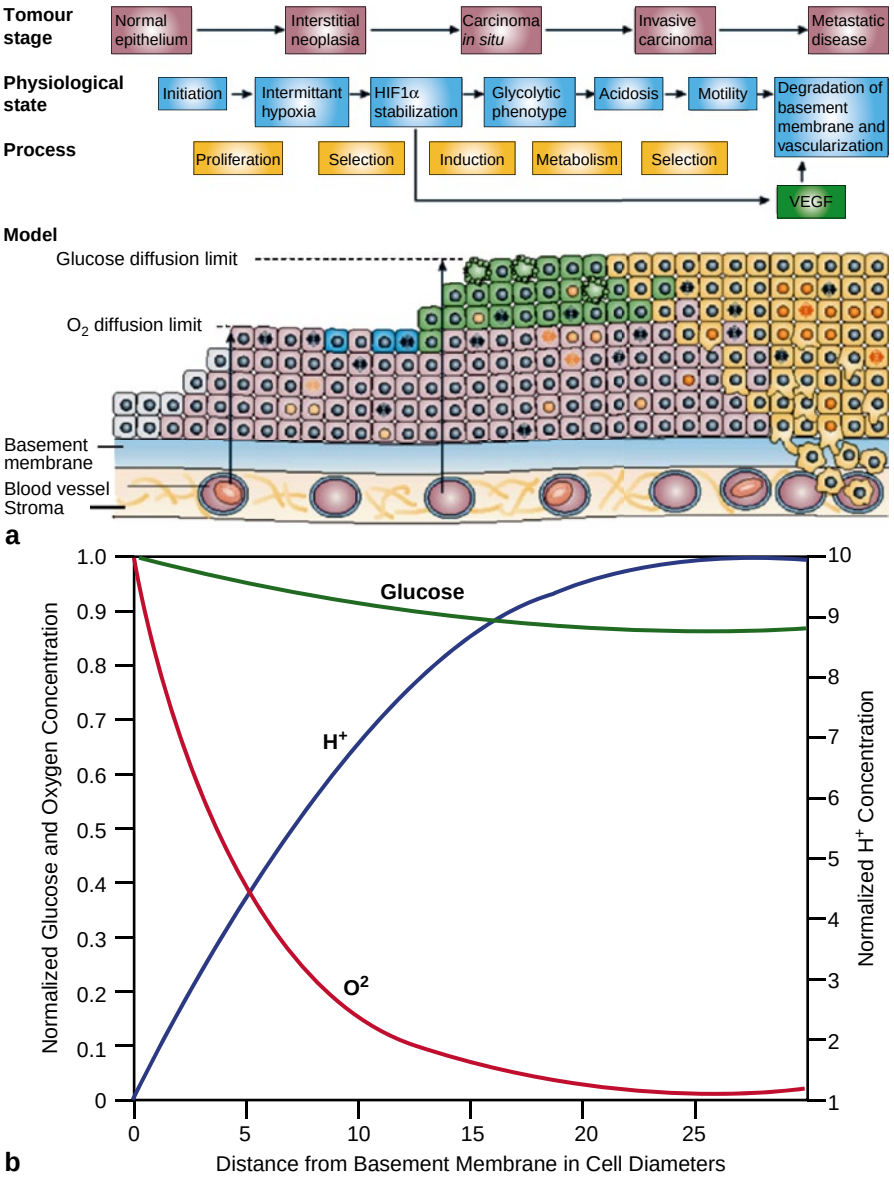


Fig. 12.5 A theoretical model of the evolutionary dynamics of carcinogenesis, based on mathematical modelling predictions showing that substrate limitation dominates the later stages of somatic evolution. A critical component of the translation of the mathematical results was recognition that *in-situ* cancers evolve on a basement membrane which separates the tumour populations from their blood supply (**a**). This requires substrate and metabolites to diffuse over increasingly long distances to the tumour cells as they proliferate into the lumen and away from the basement membrane. The resulting diffusion-reaction kinetics produces regions of hypoxia and acidosis (**b**) requiring adaptations including up-regulation of anaerobic glucose metabolism and increased H^+ export. It is proposed that this adapted phenotype is critical to subsequent transition to invasive cancer, because the cells produce an acidic environment (through up-regulated glycolysis) that is toxic to competing populations

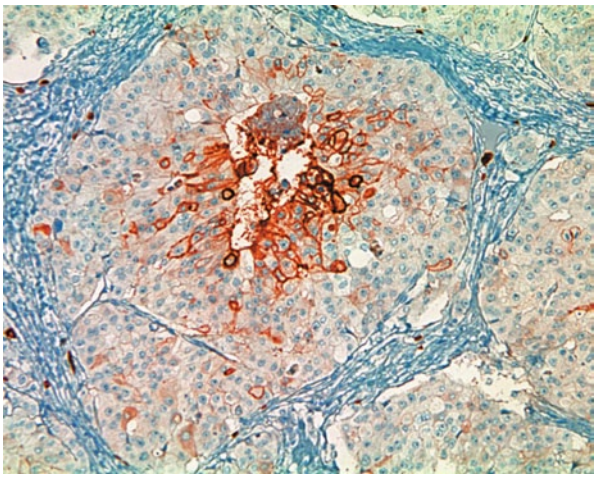


Fig. 12.6 Clinical specimen for a breast biopsy showing ductal carcinoma in situ (DCIS). Immunohistochemical stains (IHC) show increased expression of Glucose Transporter-1 (GLUT-1) in tumour cells furthest from the basement membrane, demonstrating response to hypoxia as predicted by the theoretical model in Fig. 12.1

Thus, as a result of the modelling studies, the authors concluded that there is a previously unknown era of carcinogenesis in which environmental selection forces are dominated by competition for limited substrate (Gatenby and Vincent 2003).

12.7.2 Somatic Evolution

These results demanded a re-examination of the adaptive landscape of somatic evolution, and recognition of the critical role of the anatomy and physiology of epithelial surfaces (Gatenby and Gillies 2004). As shown in Fig. 12.5, evolving *in-situ* cancer cells proliferate on the surface of the basement membrane, which maintains a separation from the epithelial cells and the underlying stroma, including blood vessels. As the tumour cells proliferate into the lumen, their distance from the blood vessels increases, and the resulting diffusion reaction kinetics results in regional hypoxia, low glucose concentrations, and acidosis.

The authors developed a theoretical model (Gatenby and Gillies 2004) of carcinogenesis proposing that, because of the anatomy and physiology of tumour growth on epithelial surfaces, some regions of *in-situ* cancer cells will be subject to cyclical hypoxia. This will promote adaptation by up-regulating glycolysis. However, the resulting increased production of lactic acid affects the adaptive landscape, replacing it with one that is dominated by the toxic effects of acidosis. This requires a second evolutionary step that allows the tumour cells to adapt to unusually acidic environments. The final outcome of this sequence produces a cellular phenotype with a profound proliferative advantage, because it can produce an acidic environment (through up-regulated glycolysis) that is toxic to its competitors but not to itself. As a result of this study, it was hypothesized that adaptation to regional hypoxia and acidosis (Fig. 12.6) was a critical component in late carcinogenesis that promoted transition from *in-situ* to invasive growth (Fig. 12.7).

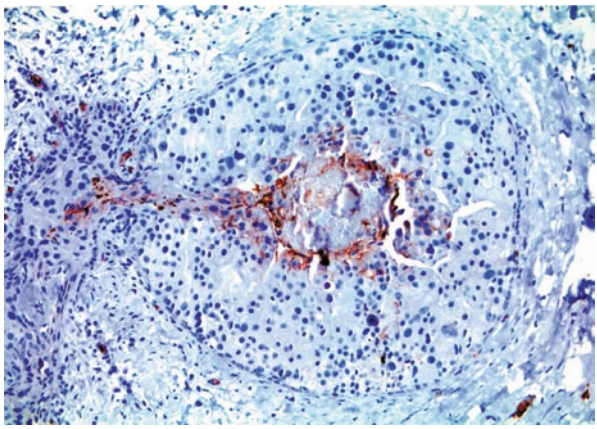


Fig. 12.7 Clinical specimen from a breast biopsy showing DCIS with a focus of micro-invasion. IHC for GLUT-1 similar to that in Fig. 12.6 shows increased expression in the central region of the intraductal tumour. However, up-regulated GLUT-1 is also present in the invasive component of the tumour as predicted by the theoretical model

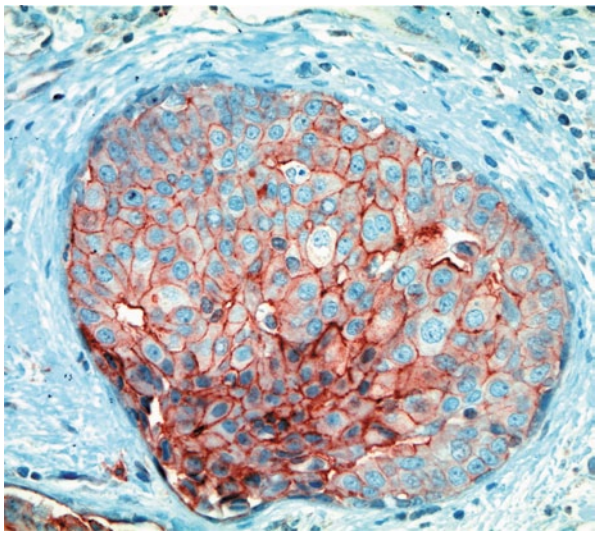


Fig. 12.8 Clinical specimen from a breast biopsy showing DCIS. In this sample, IHC was performed to evaluate expression of Na^+/H^+ (NHE-1) transporters since up regulation will indicate the presence of an acidic miroenvironment. Note that markedly increased expression is seen in a region in which the tumour is bulging out of the basement membrane prior to invasion

This theoretical result was examined (Gatenby et al. 2007) using *in-vitro* experimental methods with tumour spheroids and clinical observations. As shown in Figs. 12.6 and 12.7, this work confirmed the presence of adaptation to hypoxia in the central regions of DCIS resulting in up-regulation of glucose transporter 1. A central prediction of the theoretical model was that acid-mediated invasion promoted invasion of the glycolytic cells into normoxic regions, and facilitated breeching of the basement membrane during transition from *in-situ* to invasive tumour growth. This was supported by observations in DCIS that showed up-regulation of Na^+/H^+ pumps in tumour regions corresponding to thinning or invasion of the basement membrane (Fig. 12.8).

12.8 Conclusions

Developing mathematical models of the complex, non-linear dynamics of cancer biology is difficult. However, understanding these dynamics through purely intuitive reasoning is impossible. Establishing theoretical models that define the first principles of cancer biology will require a cadre of investigators conversant in both tumour biology and mathematical methods, to develop quantitative theoretical models designed to provide appropriate conceptual frameworks for organizing extant data, integrating new information, and guiding future experiments. Accomplishing this will necessitate the surmounting of many intellectual, social and philosophical barriers. However, it seems likely that future progress in understanding and controlling cancer will depend, as well, upon mathematical modelling.

References

1950 Mortality Data (2001) CDC/NCHS, NVSS and 2001 Mortality Data—NVSR-death final data 52(3)

Anderson AR, Quaranta V (2008) Integrative mathematical oncology. Nat Rev Cancer 8(3):227–234

Armitage P, Doll R (1954) The age distribution of cancer and a multi-stage theory of carcinogenesis. Br J Cancer 8:1–12

Byrne HM, Leeuwen IMM van, Owen MR, Alarcon T, Maini PK (2008) Multiscale modelling of solid tumour growth. Selected topics in cancer modelling. genesis, evolution, immune competition, & therapy, n/a (n/a). Birkhauser, Boston, pp 449–473. ISBN 978-0-8176-4712-4

CVIT (2009) The center for the development of a virtual tumour. https://www.cvit.org. Accessed 12 Jan 2009

Deisboeck TS, Stamatakos G (2010) Multiscale cancer modelling. CRC press. ISBN 9781439814406. To be published in 2010. http://www.crcpress.com/product/isbn/9781439814406;jsessionid=QRJa80y3kd5+0fUhumPf+Q**. Accessed 10 Aug 2010

Fearon ER, Vogelstein B (1990) A genetic model for colorectal tumorigenesis. Cell 61:759–767

Feynman RP (1965) The character of physical law. MIT Press, Cambridge

Galilei G (1623) Il Saggiatore, Rome

Gatenby RA, Gillies RJ (2004) Why do cancers have high aerobic glycolysis? Nat Rev Cancer 4(11):891–899

Gatenby RA, Maini P (2002) Modelling a new angle on understanding cancer. Nature 420:462

Gatenby RA, Maini P (2003) Mathematical oncology—cancer summed up. Nature 421:321

Gatenby RA, Vincent TL (2003) An evolutionary model of carcinogenesis. Cancer Res 63:6212–6220

Gatenby RA, Smallbone K, Maini PK, Rose F, Averill J, Nagle RB, Worrall L, Gillies RJ (2007) Cellular adaptations to hypoxia and acidosis during somatic evolution of breast cancer. Br J Cancer 97(5):646–53

Johnston MD, Edwards CM, Bodmer WF, Maini PK, Chapman SJ (2007) Mathematical modelling of cell population dynamics in the colonic crypt and in colorectal cancer PNAS 104(10):4008–4013. http://www.pnas.org/cgi/reprint/104/10/4008.pdf. Accessed 12 Jan 2009

Kepler J (1619) In Harmonice mundi (trans. Harmonies of the world) (Linz, 1619) Book 5, Chapter 3, trans. Aiton, Duncan and Field, p 411

Krogh A (1924) The anatomy and physiology of capillaries. Yale University Press, New Haven, p 196

Maini PK, Gatenby RA (2006) Some mathematical modelling challenges and approaches in cancer. In: Nagl S (ed) Cancer bioinformatics: from therapy design to treatment. Wiley, New York, pp 95–107

Multiscale Cancer Modelling (2009) Report on first transatlantic workshop on multiscale cancer modelling. http://ec.europa.eu/information_society/events/ict_bio/2008/docs/200811cancer-model-wkshp-report.pdf. Accessed 1 Dec 2009

Newton I (1846) Mathematical principles of natural philosophy. Translated by Andrew Motte, First American Edition New York

Owen MR, Alarcon T, Maini PK, Byrne HM (2009) Angiogenesis and vascular remodelling in normal and cancerous tissues. J Math Biol 58(4–5):689–721

Schrödinger E (1944) What is life? Cambridge University Press, London

The Economist (2004, 22 Jan) http://www.economist.com/node/2367273

Thompson DW (1942) On growth and form. Dover reprint of 2nd ed (1st ed, 1917).

Thoren V (1990) The lord of Uraniborg: a biography of Tycho Brahe. Cambridge University Press, Cambridge

Tiner JH (1977) Johannes Kepler: giant of faith and science. Mott Media, Milford

Tomlinson IPM, Bodmer WF (1995) Proc Natl Acad Sci USA 92:11130–11134

Tomlinson RH, Gray LH (1955) The histological structure of some human lung cancers and the possible implications for radiotherapy. Br J Cancer 9:539–549

Vincent TL, Gatenby RA (2008) An evolutionary model for initiation, promotion, and progression in carcinogenesis. Int J Oncol 32(4):729–737

Part IV
Diagnosis, Clinical and Treatment Applications

Chapter 13
Diagnostic and Prognostic Cancer Biomarkers: From Traditional to Systems Approaches

Francesca M. Buffa and Adrian L. Harris

Abstract In this chapter we provide a survey of the state of the art in cancer biomarkers. The different roles and uses of biomarkers are introduced. Several examples are provided of biomarkers that are now used in the clinical routine, with a particular emphasis on breast cancer. Furthermore, several promising new areas for research and application are highlighted.

The challenges and difficulties involved with the development of fundamental discoveries into useful clinical biomarkers are discussed; and we describe some of the current international initiatives and guidelines that have been attempting to assist this process. We present bioinformatics, system analyses and system modelling approaches, and discuss their application and relevance for biomarker effective discovery and prioritisation.

Finally, we emphasize the need for large-scale harmonization, standardization and integration of multi-level data to ensure that the potential offered by the current unprecedented availability of data and information is fully exploited, so as to develop validated and useful biomarkers.

13.1 Introduction

Biomarkers are measurable genetic, chemical or molecular indicators of disease presence, progression status, and/or potential response to therapeutic intervention. They can be used, for example, to screen large populations for cancer or to monitor effects of a specific cancer treatment and make effective plans for further therapeutic intervention. While predictive biomarkers indicate the efficacy of a specific treatment, prognostic biomarkers provide general information about disease outcome independently from the specific therapeutic intervention (Fig. 13.1).

Effective biomarker development is critical for improvements in patient-tailored clinical management of disease, and as a consequence, biomarker discovery in fields such as cancer research has expanded rapidly over recent years. This progress has been enabled by recent major advances in molecular biology, biotechnology and

F. M. Buffa (✉)
Department of Medical Oncology, Weatherall Institute of Molecular Medicine, University of Oxford, OX3 9DS Oxford, UK
e-mail: francesca.buffa@imm.ox.ac.uk

A. Cesario, F. B. Marcus (eds.), *Cancer Systems Biology, Bioinformatics and Medicine,*
DOI 10.1007/978-94-007-1567-7_13,

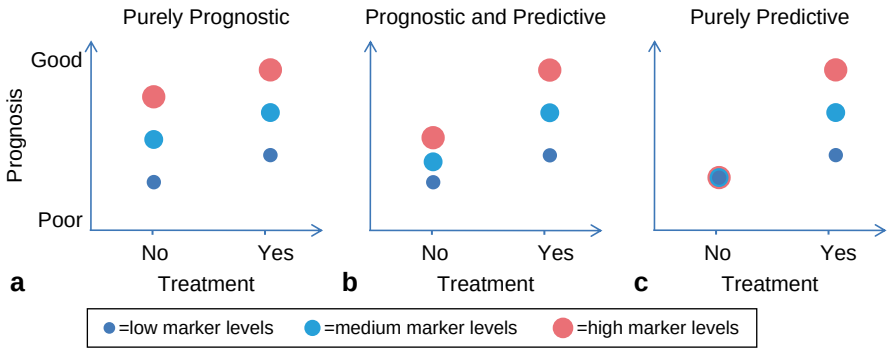


Fig. 13.1 Schematic representation of prognostic and predictive markers. a Purely prognostic marker: example of a marker whose high levels are favourable for prognosis independently of the specific treatment received. **b** Mixed prognostic and predictive marker: example of marker whose high levels are generally favourable for prognosis, but displaying a significantly higher separation between poor and good prognosis when patients receive the treatment for which the marker predicts sensitivity. **c** Purely Predictive: example of marker whose high levels are indicative of good prognosis only, provided the patients receive the treatment. (Adapted from Hayes et al. 1998)

imaging. However, study-dependent differences in methodology, both technological and analytical, have frequently produced contrasting conclusions, and thereby delayed development of fundamental discoveries into validated and useful clinical tools. Hence, it is increasingly recognized that consistent protocols, terminology and techniques should be adopted to ensure rapid and effective translation of biomarkers to the clinic. To this end, several recent initiatives have set out guidelines for consistent methodology, including the Cancer Research UK biomarker and imaging discovery and development guidelines[1], and the recently published guidelines of the National Cancer Institute (NCI, Dancey et al. 2010).

Cancer researchers, and as a consequence cancer biomarker studies, have traditionally investigated discrete molecular pathway alterations such as individual mutations. This work has generated a great amount of new knowledge and new biomarkers, but it is now recognized that cancer is a complex system of interacting molecules and pathways (Hanahan and Weinberg 2000). Multiple factors including genetic abnormality, epigenetic and post-translational regulation, and the interaction between tumour and host must all be taken into account to fully understand disease progression and prognosis. Therefore, an integrative biology approach to biomarker studies considering cancer as a complex and dynamic disease will play an important role in development of more effective biomarkers and therapies.

[1] http://science.cancerresearchuk.org/reps/pdfs/bidd_diagnostic_roadmap.pdf, http://science.cancerresearchuk.org/reps/pdfs/bidd_pharmacological_roadmap.pdf, http://science.cancerresearchuk.org/reps/pdfs/bidd_prognostic_roadmap.pdf, http://science.cancerresearchuk.org/reps/pdfs/bidd_screening_roadmap.pdf.

13.2 Role of Biomarkers

A high percentage of cancer patients are diagnosed when the disease is at an advanced stage and the probability of cure or control is lowest. New treatments frequently produce only a limited increase in overall survival, with significant benefit restricted to subgroups of patients. Biomarkers can improve disease outcome not only by enabling earlier-stage diagnosis, but also by indicating which patient subgroups, or perhaps even individuals, are likely to benefit from specific therapeutic strategies. Biomarkers play an additional role during therapy to monitor toxicity and treatment responses.

13.3 Biomarkers for Prediction of Response to Treatment

Predictive biomarkers indicate the likely efficacy of a specific anti-cancer treatment. Such markers are important as they can be used to identify patients who would benefit most from treatment, and conversely to identify patients who would not benefit; the latter are thus spared unjustified toxicity by being assigned to alternative therapy.

Examples of established predictive markers are oestrogen and progesterone receptors (ER, PR) in breast cancer, their expression status being used routinely in the clinic to select patients benefiting from adjuvant hormonal therapy (Williams et al. 2006).

13.3.1 The ErbB Family of Receptor Tyrosine Kinases: HER2 as a Predictive Marker in Breast Cancer

Examples of recently identified predictive markers successfully translated to the clinic are the ErbB family of receptor tyrosine kinases (RTKs). RTKs are cell-surface receptors that when bound to the appropriate ligand, trigger a signal-transduction cascade, ultimately leading to changes in cellular biology and physiology. Signalling pathways downstream of the ErbB family regulate diverse biological responses such as proliferation, differentiation, angiogenesis, cell motility, and cell survival. In cancer, over-expression or constitutive activation of ErbB family members has been demonstrated to promote tumour progression and spread, and to be important in prognosis (Yarden and Sliwkowski 2001).

Thus, ErbB family receptors represent interesting targets for cancer therapy, with epidermal growth factor receptor (EGFR) and herstatin (HER2) being the most prominent. Indeed, these receptors are now validated targets for anti-cancer therapy, and several targeting agents, such as antibodies binding to the ligand-binding domains (cetuximab, panitumumab and trastuzumab) or small molecules targeting

the tyrosine kinase domains (erlotinib, gefitinib, and lapatinib), have been approved for the treatment of cancer (Caponigro et al. 2005). Conversely, accurate measurements of receptor status before therapy allow selection of patients who are likely to respond to these therapies; for example, positive HER2 expression is a biomarker now used in the clinic in the management of breast cancer (Piccart et al. 2001) and three large trials have investigated the effect of adding trastuzumab (Herceptin) as adjuvant therapy to standard chemotherapy in HER2-positive early breast cancer, (Piccart-Gebhart et al. 2005; Romond et al. 2005). An improvement in outcome, measured as disease-free survival, was observed when Herceptin was added to chemotherapy (this outcome will need to be confirmed by long-term follow-up for all patients in the trials).

13.3.2 EGFR in Head and Neck, Colorectal and Non-small-cell Lung Cancers

Agents targeting EGFR are used to treat cancers such as head and neck squamous cell carcinomas (HNSCC), colorectal cancer and non-small-cell lung cancer (NSCLC). However, in these tumours the detection of positive EGFR expression does not reliably predict clinical outcome of anti-EGFR treatment (Siena et al. 2009). This has led to searching of downstream signalling pathways for alternative predictive biomarkers, and amongst these KRAS mutation has emerged as a predictive marker of resistance to anti-EGFR treatment in colorectal cancer (Benvenuti et al. 2007; Lievre et al. 2006; Linardou et al. 2008; Moroni et al. 2005) and NSCLC (Linardou et al. 2008). Also, EGFR tyrosine kinase domain mutations predict clinical response to the tyrosine kinases inhibitors erlotinib and gefitinib (Han et al. 2005; Kim et al. 2005; Lynch et al. 2004; Paez et al. 2004; Pao et al. 2004), but not response to antibody-based therapy (Mukohara et al. 2005). Although these studies have already highlighted several new predictive markers and avenues for selection of patients for ErbB family antibody-based therapies, further understanding of the mechanisms involved in intrinsic and acquired resistance might help to refine and optimize this selection (Kruser and Wheeler 2010).

Recently, EGFR positivity has also been proposed as a marker to identify patients who would benefit from accelerated radiotherapy (Bentzen et al. 2005; Eriksen et al. 2005, p. 66) or radiotherapy combined with anti-EGFR treatment (Krause and Baumann 2008). The main hypothesis has been that tumours with high expression of EGFR are more likely to experience an accelerated repopulation during radiotherapy, and thus would benefit from either a shorter radiotherapy schedule, or concomitant anti-EGFR treatment. However, mechanisms of EGFR interaction with radiotherapy, and more generally with radiation, might be more complex, and have not yet been fully clarified (Krause and Baumann 2008).

13.4 Biomarkers for Prognosis

Prognostic biomarkers provide general information regarding disease outcome, independently of the specific therapeutic intervention or in absence of treatment. They allow, for example, identification of poorly-responding subgroups of patients for whom more aggressive therapy might be useful, and situations where research into further therapies is most needed.

13.4.1 Traditional Clinical Markers—Lymph Node Involvement

In most cancer types the morphology of the tumour, its histopathological stage or grade, and spread to near nodes are routine prognostic markers. However, while this information can provide some indication regarding possible clinical outcome, it cannot be employed with any certainty for prediction of distant metastases, prognosis, and ultimately patient survival in many cancers (Ferlay et al. 2004).

An example of a clinically recognized prognostic factor for breast cancer is the absolute number of positive axillary lymph nodes (LNs). In other words, patients whose disease has spread to the LNs are more likely to have a poor clinical outcome, and this risk increases as the number of involved LNs increases. The absolute number of positive axillary LNs and other factors were added to a revised staging system for breast carcinoma by the American Joint Committee on Cancer (AJCC) (Singletary et al. 2002). This revision incorporated some of the findings from prognostic studies into clinical management guidelines. However, not all prognostic markers were included, and several promising markers were not considered at this stage for various reasons, including lack of conclusive evidence of their prognostic usefulness and lack of standardization in measurement techniques (Singletary et al. 2002). For example, histological grade, although it has been considered a prognostic factor for many years, was not included due to lack of concordance across different institutions.

13.4.2 Histological Grade and Proliferation

Histological grade (see Chap. 2) is a three-category ordinal system (low, intermediate and high) based on the morphological evaluation of cancer samples, and is likely to reflect both the proliferation and differentiation status of the cancer (Elston and Ellis 1991). Various grading systems have been devised to optimize this classification. However, even relatively advanced systems such as the Nottingham combined histological grade, a semi-quantitative evaluation combining the percentage of tubule formation, the degree of nuclear pleomorphism and accurate mitotic count in a defined field area (Elston and Ellis 1991), give varying agreement across institutions (Robbins et al. 1995). Modifications to the Nottingham combined histological

grading have been suggested, but the evidence collected to date is not yet sufficient to support them (Ignatiadis and Sotiriou 2008).

Fundamental genetic differences between grades 1 and 3 have been found by comparative genomic hybridization, suggesting that low- and high-grade breast cancers might reflect distinct diseases, rather than progression of the same disease (Roylance et al. 1999).

13.4.3 Gene Expression Grade

Recently, a gene-expression grade index (GGI) has been developed that classifies tumours into only two grades, low and high, thus questioning the need for an intermediate grade group (Sotiriou et al. 2006). Specifically, in an analysis of 597 tumours from five centres, including our institution, it was found that GGI was able to reclassify tumours with histological grade two into either grade 1 or 3 tumours, and that the clinical outcome of these subgroups was similar to the overall clinical outcome of grade 1 and 3 tumours respectively. The function of many genes in the GGI has been associated with cell cycle progression and proliferation, indicating that histological grade may largely reflect the proliferation and differentiation status of the tumour cells (Elston and Ellis 1991).

13.4.4 Proliferation Markers

The rate of tumour cell proliferation is an important contributor to outcome and may underlie the prognostic power of other biomarkers or gene signatures (Whitfield et al. 2006). An established proliferation marker is Ki-67, a protein present during all active phases of the cell cycle that is used as an indicator of the proliferating cell fraction. Ki-67 expression has been shown to be prognostic in a recent large meta-analysis (de Azambuja et al. 2007), but was not included in the AJCC review (Singletary et al. 2002) due to lack of standardization of the assay and lack of evidence that Ki-67 was prognostically independent from traditional clinical co-variates (de Azambuja et al. 2007).

13.4.5 Hypoxia Biomarkers

Another tumour trait that has been investigated in several studies as a potential prognostic marker is hypoxia. Severe or prolonged hypoxia results mostly in death of non-transformed cells. However, heterogeneous cancer cell populations have an increased ability to survive hypoxia and undergo selection for a more aggressive phenotype (for a review see Harris 2002). Several studies have addressed the role

of hypoxia in cancer prognosis; however, there is no current consensus regarding a gold-standard marker for hypoxia in the clinical setting. This is in part due to the fact that tumour response to hypoxia is characterized by both temporal and spatial heterogeneity, and that different markers have shown different biological and prognostic value. (For a recent review, see Jubb et al. 2009).

13.4.6 Global and Multi-gene Expression Profiling

Global gene expression profiling to generate tumour 'signatures' has an emerging role in prognostic marker studies, although in most cases application to the clinical management of patients is still lacking, and reproducibility across centres, protocols and technologies is not fully determined.

In breast cancer, different profiling approaches have identified several potentially prognostic signatures (Table 13.1). Comparative analyses (Desmedt et al. 2008; Fan et al. 2006; Haibe-Kains et al. 2008) found that several of these signatures, with significantly different gene lists, showed a similar prognostic significance. Two prognostic gene expression signatures, the "70-gene" and the "76-gene" signatures (Table 13.1), have been tested in a large European multicentre prospective study by the TRANSBIG Consortium (http://www.breastinternationalgroup.org/). Both signatures show prognostic ability in node negative breast cancers (Buyse et al. 2006; Desmedt et al. 2007), although they show a significant dependence on time-from-treatment, perhaps indicating that they are an index of early distant relapse. The 70-gene signature, commercialised as MammaPrint®, was approved in 2008 by the Food and Drug Administration (FDA) for use as a prognostic test in pre-menopausal breast cancer patients with early-stage, < 5 cm-diameter tumours with no evidence of nodal disease.

A third smaller signature, RS/ONC-16, which includes 16 cancer-related (Table 13.2) and five control genes, has been commercialized as Oncotype DX®, an assay applicable to formalin-fixed paraffin-embedded (FFPE) samples. This signature has been validated by prospective analysis of samples collected within clinical trials of node-negative ER-positive breast cancer patients (Mamounas et al. 2010; Paik et al. 2006). Furthermore, a very recent study (Albain et al. 2010) reported prognostic and predictive value of Oncotype DX® in postmenopausal women with node-positive ER-positive breast cancers receiving either tamoxifen alone or anthracycline-based chemotherapy (cyclophosphamide, doxorubicin, and fluorouracil—CAF); followed by tamoxifen after surgery. The results showed that Oncotype DX® was a prognostic factor for women treated with tamoxifen alone and could predict the efficacy of CAF chemotherapy. Both MammaPrint® and Oncotype DX® are currently under evaluation for patient stratification in large multicentre trials. These two trials, named MINDACT and TAYLORx, are hoping to recruit respectively 6000 patients with no evidence of nodal disease (MINDACT-http://www.eortc.be/services/unit/mindact/MINDACT_websiteii.asp), and 11,000 patients with ER-positive HER2-negative tumours with no evidence of nodal disease

Table 13.1 Summary of prognostic gene expression signatures in breast cancer

Signature name	Original publication	# samples	Study design type*
Intrinsic subtypes	(Perou et al. 2000)	65 BC samples, 42 individuals	Class discovery: hierarchical clustering of intrinsic genes
70-gene profile[a]	(van de Vijver et al. 2002)	117 BCs	Class comparison: distant metastases
RS/ONC-16[a]	(Paik et al. 2004)	668 BCs	Mixed approach, knowledge-based and class comparison
2-gene ratio	(Ma et al. 2004)	60 tamoxifen treated BCs	Class comparison: tumour recurrence
76-gene signature	(Wang et al. 2005)	115 BCs	Class comparison: distant metastases
NCH70	(Naderi et al. 2007)	135 BCs	Class comparison: survival
CON52	(Teschendorff et al. 2006)	877 BCs	Class comparison: survival, meta-analysis
p53 signature	(Miller et al. 2005)	251 BCs	Class comparison: p53 mutation
GGI	(Sotiriou et al. 2006)	64 BCs	Class comparison: histological grade
Wound response	(Chang et al. 2005)	50 fibroblast cultures	Bottom-up: derived in cell lines, used to predict class in clinical samples
Proliferation	(Whitfield et al. 2002, 2006)	Cervix cancer cell line	Bottom-up: derived in cell lines, used to predict class in clinical samples
	(Starmans et al. 2008)	NA	Bottom-up: integration of published signatures from cell line studies
	(Desmedt et al. 2008)	581 BCs	Bottom-up: data-mining using previous knowledge, meta-analysis
Hypoxia	(Chi et al. 2006)	4 cell cultures: EC, SMC, HMEC, RPTEC	Bottom-up: derived in cell lines used to predict class in clinical samples
	(Winter et al. 2007)	59 HN cancers	Bottom-up: data-mining using knowledge from previous studies
	(Buffa et al. 2010)	121 HN cancers & 1015 BCs	Bottom-up: data-mining using previous knowledge, meta-analysis

* For more details on the methods see Sect. 13.10

HN head and neck, *BC* breast cancer. *EC* endothelial cells, *SMC* smooth muscle cells, *HMEC* human mammalian epithelial cells, *RPTEC* renal proximal tubule epithelial cells.

[a] Currently in test in clinical trials: see Section "Prognostic Biomarker", "Gene expression profiling"

Table 13.2 Oncotype DX Recurrence Signature (Paik et al. 2004): genes, and related function, processes and pathways. This table excludes the five control genes

Symbol	GeneID	Refseq	Description	Molecular function* *(italics: unclassified by Panther, comment added)*	Biological process* *(italics: unclassified by Panther, comment added)*	Pathway* *(italics: unclassified by Panther, comment added)*	Chromosome
SCUBE2	57758	NM_020974	Signal peptide, CUB domain, EGF-like 2	Extracellular matrix glycoprotein	Cell adhesion-mediated signalling; vision; skeletal development	*Invasion*	11
CTSL2	1515	NM_001333	Cathepsin L2	Cysteine protease	Proteolysis	*Invasion*	9
GSTM1	2944	NM_000561	Glutathione S-transferase M1	Other transferase	Detoxification	*Drug resistance*	1
MKI67	4288	NM_002417	Antigen identified by monoclonal antibody Ki-67	*Cell cycle*	*Cell cycle*	*Proliferation*	10
MMP11	4320	NM_005940	Matrix metallopeptidase 11 (stromelysin 3)	Metalloprotease; other extracellular matrix	Proteolysis	*Invasion*	22
AURKA	6790	NM_003600	Aurora kinase A	Non-receptor serine/threonine protein kinase	Protein phosphorylation; Cytokinesis	*Mitosis*	20
ERBB2	2064	NM_004448	v-erb-b2 erythroblastic leukemia viral oncogene homologue 2, neuro/glioblastoma derived oncogene homolog (avian)	Tyrosine protein kinase receptor; growth factor; protein kinase	Protein phosphorylation; receptor protein tyrosine kinase signalling pathway; cell cycle control; cell proliferation and differentiation; oncogenesis	EGF receptor signalling pathway->EGFR; Cadherin signaling pathway->Epidermal growth factor receptor	17
BCL2	596	NM_000633	B-cell CLL/lymphoma 2	Other signalling molecule	Inhibition of apoptosis; oncogenesis	Oxidative stress response->b-cell lymphoma protein-2; Apoptosis signalling pathway->Bcl-2	18

Table 13.2 (continued)

Symbol	GeneID	Refseq	Description	Molecular function* *(italics: unclassified by Panther, comment added)*	Biological process* *(italics: unclassified by Panther, comment added)*	Pathway* *(italics: unclassified by Panther, comment added)*	Chromosome
ESR1	2099	NM_000125	Oestrogen receptor 1	Nuclear hormone receptor; Transcription factor; Nucleic acid binding	mRNA transcription regulation; steroid hormone-mediated signalling; other neuronal activity; oogenesis; cell cycle control; mitosis; cell proliferation and differentiation; cell motility	*Oestrogen Hormone Response*	6
CCNB1	891	NM_031966	Cyclin B1	Kinase activator	Cell cycle control	Cell cycle->Cyclin B; p53 pathway->Cyclin B	5
CD68	968	NM_001251	CD68 molecule	*Macrophage marker*	*Macrophage function*	*Inflammation*	17
PGR	5241	NM_000926	Progesterone receptor	Nuclear hormone receptor; transcription factor; nucleic acid binding	mRNA transcription regulation; steroid hormone-mediated signalling; other neuronal activity; other oncogenesis	*Progesterone hormone response*	11
GRB7	2886	NM_005310	Growth factor receptor-bound protein 7	Trans-membrane receptor regulatory/adaptor protein	Cell surface receptor mediated signal transduction	Angiogenesis->Growth Factor Receptor-Bound Protein 7; *HER-2 signalling*	17

Table 13.2 (continued)

Symbol	GeneID	Refseq	Description	Molecular function* *(italics: unclassified by Panther, comment added)*	Biological process* *(italics: unclassified by Panther, comment added)*	Pathway* *(italics: unclassified by Panther, comment added)*	Chromosome
MYBL2	4605	NM_002466	v-myb myeloblastosis viral oncogene homologue (avian)-like 2	Other transcription factor; nucleic acid binding	mRNA transcription regulation; inhibition of apoptosis; cell cycle control; cell proliferation and differentiation	*Cell proliferation*	20
BIRC5	332	NM_001168	Baculoviral IAP repeat-containing 5 (survivin)	Protease inhibitor	Inhibition of apoptosis	Angiogenesis-> Baculoviral IAP repeat-containing protein 5	17
BAG1	573	NM_004323	BCL2-associated athanogene	*Molecular function unclassified*	*Regulation of apoptosis*	*Unclassified*	9

* Pathway and function analysis was done using Protein ANalysis THrough Evolutionary Relationships (PANTHER) (Thomas et al. 2003, 2006)

(TAYLORx—http://www.cancer.gov/clinicaltrials/ECOG-PACCT-1). Results from randomized controlled trials (RCTs) will start to clarify the real potential of these markers and refine the indications for their clinical use.

13.4.7 New Areas for Biomarker Development—microRNA

A new area for biomarker development is microRNA (miRNA) expression measurement. The primary functions of miRNAs (short non-coding RNA molecules) in vertebrates are to both degrade mRNA transcripts and inhibit translation from them (Bartel 2004; Pillai et al. 2007). Recent experimental studies indicate a potential role for some miRNAs in cancer progression and invasion (Calin and Croce 2006; Esquela-Kerscher and Slack 2006; Nicoloso et al. 2009). For example, miR-210 has been found to be regulated by hypoxia and prognostic in breast cancer (Camps et al. 2008; Foekens et al. 2008), pancreatic adenocarcinoma (Greither et al. 2010), and in HNSCC (Gee et al. 2010). However, the modalities of application and the reproducibility of these assays have yet to be determined (Gee et al. 2008).

13.4.8 Chromosome Aberration

Other promising areas of marker development, although applications are still limited, are the use of genomic region copy-number and chromosomal instability analyses. A genomic instability index, defined as the fraction of genome altered as measured by copy number analysis in high-resolution array comparative genomic hybridization (aCGH), was found to be prognostic in ER-positive breast cancers (Chin et al. 2007b; p. 157). Furthermore, common genomic regions of alteration with strong association between copy number variant and gene expression (defined as 'hotspots') were found to be prognostic, independently from other clinical co-variates and NPI (Chin et al. 2007a). Another recent study showed that a gene expression signature associated with chromosomal instability and the total level of chromosomal aberrations was predictive of poor clinical outcome in 12 cancer datasets from 6 cancer types (Carter et al. 2006).

13.5 Biomarkers for Monitoring

Examples of biomarkers already used clinically to indicate presence of tumour or to monitor for recurrence after treatment, are: human chorionic gonadotropin (hCG), the expression of which is specific to trophoblastic neoplasms (Cole and Muller 2010); prostate-specific antigen (PSA) in prostate cancer; and cancer antigen 125

(ca125) in ovarian cancer, although there is some debate on their specificity and usefulness in large screenings (Hensley and Spriggs 2004; Ito 2009).

13.5.1 DNA Methylation

Another promising marker for monitoring the tumour before or during therapy is methylated DNA released in the blood circulation. DNA methylation is a common epigenetic change in cancer, and many genes are methylated in several cancer types (Jones and Baylin 2002). In several studies and cancer types, methylated DNA released in the circulation has been found associated with prognosis or response to treatment (Bastian et al. 2005; Brock et al. 2008; Fiegl et al. 2005; Mori et al. 2006; Ramirez et al. 2005; Wallner et al. 2006; Wei et al. 2008; Widschwendter et al. 2004). Although each of these studies has been relatively small, and the methodologies quite different, they provide strong cumulative evidence that circulating methylated DNA is a promising biomarker. However, in terms of clinical use it is still in early development, with many reproducibility and specificity issues yet to be addressed.

13.5.2 Mutated Plasma DNA

Mutated DNA is another circulating marker that has been suggested as a possibility for future monitoring (Diehl et al. 2008a) and also for early detection (Diehl et al. 2008b) in colorectal cancer. However, the sensitivity and reliability of methodology designed to detect mutated DNA at very early stages of the disease, where its concentration is lowest, has yet to be tested.

Recent genome-wide studies in cancer have suggested single-nucleotide polymorphisms (SNPs) as a risk factor in different cancers. Although these results constitute landmarks for cancer and genomic research, the practical applicability and usefulness of SNPs as screening or monitoring biomarkers is not yet clear (for a review see (Offit 2009)).

13.6 Measurement and Analysis of Biomarkers

Bio-banking initiatives are assisting in the collection of material, and there has been an exponential increase in possibilities for biomarker research (see Chap. 4 and 6). However, the technologies involved in biomarker studies and the type of data are constantly evolving, and this requires a constant effort to set and update guidelines that incorporate standards for the previous and newly developed techniques.

13.6.1 Key Measurement Technologies

Biomarker research is evolving rapidly and the number of techniques and applications has been increasing exponentially. Numerous biological platforms (See Chap. 3) have been developed in the past decades, and are now available for biomarker studies. Some of these, such as molecular and functional imaging (see e.g. (Brindle 2008; Harry et al. 2010)), allow a non-invasive *in-vivo* diagnosis or monitoring of the tumour before, during and/or after treatment. Others require blood samples or biopsy material (for a recent general review see e.g. http://www.nature.com/nrc/focus/biomarkers/index.html); these range from immunohistochemistry (Camp et al. 2008)), to genomics (van't Veer et al. 2005), genetics and epigenetics (Teschendorff et al. 2009), mass spectrometry and proteomics technologies (Latterich et al. 2008; Rodriguez et al. 2010).

We will focus in the following sub-paragraphs on two of the technologies that have been most extensively applied to biomarker studies to date. *In-vivo* imaging biomarkers will be discussed in the 'Pharmacokinetics and Pharmacodynamics—PD/PK' section of this chapter.

13.6.2 Tissue Arrays

A large proportion of biomarker research and clinical application has been based on immunohistochemistry (IHC) measurements, and with the advent of tissue microarray (TMA) techniques, IHC constitutes a widespread and powerful tool in clinical research studies (Bentzen et al. 2008; Camp et al. 2008). IHC provides a semi-quantitative measurement of protein expression by detecting antigens in histopathological samples using antibodies. A useful feature of this technique is that it provides a direct link between expression and location of the protein, both in terms of cell type (e.g. expressed cancer versus stroma cells) and the sub-cellular compartment.

One of the limitations in IHC studies is that antigenicity in FFPE samples is retained for many years, but not indefinitely. A recent study on 522 breast cancer patients, using a panel of well known biomarkers (e.g. ER, HER2, Ki-67), highlighted the risk of correlation between the intensity of the staining and sample collection date. Although most significant associations between IHC intensity and clinico-pathological characteristics were retained between analysis of fresh sections and old sections, a significant decrease in the number of samples classified as positive was observed in FFPE samples with a longer length of archival (Mirlacher et al. 2004). This highlights the importance of well-processed samples, but also the need for quality control studies to ensure that length of the storage does not compromise the staining results. Sample quality, antibody validation, and standardization of processing procedures are central to obtain good quality IHC data, and this is even more true with the advent of TMAs (for recent reviews see Bentzen et al. 2008; Camp et al. 2008).

Various techniques have been reported for constructing TMAs (Camp et al. 2008), but the basic concept is to collect cores from several FFPE samples on one paraffin block, so that any processing, such as IHC staining, can be performed si-

multaneously for hundreds of patients. A problem with studying the small cores present in TMAs, rather than whole-sample slides as done traditionally, is that the former might not be representative of the tumour variability in marker expression. Although reasonable concordance has been observed in studies comparing the two techniques, this requires marker-specific quality control, and it might be addressed by including multiple cores from the same sample (Camp et al. 2008).

Quantification of IHC considers several characteristics such as staining intensity, percentage of stained cells, and staining pattern, at various sub-cellular locations. Traditionally this has consisted of a qualitative scoring from human visual inspection, producing a semi-quantitative ordinal or nominal classification (e.g. weak, moderate and strong for intensity; organized, random for pattern, etc.). However, in some cases, approximately continuous scores have been adopted, e.g. for the percentage of stained cells. In the past decade, automated methods for scan and analysis of TMAs have been suggested (Bova et al. 2001; Jubb et al. 2003; Liu et al. 2002) and commercial software has been developed, such as the Aperio Digital Pathology Environment (http://www.aperio.com; for reviews see Camp et al. 2008; Giltnane and Rimm 2004). The adoption of more objective and automated methods may help to set more standardized scoring criteria.

Most TMA studies to date have used IHC; however, other methods such as *in situ* hybridisation (ISH) have been used (Braunschweig et al. 2004). Promising examples of use of this technology are the detection of copy number amplification, such as HER2 amplification in breast cancer (Penault-Llorca et al. 2009), and the detection of short nucleotide sequences such as miRNAs (Nuovo 2008; Nuovo et al. 2009).

13.6.3 Microarrays

Microarray technology has been successfully applied in several areas ranging from gene expression RNA (Shi et al. 2006) and miRNA analyses (Blenkiron and Miska 2007), to DNA analyses including those of copy number (Carter 2007), DNA methylation (Zilberman and Henikoff 2007) and SNP genotype, see e.g. (Ding and Jin 2009). In general, this technology has been seen as providing major potential for biomarker research, as it allows simultaneous genome-wide measurements. However, some of these applications are still relatively young. The only one that has been used widely to date in biomarker research is gene expression microarray.

13.6.4 RNA Analysis

Several platforms based on different technologies have been developed and widely employed in laboratory and biomarker research. However, the promising field of genomic expression has yet to deliver clinically validated and useful biomarkers, and translation to the clinic has proved challenging. A reproducibility problem may

arise from the heterogeneity of the technologies employed to measure gene expression. Several comparative studies have been looking into these problems; examples are the studies carried out by the MicroArray Quality Control (MAQC) project (Shi et al. 2006). The first MAQC study comprised 51 organizations, including platform providers, and was based on analysis of highly calibrated reference RNA pools. Although the overall conclusion suggested relatively high inter- and intra-platform reproducibility of gene expression measurements, not all platforms showed the same high agreement and reproducibility. Furthermore, this study was performed using very controlled RNA material, which is generally not available in clinical trials, and problems of reproducibility of these platforms in a clinical laboratory context were not addressed. Finally, the question of the consequences of this variability and the different genomic coverage for classification performance of existing signatures was not addressed, nor was the difference in platform performance in signature development. These problems are now being addressed in MAQC phase II (http://www.fda.gov/).

Further to global profiling platforms, several smaller platform assays based on real-time polymerase chain reaction (real-time PCR) are now also available, and are being evaluated in clinical studies. These might provide both an effective validation of signatures derived in global profiling studies, and in the long term, a more manageable assay for clinical application. An example is the TaqMan® Gene Sets card (https://products.appliedbiosystems.com), a 384-well micro fluidic card that enables simultaneous performance of gene expression assays in one to eight samples using a simple real-time PCR system, with minimal amounts of sample.

After the first enthusiastic wave of gene expression studies producing prognostic signatures, re-analyses of existing datasets have raised issues of reproducibility of results, and lack of consistency in methods and designs (Ein-Dor et al. 2005; Ioannidis et al. 2009; Michiels et al. 2005). Consequently, only a small proportion of gene expression signatures is currently being tested in clinical trials (examples for breast cancer in Table 13.1). These studies have stimulated researchers to look into reproducibility and robustness of the analyses and results (see Chap. 5). For example, several initiatives have proposed adoption of standardized processing, annotation (Brazma et al. 2001) and analysis (Ramasamy et al. 2008). Furthermore, genomic signatures are now more frequently derived and/or evaluated in large cohorts of patients or meta-analyses (see Table 13.1). However, it is now recognized that in order to transform genomic signatures into validated and useful clinical tools, more effort needs to be put into the design of good and efficient prospective studies and trials (Ahmed and Brenton 2005; Weigelt et al. 2010).

13.7 Identification, Standardization and Validation of Effective Biomarkers

Each cancer type represents a heterogeneous disease, often with several different genetic or molecular subtypes. It is crucial therefore to determine the degree of specificity of a biomarker to a given cancer, its variability between subtypes and

Table 13.3 Probability of obtaining a false positive or false negative result when performing a single statistical test

Decision / Real status	"No significant difference "	"Significant difference"
H0 (null hypothesis) is true **(e.g. no difference in relapse for patients with low or high marker level)**	Correct decision (TRUE NEGATIVE) P=1-α	**Type I error** (FALSE POSITIVE) P=α
Ha (alternative hypothesis) is true **(e.g. relapse is different between patients with low and high marker level)**	**Type II error** (FALSE NEGATIVE) P=β	Correct decision (TRUE POSITIVE) P=1-β

P=probability

α=confidence level; the probability of having observed the data, or more extreme data, given the null hypothesis H0 is true (typically α=0.05 if one test is performed)

1-β=power of the test; the probability of detecting a real difference as specified by Ha (β=10-20% gives a 80-90% power)

its evolution during cancer progression (Hutchinson and DeVita 2009). The clinical usefulness of a marker is heavily dependent on the reproducibility and specificity of its detection. Sensitivity of a test determines the proportion of true positives (Table 13.3), that is, the proportion of patients with a positive test result amongst the ones who will have cancer or will relapse. A test with high sensitivity will maximize true positives.

The specificity of a test determines the proportion of true negatives (Table 13.3), that is the proportion of patients with a negative test result amongst the ones who will not have cancer or will not relapse. A test with high specificity will maximize true negatives.

Sensitivity and specificity must be validated and optimized for each intended use of a biomarker. The development of a reproducible biomarker assay with high sensitivity and specificity is a multi-step process, requiring not only efficient study design but also standardization of methodologies and procedures at all levels, from discovery studies to prospective trials.

Although, as discussed above, most biomarker research is still some distance away from achieving such goals, important steps have been made towards improving the whole process. The WHO Good Clinical and Laboratory Practice (GCLP) guidelines have provided a framework for a quality system in analysis of clinical trial samples, ensuring Good Clinical Practice (GCP) compliance of processes and results (Stiles et al. 2003; WHO 2009). Specifically for cancer bio-

Table 13.4 Roadmap for development of diagnostic biomarkers recommended by the Cancer Research UK Biomarkers & Imaging Discovery & Development Committee. Adapted from http://science.cancerresearchuk.org/reps/pdfs/bidd_diagnostic_roadmap.pdf

Rationale	What is the clinical need?		
	Is retrospective material available from clinical trial or cohort?		
	Is a BM assay available?	Is it accurate a reproducible?	Assay Development Stage 1 (ADS1)
BM discovery and assay development	Define biomarker (BM) distribution in retrospective samples representative of target population		BM Discovery Stage1 (BMDS1)
	If BMDS1 indicates potential clinical utility	Refine assay and define Standard Operating Procedures (SOPs)	Assay Development Stage 2 (ADS2)
	How does BM compare with gold standard diagnostic tests in retrospective material?		BM Discovery Stage2 (BMDS2)
	If BMDS2 reveals improved diagnostic accuracy	Assay development to appropriate clinical standards	Assay Development Stage 3 (ADS3)
BM quantification	How does BM compare with gold standard diagnostic tests in a prospective study?		BM Quantification Stage 1 (BMQS1)
	If BMQS1 reveals improved diagnostic accuracy	Transfer to routine clinical practice	

markers research, the five levels of evidence (LOEs) of the American Society of Clinical Oncology (Hayes et al. 1996) offer a guideline for design and conduct of biomarker studies. Numerous initiatives have also begun to address and provide standards, including the Biomarkers Definitions Working Group (Group 2001); ASCO guidelines for use of biomarkers of breast cancer (Harris et al. 2007) and gastrointestinal cancer (Locker et al. 2006); and the FDA Harmonization initiatives and documents (http://www.fda.gov/RegulatoryInformation/Guidances/ucm129286.htm).

Furthermore, initiatives such as the Cancer Research UK roadmaps for biomarker discovery (Tables 13.4 and 13.5) provide summary guidance of the steps that should be taken for successful transfer of biomarker discovery into clinical practice. A similar initiative by the NCI recently published guidelines for biomarker development and use in early clinical trials (Dancey et al. 2010).

Finally, it is increasingly recognized that accurate reporting of marker studies aids in smoothing this process, and initial guidelines such as the REMARK criteria have been published to provide basic reporting requirements (McShane et al. 2005).

Table 13.5 Roadmap for prognostic/predictive biomarker development recommended by the Cancer Research UK Biomarkers & Imaging Discovery & Development Committee. Adapted from http://science.cancerresearchuk.org/reps/pdfs/bidd_prognostic_roadmap.pdf

Rationale	What is the clinical need?		
	Is retrospective material available from clinical trial or cohort?		
	Is a BM assay available?	Is it accurate a reproducible?	Assay Development Stage 1 (ADS1)
BM discovery and assay development	Define biomarker (BM) distribution in retrospective samples representative of target population		BM Discovery Stage1 (BMDS1)
	If BMDS1 indicates potential clinical utility	Refine assay and define Standard Operating Procedures (SOPs)	Assay Development Stage 2 (ADS2)
	Is BM associated with clinical outcome in retrospective material?		BM Discovery Stage2 (BMDS2)
	If BMDS2 reveals association with outcome	Assay development to appropriate clinical standards	Assay Development Stage 3 (ADS3)
BM quantification	Is BM associated with outcome in a prospective study (or in a prospective analysis of retrospective material)?		BM Quantification Stage 1 (BMQS1)
	If BMQS1 reveals robust association with outcome	Randomised Clinical Trial with stratification based upon BM*	BM Quantification Stage 2 (BMQS2)*
	If BMQS2 reveals improved clinical outcome*	Transfer to routine clinical practice	

*Optional steps

13.8 Annotated High-quality Clinical Samples

Biomarker research, and translational cancer research in general, rely heavily on large clinical series of well annotated and high-quality samples. In this respect, GCLP guidelines help to ensure good practice during the process of collection, annotation and analysis of samples (Stiles et al. 2003; WHO 2009).

Clinical trials often provide a good starting point for biomarker research, as patients are followed up regularly to investigate the treatment effects, and frequent samples are collected which can be used for further research. In fact, an increasing number of translational studies are conducted using *de-novo* clinical trials, so that

problems in data collection and annotation can be addressed at the setting-up stage of the trial. However, material from clinical trials, together with existing samples from retrospective series of patients, should be well annotated using standardized procedures to enable future use. Broader biobanking initiatives in various clinical and laboratory domains, including clinical trials, have come into being in the past decade. This effort implies a requirement to integrate, standardize and optimize several tasks and competences, ranging from collection of very diverse clinical and laboratory data, to efficient annotation, integration and management (see Chap. 2, 4, and 6). A recent initiative directed toward integrating and harmonizing existing bio-banking resources and technologies is the Biobanking and Biomolecular Resources Research Infrastructure (BBMRI). This is a European Union infrastructure (http://www.bbmri.eu/) with hundreds of members, including existing biobanking organisations, from cell line collections (e.g. the European Collection of Cell Cultures, http://www.hpacultures.org.uk/) to disease related biobanking infrastructures (e.g. the UK Confederation of Cancer Biobanks, http://www.oncoreuk.org).

13.9 Analyses and Simulations to Predict and Identify Biomarkers

A number of methods have been applied to the analysis of information-rich multiple biomarker data sets. For simplicity, they have historically been divided into two main branches: “class discovery” or “unsupervised”; and “class comparison” or “supervised” (Fig. 13.2; see paragraph 13.10 for extended definitions). However, complex and mixed approaches that go beyond this categorization have been suggested and are beginning to be applied to biomarker research; these will be discussed in the following sections.

13.10 Approaches to Data Analyses in Genomic Studies

13.10.1 Class Discovery and Class Prediction

Class discovery is an unsupervised approach that has been used increasingly in cancer and biomarker research over the past decade. In general, a classifier is developed that enables patients to be assigned to a given group based on their similarity with respect to a specific biomarker, such as gene expression profile. Clinical outcome is not considered during formation of the classifier, and the groups are compared with prognosis or other clinical variables only after the classifier is developed. The assumption behind this approach is that tumours with a similar pathological, molecular or genetic phenotype are likely to have not only a similar prognosis,

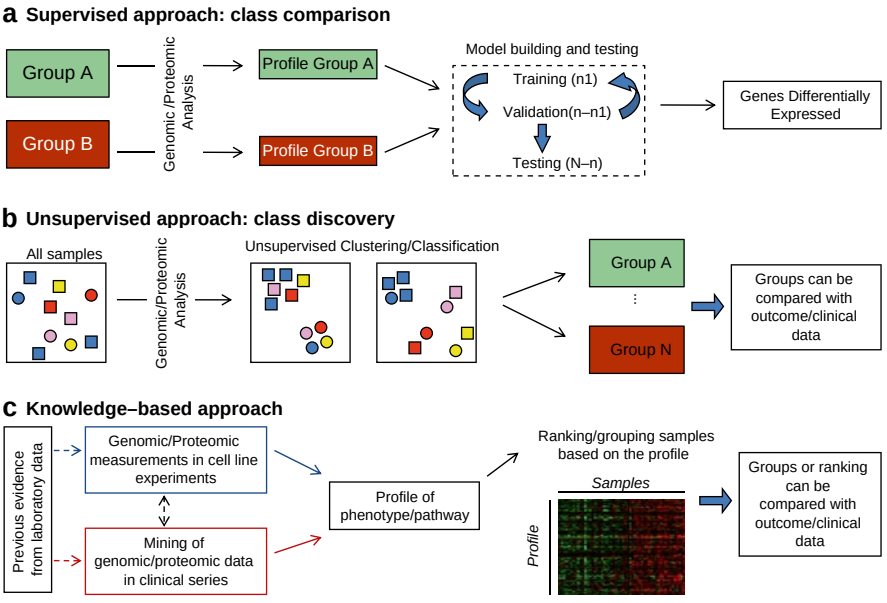


Fig. 13.2 Schematic representation of different approaches to the analysis of information-rich multiple biomarker datasets. a Class comparison is a supervised analysis approach where samples or patients are assigned to groups depending on their association with clinical outcome or other clinical data; the model identifies markers that maximize the contrast in outcome between classes. Because this approach could provide biased results and a high number of false discoveries, it is important to validate and test the classifier in independent cohorts, or to use cross-validation and leave-one-out methodologies (*dotted box*, see this section for details). **b** Class discovery is an unsupervised approach. A classifier is developed that assigns patients to a given group based on their similarity with respect to the specific biomarker considered (e.g. gene expression profile). Clinical outcome is not considered in the formation of the classifier; however, the groups can be compared to prognosis or other clinical variables after the classifier is developed. **c** Knowledge-based bottom-up approaches use different pieces of pre-existing knowledge, deriving for example from cell line systems, to develop a classifier. This knowledge can be used directly, e.g. a signature derived from experiments using cell lines can be applied directly to classify patients (*blue box*), or can be incorporated into data-mining of clinical samples (*red box*) and then applied to classify patients

but also a similar benefit for a specific treatment. Furthermore, grouping tumours into subtypes might help in identifying more specific therapeutic targets. One advantage of the class discovery approach is that it can be very efficient when there are large causal gaps between marker, phenotype, and prognosis. Another important advantage is that an eventual association with prognosis will be unbiased, provided that the groups are built using unbiased clustering methods; for a review see for example (Hastie et al. 2001). However, subtle and/or complex relationships with clinical outcome and therapy efficacy might not be detected.

Class comparison has been the more traditional approach to identifying prognostic markers. It is a supervised analysis approach where the model or classifier

is constructed by seeking direct association with the outcome data or other clinical data. This means that the model identifies markers that maximize contrast in outcome between classes. However, this approach can provide somewhat biased results in that when a large number of co-variates are present, relative to the number of outcome events it can result in false associations and a high number of false discoveries (for a review see Hastie et al. 2001). Thus, it is very important either that the model is validated and tested in independent cases, or that cross-validation and leave-one-out methodologies are used.

Cross-validation consists of partitioning the datasets into subsets in a recursive, iterative manner, so that the model is cyclically trained on one subset and validated on the other. For example, when using N-fold cross-validation, the datasets are partitioned into N subsets; N-1 of them forms the training set, 1 forms the validation set (Fig. 13.2a, *dotted box*). This process is repeated N times so that each of the subsets will be used once for the validation; a final average is taken over the subsets as a whole.

A **leave-one-out** approach consists of removing one case (that is a single sample) at each iteration from the model building; fitting the model to the remaining patients, for example using N-fold cross-validation; and then predicting treatment response for the left-out patient. This operation is repeated, permitting the sensitivity and specificity of the model to be estimated.

Class discovery and class comparison have been used extensively in bioinformatics analyses in biomarker studies. Specifically, in breast cancer these approaches have been used to derive gene expression signatures that have shown prognostic potential (Table 13.1). However, other approaches have been suggested that either complement or combine these two strategies.

13.10.2 Gene and Protein Networks

A data-mining approach that has been extensively applied to genomic and proteomic studies is the theory and analysis of networks. For example, protein interaction networks have been defined in which proteins are ‘nodes’ and predicted or validated protein-protein interactions are ‘edges’ (for a review see Walhout and Vidal 2001). Another application has been to define gene networks in which nodes represent genes, and edges can represent features such as regulation or co-expression. These edges are created between genes when their expression patterns are significantly correlated, for example, in a dataset of clinical samples (for a review see Bower and Bolouri 2001). Although protein and gene networks have not hitherto been used directly as biomarkers to classify patients, a recent study has shown that differences in structure of the human protein interaction network in breast cancer might be associated with poorer prognosis (Taylor et al. 2009). Furthermore, protein and gene networks can be used to formulate functional and clinical hypotheses which might then enable development and/or identification of biomarkers (Butte and Kohane 2003; Hahn and Kern 2005; Wolfe et al. 2005).

13.10.3 Knowledge-based Class Comparison

In the past few decades, a technological revolution in cancer research has enabled rapid collection of a large amount of information on different biological systems at a scale not seen before. This has encouraged the use of bottom-up approaches, where different components of pre-existing knowledge are integrated to inform the data analysis (Fig. 13.2). The simplest and most widely applied bottom-up approach consists of analysis of over-represented pathways or functional categories in gene lists obtained using class comparison analysis. Several tools have been published for this type of analysis, including the Protein ANalysis THrough Evolutionary Relationships (PANTHER) classification system and related tools (Thomas et al. 2003), Genecodis annotation and analysis tools (Carmona-Saez et al. 2007), and Ingenuity Pathway analysis tools (http://www.ingenuity.com). The concept behind this approach is that single genes in a given relevant pathway can be missed in a class comparison analysis due to the large number of genes analysed and the need to correct for multiple testing. However, the overall number of genes that are statistically significant in the class comparison and also involved in the pathway will be higher than observed by chance. Also, a pathway analysis might provide some functional characterizations of the classes in question. An increasingly popular example of this application is gene-set enrichment analysis (GSEA) (Subramanian et al. 2005). By this method, a class comparison is conducted to rank objects such as single gene expressions according to their usefulness in discriminating between classes. Previously defined gene sets of interest are then mapped onto this, the goal being to identify sets containing a statistically significant number of highly ranked genes. In general, such methods are relatively powerful and conceptually simple, but since they rely on class comparison, they are not suitable for defining phenotypes in heterogeneous clinical data sets.

13.10.4 Knowledge-based Class Prediction and Mining of Genomic Data

Another bottom-up approach employs cell-line derived gene expression signatures to classify clinical samples with respect to a given phenotype (Table 13.1). Processes such as wound healing or hypoxia have long been known to be involved in cancer progression. Gene expression signatures for these processes, derived from cell line experiments using simple class comparison designs, have been mapped to expression data from retrospective cancer studies and could stratify patients into distinct prognostic groups (Chang et al. 2005; Chi et al. 2006).

Despite cell-line diversity, derivation of process signatures in these model systems can be powerful, as many fundamental processes are conserved, and clean experimental design can be easily applied. In contrast, the *in-vivo* tumour system requires consideration of multiple cell types, microenvironmental changes, and

three-dimensional complexity, so that each system type provides information complementary to that of the others. Approaches that integrate knowledge about gene function from *in-vitro* experiments together with the analysis of pathway interactions *in-vivo* may deliver signatures that are more representative of gene expression and pathway activation occurring in cancer.

Such an approach, developed within the context of gene networks, has been used, for example to derive a hypoxia signature in head-and-neck cancer (Winter et al. 2007), and in a large meta-analysis of head-and-neck and breast cancer (Buffa et al. 2010). Multiple hypoxia prototype genes (or seeds) were considered; these are genes known from previous studies to be related with hypoxia. These prototypes were used as a starting point from which to build a co-expression network. This approach defined a compact hypoxia signature which was prognostic in multiple cancer types and resulted more efficient than the direct application of signatures derived from cell line experiments (Buffa et al. 2010). A similar approach was used in a recent study to derive signatures of major pathways (or processes) known to affect breast cancer progression, such as proliferation, tumour invasion, immune response, angiogenesis, apoptosis, ER and HER2 signalling (Desmedt et al. 2008; Wirapati et al. 2008). These studies used one prototype per given biological process; signatures for each process were derived by identifying genes specifically co-expressed with each prototype gene. This method identified prognostic signatures in a large cohort of breast cancer patients. Combined results from these studies show that the use of multiple prototypes, with respect to a single prototype gene, delivered an improved signature. However, the initial choice of prototypes can affect the signature derived; hence continuous optimization of the prototypes, based on newly acquired knowledge, is crucial.

13.10.5 Literature Data-mining and Data Repositories

Other knowledge-based approaches have been suggested where the existing literature is mined directly to obtain functional and clinical information on given biomarkers or biological processes that aid in the prioritization of biomarkers. An example of this type of application is Biovista's Adverse Event Analysis technology (http://www.biovista.com/) which analyses mechanisms of action-based correlations between drugs, disease efficacy and adverse events, by cross-correlating existing databases. This technology has recently been used in collaboration with the FDA to explore the possibility of predicting which patients may be at risk of adverse reactions for a given therapy.

It is clear that knowledge-based bottom-up approaches rely heavily on well-annotated genomic/proteomic databases, and pathway/gene function resources. Several such resources exist, with different information and different levels of data curation. These include the repositories and tools of the National Center for Biotechnology Information (NCBI—http://www.ncbi.nlm.nih.gov/); the European Bioinformatics Institute (EBI—http://www.ebi.ac.uk/); and the UCSC and ENCODE consortium 9http://genome.ucsc.edu/). Although an increased effort has been put into harmo-

nizing, standardizing and integrating annotation and data from multiple repositories, this task requires a continuous effort, especially in view of the fast discovery rate of new technologies and dimensions of knowledge.

13.11 Meta-analyses of Biomarker Studies

In any statistical analysis, for any given significance level, the larger the sample size (number of cases), the greater will be the power to detect a given difference (Table 13.3). Thus the first step in design of a clinical trial is to estimate the number of cases needed to achieve the desired power, given a predefined acceptable significance level.

In biomarker research, however, small study size and thus low power are frequent, particularly in the case of studies performed on retrospective series. Although the use of carefully collected retrospective material is crucial to the biomarker discovery phase, often the sample collection has not been designed with the biomarker study question in mind so that it can be undersized for the specific tests employed. There has been an unfortunate lack of prospective validation studies designed to address the utility of specific previously identified biomarkers, mostly in consideration of the size such studies require and the costs involved. Another way of addressing this problem, already in use, is to conduct large meta-analyses, gaining power by pooling existing undersized retrospective studies.

There are several examples where such meta-analyses have been able to identify and prioritize promising markers, or to exclude markers showing insufficient evidence for continued study. However, it has proved difficult to design and use meta-analysis, and there are few large biomarker studies using meta-analytical methods, compared with the far greater number of available retrospective studies.

Reasons for these difficulties include incomple or non-existent reporting of biomarker studies; difficulty in accessing original data and material; and sometimes large differences in biomarker measurement protocols or the clinical data recorded. Most retrospective studies were conducted, or the material for them was collected, before initiatives aiming to standardize laboratory and clinical procedures, such as GCLP, and to study reporting, such as REMARK (McShane et al. 2005), were brought into place. In the future, large integrated and well-annotated biobank resources should assist meta-analysis studies.

13.12 Quantitative Simulations of Major Pathways Leading to Biomarker Development

In-silico modelling and simulations have been applied to medicine, and cancer, for many years (for a review see Wheldon 1988). The range of models covers such topics as cancer proliferation growth, drug response and radiation and

radiotherapy response. This type of approach has contributed greatly to our understanding of cancer behaviour, to formulating new hypotheses, and to devising new treatment approaches capable of being tested in laboratory or preclinical systems. Although examples abound of the application of these approaches to cell biology and physiology, their application to the study of the dynamics of molecular pathways is more recent. Furthermore, now that it has been recognized that tumour behaviour and growth is determined by the interaction between tumour cells and the tumour environment, it is important for the discovery of biomarkers to address this heterogeneity and to be able to quantify these components, as well as their crosstalk in time, by modelling cellular and extra-cellular pathways.

13.12.1 Simulation of Cancer Pathways: The EGFR Pathway

Study of the pathways involved in cancer progression has identified many potentially useful prognostic and predictive markers and related drug targets. Examples discussed in the previous sections are the receptor tyrosine kinases (RTKs); RTKs are one of the most studied and better characterised pathways in cancer, and several drugs have been developed for the purpose of directly or indirectly targeting these kinases. Examples with demonstrated potential are targeting of the EGFR and HER2 pathways (see Sect. 13.3 Biomarkers for prediction). A better and more profound understanding of cancer progression pathways is crucial not only to cancer research, but also to biomarker and drug development. The technological revolution in molecular biology methods over the past few decades has introduced high-throughput techniques facilitating large-scale study of molecular pathways using genomics and proteomics, as discussed above. In parallel, the development of new imaging techniques has enormously increased our ability to acquire *in-vivo* information and to quantify the dynamics of molecular and chemical pathways. This is opening up exciting possibilities for developing and using *in-silico* systems to study the dynamics of molecular and chemical pathways; recent studies of the EGFR signalling pathway are a good example. This constitutes one of the natural starting points for *in-silico* models, as it is a key pathway in cancer progression and prognosis, and as such has been highly studied and better characterized than other cancer progression pathways. It is, however, a very complex pathway and *in-silico* studies have used simulation in an attempt to improve our understanding (Blinov et al. 2006; Brown et al. 2004; Hornberg et al. 2005; Orton et al. 2009; Samaga et al. 2009; Sasagawa et al. 2005; Schoeberl et al. 2002; Wang et al. 2009). Although the direct application of these methods to drug and biomarker development studies is still limited, some examples exist that go some way towards this goal; one such is the modelling of RTK pathways (for a review see Amit et al. 2007) and their interaction with kinase inhibitor treatments (Araujo et al. 2005).

13.12.2 Databases and Repositories of Models

The simulations described above have been increasingly used to formulate hypotheses, design experiments and study the effect of perturbing various elements of a given pathway (for a review see Kholodenko 2006). To facilitate this, repositories and databases of models have been created, such as the European Bioinformatics Institute (EBI) BioModels database (Le Novere et al. 2006). This database stores published bio-mathematical models and permits searching and retrieving of these models; models are also annotated and linked to other resources such as databases of compounds and pathways, or the gene ontology database. Furthermore, various approaches and associated implementation of tool repositories have been developed to facilitate the construction of models and simulations of biological systems (Adalsteinsson et al. 2004; Dematte et al. 2008; Dill et al. 2007; Gilmore and Hillston 1994; Hoops et al. 2006; Hucka et al. 2001; Ramsey et al. 2005; for a review, see Fisher and Henzinger 2007).

13.13 Pharmacokinetics and Pharmacodynamics

The practical possibility of performing functional and molecular imaging in the clinical setting and in clinical trials, using technologies such as Positron Emission Tomography (PET) or Magnetic Resonance Imaging (MRI), has significantly increased in the past decades. Quantitative approaches and tools for advanced analyses of imaging data are being continuously developed (Flower and Webb 2010; Harry et al. 2010). This has facilitated an increase in the application of pharmacokinetic (PK) and pharmacodynamic (PD) approaches to biomarker research and drug development. PK approaches analyse and quantify the kinetics of drugs or compounds in the body, and thus their absorption, distribution and elimination; PD approaches study their action and physiological effect. In the context of cancer, PK enables us to closely monitor drug exposure and PD to measure endpoints indicative of the effect of the therapy. One such PK-PD model is used in the study assessing the effect of vascular endothelial growth factor (VEGF) drug inhibitors by adopting magnetic resonance imaging (MRI).

VEGF is a growth factor that has been shown to promote angiogenesis (for a review see Fox et al. 2007), tumour growth and invasion. Several agents, such as Avastin (bevacizumab), have been developed which inhibit VEGF; however, their usefulness, and the subgroups of patients who might benefit from these drugs, has not yet been fully determined. PK and PD parameters related to MRI-based studies have been applied, along with other biomarkers, to studying individual patient response to drugs targeting VEGF (for a recent review see Murukesh et al. 2010).

A systematic use of this type of analysis could facilitate the optimization of drug regimens, the selection of patients for specific treatments, and the selection in small early-phase trials of drugs to be entered in larger late-phase clinical trials (for re-

views see Sarker and Workman 2007; Walko and McLeod 2009; West et al. 2004). However, the application of these approaches in clinical practice and in clinical trials has so far been limited. This has been partly due to lack of standardization of procedures and in guidelines for PK measurements or PD endpoints to be considered, particularly in drug development and early-phase clinical trials, or in preclinical testing. Recent publications and initiatives are suggesting possible ways to begin to address this (Kinders et al. 2007; Takimoto 2009).

However, the method using Positron Emission Tomography plus 2-[fluorine-18] fluoro-2-deoxy-d-glucose (FDG PET) has been shown to be useful in assessing the tumour response to Imatinib (Glivec), a tyrosine kinase inhibitor, and has now been included in the revised Response Evaluation Criteria in Solid Tumours (RECIST) to complement computed tomography (CT) in determining disease progression. RECIST was published in 2000 and has provided guidelines for measuring tumour response to treatment (Therasse et al. 2000). RECIST is intended to be used in clinical trials where the primary endpoint is the objective response of tumours to treatment. A one-dimensional assessment of all lesions is made, and the tumour burden is assessed by summing the longest diameter of measurable lesions. There are four categories of responses:

- complete response (CR): disappearance of all measurable lesions;
- partial response (PR): 30% or larger decrease from baseline in the sum of the longest tumour diameter of all measurable lesions;
- progressive disease (PD): 20% or more increase from nadir (lowest point) in the sum of the diameter;
- stable disease (SD): everything between 30% decrease and 20% growth.

Although these criteria have been used by various groups and in various trials, several limitations have been identified, one of them being that functional imaging modalities such as PET and MRI were not considered (Benjamin et al. 2007; Choi 2008). A revised version has therefore been published (Eisenhauer et al. 2009) which addresses some of the issues raised, including the use of functional imaging techniques.

13.14 Integrated Approaches to Biomarker Discovery and Development

As highlighted above, thanks to a vast growth in the number of high-throughput proteomics and genomic platforms and technologies, the availability of new imaging techniques, and the unprecedented availability of multi-level large-scale data, we now have greatly enhanced possibilities for changing the way we do cancer research, and for dramatically increasing the speed at which new findings are translated to the clinic. Although this promise has not yet been fulfilled in terms of the number, efficacy or availability of new biomarkers and treatments, real improve-

ments may soon become achievable through better methods of standardization, harmonization and integration of assays and technologies.

Integration has been already applied in specific areas, such as the use of multigene signatures providing information on one or multiple pathways, for breast cancer prognosis (see Table 13.1), for which Oncotype DX® and Mammaprint® are now being tested in clinical trials (see *Prognostic Markers* Section). Furthermore, the combination of different signatures has been shown to improve breast cancer stratification with respect to outcome (for a review see van't Veer et al. 2005; Ignatiadis and Sotiriou 2008). Similarly, in the context of breast cancer response to neoadjuvant therapy (see Chap. 2), an international expert panel recently highlighted the possibility that a combination of different markers and the establishment of gene-expression profiles, rather than a single biomarker, might provide a more accurate prediction (Kaufmann et al. 2007).

Furthermore, the combination of technological platforms providing complementary genomic information has shown great potential for improving our understanding of disease, classification and prognostication. For example, a combination of copy number variant information and gene expression measurements has resulted in improved classification and outcome stratification of breast cancer (Chin et al. 2007b, Chin et al. 2006). Also, integration of miRNA with gene expression profiles has served to identify and characterize potential prognostic markers in breast cancer (Buffa et al. 2011, under revision). On another level, integration of genomic information with functional imaging data might provide a more specific assessment of tumour response to therapy, with the possibility of targeting pathways involved in tumour escape or recovery from treatment.

This integration effort has been pursued by international initiatives devoted to developing tools and platforms to facilitate effective and rapid exchange of knowledge and information between different knowledge and research domains, with the aim of accelerating discovery and translation of new findings to the clinic. Although different approaches and technologies are used, examples of initiatives pursuing this goal in cancer research are the cancer Biomedical Informatics Grid (caBIG, https://cabig.nci.nih.gov/overview/); Advancing Clinico Genomic Trials in cancer (ACGT, http://www.eu-acgt.org/); CancerGrid UK (http://www.cancergrid.org/); and the NCRI and the Oncology Information Exchange platform (ONIX, http://www.ncri-onix.org.uk). In the specific domain of breast cancer clinical and translation research, the Breast International Group (BIG) (http://www.breastinternationalgroup.org/) is an example of an initiative aimed towards enhancing collaborations and harmonisation of research and clinical trials in breast cancer.

However, the move towards large-scale harmonization, standardization and integration of multi-level data, within and between research and clinical communities, is still in its infancy. A continuous effort is needed to ensure that the potential offered by this hitherto unimaginable availability of novel data types and information is fully exploited. Conversely, this integration effort will also facilitate comparison of new biomarkers with current treatment selection criteria, hopefully elucidating those cases for which the new biomarkers offer a significant benefit with respect to current clinical practice, and how they should be applied.

References

Adalsteinsson D, McMillen D, Elston TC (2004) Biochemical Network Stochastic Simulator (BioNetS): software for stochastic modeling of biochemical networks. BMC Bioinformatics 5:24

Ahmed AA, Brenton JD (2005) Microarrays and breast cancer clinical studies: forgetting what we have not yet learnt. Breast Cancer Res 7:96–99

Albain KS, Barlow WE, Shak S, Hortobagyi GN, Livingston RB, Yeh IT, Ravdin P, Bugarini R, Baehner FL, Davidson NE, Sledge GW, Winer EP, Hudis C, Ingle JN, Perez EA, Pritchard KI, Shepherd L, Gralow JR, Yoshizawa C, Allred DC, Osborne CK, Hayes DF (2010) Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. Lancet Oncol 11:55–65

Amit I, Wides R, Yarden Y (2007) Evolvable signaling networks of receptor tyrosine kinases: relevance of robustness to malignancy and to cancer therapy. Mol Syst Biol 3:151

Araujo RP, Petricoin EF, Liotta LA (2005) A mathematical model of combination therapy using the EGFR signaling network. Biosystems 80:57–69

Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116:281–297

Bastian PJ, Ellinger J, Wellmann A, Wernert N, Heukamp LC, Muller SC, Ruecker A von (2005) Diagnostic and prognostic information in prostate cancer with the help of a small set of hypermethylated gene loci. Clin Cancer Res 11:4097–4106

Benjamin RS, Choi H, Macapinlac HA, Burgess MA, Patel SR, Chen LL, Podoloff DA, Charnsangavej C (2007) We should desist using RECIST, at least in GIST. J Clin Oncol 25:1760–1764

Bentzen SM, Atasoy BM, Daley FM, Dische S, Richman PI, Saunders MI, Trott KR, Wilson GD (2005) Epidermal growth factor receptor expression in pretreatment biopsies from head and neck squamous cell carcinoma as a predictive factor for a benefit from accelerated radiation therapy in a randomized controlled trial. J Clin Oncol 23:5560–5567

Bentzen SM, Buffa FM, Wilson GD (2008) Multiple biomarker tissue microarrays: bioinformatics and practical approaches. Cancer Metastasis Rev 27:481–494

Benvenuti S, Sartore-Bianchi A, Di Nicolantonio F, Zanon C, Moroni M, Veronese S, Siena S, Bardelli A (2007) Oncogenic activation of the RAS/RAF signaling pathway impairs the response of metastatic colorectal cancers to anti-epidermal growth factor receptor antibody therapies. Cancer Res 67:2643–2648

Blenkiron C, Miska EA (2007) miRNAs in cancer: approaches, aetiology, diagnostics and therapy. Hum Mol Genet 16 Spec No 1:R106–R113

Blinov ML, Faeder JR, Goldstein B, Hlavacek WS (2006) A network model of early events in epidermal growth factor receptor signaling that accounts for combinatorial complexity. Biosystems 83:136–151

Bova GS, Parmigiani G, Epstein JI, Wheeler T, Mucci NR, Rubin MA (2001) Web-based tissue microarray image data analysis: initial validation testing through prostate cancer Gleason grading. Hum Pathol 32:417–427

Bower JM, Bolouri H (2001) Computational modeling of genetic and biochemical networks. The MIT Press, Cambridge

Braunschweig T, Chung JY, Hewitt SM (2004) Perspectives in tissue microarrays. Comb Chem High Throughput Screen 7:575–585

Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet 29:365–371

Brindle K (2008) New approaches for imaging tumour responses to treatment. Nat Rev Cancer 8:94–107

Brock MV, Hooker CM, Ota-Machida E, Han Y, Guo M, Ames S, Glockner S, Piantadosi S, Gabrielson E, Pridham G, Pelosky K, Belinsky SA, Yang SC, Baylin SB, Herman JG (2008) DNA methylation markers and early recurrence in stage I lung cancer. N Engl J Med 358:1118–1128

Brown KS, Hill CC, Calero GA, Myers CR, Lee KH, Sethna JP, Cerione RA (2004) The statistical mechanics of complex signaling networks: nerve growth factor signaling. Phys Biol 1:184–195

Buffa FM, Camps C, Winchester L, Snell CE, Gee HE, Sheldon H, Taylor M, Harris AL, Ragoussis J (2011) microRNA-associated progression pathways and potential therapeutic targets identified by integrated mRNA and microRNA expression profiling in breast cancer. Cancer Res. In Press

Buffa FM, Harris AL, West CM, Miller CJ (2010) Large meta-analysis of multiple cancers reveals a common, compact and highly prognostic hypoxia metagene. Br J Cancer 102:428–435

Butte AJ, Kohane IS (2003) Relevance networks: a first step towards finding genetic regulatory networks within microarray data. In: Parmigiani G, Gar-rett ES, Irizarry RA, Zeger S (eds) The analysis of gene expression data, Springer, New York

Buyse M, Loi S, van't Veer L, Viale G, Delorenzi M, Glas AM, d'Assignies MS, Bergh J, Lidereau R, Ellis P, Harris A, Bogaerts J, Therasse P, Floore A, Amakrane M, Piette F, Rutgers E, Sotiriou C, Cardoso F, Piccart MJ (2006) Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. J Natl Cancer Inst 98:1183–1192

Calin GA, Croce CM (2006) MicroRNA signatures in human cancers. Nat Rev Cancer 6:857–866

Camp RL, Neumeister V, Rimm DL (2008) A decade of tissue microarrays: progress in the discovery and validation of cancer biomarkers. J Clin Oncol 26:5630–5637

Camps C, Buffa FM, Colella S, Moore J, Sotiriou C, Sheldon H, Harris AL, Gleadle JM, Ragoussis J (2008) hsa-miR-210 Is induced by hypoxia and is an independent prognostic factor in breast cancer. Clin Cancer Res 14:1340–1348

Caponigro F, Basile M, Rosa V de, Normanno N (2005) New drugs in cancer therapy, National Tumor Institute, Naples, 17–18 June 2004. Anticancer Drugs 16:211–221

Carmona-Saez P, Chagoyen M, Tirado F, Carazo JM, Pascual-Montano A (2007) GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. Genome Biol 8:R3

Carter NP (2007) Methods and strategies for analyzing copy number variation using DNA microarrays. Nat Genet 39:S16–S21

Carter SL, Eklund AC, Kohane IS, Harris LN, Szallasi Z (2006) A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. Nat Genet 38:1043–1048

Chang HY, Nuyten DS, Sneddon JB, Hastie T, Tibshirani R, Sorlie T, Dai H, He YD, van't Veer LJ, Bartelink H, Rijn M van de, Brown PO, Vijver MJ van de (2005) Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. Proc Natl Acad Sci USA 102:3738–3743

Chi JT, Wang Z, Nuyten DS, Rodriguez EH, Schaner ME, Salim A, Wang Y, Kristensen GB, Helland A, Børresen-Dale AL, Giaccia A, Longaker MT, Hastie T, Yang GP, Vijver MJ van de, Brown PO (2006) Gene expression programs in response to hypoxia: cell type specificity and prognostic significance in human cancers. PLoS Med 3:e47

Chin SF, Teschendorff AE, Marioni JC, Wang Y, Barbosa-Morais NL, Thorne NP, Costa JL, Pinder SE, Wiel MA van de, Green AR, Ellis IO, Porter PL, Tavare S, Brenton JD, Ylstra B, Caldas C (2007a) High-resolution aCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. Genome Biol 8:R215

Chin SF, Wang Y, Thorne NP, Teschendorff AE, Pinder SE, Vias M, Naderi A, Roberts I, Barbosa-Morais NL, Garcia MJ, Iyer NG, Kranjac T, Robertson JF, Aparicio S, Tavare S, Ellis I, Brenton JD, Caldas C (2007b) Using array-comparative genomic hybridization to define molecular portraits of primary breast cancers. Oncogene 26:1959–1970

Choi H (2008) Response evaluation of gastrointestinal stromal tumors. Oncologist 13 (Suppl 2):4–7

Cole LA, Muller CY (2010) Hyperglycosylated hCG in the management of quiescent and chemorefractory gestational trophoblastic diseases. Gynecol Oncol 116:3–9

Dancey, JE, Dobbin KK, Groshen S, Jessup JM, Hruszkewycz AH, Koehler M, Parchment R, Ratain MJ, Shankar LK, Stadler WM, True LD, Gravell A, Grever MR (2010) Guidelines for the development and incorporation of biomarker studies in early clinical trials of novel agents. Clin Cancer Res 16(6):1745–1755

de Azambuja E, Cardoso F, Castro G de Jr, Colozza M, Mano MS, Durbecq V, Sotiriou C, Larsimont D, Piccart-Gebhart MJ, Paesmans M (2007) Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12,155 patients. Br J Cancer 96:1504–1513

Dematte L, Priami C, Romanel A (2008) The Beta Workbench: a computational tool to study the dynamics of biological systems. Brief Bioinform 9:437–449

Desmedt C, Piette F, Loi S, Wang Y, Lallemand F, Haibe-Kains B, Viale G, Delorenzi M, Zhang Y, d'Assignies MS, Bergh J, Lidereau R, Ellis P, Harris AL, Klijn JG, Foekens JA, Cardoso F, Piccart MJ, Buyse M, Sotiriou C (2007) Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. Clin Cancer Res 13:3207–3214

Desmedt C, Haibe-Kains B, Wirapati P, Buyse M, Larsimont D, Bontempi G, Delorenzi M, Piccart M, Sotiriou C (2008) Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. Clin Cancer Res 14:5158–5165

Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, Thornton K, Agrawal N, Sokoll L, Szabo SA, Kinzler KW, Vogelstein B, Diaz LA Jr (2008a) Circulating mutant DNA to assess tumor dynamics. Nat Med 14:985–990

Diehl F, Schmidt K, Durkee KH, Moore KJ, Goodman SN, Shuber AP, Kinzler KW, Vogelstein B (2008b) Analysis of mutations in DNA isolated from plasma and stool of colorectal cancer patients. Gastroenterology 135:489–498

Dill DL, Knapp MA, Gage P, Talcott C, Laderoute K, Lincoln P (2007) The pathalyzer: a tool for analysis of signal transduction pathways. Systems biology and regulatory genomics. Springer, Berlin

Ding C, Jin S (2009) High-throughput methods for SNP genotyping. Methods Mol Biol 578:245–254

Ein-Dor L, Kela I, Getz G, Givol D, Domany E (2005) Outcome signature genes in breast cancer: is there a unique set? Bioinformatics 21:171–178

Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M, Rubinstein L, Shankar L, Dodd L, Kaplan R, Lacombe D, Verweij J (2009) New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 45:228–247

Elston CW, Ellis IO (1991) Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology 19:403–410

Eriksen JG, Steiniche T, Overgaard J (2005) The influence of epidermal growth factor receptor and tumor differentiation on the response to accelerated radiotherapy of squamous cell carcinomas of the head and neck in the randomized DAHANCA 6 and 7 study. Radiother Oncol 74:93–100

Esquela-Kerscher A, Slack FJ (2006) Oncomirs—microRNAs with a role in cancer. Nat Rev Cancer 6:259–269

Fan C, Oh DS, Wessels L, Weigelt B, Nuyten DS, Nobel AB, van't Veer LJ, Perou CM (2006) Concordance among gene-expression-based predictors for breast cancer. N Engl J Med 355:560–569

Ferlay J, Bray F, Pisani P, Parkin D (2004) In: Press I (ed) GLOBOCAN 2002: cancer incidence, mortality and prevalence worldwide. Lyon, France

Fiegl H, Millinger S, Mueller-Holzner E, Marth C, Ensinger C, Berger A, Klocker H, Goebel G, Widschwendter M (2005) Circulating tumor-specific DNA: a marker for monitoring efficacy of adjuvant therapy in cancer patients. Cancer Res 65:1141–1145

Fisher J, Henzinger TA (2007) Executable cell biology. Nat Biotechnol 25:1239–1249

Flower MA, Webb S (2010) The physics of medical imaging. Taylor and Francis, Abingdon

Foekens JA, Sieuwerts AM, Smid M, Look MP, Weerd V de, Boersma AW, Klijn JG, Wiemer EA, Martens JW (2008) Four miRNAs associated with aggressiveness of lymph node-negative, estrogen receptor-positive human breast cancer. Proc Natl Acad Sci U S A 105:13021–13026

Fox SB, Generali DG, Harris AL (2007) Breast tumour angiogenesis. Breast Cancer Res 9:216

Gee HE, Camps C, Buffa FM, Colella S, Sheldon H, Gleadle JM, Ragoussis J, Harris AL (2008) MicroRNA-10b and breast cancer metastasis. Nature 455:E8–E9; author reply E9

Gee HE, Camps C, Buffa FM, Patiar S, Winter SC, Betts G, Homer J, Corbridge R, Cox G, West CML, Ragoussis J, Harris AL (2010) hsa-mir-210 is a marker of tumor hypoxia and a prognostic factor in head and neck cancer. Cancer 116(9):2148–2158

Gilmore S, Hillston J (1994) The PEPA workbench: a tool to support a process algebra-based approach to performance modelling. In: Proceedings of the seventh international conference on modelling techniques and tools for computer performance evaluation, vol 794. Springer, Vienna, pp 353–368

Giltnane JM, Rimm DL (2004) Technology insight: identification of biomarkers with tissue microarray technology. Nat Clin Pract Oncol 1:104–111

Greither T, Grochola LF, Udelnow A, Lautenschlager C, Wurl P, Taubert H (2010) Elevated expression of microRNAs 155, 203, 210 and 222 in pancreatic tumors is associated with poorer survival. Int J Cancer 126:73–80

Group BDW (2001) Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther 69:89–95

Hahn MW, Kern AD (2005) Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. Mol Biol Evol 22:803–806

Haibe-Kains B, Desmedt C, Piette F, Buyse M, Cardoso F, Van't Veer L, Piccart M, Bontempi G, Sotiriou C (2008) Comparison of prognostic gene expression signatures for breast cancer. BMC Genomics 9:394

Han SW, Kim TY, Hwang PG, Jeong S, Kim J, Choi IS, Oh DY, Kim JH, Kim DW, Chung DH, Im SA, Kim YT, Lee JS, Heo DS, Bang YJ, Kim NK (2005) Predictive and prognostic impact of epidermal growth factor receptor mutation in non-small-cell lung cancer patients treated with gefitinib. J Clin Oncol 23:2493–2501

Hanahan D, Weinberg RA (2000) The hallmarks of cancer. Cell 100:57–70

Harris AL (2002) Hypoxia—a key regulatory factor in tumour growth. Nat Rev Cancer 2:38–47

Harris L, Fritsche H, Mennel R, Norton L, Ravdin P, Taube S, Somerfield MR, Hayes DF, Bast RC Jr (2007) American society of clinical oncology 2007 update of recommendations for the use of tumor markers in breast cancer. J Clin Oncol 25:5287–5312

Harry VN, Semple SI, Parkin DE, Gilbert FJ (2010) Use of new imaging techniques to predict tumour response to therapy. Lancet Oncol 11:92–102

Hastie T, Tibshirani R, Friedman J (2001) The elements of statistical learning. Springer, New York

Hayes DF, Bast RC, Desch CE, Fritsche H Jr, Kemeny NE, Jessup JM, Locker GY, Macdonald JS, Mennel RG, Norton L, Ravdin P, Taube S, Winn RJ (1996) Tumor marker utility grading system: a framework to evaluate clinical utility of tumor markers. J Natl Cancer Inst 88:1456–1466

Hayes DF, Trock B, Harris AL (1998) Assessing the clinical impact of prognostic factors: when is "statistically significant" clinically useful? Breast Cancer Res Treat 52:305–319

Hensley ML, Spriggs DR (2004) Cancer screening: how good is good enough? J Clin Oncol 22:4037–4039

Hoops S, Sahle S, Gauges R, Lee C, Pahle J, Simus N, Singhal M, Xu L, Mendes P, Kummer U (2006) COPASI—a COmplex PAthway SImulator. Bioinformatics 22:3067–3074

Hornberg JJ, Binder B, Bruggeman FJ, Schoeberl B, Heinrich R, Westerhoff HV (2005) Control of MAPK signalling: from complexity to what really matters. Oncogene 24:5533–5542

Hucka M, Finney A, Sauro H, Bolouri H, Doyle J, Kitano H (2001) The ERATO systems biology workbench: an integrated environment for multiscale and multitheoretic simulations in systems biology. In: Kitano H (ed) Foundations of systems biology. MIT Press, Cambridge

Hutchinson L, DeVita VT Jr (2009) The Holy Grail of biomarkers. Nat Rev Clin Oncol 6:553

Ignatiadis M, Sotiriou C (2008) Understanding the molecular basis of histologic grade. Pathobiology 75:104–111

Ioannidis JP, Allison DB, Ball CA, Coulibaly I, Cui X, Culhane AC, Falchi M, Furlanello C, Game L, Jurman G, Mangion J, Mehta T, Nitzberg M, Page GP, Petretto E, Noort V van (2009) Repeatability of published microarray gene expression analyses. Nat Genet 41:149–455

Ito K (2009) Prostate-specific antigen-based screening for prostate cancer: evidence, controversies and future perspectives. Int J Urol 16:458–464

Jones PA, Baylin SB (2002) The fundamental role of epigenetic events in cancer. Nat Rev Genet 3:415–428

Jubb AM, Landon TH, Burwick J, Pham TQ, Frantz GD, Cairns B, Quirke P, Peale FV, Hillan KJ (2003) Quantitative analysis of colorectal tissue microarrays by immunofluorescence and in situ hybridization. J Pathol 200:577–588

Jubb AM, Buffa FM, Harris AL (2009) Assessment of tumor hypoxia for prediction of response to therapy and cancer prognosis. J Cell Mol Med 14(1-2):18–29

Kaufmann M, Minckwitz G, von Bear HD, Buzdar A, McGale P, Bonnefoi H, Colleoni M, Denkert C, Eiermann W, Jackesz R, Makris A, Miller W, Pierga JY, Semiglazov V, Schneeweiss A, Souchon R, Stearns V, Untch M, Loibl S (2007) Recommendations from an international expert panel on the use of neoadjuvant (primary) systemic treatment of operable breast cancer: new perspectives 2006. Ann Oncol 18:1927–1934

Kholodenko BN (2006) Cell-signalling dynamics in time and space. Nat Rev Mol Cell Biol 7:165–176

Kim KS, Jeong JY, Kim YC, Na KJ, Kim YH, Ahn SJ, Baek SM, Park CS, Park CM, Kim YI, Lim SC, Park KO (2005) Predictors of the response to gefitinib in refractory non-small cell lung cancer. Clin Cancer Res 11:2244–2251

Kinders R, Parchment RE, Ji J, Kummar S, Murgo AJ, Gutierrez M, Collins J, Rubinstein L, Pickeral O, Steinberg SM, Yang S, Hollingshead M, Chen A, Helman L, Wiltrout R, Simpson M, Tomaszewski JE, Doroshow JH (2007) Phase 0 clinical trials in cancer drug development: from FDA guidance to clinical practice. Mol Interv 7:325–334

Krause M, Baumann M (2008) Clinical biomarkers of kinase activity: examples from EGFR inhibition trials. Cancer Metastasis Rev 27:387–402

Kruser TJ, Wheeler DL (2010) Mechanisms of resistance to HER family targeting antibodies. Exp Cell Res 316(7):1083–1100

Latterich M, Abramovitz M, Leyland-Jones B (2008) Proteomics: new technologies and clinical applications. Eur J Cancer 44:2737–2741

Le Novere N, Bornstein B, Broicher A, Courtot M, Donizelli M, Dharuri H, Li L, Sauro H, Schilstra M, Shapiro B, Snoep JL, Hucka M (2006) Biomodels database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. Nucleic Acids Res 34:D689–D691

Lievre A, Bachet JB, Le Corre D, Boige V, Landi B, Emile JF, Cote JF, Tomasic G, Penna C, Ducreux M, Rougier P, Penault-Llorca F, Laurent-Puig P (2006) KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer. Cancer Res 66:3992–3995

Linardou H, Dahabreh IJ, Kanaloupiti D, Siannis F, Bafaloukos D, Kosmidis P, Papadimitriou CA, Murray S (2008) Assessment of somatic k-RAS mutations as a mechanism associated with resistance to EGFR-targeted agents: a systematic review and meta-analysis of studies in advanced non-small-cell lung cancer and metastatic colorectal cancer. Lancet Oncol 9:962–972

Liu CL, Prapong W, Natkunam Y, Alizadeh A, Montgomery K, Gilks CB, Rijn M van de (2002) Software tools for high-throughput analysis and archiving of immunohistochemistry staining data obtained with tissue microarrays. Am J Pathol 161:1557–1565

Locker GY, Hamilton S, Harris J, Jessup JM, Kemeny N, Macdonald JS, Somerfield MR, Hayes DF, Bast RC Jr (2006) ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. J Clin Oncol 24:5313–5327

Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG, Louis DN, Christiani DC, Settleman J, Haber DA (2004) Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med 350:2129–2139

Ma XJ, Wang Z, Ryan PD et al (2004) A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. Cancer Cell 5:607–616

Mamounas EP, Tang G, Fisher B et al (2010) Association Between the 21-Gene Recurrence Score Assay and Risk of Locoregional Recurrence in Node-Negative, Estrogen Receptor-Positive Breast Cancer: Results From NSABP B-14 and NSABP B-20. J Clin Oncol 28(10):1677–1683

McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM (2005) Reporting recommendations for tumor marker prognostic studies (REMARK). J Natl Cancer Inst 97:1180–1184

Michiels S, Koscielny S, Hill C (2005) Prediction of cancer outcome with microarrays: a multiple random validation strategy. Lancet 365:488–492

Miller LD, Smeds J, George J et al (2005) An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. Proc Natl Acad Sci U S A 102:13550–13555

Mirlacher M, Kasper M, Storz M, Knecht Y, Durmuller U, Simon R, Mihatsch MJ, Sauter G (2004) Influence of slide aging on results of translational research studies using immunohistochemistry. Mod Pathol 17:1414–1420

Mori T, Martinez SR, O'Day SJ, Morton DL, Umetani N, Kitago M, Tanemura A, Nguyen SL, Tran AN, Wang HJ, Hoon DS (2006) Estrogen receptor-alpha methylation predicts melanoma progression. Cancer Res 66:6692–6698

Moroni M, Veronese S, Benvenuti S, Marrapese G, Sartore-Bianchi A, Di Nicolantonio F, Gambacorta M, Siena S, Bardelli A (2005) Gene copy number for epidermal growth factor receptor (EGFR) and clinical response to antiEGFR treatment in colorectal cancer: a cohort study. Lancet Oncol 6:279–286

Mukohara T, Engelman JA, Hanna NH, Yeap BY, Kobayashi S, Lindeman N, Halmos B, Pearlberg J, Tsuchihashi Z, Cantley LC, Tenen DG, Johnson BE, Janne PA (2005) Differential effects of gefitinib and cetuximab on non-small-cell lung cancers bearing epidermal growth factor receptor mutations. J Natl Cancer Inst 97:1185–1194

Murukesh N, Dive C, Jayson GC (2010) Biomarkers of angiogenesis and their role in the development of VEGF inhibitors. Br J Cancer 102:8–18

Naderi A, Teschendorff AE, Barbosa-Morais NL et al (2007) A gene-expression signature to predict survival in breast cancer across independent data sets. Oncogene 26:1507–1516

Nicoloso MS, Spizzo R, Shimizu M, Rossi S, Calin GA (2009) MicroRNAs—the micro steering wheel of tumour metastases. Nat Rev Cancer 9:293–302

Nuovo GJ (2008) In situ detection of precursor and mature microRNAs in paraffin embedded, formalin fixed tissues and cell preparations. Methods 44:39–46

Nuovo GJ, Elton TS, Nana-Sinkam P, Volinia S, Croce CM, Schmittgen TD (2009) A methodology for the combined in situ analyses of the precursor and mature forms of microRNAs and correlation with their putative targets. Nat Protoc 4:107–115

Offit K (2009) Breast cancer single-nucleotide polymorphisms: statistical significance and clinical utility. J Natl Cancer Inst 101:973–975

Orton RJ, Adriaens ME, Gormand A, Sturm OE, Kolch W, Gilbert DR (2009) Computational modelling of cancerous mutations in the EGFR/ERK signalling pathway. BMC Syst Biol 3:100

Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ, Naoki K, Sasaki H, Fujii Y, Eck MJ, Sellers WR, Johnson BE, Meyerson M (2004) EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 304:1497–1500

Paik S, Shak S, Tang G et al (2004) A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 351:2817–2826

Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, Cronin M, Baehner FL, Watson D, Bryant J, Costantino JP, Geyer CE Jr, Wickerham DL, Wolmark N (2006) Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. J Clin Oncol 24:3726–3734

Pao W, Miller V, Zakowski M, Doherty J, Politi K, Sarkaria I, Singh B, Heelan R, Rusch V, Fulton L, Mardis E, Kupfer D, Wilson R, Kris M, Varmus H (2004) EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. Proc Natl Acad Sci USA 101:13306–13311

Penault-Llorca F, Bilous M, Dowsett M, Hanna W, Osamura RY, Ruschoff J, Vijver M van de (2009) Emerging technologies for assessing HER2 amplification. Am J Clin Pathol 132:539–548

Perou CM, Sorlie T, Eisen MB et al (2000) Molecular portraits of human breast tumours. Nature 406:747–52

Piccart M, Lohrisch C, Di Leo A, Larsimont D (2001) The predictive value of HER2 in breast cancer. Oncology 61(Suppl 2):73–82

Piccart-Gebhart MJ, Procter M, Leyland-Jones B, Goldhirsch A, Untch M, Smith I, Gianni L, Baselga J, Bell R, Jackisch C, Cameron D, Dowsett M, Barrios CH, Steger G, Huang CS, Andersson M, Inbar M, Lichinitser M, Lang I, Nitz U, Iwata H, Thomssen C, Lohrisch C, Suter TM, Ruschoff J, Suto T, Greatorex V, Ward C, Straehle C, McFadden E, Dolci MS, Gelber RD (2005) Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. N Engl J Med 353:1659–1672

Pillai RS, Bhattacharyya SN, Filipowicz W (2007) Repression of protein synthesis by miRNAs: how many mechanisms? Trends Cell Biol 17:118–126

Ramasamy A, Mondry A, Holmes CC, Altman DG (2008) Key issues in conducting a meta-analysis of gene expression microarray datasets. PLoS Med 5:e184

Ramirez JL, Rosell R, Taron M, Sanchez-Ronco M, Alberola V, Las Penas R de, Sanchez JM, Moran T, Camps C, Massuti B, Sanchez JJ, Salazar F, Catot S (2005) 14-3-3sigma methylation in pretreatment serum circulating DNA of cisplatin-plus-gemcitabine-treated advanced non-small-cell lung cancer patients predicts survival: The Spanish Lung Cancer Group. J Clin Oncol 23:9105–9112

Ramsey S, Orrell D, Bolouri H (2005) Dizzy: stochastic simulation of large-scale genetic regulatory networks. J Bioinform Comput Biol 3:415–436

Robbins P, Pinder S, Klerk N de, Dawkins H, Harvey J, Sterrett G, Ellis I, Elston C (1995) Histological grading of breast carcinomas: a study of interobserver agreement. Hum Pathol 26:873–879

Rodriguez H, Tezak Z, Mesri M, Carr SA, Liebler DC, Fisher SJ, Tempst P, Hiltke T, Kessler LG, Kinsinger CR, Philip R, Ransohoff DF, Skates SJ, Regnier FE, Anderson NL, Mansfield E (2010) Analytical validation of protein-based multiplex assays: a workshop report by the NCI-FDA interagency oncology task force on molecular diagnostics. Clin Chem 56:237–243

Romond EH, Perez EA, Bryant J, Suman VJ, Geyer CE Jr, Davidson NE, Tan-Chiu E, Martino S, Paik S, Kaufman PA, Swain SM, Pisansky TM, Fehrenbacher L, Kutteh LA, Vogel VG, Visscher DW, Yothers G, Jenkins RB, Brown AM, Dakhil SR, Mamounas EP, Lingle WL, Klein PM, Ingle JN, Wolmark N (2005) Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. N Engl J Med 353:1673–1684

Roylance R, Gorman P, Harris W, Liebmann R, Barnes D, Hanby A, Sheer D (1999) Comparative genomic hybridization of breast tumors stratified by histological grade reveals new insights into the biological progression of breast cancer. Cancer Res 59:1433–1436

Samaga R, Saez-Rodriguez J, Alexopoulos LG, Sorger PK, Klamt S (2009) The logic of EGFR/ErbB signaling: theoretical properties and analysis of high-throughput data. PLoS Comput Biol 5:e1000438

Sarker D, Workman P (2007) Pharmacodynamic biomarkers for molecular cancer therapeutics. Adv Cancer Res 96:213–268

Sasagawa S, Ozaki Y, Fujita K, Kuroda S (2005) Prediction and validation of the distinct dynamics of transient and sustained ERK activation. Nat Cell Biol 7:365–373

Schoeberl B, Eichler-Jonsson C, Gilles ED, Muller G (2002) Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. Nat Biotechnol 20:370–385

Shi L, Reid LH, Jones WD, Shippy R, Warrington JA, Baker SC, Collins PJ, Longueville F de, Kawasaki ES, Lee KY, Luo Y, Sun YA, Willey JC, Setterquist RA, Fischer GM, Tong W, Dragan YP, Dix DJ, Frueh FW, Goodsaid FM, Herman D, Jensen RV, Johnson CD, Lobenhofer EK, Puri RK, Schrf U, Thierry-Mieg J, Wang C, Wilson M, Wolber PK, Zhang L, Amur S, Bao W, Barbacioru CC, Lucas AB, Bertholet V, Boysen C, Bromley B, Brown D, Brunner A, Canales R, Cao XM, Cebula TA, Chen JJ, Cheng J, Chu TM, Chudin E, Corson J, Corton JC, Croner LJ, Davies C, Davison TS, Delenstarr G, Deng X, Dorris D, Eklund AC, Fan XH, Fang H, Fulmer-Smentek S, Fuscoe JC, Gallagher K, Ge W, Guo L, Guo X, Hager J, Haje PK, Han J, Han T, Harbottle HC, Harris SC, Hatchwell E, Hauser CA, Hester S, Hong H, Hurban P, Jackson SA, Ji H, Knight CR, Kuo WP, LeClerc JE, Levy S, Li QZ, Liu C, Liu Y, Lombardi MJ, Ma Y, Magnuson SR, Maqsodi B, McDaniel T, Mei N, Myklebost O, Ning B, Novoradovskaya

N, Orr MS, Osborn TW, Papallo A, Patterson TA, Perkins RG, Peters EH, Peterson R, Philips KL, Pine PS, Pusztai L, Qian F, Ren H, Rosen M, Rosenzweig BA, Samaha RR, Schena M, Schroth GP, Shchegrova S, Smith DD, Staedtler F, Su Z, Sun H, Szallasi Z, Tezak Z, Thierry-Mieg D, Thompson KL, Tikhonova I, Turpaz Y, Vallanat B, Van C, Walker SJ, Wang SJ, Wang Y, Wolfinger R, Wong A, Wu J, Xiao C, Xie Q, Xu J, Yang W, Zhong S, Zong Y, Slikker W Jr (2006) The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. Nat Biotechnol 24:1151–1161

Siena S, Sartore-Bianchi A, Di Nicolantonio F, Balfour J, Bardelli A (2009) Biomarkers predicting clinical outcome of epidermal growth factor receptor-targeted therapy in metastatic colorectal cancer. J Natl Cancer Inst 101:1308–1324

Singletary SE, Allred C, Ashley P, Bassett LW, Berry D, Bland KI, Borgen PI, Clark G, Edge SB, Hayes DF, Hughes LL, Hutter RV, Morrow M, Page DL, Recht A, Theriault RL, Thor A, Weaver DL, Wieand HS, Greene FL (2002) Revision of the American Joint Committee on Cancer staging system for breast cancer. J Clin Oncol 20:3628–3636

Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J, Nordgren H, Farmer P, Praz V, Haibe-Kains B, Desmedt C, Larsimont D, Cardoso F, Peterse H, Nuyten D, Buyse M, Van de Vijver MJ, Bergh J, Piccart M, Delorenzi M (2006) Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. J Natl Cancer Inst 98:262–272

Starmans MH, Krishnapuram B, Steck H et al (2008) Robust prognostic value of a knowledge-based proliferation signature across large patient microarray studies spanning different cancer types. Br J Cancer 99:1884–1890

Stiles T, Grant V, Mawbey N (2003) Good clinical laboratory practice (GCLP): A quality system for laboratories which undertake the analyses of samples from clinical trials. Ipswich (UK): Brit Assoc Res Qual Assur 1–17. ISBN 1-904610-00-5

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102:15545–15550

Takimoto CH (2009) Pharmacokinetics and pharmacodynamic biomarkers in early oncology drug development. Eur J Cancer 45(Suppl 1):436–48

Taylor IW, Linding R, Warde-Farley D, Liu Y, Pesquita C, Faria D, Bull S, Pawson T, Morris Q, Wrana JL (2009) Dynamic modularity in protein interaction networks predicts breast cancer outcome. Nat Biotechnol 27:199–204

Teschendorff AE, Naderi A, Barbosa-Morais NL et al (2006) A consensus prognostic gene expression classifier for ER positive breast cancer. Genome Biol 7:R101

Teschendorff AE et al (2009) An epigenetic signature in peripheral blood predicts active ovarian cancer. PLoS One 4(12):e8274

Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, Verweij J, Van Glabbeke M, Oosterom AT van, Christian MC, Gwyther SG (2000) New guidelines to evaluate the response to treatment in solid tumors. European organization for research and treatment of cancer, national cancer institute of the United States, national cancer institute of Canada. J Natl Cancer Inst 92:205–216

Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A (2003) PANTHER: a library of protein families and subfamilies indexed by function. Genome Res 13:2129–2141

Thomas PD, Kejariwal A, Guo N et al (2006) Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. Nucleic Acids Res 34:W645–650

van't Veer LJ, Paik S, Hayes DF (2005) Gene expression profiling of breast cancer: a new tumor marker. J Clin Oncol 23:1631–1635

van de Vijver MJ, He YD, van't Veer LJ et al (2002) A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 347:1999–2009

Walhout AJ, Vidal M (2001) Protein interaction maps for model organisms. Nat Rev Mol Cell Biol 2:55–62

Walko CM, McLeod H (2009) Pharmacogenomic progress in individualized dosing of key drugs for cancer patients. Nat Clin Pract Oncol 6:153–162

Wallner M, Herbst A, Behrens A, Crispin A, Stieber P, Goke B, Lamerz R, Kolligs FT (2006) Methylation of serum DNA is an independent prognostic marker in colorectal cancer. Clin Cancer Res 12:7347–7352

Wang Y, Klijn JG, Zhang Y et al (2005) Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 365:671–679

Wang DY, Cardelli L, Phillips A, Piterman N, Fisher J (2009) Computational modeling of the EGFR network elucidates control mechanisms regulating signal dynamics. BMC Syst Biol 3:118

Wei Y, Xia W, Zhang Z, Liu J, Wang H, Adsay NV, Albarracin C, Yu D, Abbruzzese JL, Mills GB, Bast RC Jr, Hortobagyi GN, Hung MC (2008) Loss of trimethylation at lysine 27 of histone H3 is a predictor of poor outcome in breast, ovarian, and pancreatic cancers. Mol Carcinog 47:701–706

Weigelt B, Baehner FL, Reis-Filho JS (2010) The contribution of gene expression profiling to breast cancer classification, prognostication and prediction: a retrospective of the last decade. J Pathol 220:263–280

West CM, Jones T, Price P (2004) The potential of positron-emission tomography to study anticancer-drug resistance. Nat Rev Cancer 4:457–469

Wheldon TE (1998) Mathematical Models in Cancer Research. Bristol (UK). Adam Hilger. ISBN-13: 9780852742914

Whitfield ML, Sherlock G, Saldanha AJ et al (2002) Identification of genes periodically expressed in the human cell cycle and their expression in tumors. Mol Biol Cell 13:1977–2000

Whitfield ML, George LK, Grant GD, Perou CM (2006) Common markers of proliferation. Nat Rev Cancer 6:99–106

WHO (2009) Good Clinical Laboratory Practice (GCLP). Special Programme for Research & Training in Tropical Diseases (TDR). World Health Organization

Widschwendter A, Muller HM, Fiegl H, Ivarsson L, Wiedemair A, Muller-Holzner E, Goebel G, Marth C, Widschwendter M (2004) DNA methylation in serum and tumors of cervical cancer patients. Clin Cancer Res 10:565–571

Williams C, Brunskill S, Altman D, Briggs A, Campbell H, Clarke M, Glanville J, Gray A, Harris A, Johnston K, Lodge M (2006) Cost-effectiveness of using prognostic information to select women with breast cancer for adjuvant systemic therapy. Health Technol Assess 10:iii-iv, ix-xi, 1–204

Winter SC, Buffa FM, Silva P, Miller C, Valentine HR, Turley H, Shah KA, Cox GJ, Corbridge RJ, Homer JJ, Musgrove B, Slevin N, Sloan P, Price P, West CM, Harris AL (2007) Relation of a hypoxia metagene derived from head and neck cancer to prognosis of multiple cancers. Cancer Res 67:3441–3449

Wirapati P, Sotiriou C, Kunkel S, Farmer P, Pradervand S, Haibe-Kains B, Desmedt C, Ignatiadis M, Sengstag T, Schutz F, Goldstein DR, Piccart M, Delorenzi M (2008) Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. Breast Cancer Res 10:R65

Wolfe CJ, Kohane IS, Butte AJ (2005) Systematic survey reveals general applicability of "guilt-by-association" within gene coexpression networks. BMC Bioinformatics 6:227

Yarden Y, Sliwkowski MX (2001) Untangling the ErbB signalling network. Nat Rev Mol Cell Biol 2:127–137

Zilberman D, Henikoff S (2007) Genome-wide analysis of DNA methylation patterns. Development 134:3959–3965

Chapter 14
Systems Biology Approaches to Cancer Drug Development

Christopher Snell, David Orrell, Eric Fernandez, Christophe Chassagnole and David Fell

Abstract New approaches are currently being investigated in drug development to improve the large failure rate seen in many clinical trials. Systems biology is one area which shows promise for changing the way we think about disease and drug development. Over the last decade, several biotechnology companies have been set up with the aim of incorporating systems biology into the drug development process. Both descriptive and predictive models are used in order to provide the right approach for different situations. The development of a 'virtual tumour' model, coupled with a pharmacokinetic one, which together is capable of designing optimal drug schedules and combinations, is often the focus at the industrial level. This chapter describes a typical modelling approach, and shows how it is being used to aid the drug development process.

14.1 Introduction

As more and more drugs fail clinical trials, it is becoming increasingly important for changes to occur in the drug development process to reduce these attrition rates. For oncology, in particular, success rates are very low (Kola and Landis 2004). The drug development process needs to embrace new technologies with the aim of integrating all knowledge and data. Systems biology can provide valuable insights into the major causes of attrition in the clinic, such as lack of drug efficacy. It provides a much needed broader view of the system and the molecular mechanisms of disease (Cho et al. 2006). Without a systems-level understanding, research efficiency and productivity can be severely limited, especially in cancer.

14.1.1 The Systems View of Drug Action

It is self-evident that administering a drug to a human sets in train a complicated sequence of events. Firstly there is the dispersion and absorption of the drug in the

C. Snell (✉)
Physiomics Plc, Magdalen Centre, Oxford Science Park, Oxford, OX4 4GA, UK
e-mail: chris.m.snell@gmail.com

A. Cesario, F. B. Marcus (eds.), *Cancer Systems Biology, Bioinformatics and Medicine,*
DOI 10.1007/978-94-007-1567-7_14, © Springer Science+Business Media B.V. 2011

body, its distribution in different tissues, and its metabolism and eventual excretion. At some point it interacts with cells, possibly even penetrating them. The interaction of the drug with its target is a molecular event that affects the properties of the target, which in turn affects the functioning of the cellular sub-systems of which the target is a component. This altered functioning leads to changed cellular behaviour, resulting in consequences at the level of first tissue, then organism. The current process of drug development, to a significant degree, explores these steps in isolation, presumably for historical and pragmatic reasons, the latter including the fact that they fall within the domains of different disciplines. It is not obvious, however, that this modularization is either intrinsically valid or appropriate, and systems biology is emerging as a different approach.

The aim of systems biology is to understand biological phenomena at multiple scales, i.e. determining which properties and interactions of the components account for the functional properties of the systems to which they belong, and how the functioning of the system influences the environment that the components experience (see Chap. 9). But this implies a nested description, where a system at one scale is a component of another system on a larger scale, and where scale-crossing explanations can extend from the atomic to the ecosystem level. In fact, most systems biology projects can then be regarded as 'middle-out' (Noble 2008), where the middle can be defined at different levels as befits to the questions to be addressed (We are not here considering another contrasting view of systems biology which presents it as an approach to identifying and cataloguing components of systems and their interactions using high-throughput techniques; this has been referred to as 'systematic biology'). These characteristics of systems biology mean that it crosses conventional disciplinary boundaries, and this, with the mix of theory, computation and experiment required to predict the overall behaviour of systems consisting of interacting components, results in the need for interdisciplinary teams to carry out the research. Understanding and predicting the organism-level effects of a drug that interacts with a specific molecular target appears to fall exactly within the ambit of systems biology.

A simple example of how important is the systems context comes from the well-studied effect of an enzyme inhibitor (drug) on a metabolic pathway (a system of many enzymes). A systems approach to metabolic control (Metabolic Control Analysis, reviewed by Fell 1997) had established that it would be unusual for any enzyme to exert proportional control over the rate of metabolism; i.e. a 1% change in the activity of an enzyme would in most cases cause a change in metabolic rate smaller than 1%, even if the enzyme is essential to the pathway when completely inactivated. This contradicted the previously prevailing dogma of control by a single rate-limiting step, and in particular the belief that feedback-inhibited enzymes would determine the rate of metabolic pathways. It follows that if an inhibitor is applied that reduces an enzyme activity by 1%, the response of the metabolic system is a reduction in rate of less than 1%. In other words, the titration curves of the enzyme activity and the functional system response against inhibitor concentration are divergent from the start. Numerous experimental in-

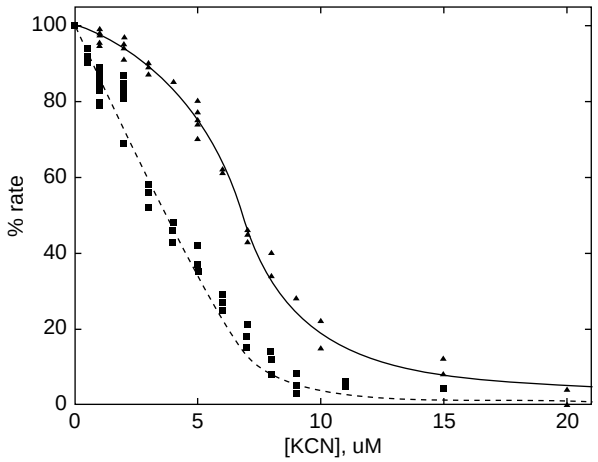


Fig. 14.1 Inhibition of oxidative phosphorylation by a 'drug' (cyanide). The percentage rate of the system property (mitochondrial oxidative phosphorylation) relative to uninhibited is plotted as solid triangles. The enzyme activity of cyanide's target is plotted as a percentage of uninhibited as solid squares. Note the difference in concentration of cyanide required to achieve 50% inhibition. The experiment is described in Rossignol et al. (2000); data was supplied by T. Letellier

stances of this are known; an example is shown in Fig. 14.1. The exact shape of the metabolic response curve depends on system properties (in principle the kinetic characteristics of all the enzymes involved) and the type of inhibition. However, it is inevitable that a higher inhibitor concentration will be required to cause a 50% reduction in metabolic rate, than to cause 50% inhibition of the enzyme. In general, the best results, i.e. the least difference between the two curves, are likely to be obtained with irreversible inactivators or uncompetitive inhibitors, while a poorer result is to be expected with competitive inhibitors (Eisenthal and Cornish-Bowden 1998).

Though this result is well-known in the field of metabolic control analysis, it appears to be little appreciated outside it. The result implies the existence of inherent system reasons why the inhibition of function will be less strong than the inhibition of the drug target. It is true that the discrepancy will be less marked, the greater the sensitivity of the biological function to the activity of the target; but the corollary is that the discrepancy is larger if the sensitivity of the function to the target activity is weak. Though the experimental examples and theory have been developed within metabolism, the same principles apply in other cellular networks such as signalling.

It is important to emphasize that the origin of the differential response is in the system itself. Reducing the inhibition constant of the drug will not abolish it. Similarly, poor access of the drug to the target may be another factor causing poor inhibition of the target function, but that is in addition to the intrinsic system effect, which will still persist if drug access to the target is improved. This emphasises the under-used potential for theoretical and/or experimental assessment of the sensitiv-

ity of a cellular process to modulation of the amounts or activities of its molecular components, before one of them is selected as a potential drug target.

14.1.2 Introducing Systems Biology into Drug Development

Several biotechnology companies, including ours (Physiomics plc, http://www.physiomics-plc.com), now use systems biology approaches to aid drug development and clinical trials. Computational models are being developed to predict and understand cancer drug efficacy where the modelling methods chosen are adapted to the industrial and clinical partners' requirements, and crucially to the questions posed and answers required. In this chapter, we shall use examples (drawn from the work within our company) that illustrate the role of modelling at different scales for different aspects of the development process. The first example uses modelling of sub-cellular networks and explores the relationship between the sensitivity of multiple targets to drug candidates and the response of cells. Next, modelling the drug responses of cell populations, and how this can connect to experimentation at the cellular level in drug development, is illustrated in detail. Finally we will mention mouse xenograft models, where the cell population is structured and drug exposure is described by pharmacokinetic models.

14.2 Model Building

Prior to model building, a modelling strategy is chosen depending on the questions posed and expected answers. It is also important to keep in mind that the systems biology models selected must be capable of integration into the industrial and clinical partners' existing drug development processes. An essential prerequisite is a thorough understanding of the literature and databases, in order to construct a model of the network. For this reason the models are constructed by a team that includes experts in biology, mathematics and computer science. This model is then parameterized using data from our partners and the literature. Predictions can then be made using the model to aid future experimental design and drug discovery.

14.2.1 Linking Data to the Models

The challenge with using systems biology within the drug development process is how to integrate many different levels of data using the computational models. On the enzyme level it is common to measure the IC_{50} or K_i values of a compound against the proposed targets. It is important to study the specificity of a drug to-

wards cellular targets using a systems biology approach, because precise specificity against a single target may actually be undesirable, compared to multiple inhibition modes. These K_i values are used alongside IC_{50} values from cell cytotoxicity assays in a range of cell lines (Chassagnole et al. 2006). A systems biology model at the level of the genes and proteins can be used to investigate how inhibition of one or many enzymes leads to the effects seen on the cellular level. It can help to explain unexpected behaviour seen in cell cytotoxicity assays which cannot be explained by studying the genes or proteins alone.

On the drug level, it is important to have an understanding of the behaviour of the drug once it has been administered. Pharmacokinetic (PK) models are an established and extensively used discipline within the drug development process (Urso et al. 2002; Theil et al. 2003). The data derived from PK experiments is used to construct a model which can then simulate the uptake of the drug and its dispersion through the relevant tissues in the body. This greatly aids in the determination of optimal dosage levels. To simulate the effect of the drug on the tumour, it is vital to incorporate a pharmacokinetic model with a model at the cell population level, producing a combined PK/PD model.

Decision-making in cancer drug development could be greatly improved by the discovery and validation of effective biomarkers. Systems biology models can aid in the development and interpretation of these biomarkers (Aebersold et al. 2009). To simulate cell populations, biomarkers based on required measurements such as tumour growth are generally used; these give vital indications on the progression of cancer and responses to treatment, and are crucial for calibrating the models. In fact they often represent the main measurable outcomes of the experimental investigations, even if not directly related to the variables being simulated.

Clinical trials currently make use of many biomarkers which can be exploited by biotech companies. In particular, ‘adaptive’ trials could greatly benefit from the integration of systems biology approaches. These clinical trials have a review period halfway through the trial where early data is analysed. The second half of the trial can then be adapted to focus on patients who have shown promising results during the early phase of the trial. These trials are becoming increasingly popular as pharmaceutical companies look for ways to limit the failure rate. If the data from the first half of the trial is fed into a computer model, predictions could be made about how best to perform the rest of the trial (Ledford 2010).

Cell cycle biomarkers are commonly used, in particular biomarkers measuring the different cell cycle phases, e.g. pRb biomarker for the G1 phase. Apoptosis biomarkers such as Caspase 3 or PARP are also used to measure cell death, which can then be used to parameterize the models. Systems biology approaches such as sensitivity analysis can be used to investigate the effectiveness and validation of potential biomarkers. Flow cytometry (FACs) data is used to study changes in the cell cycle phases. This data can be used to monitor changes in cell populations following drug administration. Models can also be built which can display an output similar to FACs data, allowing direct comparison with experimental data. In this way the model can provide a much deeper understanding than laboratory data alone.

14.3 Case Studies of Modelling Cellular Networks

Systems biology models of the cell cycle have provided valuable insights into the effects of perturbations in the pathway and architecture of the networks of normal and cancer cells. Traditionally, cell cycle models have been built to investigate systems properties of the pathways in question, such as bi-stability and robustness, rather than to directly investigate anti-cancer drug development (Tyson et al. 2002; Haberichter et al. 2007; Csikasz-Nagy 2009; Conradie et al. 2010). Recent systems biology models have been built in parallel with experimental data to validate the models, as it is becoming increasingly important to demonstrate their thrustworthiness (Alfieri et al. 2009). The models still require further development and correlation with clinical biomarkers in order to become significantly useful for drug development (Clyde et al. 2006).

14.3.1 Using Cellular Networks in Drug Development

Detailed pathway modelling is typically used to initiate the investigation of aspects of drug development such as drug target validation, lead compound selection, and mechanisms of action. A descriptive model is used to study pathway interactions and how the pathway is affected through drug inhibition. It allows monitoring of the changes in the concentrations of the pathways components following any drug inhibition. For instance, using pathway modelling it is possible to aid lead selection by decoupling and evaluating dual kinase inhibition effects (Fig. 14.2). This involves assessing the balance between two mechanisms of action and their effects on the cell cycle and apoptosis, thus facilitating investigation of why one drug may be favoured over another.

14.3.2 Modelling the Cellular Action of Seliciclib and Other cdk2 Inhibitors

To demonstrate the potential of systems biology in drug development at the level of cellular networks, our company developed a detailed cell cycle model to test data on five preclinical compounds and a compound in phase II clinical trials, Seliciclib. These compounds were being developed by Cyclacel Pharmaceutical, Ltd. For each compound, K_i data for targets within the model was provided. The only other piece of data used was the IC_{50} cytotoxicity values against colon (HT29) and lung (A549) cancer cells (Chassagnole et al. 2006). The model was based on ordinary differential equations (ODEs) and constructed in the biochemical simulator JARNAC (Sauro et al. 2003). It contained over 60 components and was able to reproduce the timing and events of the normal cell cycle. Inhibition

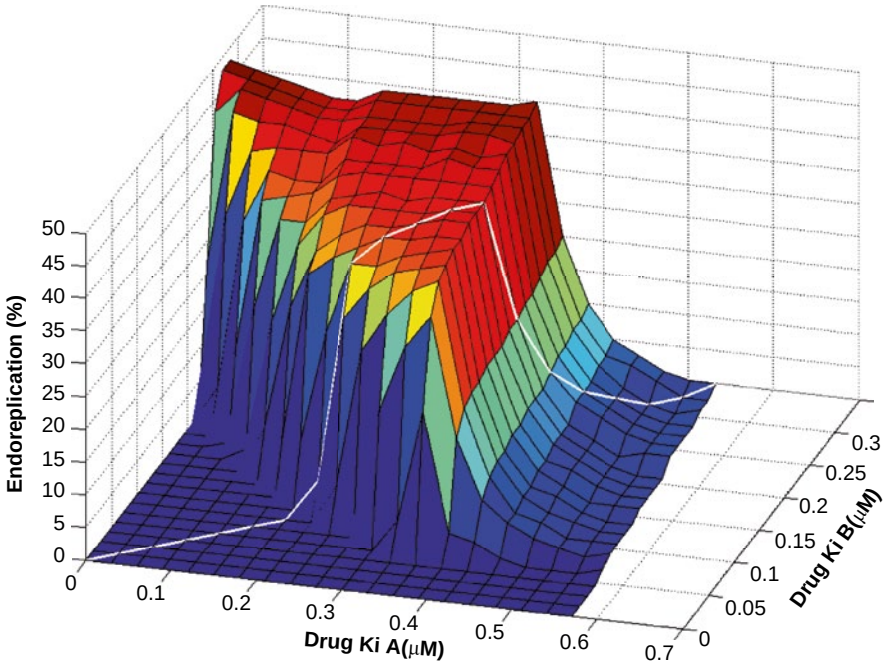


Fig. 14.2 Mechanisms of action can be decoupled. By modelling the mechanism of action of a family of dual target inhibitors, it is possible to predict the overall effect on the cell cycle as a function of the inhibition level of each individual target, A and B. The effect on the cell cycle is shown as the probability of the cell undergoing endoreplication, the completion of S phase but no subsequent cytokinesis (depending on the value of the two dissociation constants KiA and KiB). This helps in determining the best ratio and selecting a lead compound

terms, using the IC_{50} values provided, were then integrated into the kinetic rates of the model to allow modelling of the effects of CDK inhibitors. The pattern of cytotoxicity for six compounds was reproducible by the model, without any alteration of the model parameters. The model's serviceability was further demonstrated by plotting the Seliciclib dose-responsive curve, in order to investigate the effective concentrations of drug needed for efficacy. The model was then used to show the change which occurs in the specific species of the cell cycle during inhibition from Seliciclib, and to point out which aspect of Seliciclib inhibition has the greatest effect on the cell cycle. It was demonstrated that the model can also indicate good targets for second- and third-generation compounds (Chassagnole et al. 2006).

14.3.3 Apoptosis and Signal Transduction Pathways

Systems biology models have also been constructed of signal transduction pathways such as the MAPK pathway (Kholodenko et al. 1999; Brightman and Fell

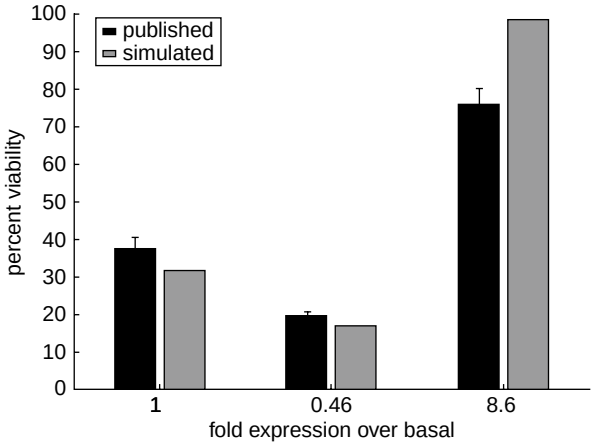


Fig. 14.3 Altering the expression of potential drug target bcl-xL greatly impacts upon the effect of mitoxantrone-induced cell death

2000; Schoeberl et al. 2002; Hornberg et al. 2005). Recently, a computational model of the ErbB pathway was utilized within the drug development process to predict the effects of a monoclonal antibody targeting the pathway (Schoeberl et al. 2009). The apoptosis pathway leading to cell death (see Chap. 10) has also been widely studied using systems biology (Hua et al. 2005; Legewie et al. 2006; Bagci et al. 2006). The majority of these models were not constructed to investigate the effects of anticancer drugs, but to study the dynamics of the pathway in question. We have investigated both signal transduction and apoptosis pathways by building detailed models. Apoptosis models have been coupled with the cell cycle model to investigate the links between the two pathways and the effects of drugs targeting components of either pathway. For example, a detailed apoptosis model was used to show the effect of altering the expression of bcl-xL on prostate (LNCaP) cancer cell viability after treatment with an antineoplastic agent, mitoxantrone (Fig. 14.3).

14.3.4 Difficulties with Detailed Modelling

Although detailed pathway models can be very valuable they can often face difficulties as the models become larger, making them difficult to calibrate against experimental data. The most common problem is that parameters are not available or are very difficult, if not impossible, to measure experimentally (Cho and Wolkenhauer 2003). Data is often only available for certain species, such as those used as biomarkers. Biological data can also be very noisy and variable. Consequently, many parameter values are arbitrary values assigned to tune the model and fit it to experimental data. This reduces the reliability of the model results. Large models can also display unexpected properties which cannot be predicted until the models are extensively tested. There is therefore a need to develop models which are predictive and can be parameterized to experimental data.

14.4 Modelling at Cellular Scales

14.4.1 'Virtual Tumour' Model as a Simpler Approach

We have adopted a two-track approach in an effort to build predictive as well as descriptive models. The relevant features are extracted from the detailed models to construct simpler agent-based models of cell populations. An agent-based approach is therefore adopted to create a 'virtual tumour' model. Separate software agents describe each cell in the tumour. Unlike the detailed models which capture the precise processes within each cell, the 'virtual tumour' model simulates a cell population and captures the overall effect of the drug on the population. This allows us to simulate xenograft experiments and the effects of anticancer drugs. Several simplifications are made to allow the model to be parameterized with available data, including PK data for the drug and xenograft growth measurements. The central core of the tumour is also simplified, with the peripheral layer being modelled in detail, as this is the only layer of the tumour where cells are growing.

The 'virtual tumour' model is coupled to a PK model which models the PK properties of the drug. This is a crucial component of modelling drug action, since the effect on the tumour is ultimately controlled by how much of the drug reaches the vulnerable tumour cells. Generally there are three levels of PK modelling which are commonly used: non-compartmental analysis, compartmental models, and physiologically-based pharmacokinetic (PBPK) models. The type of PK modelling approach needs to be taken into account before building the model. We commonly build compartmental models using ODEs to simulate the PK effects of the drug. The key outcome of the model is to find the time course of the drug concentration within the tumour compartment, and to use this as input for the PD model.

14.4.2 Modelling Schedules and Combinations

One of the main applications of this modelling approach is designing drug schedules. The effectiveness of schedule variation in combination therapy has been demonstrated in numerous preclinical and some clinical trials. A phase II study found that a much greater response rate and lower toxicity can be achieved by adding a 12-h delay between two commonly used anti-cancer drugs, compared with administering the drugs around the same time. One mechanism for this could be that one anti-cancer drug synchronizes the population of tumour cells, which then enables the second drug to be added at the time when it will be most effective. This is particularly true for drugs targeting the cell cycle (Shah and Schwartz 2001; Schwartz and Shah 2005). It is this kind of study that indicates that systems biology can provide valuable insights into problems of combinations and schedules.

When considering timing-related effects of drugs, the number of possible drugs, doses and administration schedules tends to explode, so that it becomes impossible to

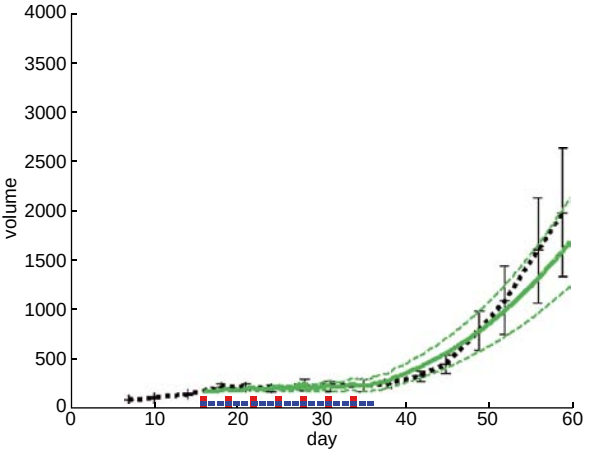


Fig. 14.4 (Parallel administration).The change in tumour volume can be predicted for drug combinations. Our predictions (Physiomics, proprietary data) are shown in *green*, along with estimated upper and lower bounds. The *black lines* show the experimental average xenograft growth, along with 5 and 95 percentile error bounds. Schedules for the two drugs are indicated in red and blue on the bottom axis

test them all in the laboratory or during clinical trials. Accordingly, a computational approach is needed. The dynamic modelling of cell populations, as performed in our company and within the drug development process, has the potential to predict the optimal dosing schedules for a wide class of anti-cancer drugs—alone or in combination—to enhance efficacy and lower side effects. The best schedules can be further investigated and chosen for verification *in vivo*. This avoids the costly trial-and-error approaches hitherto necessitated for determining the best administration schedules.

14.4.3 Predicting Schedules in Drug Development

As with the detailed model, it is important to test the models with data provided by partners, and to demonstrate their predictive potential. Data was therefore provided by a major pharmaceutical company to use as input to the models and provide predictive schedules. The data consisted of individual xenograft growth and biomarker information for two drugs administered individually. In a single-blind test, we made predictions for xenograft growth and compared these with experimental data (Figs. 14.4 and 14.5). These predictions were for two different combinations, using the data for the two separate drugs. The predictions were in good agreement with the experimental data. One prediction was made for the drugs given in parallel (Fig. 14.4), and another prediction for the drugs administered sequentially (Fig. 14.5).

14.4.4 Chronotherapy and the TEMPO Project

Chronotherapy (see Chap. 15) is another area of timing-related drug delivery which can benefit from systems biology. It is the practice of guiding drug administra-

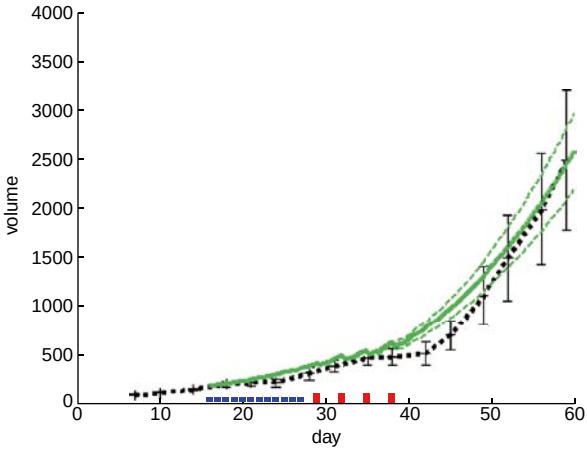


Fig. 14.5 (Sequential administration). The drugs were administered sequentially. Legend as before

tion schedules using circadian time. The response to medicinal treatments has been shown to be substantially affected by circadian time (Smolensky and Peppas 2007). In particular, chronotherapy has been extensively investigated in oncology. Adopting a chronotherapeutic regimen for patients with metastatic colorectal carcinoma was shown to give a significant improvement in tolerability for 5-FU/leucovorin and oxaliplatin administration (Levi et al. 2007; Mormont and Levi 2003).

We recently participated in a European Commission funded project, TEMPO (contract LSHG-ct-2006-037543), the goal of which was to integrate systems biology with chronotherapy. It aimed to provide further evidence supporting the use of cancer chronotherapy, which is becoming increasingly important in so far as it still remains relatively underexploited in the clinic, largely because of the difficulty of determining optimal schedules. Previously, this was achieved primarily through 'trial-and-error' based approaches in mice, which are expensive and provide only poor extrapolation to humans. By using systems biology, we aimed to improve this performance. This was demonstrated by designing an optimal dosing schedule for Seliciclib, a drug whose response had previously been shown to be possibly governed according to circadian time (Iurisci et al. 2006). Experimental and mathematical modelling was combined in order to predict both the optimal and worst chronotherapeutic dosing schedule in mice. Clinical trials are currently ongoing to further demonstrate the potential of chronotherapy.

14.5 Technologies Typically Used at a Biotech Company

14.5.1 Computing Requirements

Single instances of models of sub-cellular networks can be adequately simulated with PC-based desktop workstations in an acceptable time. However, if such simu-

lations are embedded in a cell population, or repeated many times during optimization of a drug combination, the computing time starts to limit productivity. A High-Performance Computer (HPC) is used to rapidly perform many simulations in a short space of time. It is possible to perform thousands of simulations to find the best treatment regime. This is often necessary when dealing with multiple drugs, doses and administration schedules.

New technologies are continuously being utilized and developed, at our company, among others. This is driven by advancements in biology, cancer research, computing and requirements from our partners. Several tools are used at various stages. These include open-source tools such as the JARNAC, CellDesigner and tools built in-house (Funahashi et al. 2003). Matlab is also extensively used to perform modelling and statistical analysis.

14.5.2 Model Database and Reports

Databases have become increasingly vital in the cancer research field to maintain the data generated by high-throughput techniques such as expression arrays (Aebersold et al. 2009). It is crucial that systems biology companies are able to make the most from this data. Our company makes use of both published data in the literature and unpublished data produced by partners. A database of the clinical reactions and biological background included in our models is maintained throughout projects. The database allows modellers to readily access relevant information on the reactions, such as references, discussions and comments. The database software also allows reports to be generated and transferred to partners.

14.5.3 One Operational Example: Delivering the Outputs with ModelPlayer™

ModelPlayer™ (http://www.physiomics-plc.com/services/modelplayer-and-database/) is a modelling platform that facilitates the simulation of cell population models and can be used to design drug schedules and combinations. It has many features crucial for investigating the properties of the model and perturbations by drug addition. These include the ability to perform parameter scans, sensitivity plots, parameter optimisation and automatic model initialization.

14.6 Conclusion

It is important to integrate modelling on different scales in order to address different questions within the drug development process. Data at the cellular and tumour level is integrated to create a more coherent picture of cancer and the effects of

anti-cancer drugs. The models can greatly assist in the testing of schedules and combinations, an area of drug development which could greatly benefit from a computational approach. Other areas where a systems biology approach could be beneficial include drug toxicity, clinical trial design, and personalized medicine. As with other branches of science, systems biology has the potential to make the drug development process more efficient.

References

Aebersold R, Auffray C, Baney E, Barillot E, Brazma A, Brett C, Brunak S, Butte A, Califano A, Celis J, Cufer T, Ferrell J, Galas D, Gallahan D, Gatenby R, Goldbeter A, Hace N, Henney A, Hood L, Iyengar R, Jackson V, Kallioniemi O, Klingmuller U, Kolar P, Kolch W, Kyriakopoulou C, Laplace F, Lehrach H, Marcus F, Matrisian L, Nolan G, Pelkmans L, Potti A, Sander C, Seljak M, Singer D, Sorger P, Stunnenberg H, Superti-Furga G, Uhlen M, Vidal M, Weinstein J, Wigle D, Williams M, Wolkenhauer O, Zhivotovsky B, Zinovyev A, Zupan B (2009) Report on EU-USA workshop: how systems biology can advance cancer research (27 October 2008). Mol Oncol 3:9–17

Alfieri R, Barberis M, Chiaradonna F, Gaglio D, Milanesi L, Vanoni M, Klipp E, Alberghina L (2009) Towards a systems biology approach to mammalian cell cycle: modeling the entrance into S phase of quiescent fibroblasts after serum stimulation. BMC Bioinformatics 10(Suppl 12):S16

Bagci EZ, Vodovotz Y, Billiar TR, Ermentrout GB, Bahar I (2006) Bistability in apoptosis: roles of bax, bcl-2, and mitochondrial permeability transition pores. Biophys J 90:1546–1559

Brightman FA, Fell DA (2000) Differential feedback regulation of the MAPK cascade underlies the quantitative differences in EGF and NGF signalling in PC12 cells. FEBS Lett 482:169–174

Chassagnole C, Jackson RC, Hussain N, Bashir L, Derow C, Savin J, Fell DA (2006) Using a mammalian cell cycle simulation to interpret differential kinase inhibition in anti-tumour pharmaceutical development. Biosystems 83:91–97

Cho KH, Wolkenhauer O (2003) Analysis and modelling of signal transduction pathways in systems biology. Biochem Soc Trans 31:1503–1509

Cho CR, Labow M, Reinhardt M, Van Oostrum J, Peitsch MC (2006) The application of systems biology to drug discovery. Curr Opin Chem Biol 10:294–302

Clyde RG, Bown JL, Hupp TR, Zhelev N, Crawford JW (2006) The role of modelling in identifying drug targets for diseases of the cell cycle. J R Soc Interface 3:617–627

Conradie R, Bruggeman FJ, Ciliberto A, Csikasz-Nagy A, Novak B, Westerhoff HV, Snoep JL (2010) Restriction point control of the mammalian cell cycle via the cyclin E/Cdk2:p27 complex. FEBS J 277:357–367

Csikasz-Nagy A (2009) Computational systems biology of the cell cycle. Brief Bioinform 10:424–434

Eisenthal R, Cornish-Bowden A (1998) Prospects for antiparasitic drugs: the case of trypanosoma brucei, the causative agent of African sleeping sickness. J Biol Chem 273:5500–5505

Fell DA (1997) Understanding the control of metabolism. Portland Press, London

Funahashi A, Morohashi M, Kitano H (2003) CellDesigner: a process diagram editor for gene-regulatory and biochemcial networks. BIOSILICO 1:159–162

Haberichter T, Madge B, Christopher RA, Yoshioka N, Dhiman A, Miller R, Gendelman R, Aksenov SV, Khalil IG, Dowdy SF (2007) A systems biology dynamical model of mammalian G1 cell cycle progression. Mol Syst Biol 3:84

Hornberg JJ, Bruggeman FJ, Binder B, Geest CR, De Vaate AJ, Lankelma J, Heinrich R, Westerhoff HV (2005) Principles behind the multifarious control of signal transduction. ERK phosphorylation and kinase/phosphatase control. FEBS J 272:244–258

Hua F, Cornejo MG, Cardone MH, Stokes CL, Lauffenburger DA (2005) Effects of Bcl-2 levels on Fas signaling-induced caspase-3 activation: molecular genetic tests of computational model predictions. J Immunol 175:985–995

Iurisci I, Filipski E, Reinhardt J, Bach S, Gianella-Borradori A, Iacobelli S, Meijer L, Levi F (2006) Improved tumor control through circadian clock induction by Seliciclib, a cyclin-dependent kinase inhibitor. Cancer Res 66:10720–10728

Kholodenko BN, Demin OV, Moehren G, Hoek JB (1999) Quantification of short term signaling by the epidermal growth factor receptor. J Biol Chem 274:30169–30181

Kola I, Landis J (2004) Can the pharmaceutical industry reduce attrition rates? Nat Rev Drug Discov 3:711–715

Ledford H (2010) Clinical drug tests adapted for speed. Nature 464:1258

Legewie S, Bluthgen N, Herzel H (2006) Mathematical modeling identifies inhibitors of apoptosis as mediators of positive feedback and bistability. PLoS Comput Biol 2:e120

Levi et al (2007) Implications of circadian clocks for the rhythmic delivery of cancer therapeutics. Adv Drug Deliv Rev 59 (9–10):1015–1035

Mormont MC, Levi F (2003) Cancer chronotherapy: principles, applications, and perspectives. Cancer 97:155–169

Noble D (2008) The music of life: biology beyond genes. Oxford University Press, Oxford

Rossignol R, Letellier T, Malgat M, Rocher C, Mazat JP (2000) Tissue variation in the control of oxidative phosphorylation: implication for mitochondrial diseases. Biochem J 347:45–53

Sauro HM, Hucka M, Finney A, Wellock C, Bolouri H, Doyle J, Kitano H (2003) Next generation simulation tools: the systems biology workbench and BioSPICE integration. Omics 7:355–372

Schoeberl B, Eichler-Jonsson C, Gilles ED, Muller G (2002) Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. Nat Biotechnol 20:370–375

Schoeberl B, Pace EA, Fitzgerald JB, Harms BD, Xu L, Nie L, Linggi B, Kalra A, Paragas V, Bukhalid R, Grantcharova V, Kohli N, West KA, Leszczyniecka M, Feldhaus MJ, Kudla AJ, Nielsen UB (2009) Therapeutically targeting ErbB3: a key node in ligand-induced activation of the ErbB receptor-PI3K axis. Sci Signal 2:ra31

Schwartz GK, Shah MA (2005) Targeting the cell cycle: a new approach to cancer therapy. J Clin Oncol 23:9408–9421

Shah MA, Schwartz GK (2001) Cell cycle-mediated drug resistance: an emerging concept in cancer therapy. Clin Cancer Res 7:2168–2181

Smolensky MH, Peppas NA (2007) Chronobiology, drug delivery, and chronotherapeutics. Adv Drug Deliv Rev 59:828–851

Theil FP, Guentert TW, Haddad S, Poulin P (2003) Utility of physiologically based pharmacokinetic models to drug development and rational drug discovery candidate selection. Toxicol Lett 138:29–49

Tyson JJ, Csikasz-Nagy A, Novak B (2002) The dynamics of cell cycle regulation. Bioessays 24:1095–1109

Urso R, Blardi P, Giorgi G (2002) A short introduction to pharmacokinetics. Eur Rev Med Pharmacol Sci 6:33–44

Chapter 15
Circadian Rhythms and Cancer Chronotherapeutics

Francis Lévi, Atilla Altinok and Albert Goldbeter

Abstract The Circadian Timing System (CTS) controls cellular proliferation and drug metabolism over a 24-h period through molecular clocks in each cell. These cellular clocks are coordinated by a hypothalamic pacemaker, the suprachiasmatic nuclei, which generate or control circadian physiology. The CTS down-regulates malignant growth in experimental models and in cancer patients. It also generates large and predictable 24-h changes in toxicity and efficacy of experimental and clinical anticancer treatments, which have been validated in randomized studies. Modelling of the interactions between circadian clocks, cell division cycle and pharmacology pathways reveals why the same circadian timing jointly optimizes the tolerability and efficacy of a given anticancer drug, both in experimental models and in cancer patients. Thus, an automaton model for the cell cycle shows the critical roles of variability in circadian entrainment, cell cycle length, and phase durations, which determine the success of cancer chronotherapeutics. Stochastic and deterministic models further confirm the poor therapeutic value of constant-rate infusion or wrongly-timed chronomodulated infusion. The integration of the circadian clock into the algorithms of anticancer treatments represents a critical step towards the tailoring of optimal chronotherapeutic delivery. The adjustment of mathematical models of circadian and cell cycle clocks to relevant clinical factors such as circadian biomarkers and gender constitutes an innovative and promising approach for personalized cancer medicine.

15.1 Circadian Rhythms in Health and Diseases

Circadian rhythms (with an approximately 24-h periodicity) have been demonstrated for most biological variables and in many living organisms, including cyanobacteria, plants, flies, rodents and humans. Rhythms on other timescales also charac-

F. Lévi (✉)
INSERM, U776, Rythmes Biologiques et Cancers, Villejuif, F-94807, France

Université Paris-Sud, UMR-S0776, Orsay, F-94807, France

Assistance Publique-Hôpitaux de Paris, Unité de Chronothérapie, Département de Cancérologie, Hôpital Paul Brousse, Villejuif, F-94807, France
e-mail: francis.levi@inserm.fr

A. Cesario, F. B. Marcus (eds.), *Cancer Systems Biology, Bioinformatics and Medicine,*
DOI 10.1007/978-94-007-1567-7_15,

terize biological functions, such as ultradian hourly rhythms in pituitary hormonal secretions or NF-κB cellular signalling pathways, and yearly rhythms in the reproductive behaviour of mammals. In this chapter, we focus on circadian rhythms, where their basic mechanisms have been investigated, modelled and examined for relevance to cancer processes and treatments. Systems approaches to circadian rhythms and cancer chronotherapy are not only highly important and relevant fields in themselves, but also serve as paradigms for modelling other body systems and treatment modalities.

15.1.1 Biological Evidence

15.1.1.1 The Circadian Timing System

The Circadian Timing System (CTS) coordinates physiology and cellular functions over a 24-h period. Environmental synchronizers, such as the alternation of days and nights, socio-professional routines, and meal times, entrain and calibrate the CTS at precisely 24 h, the endogenous period of the oscillators which constitute the CTS (Lévi et al. 2010, and see Fig. 15.1). These oscillators can display endogenous circadian periods differing from precisely 24 h in the absence of time cues (Dibner et al. 2010). However, the usual timings of light-dark, socio-professional routine, and meals timing synchronize the CTS of humans. As a result, motor activity is high in daytime and low at night; body temperature reaches a maximum in the early evening; cortisol secretion by the adrenal gland rapidly rises from a nadir near 2:00 a.m., to a maximum near 8:00 a.m. in the morning; and melatonin secretion by the pineal gland mostly occurs at night, with a maximum near 2:00 a.m. in healthy humans living in a normal routine environment (Lévi and Schibler 2007; Dibner et al. 2010). This circadian physiology is generated or controlled by a central pacemaker, the suprachiasmatic nuclei (SCN) in the hypothalamus. The circadian period of the SCN neurons is calibrated to 24 h through the perception of synchronization signals, namely light and darkness via the retino-hypothalamic tract using glutamate and pituitary-adenylate-cyclase-activating peptide (PACAP) as neuromediators, and other brain areas via neuropeptide Y fibres. The SCN generates circadian physiology through diffusible signals, including transforming growth factor α, epidermal growth factor (EGF), prokineticin-2 (PK-2), cardiotrophin-like cytokine, and neuroanatomic sympathetic and parasympathetic pathways (Lévi et al. 2010; Dibner et al. 2010). Circadian physiology, and other signals directly or indirectly emanating from the SCN, coordinate molecular clocks in each cell (Takahashi et al. 2008). In turn, the molecular clock rhythmically controls many cellular functions that are relevant for cancer treatment, including drug metabolism and detoxification as well as cellular proliferation, DNA damage sensing and repair, apoptosis, and angiogenesis (Lévi et al. 2010).

The periodic resetting of the circadian time structure by external 24 h cycles allows for the prediction of times of the peaks and troughs of circadian rhythms in

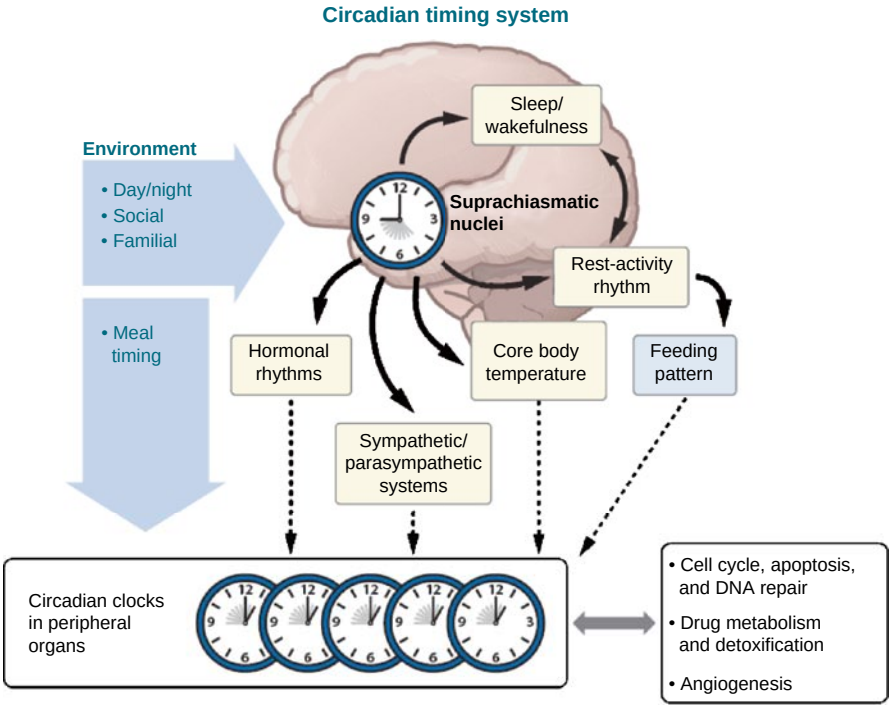


Fig. 15.1 Schematic representation of the Circadian Timing System (CTS). The CTS is composed of (a) a hypothalamic pacemaker, the suprachiasmatic nuclei SCN, (b) an array of SCN-generated circadian physiology outputs, and (c) molecular clocks in the cells of all peripheral tissues. Molecular clocks rhythmically control xenobiotic metabolism and detoxification, cell cycle, apoptosis, DNA repair, and angiogenesis over a 24 h period. The CTS is synchronized with time cues provided by light-dark cycles and other environmental factors. Circadian physiology outputs can also serve as CTS biomarkers. (Reproduced with permission from the Annual Review of Pharmacology and Toxicology, Volume 50 © 2010 by Annual Reviews, http://www.annualreviews.org.)

rodents and in humans. This applies to the rhythms that regulate anticancer drug pharmacology and cellular proliferation (Lévi et al. 2010). Conversely, a lack of external synchronizers, that is, a defect in the perception of environmental time cues through blindness, for instance, or an alteration of the circadian physiology, molecular clock, or clock-controlled pathways, results in the deregulation of the circadian time structure (Lévi et al. 2010; Dibner et al. 2010). In turn, relevant 24-h rhythms become damped, ablated, or phase-shifted, with an unpredictable timing of the peaks and troughs, if the circadian period is lengthened, shortened, or shifted. In such cases, melatonin, glucocorticoids, or other chronobiotic agents can restore proper circadian coordination (Lévi and Schibler 2007; Lévi et al. 2010).

Healthy human subjects can display different CTS phasing, despite exposure to the same environmental synchronizers. Such distinct chronotypes are defined with questionnaires on living habits, which reflect distinct timing of circadian behaviour, physiology, and clock gene expression patterns (Dibner et al. 2010).

15.1.1.2 Molecular Mechanisms of Circadian Clocks and Relevant Clock-controlled Pathways

Nearly 15 genes constitute the core of the molecular clock in mammals (Fig. 15.2). These clock genes are involved in transcriptional and post-transcriptional activation and inhibition regulatory loops that result in the generation of the circadian oscillation in individual mammalian cells. In particular, the CLOCK-BMAL1 or NPAS2-BMAL1 protein dimers play a key role in the molecular clock through the activation of the transcription of the clock genes *Per* and *Cry* (Takahashi et al. 2008; Dibner et al. 2010; Lévi et al. 2010). The functionality of the molecular clock in peripheral tissues, including malignant tumours, can be estimated through the relative phase relations of circadian expression patterns of three core clock genes, whose transcription is regulated by one another: *Rev-erbα* down-regulates *Bmal1, Bmal1* up-regulates *Rev-erbα* and *Per2*, and *Per2* down-regulates *Rev-erbα* and its own transcription (Lévi et al. 2010).

The CLOCK-BMAL1 transactivation complex also rhythmically controls the mRNA transcription of proline-acidic amino acid-rich basic leucine zipper (PAR bZip) transcription factors, including albumin D-binding protein (DBP), thyrotroph embryonic factor (TEF), and hepatic leukaemia factor (HLF) (Dibner et al. 2010). These transcription factors regulate most pathways that handle xenobiotic metabolism and detoxification in liver, intestine, and kidney through the rhythmic control of the C-androstane receptor, P450 oxydo-reductases, and 5-amino-δ-levulinic acid synthetase (*Alas1*) (Lévi and Schibler 2007). Furthermore, post-translational modifications regulate the ticking of the molecular clock (Dibner et al. 2010).

The CLOCK-BMAL1 dimer also gates cell cycle phase transitions through the repression of *c-Myc* and *p21*, two important players in cellular proliferation and apoptosis, the activation of *p53*, a proapoptotic gene, and that of *Wee1*, whose protein gates transition from G_2 to mitosis (Fu and Lee 2003; Filipski et al. 2005; Matsuo et al. 2003; Gréchez-Cassiau et al. 2008). Circadian clocks further regulate apoptosis through the rhythmic expressions of antiapoptotic BCL-2 protein and proapoptotic BAX protein (Okyar and Lévi 2008). The clock proteins PERs, CRYs and TIM also seem to control DNA damage sensing through molecular interactions with Ataxia telangiectasia mutated (ATM) and rad3-related interacting proteins (ATRIP) (Okyar and Lévi 2008). Rhythmic DNA repair potential further results from circadian changes in activities or levels of O_6-methylguanine DNA methyltransferase, a protein that excises lethal DNA-alkylated lesions produced by nitrosoureas, as well as *Tip60*, *Xpa*, and possibly *Ercc1*, which repair platinum-induced DNA adducts (Lévi et al. 2010). Furthermore, DNA damage signalling can itself reset circadian clocks (Oklejewicz et al. 2008).

The intrinsic sustainability of molecular clocks has been demonstrated in synchronized cell cultures. Thus, cell lines are potential models for *in vitro* studies of circadian clocks and clock-controlled pathways (Takahashi et al. 2008; Dibner et al. 2010). A 2-h exposure of cultured cells to 50% horse serum, dexamethasone, or other compounds, synchronizes the circadian clocks in cultured cells whose internal timing is otherwise drifting at different paces. Circadian transcription has been dem-

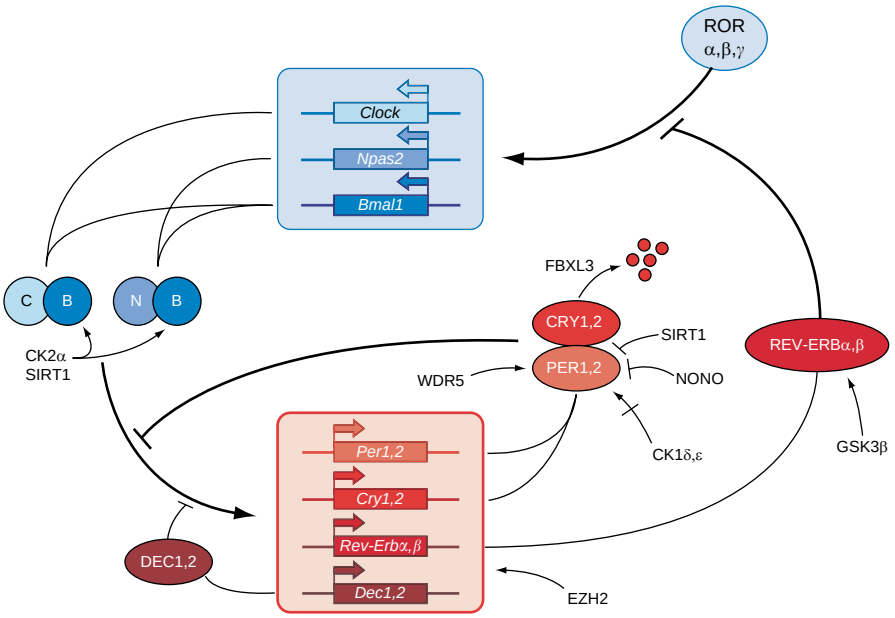


Fig. 15.2 Simplified hypothetical mammalian circadian clock. The molecular oscillator is thought to be based on molecular feedback loops within a positive limb (CLOCK, NPAS2, BMAL1) and a negative limb (PER and CRY) that are interconnected via the nuclear orphan receptor REV-ERBα. The transcription of *Per* and *Cry* genes is activated by heterodimers between BMAL1 (B) and either of the two related proteins CLOCK (C) or NPAS2 (N). The polycomb protein EZH2, together with casein kinase 2 (CK2) and the silencing information regulator SIRT1, interact with these heterodimers, and thereby facilitate their action. The accumulation and activity of PER and CRY proteins are also influenced by phosphorylation by protein kinases (CK1δ,ε); by ubiquitination via a complex containing the F-box protein FBXL3 (specific for CRYs); by the histone methyl-transferase-binding protein WDR5; and by NONO, an RNA- and DNA-binding protein. DEC1 and DEC2 compete with BMAL1-CLOCK/NPAS2 heterodimers for E-box binding, and thereby reduce E-box-mediated transactivation. An accessory feedback loop, employing the nuclear orphan receptors RORα, RORβ, and RORγ as activators, and REV-ERBα and REV-ERBβ as repressors, regulates the circadian transcription of *Bmal1*. (Reproduced with permission from the Annual Review of Pharmacology and Toxicology, Volume 50 © 2010 by Annual Reviews, http://www.annualreviews.org.)

onstrated for at least three full periods in synchronized cultures of cell lines and in *ex vivo* cellular preparations or tissue explants from rodents or humans, including CNS, liver, lung, kidney, intestine, and adipose tissue. The use of a PERIOD2-LUCIFERASE fusion protein as a real-time reporter of circadian dynamics demonstrates that peripheral tissues from mice self-sustain circadian oscillations for >20 cycles in isolation, with tissue-specific differences in circadian period and phase (Takahashi et al. 2008). Repeat serum shocks at 3-day intervals or 24-h cycles in external temperature avoid the desynchronization of *in vitro* transcription circadian rhythms (Dibner et al. 2010). The properties of synchronized cell cultures thus support their recent use as potential models for cellular chronopharmacology (Lévi et al. 2010).

15.1.2 Experimentally-based Computational Models

15.1.2.1 Modelling the Mammalian Circadian Clock

Detailed computational models for the mammalian circadian clock based on the intertwined positive and negative regulatory loops involving the *Per*, *Cry*, *Bmal1*, *Clock*, and *Rev-erbα* genes have been proposed (Leloup and Goldbeter 2004; Forger and Peskin 2003; Mirsky et al. 2009). The variables in the models are the concentrations of the clock proteins and their mRNAs. The time evolution of these variables is governed by a set of kinetic equations which can be integrated numerically. The effect of changes in experimental conditions (e.g. light or dark phase) and in parameter values (e.g. maximum rate of PER phosphorylation) can be determined by means of such numerical simulations, which show how the changes affect the period, amplitude, phase or the very occurrence of the oscillations.

We shall briefly recall here the interesting predictions of one model for the mammalian clock (Leloup and Goldbeter 2003, 2004). In agreement with experimental observations, the model gives rise to sustained circadian oscillations in continuous darkness (DD), characterized by an antiphase relationship between *Per/Cry/Rev-erbα* and *Bmal1* mRNAs. Sustained oscillations correspond to the rhythms autonomously generated by the suprachiasmatic nuclei or peripheral tissues. For other parameter values, damped oscillations can also be obtained in the model. When incorporating the light-induced expression of the *Per* gene, the model accounts for entrainment of the oscillations by light-dark (LD) cycles. Simulations show that the phase of the oscillations can then vary by several hours, with relatively minor changes in parameter values. Such a lability of the phase could account for physiological disorders related to circadian rhythms in humans, such as sleep-related syndromes (see below). The model uncovers the possible existence of multiple sources of oscillatory behaviour in the circadian clock regulatory network. Thus, in conditions where the indirect negative autoregulation of *Per* and *Cry* expression is inoperative, sustained oscillations might still arise from the negative autoregulation of *Bmal1* expression.

15.1.2.2 Computational Approaches to Circadian Clock-related Disorders

Genetic studies indicate that dysfunctions of the circadian clock in humans are associated with physiological disorders of the sleep-wake cycle (Toh et al. 2001). A phase advance or a delay of circadian rhythms is associated with the Familial Advanced Sleep Phase Syndrome (FASPS) or with the Delayed Sleep Phase Syndrome (DSPS), respectively. The computational model for the mammalian circadian clock was used to investigate the dynamical bases of such circadian disorders (Leloup and Goldbeter 2003, 2008). FASPS is associated with a decrease in the phosphorylation of the PER protein by casein kinase I (Toh et al. 2001). The model shows that a decrease in PER phosphorylation by this enzyme can indeed lead both to a decrease in the autonomous period of circadian oscillations in DD, and to a phase advance in LD, as observed for FASPS.

The model further permits conditions to be identified in which circadian oscillations systematically fail to entrain to the LD cycle. This situation can be associated with the non-24-h sleep-wake syndrome, where the sleep-wake cycle never settles to a fixed phase. Such a phenomenon is common in blind subjects, but simulations indicate that it may also occur, this time for dynamical reasons, in sighted subjects. The circadian clock model was used to clarify the mechanism that leads to the absence of entrainment, and to predict biological systemic approaches to restore normal circadian periodicity (Leloup and Goldbeter 2008). The model also uncovers the possibility of progressive drifts of the phase, with very long periods that can go up to months or even a few years, when the circadian clock has an autonomous period close to 24 h, but fails to be entrained by the LD cycle. It will be interesting to see whether such dynamic phenomena have a counterpart in clinical observations of circadian disorders.

To properly address the dynamical bases of FASPS, however, one needs to extend the model to take into account the multiple phosphorylations of PER by CKIε (Xu et al. 2007). A first phosphorylation of PER leads to increased protein degradation, while a second phosphorylation is associated with reduced nuclear clearance of the protein (Vanselow et al. 2006) and altered interaction with CLOCK–BMAL1 leading to enhanced *Per* transcription (Xu et al. 2007). When these two phosphorylations are incorporated, the mammalian clock model can account for their two opposite effects: decrease in period and phase advance upon increasing the first phosphorylation by CK1, as observed for the Tau mutation (Gallego et al. 2006), and upon decreasing the second phosphorylation of PER by CK1, as observed for FASPS (Leloup and Goldbeter 2010).

15.1.2.3 Theoretical Models for Investigating Circadian Timing System Dynamics During Chronic Jet Lag

Jet lag associated with rapid travelling across time zones represents one of the most common perturbations of circadian rhythmicity. Many work patterns involve shifting schedules chronically. To study the pathophysiological implications of such perturbation of the circadian clock, an animal model for chronic jet lag has been developed. One of the most drastic schedules used in experiments in mice considers a phase advance of 8 h every 2 d, i.e. one D phase out of two is reduced from 12 to 4 h, while the L phase keeps a constant duration of 12 h. Experiments indicate that endogenous circadian rhythms are lost, and grafted tumours develop more rapidly, in mice subjected to such chronic jet lag schedules (Filipski et al. 2005). The model for the mammalian circadian clock allows us to investigate the nature of the dynamical behaviour of the circadian system subjected to chronic jet lag. Simulations from the model indicate that in these conditions, the behaviour becomes chaotic or quasiperiodic (Leloup and Goldbeter 2008). In either case, the circadian network ceases to oscillate periodically. These results show how models and computer simulations can help in characterizing the effect of chronic perturbations of the circadian clock.

15.2 Chronopharmacology, Chronotolerance and Chronoefficacy of Anticancer Drugs

15.2.1 Experimental Evidence and Mechanisms

15.2.1.1 Circadian Timing for Improving Anticancer Drug Tolerability

Circadian timing modifies 2- to 10-fold the extent of toxicity of 40 anticancer drugs, including cytostatics, cytokines, and targeted biological agents, in mice or rats (Lévi et al. 2010). Chronotolerance patterns are found irrespective of delivery route or schedule. They persist in constant darkness or in constant light, which demonstrates their endogenicity. The optimal circadian timing of different anticancer drugs is staggered along the 24 h period, and cannot as yet be predicted using the knowledge of pharmacologic class or of main target organs for toxicity. Even when combined, chemotherapeutic agents display the least toxicity near their respective times of best tolerability as single agents, as shown for doxorubicin-cisplatin, irinotecan-oxaliplatin, gemcitabine-cisplatin, and docetaxel-doxorubicin (Lévi et al. 2010). These findings support the persistence of the circadian control of anticancer drug metabolism and molecular targets after exposure to the first anticancer agent, at least when the latter is given near the time of best tolerability.

Mechanisms of chronotolerance involve the CTS control of both phase I metabolism and phase II detoxification and elimination of anticancer drugs, through redundant processes involving rhythmic physiology and circadian clock signalling (Fig. 15.3) (Lévi et al. 2010). Indeed, circadian changes have been reported for the mRNA and/or protein and/or activity of Cyp isoenzymes, cytochrome P-450 oxidoreductase, dihydropyrimidine dehydrogenase (DPYD), cytidine deaminase, carboxylesterases, as well as reduced glutathione (GSH) and UGT1A. Circadian clocks further control the transcription of several ATP Binding Cassette family members, including *Abcb1a* and *Abcb1b* (*Mdr1*), *Abcc2* (*Mrp2*), and *Abcb4* (*Mdr2*), a fact which accounts for the cellular uptake and efflux of many anticancer drugs (Lévi et al. 2010). The multiple circadian controls of drug absorption, distribution, metabolism, and elimination (ADME) result in specific drug chronopharmacokinetic patterns that have been determined for 15 anticancer drugs in experimental models (Lévi et al. 2010). The lack of consistent relationships between blood chronopharmacokinetics and toxicity raises doubts about the relevance of chronopharmacokinetics as the main mechanism responsible for chronotolerance.

Rather, experimental data supports chronopharmacodynamics as a critical mechanism of dosing time-dependent effects of anticancer drugs. The relevance of the control of cell-cycle-related events by the circadian clock plays an important role, since most anticancer drugs interfere with cell cycle mechanisms, DNA repair and/or apoptosis. Thus, cell cycle events are coordinated along the 24-h period in healthy bone marrow, gut, and skin, three frequent targets for the toxicity of cancer treatments (Bjarnason and Jordan 2000; Lévi et al. 2010; Innominato et al. 2010).

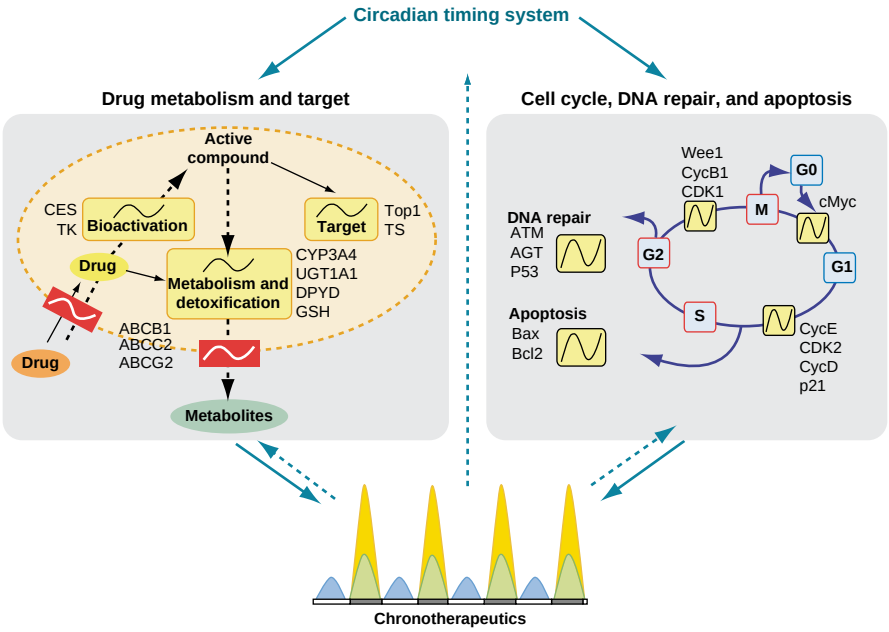


Fig. 15.3 Main cellular determinants of cancer chronotherapeutics. The CTS (*top*) determines the optimal circadian timing of anticancer medications (*bottom*). The CTS controls drug transport, bioactivation, detoxification, metabolism, targets, and elimination, which accounts for the chronopharmacology of anticancer agents at cellular, tissue, and whole organism levels (*left panel*). The CTS also regulates several cell-cycle-related events that gate G1/S or G2/M transitions, as well as DNA repair and apoptosis, which accounts for the chronopharmacodynamics of anticancer drugs (*right panel*). The relations between chronopharmacokinetics and chronopharmacodynamics help construct optimal chronomodulated drug delivery schedules, with proper parameters. (Reproduced with permission from the Annual Review of Pharmacology and Toxicology, Volume 50 © 2010 by Annual Reviews, http://www.annualreviews.org.)

The proportions of S- and G2/M-phase cells increase by ~50% in the second half of darkness, whereas G0-G1 phase cells predominate during light in mouse bone marrow (Granda et al. 2005). In this tissue, BCL2 anti-apoptotic protein expression triples over the 24 h period, with a maximum at early light. An opposite pattern characterizes pro-apoptotic BAX, with a fivefold circadian change (Granda et al. 2005). The temporary arrest of cycling cells in G0-G1, the high BCL2, and the low BAX expression during the light span when mice are resting help explain the best circadian timing for the tolerability of 5-FU, gemcitabine, irinotecan, and docetaxel in male B6D2F1 mice (Lévi et al. 2010). However, the circadian control of drug metabolism and detoxification also profoundly modifies the cellular exposure to these medications, whose molecular targets are usually clock-regulated (Lévi and Schibler 2007). For instance, the increased detoxification of 5-FU during the light span results from the circadian peak in DPYD activity in liver and other healthy cells. This implies that the reduced proportion of S-phase cells in bone marrow, gut, and skin, is a mechanism for improved circadian tolerability (Lévi

and Schibler 2007). Both transcription and activity of thymidilate synthetase (TS), which provide the unique *de novo* source of thymidilate, are linked to early S-phase in proliferating tissues. Consistently, bone marrow TS activity peaks near mid-dark in coincidence with the greatest hematologic toxicity of 5-FU in mice (Lévi and Schibler 2007).

The CLOCK:BMAL1 dimer in the molecular clock directly or indirectly controls several molecular mechanisms involved in the chronotolerance of cyclophosphamide and cisplatin, two alkylating agents, as well as mitoxantrone and irinotecan, which respectively inhibit Topoisomerase 2 and 1 (Gorbacheva et al. 2005; Gachon et al. 2006; Lévi et al. 2010).

15.2.1.2 Relevance of Drug Timing for Treatment Efficacy

Circadian timing also critically affects the antitumoral efficacy of 28 anticancer medications, including cytostatics, anti-angiogenic agents, and cell cycle or Cox2 inhibitors in rodents with various kinds of malignancies (Lévi et al. 2010). Appropriately circadian-timed and dosed chemotherapy with one or several drugs at least halves tumour growth rate, and/or significantly increases life span in tumour-bearing mice (Lévi et al. 2010). Strikingly, the circadian pattern in chronoefficacy usually coincides with that in chronotolerance. This holds true for cytostatics, interferons, anti-angiogenic agents, and cell cycle inhibitors, as well as for combination chemotherapy, such as irinotecan-oxaliplatin, gemcitabine-cisplatin, and docetaxel-doxorubicin, three widely used clinical regimens (Lévi et al. 2010).

The chronoefficacy of anticancer medications partly results from circadian changes in tumour drug uptake and/or from the circadian control of drug pharmacodynamics and/or vascular endothelial growth factor-mediated neo-angiogenesis in tumours.

However, circadian disruption frequently adds to cell cycle disruption as a hallmark of cancer, at least in rapidly growing malignancies and at an advanced stage of tumour evolution. Clock gene transcription is no longer circadian in advanced GOS or pancreatic adenocarcinoma P03 (Filipski et al. 2005; Iurisci et al. 2006; Li et al. 2010). No circadian organization is found for S-phase cells in GOS or mammary carcinoma MA13C, for BCL2 protein expression in MA13C, or for GSH content in P03 (Granda et al. 2005; Lévi et al. 2010). Nevertheless, chronoefficacy remains robust in these experimental tumours, possibly because (*a*) the CTS of the host determines the chronoefficacy of anticancer medications, and/or (*b*) an adequate resetting of tumour circadian clocks by anticancer medications critically contributes to their efficacy (Lévi et al. 2010).

15.2.1.3 The Circadian Timing System as a Target for Anticancer Drugs

Circadian biomarkers such as 24-h rhythms in rest-activity, core body temperature, and urinary or blood variables, can be severely disrupted by anticancer drugs of any

pharmacologic class. Anticancer agents can also impair molecular circadian clocks in the CNS and/or in peripheral organs (Ohdo et al. 2001; Iurisci et al. 2009). The extent and duration of circadian disruption depends upon both dose and dosing time, as shown for 12 anticancer drugs, including irinotecan, oxaliplatin, vinorelbine, interferon-α, or seliciclib (Ohdo et al. 2001; Lévi et al. 2010). Treatment at the circadian time associated with fewest toxicities usually best spares the CTS of the host, irrespective of the underlying toxicity mechanisms or target tissues. DNA repair elicited by damage from γ-radiations and possibly anticancer drugs resets free-running host circadian clocks via ATM-mediated damage signalling (Oklejewicz et al. 2008). On the other hand, tumours frequently escape from circadian coordination. Thus, no circadian expression pattern is found for *Per2*, *Bmal1*, and *Rev-erbα* in two advanced experimental tumours with different growth rates (Iurisci et al. 2006; Li et al. 2010). However, the amplification of the circadian rhythm in body temperature through meal timing re-programmes the tumour circadian transcriptome. This effect involves the entrainment of temperature-sensitive stress proteins during the 24 h period. As a result, a 24-h rhythmic expression of cell cycle and pharmacology determinants is induced in tumours, which translates into nearly halving their progression rate (Li et al. 2010). Such circadian induction effects can be achieved with seliciclib, a cyclin-dependent kinase inhibitor, which also inhibits casein kinase Iδ/ε, a key enzymatic regulator of the circadian period. Tumour clock induction by seliciclib is only achieved if this drug is administered during the light span, when its antitumoral efficacy is twice as high as that found following dosing during the dark span, when the mice display high locomotor activity (Iurisci et al. 2006).

15.2.2 Clinical Cancer Chronotherapeutics

Although clinical trials involving circadian timing have been performed in patients with ovarian, breast, lung, kidney, head-and-neck and pancreatic cancer, we will focus here on metastatic colorectal cancer, where the largest international experience has been gathered (Innominato et al. 2010).

The development of the chronotherapeutics of metastatic colorectal cancer has involved the design of chronomodulated drug delivery schedules, based upon preclinical and human studies. The theoretical treatment schedules involved the modelling of chronotherapeutic effects with a 24-h periodic cosine function, using midnight as a phase reference (Lévi et al. 1994, 2007, 2010). The delivery of a given drug was continuous or intermittent, especially whenever drug combinations were administered, and extended up to several weeks. These chronomodulated treatment schedules were implemented in single institution trials using dedicated programmable-in-time drug delivery technology. Prospective validation was undertaken in multicentre randomized comparisons of a chronotherapeutic schedule, with constant-rate infusions using the same initial doses over the same treatment duration.

Indeed, constant-rate infusion over at least 24-h eliminates any circadian timing hypothesis for drug administration. Both of these drug delivery schedules were designed for assessing tolerability and efficacy of the combination of 5-FU-leucovorin and oxaliplatin in patients with metastatic colorectal cancer.

15.2.2.1 Chronotolerance

The design of the first chronotherapeutic schedule was based on experiments in male mice, where the times of least toxicity were located near mid-activity for oxaliplatin and near mid-rest for 5-FU (Lévi et al. 2007). These circadian times were extrapolated to cancer patients, with the chronomodulated schedule combining the daily delivery of oxaliplatin over 11.5 h with peak flow rate at 4:00 p.m., and that of 5-FU-leucovorin over 11.5 h with peak flow rate at 4:00 a.m., for five consecutive days (chronoFLO5). The other cohort of patients received the same doses of the same three drugs, at a constant rate over the same five-day span. In two international randomized phase III trials involving 278 patients with metastatic colorectal cancer, chronomodulated delivery reduced the incidence of grade 3–4 mucositis by fivefold, and halved the incidence of peripheral sensory neuropathy (Table 15.1) (Lévi et al. 1994, 1997). The largest trial also reported a threefold reduction in the rate of hospitalizations for toxic events with chronomodulated infusions (Lévi et al. 1997). A subsequent study involved the comparison of time-lagged chronomodulated infusion profiles, in order to better define the characteristics of optimal chronotherapeutic delivery. Two kinds of multiple-arm chronotherapeutic trials addressed the issue of tolerability as the main endpoint. In the first design, peak times of chronomodulated infusions were shifted by a few hours over 24 h, yet with fixed intervals between the chronomodulated delivery patterns of the drugs in the combination. In 114 patients with metastatic colorectal cancer, peak times of oxaliplatin and 5-FU-leucovorin shifted by 3, 6, 9 or 12 h were compared with the reference profile, where delivery rate peaks at 4:00 p.m. for oxaliplatin and at 4:00 a.m. for 5-FU-leucovorin. This design assumed that it was important to maintain a fixed 12-h interval between the peak delivery rates of oxaliplatin and 5-FU-leucovorin. Severe toxicity occurred in 16.7% of the patients on the reference chronoFLO4 schedule and in 80% of those on the opposite chronomodulated modality (Table 15.1) (Lévi et al. 2007). The optimal times of peak delivery rate were defined with their 90% C. I., both for 5-FU-leucovorin and for oxaliplatin, with subsequent validation of an optimal timing for carboplatin similar to that of oxaliplatin (Lévi et al. 2007). Another design to find optimal times of administration involves staggering the peak times of chronomodulated delivery of the single drug of interest every 3 or 4 h over 24 h. In 90 patients with metastatic breast cancer, the peak delivery time of vinorelbine was shifted by 3, 6, 9, 12, 15, 18, or 21 h, whereas peak delivery time of chronomodulated 5-FU was fixed at 4:00 a.m. The least leukopenia corresponded to peak vinorelbine delivery at 5:15 p.m. (2:12 to 8:08 p.m.), a result in good agreement with chronotolerance in female mice (Cou-

Table 15.1 Main toxicity and efficacy outcomes in cancer patients receiving infusional 5-fluorouracil, leucovorin and oxaliplatin with chronomodulated or constant delivery rate

Delivery schedule	278 patients without any previous chemotherapy		114 patients failing prior chemotherapy	
	ChronoFLO	Constant rate	ChronoFLO	'Opposite' chronoFLO
Severe toxicity (grade 3–4)	14%[a]	76%*	16%	80%*
Major tumour responses	51%	30%*	30%	12%*

* ChronoFLO vs. other schedules, $p < 0.05$

[a] Percentage of patients.

The reference chronoFLO schedule depicted in Figure 15.3 was compared to constant rate infusion of the same 3 drugs in Phase III trials. Main results revealed that oral mucosa tolerability was improved 5-fold by chronoFLO as compared with constant infusion, while antitumour efficacy, as assessed by tumour response rate was nearly twice as high with chronoFLO5. Similar differences in tolerability and antitumour efficacy were noticed between the reference chronoFLO and chronomodulated schedules with peak times of delivery rate occurring at 4 p.m. for 5-FU-LV and at 4 a.m. for oxaliplatin in patients refractory to a first conventional chemotherapy regimen (After Lévi et al. 2008).

dert et al. 2008). Fewer dose reductions and/or treatment delays also occurred for peak vinorelbine delivery in the late evening hours (Coudert et al. 2008). Both trial designs assume that the CTS and the clock-controlled pharmacologic pathways remain stable after being challenged by the first medication studied. However, vinorelbine can induce circadian disruption in mice as a function of dose and dosing time (Lévi et al. 2007).

15.2.2.2 Chronoefficacy

The relevance of a validated chronomodulated delivery regimen for antitumoral efficacy was shown in two consecutive European randomized trials involving 278 patients with metastatic colorectal cancer. ChronoFLO5 was compared with constant-rate infusion over 5 d every 3 weeks. The percentage of patients whose metastases regressed by ≥50% was 29% on constant-rate infusion, and 51% on chronomodulated delivery ($p < 0.001$) (Table 15.1). However, overall survival did not significantly differ according to treatment schedule, an issue further discussed in Sect. 15.3 (Lévi et al. 1994, 1997, 2007). Moreover, the antitumoral activity of the least toxic chronomodulated schedules was also greater than that of the most toxic schedule (Table 15.1; Lévi et al. 2008). Critical progress in the management of patients with colorectal cancer metastases was further brought about by the chronotherapy trials: it was demonstrated that previously unresectable metastases in liver or lung could become resectable as a result of their downsizing through effective treatment, thereby leading to prolonged survival beyond five years, and even cure (Giacchetti et al. 1999; Adam et al. 2009).

15.2.3 *Probing Circadian Patterns of Anticancer Drug Delivery* in silico

Assessing the effectiveness of various temporal schedules of drug delivery is central to cancer chronotherapeutics. Modelling tools can help to optimize time-patterned drug administration to increase effectiveness and reduce toxicity (Goldbeter and Claude 2002). Probing the effect of circadian delivery of anticancer drugs by means of modelling and numerical simulations requires a model for the cell cycle. More or less detailed kinetic models have been proposed for the embryonic and yeast cell cycles and for the mammalian cycle (Qu et al. 2003; Swat et al. 2004: Novak and Tyson 2004; Csikasz-Nagy et al. 2006; Gérard and Goldbeter 2009). An alternative approach, however, is to rely on a simple phenomenological description of the cell cycle in terms of an automaton that switches between sequential states corresponding to the successive phases of the cell cycle. Not based on molecular details, this cell cycle automaton (CCA) model provides a simple phenomenological description of the cell cycle (Altinok et al. 2007a). The presence of anticancer drugs leads to probabilistic exit from the cell cycle progression, according to drug concentration. A random component is introduced in the cell cycle automaton to take into account the variability of transitions between the phases in a proliferating cell population. The cell cycle automaton can readily be used to investigate the impacts of different temporal patterns of drug administration on cell proliferation (Altinok et al. 2007a, b, 2009).

Anticancer medications generally exert their effect by interfering with the cell division cycle. The major effect of anticancer drugs interfering with the cell cycle is thus to block cells in a specific phase before cell death. Thus, the antimetabolite 5-FU disorganizes pyrimidine metabolism in cells undergoing DNA replication, and is therefore toxic for cells in S phase. Conversely, alkylating agents such as oxaliplatin exert their effects in all phases of the cell cycle.

To illustrate the use of computer simulations in searching for optimal patterns of circadian administration of anticancer drugs, we focus here on the chronotherapeutic scheduling of 5-fluorouracil (5-FU), a reference drug for treating gastrointestinal, breast and various other cancers. The half-life of this medication is 10–20 min; the exposure pattern will therefore be the only one considered here, since it matches rather well with the corresponding chronotherapeutic drug-delivery schedule. The CCA model has also been used for oxaliplatin, with results that differ from those obtained for 5-FU (see below).

15.2.3.1 Automaton Model for the Cell Cycle

The automaton model for the cell cycle is based on the following assumptions. The cell cycle consists of four successive phases along which the cell progresses: G1, S, G2, and M. Upon completion of the M phase, the cell transforms into two cells, which immediately enter a new cycle in G1 (the possibility of temporary arrest in

a G0 phase is not considered here, as we are focusing on a population of proliferating cells). Each phase is characterized by a mean duration D and a variability V. As soon as the prescribed duration of a given phase is reached, the transition to the next phase of the cell cycle occurs. The time at which the transition takes place varies in a random manner according to a distribution of durations of cell cycle phases. In the case of a uniform probability distribution, the duration varies in the interval [D (1−V), D (1+V)]. At each time step in each phase of the cycle, the cell has a certain probability to be marked for exiting the cycle and dying at the nearest G1/S or G2/M transition. To allow for homeostasis, which corresponds to the maintenance of the total cell number within a range in which it can oscillate, cell death must counterbalance cell replication at mitosis. These rules were used to simulate the dynamic behaviour of the cell cycle automaton in a variety of conditions, with or without entrainment by the circadian clock.

Entrainment by the circadian clock can be included in the automaton model by considering that the protein WEE1 undergoes circadian variation due to induction by the circadian clock proteins CLOCK and BMAL1 of the expression of the *Wee1* gene. Wee1 is a kinase that phosphorylates, and thereby inactivates, the cyclin-dependent kinase Cdk1 that controls the transition G2/M. In agreement with observations in human cells (Bjarnason and Jordan 2000), we consider that Wee1 peaks at 10 p.m. The decline in Wee1 activity is followed by a rise in the activity of the kinase Cdk1, which enhances the probability of transition to the M phase. We thus consider that the rise in Wee1 is immediately followed by a similar rise in Cdk1 kinase. Upon entrainment by the circadian clock, cells become more synchronized than in the absence of entrainment. In contrast to the progressive dampening of the oscillations in the absence of entrainment, oscillations are sustained when the cell cycle automaton is driven by the circadian clock (Altinok et al. 2007a).

15.2.3.2 Optimizing Temporal Patterns of 5-FU Delivery

The effect of 5-FU can be incorporated into the automaton model by assuming that cells exposed to 5-FU while in S phase have an enhanced propensity to quit the cycle at the next G2-M transition. This propensity to exit the cycle is taken as being proportional to the amount of 5-FU. Two kinds of temporal profile of 5-FU were compared. Either 5-FU remains constant in time, or it is delivered in a circadian semi-sinusoidal manner, with a peak time that will vary along a 24-h span. Then, over a period of 24 h, no 5-FU is administered for 12 h, while a semi-sinusoidal delivery occurs over the remaining 12 h. As seen above, this schedule corresponds to the temporal pattern used clinically when the peak circadian delivery occurs at 4 a.m. (Lévi et al. 1994, 1997). A similar pattern was used to determine the effect of changing the time for maximum drug delivery, so as to compare the predictions with clinical observations (Lévi et al. 2007).

The numerical simulations of the cell cycle automaton (CCA) model show that the cytotoxic effect of the anticancer drug markedly depends on the peak time of its circadian administration. Thus, cytotoxicity is minimal when the peak time is

around 3–4 a.m., and maximal when the peak time is in the range 10 a.m. to 6 p.m. These results are obtained when the cell cycle length is about 22 h. As will be discussed below, the outcome of the numerical simulations depends on the cell cycle duration.

The model provides an explanation for the change in cytotoxicity according to the peak time of circadian 5-FU delivery. When the CCA model is entrained by the circadian clock, the fraction of cells in S phase passes through a maximum during the light phase, and through a minimum during the dark phase. The phase of the entrained cell cycle is determined by the time at which the peaks in WEE1 and CDK1 occur; these peak times are in turn set by the time at which BMAL1 reaches its maximum during the day. When 5-FU delivery peaks at 4 a.m., the peak in 5-FU occurs at a time where the fraction of cells in S phase is minimum (Fig. 15.4a). Cytotoxicity of the drug is at that time weak, since relatively few cells are in the phase sensitive to 5-FU. In contrast, cytotoxicity is much larger when 5-FU peaks at 4 p.m. (Fig. 15.4c), because maximum drug exposure occurs precisely when the fraction of cells in S phase passes through a maximum. Intermediate cytotoxicity is observed when the peak times of 5-FU exposure occur, for example, at 10 a.m. (Fig. 15.4b) or 10 p.m. (Fig. 15.4d). Indeed, the peak in 5-FU then partly overlaps with the peak in the fraction of cells in S phase. When the infusion of 5-FU becomes continuous, there is always an overlap with the fraction of cells in S phase (Fig. 15.4e). The cytotoxicity of the drug is then close to that observed for the most cytotoxic circadian pattern, i.e. with peak at 4 p.m.

15.2.3.3 Optimizing Temporal Patterns of Oxaliplatin Delivery

Oxaliplatin (l-OHP) damages cells by irreversibly binding to DNA and forming inter- and intra-strand bridges. In contrast to 5-FU, the cytotoxicity of l-OHP is not specific to any particular cell cycle phase. What is specific, however, is the capacity of l-OHP to form complexes with compounds such as plasma and cellular reduced thiols (PSH) and gluthathione (GSH). Once complexed with either PSH or GSH, l-OHP becomes inactive. Both PSH and GSH display circadian variations.

The model shows that in contrast to the case of 5-FU, the cytotoxicity is more pronounced when peak circadian delivery of l-OHP occurs at 4 a.m. rather than at 4 p.m. even though l-OHP affects all phases of the cell cycle in a similar way (Altinok et al. 2009). Such a dependence of cytotoxicity on the temporal pattern of l-OHP delivery stems from the existence of circadian variations in plasma thiols and cellular glutathione, with which l-OHP forms inactive complexes. As a result, the amount of oxaliplatin in the effective cytotoxic form is smaller at 4 p.m. than at 4 a.m. This difference results from the circadian profiles of PSH and GSH. Because the effect of plasma thiols seems predominant compared to that of gluthathione, the remaining free-form l-OHP is more abundant when l-OHP peaks at 4 a.m. This is conceivably the reason why the circadian delivery pattern peaking at 4 a.m. is more cytotoxic than the pattern peaking at 4 p.m. (Altinok et al. 2009).

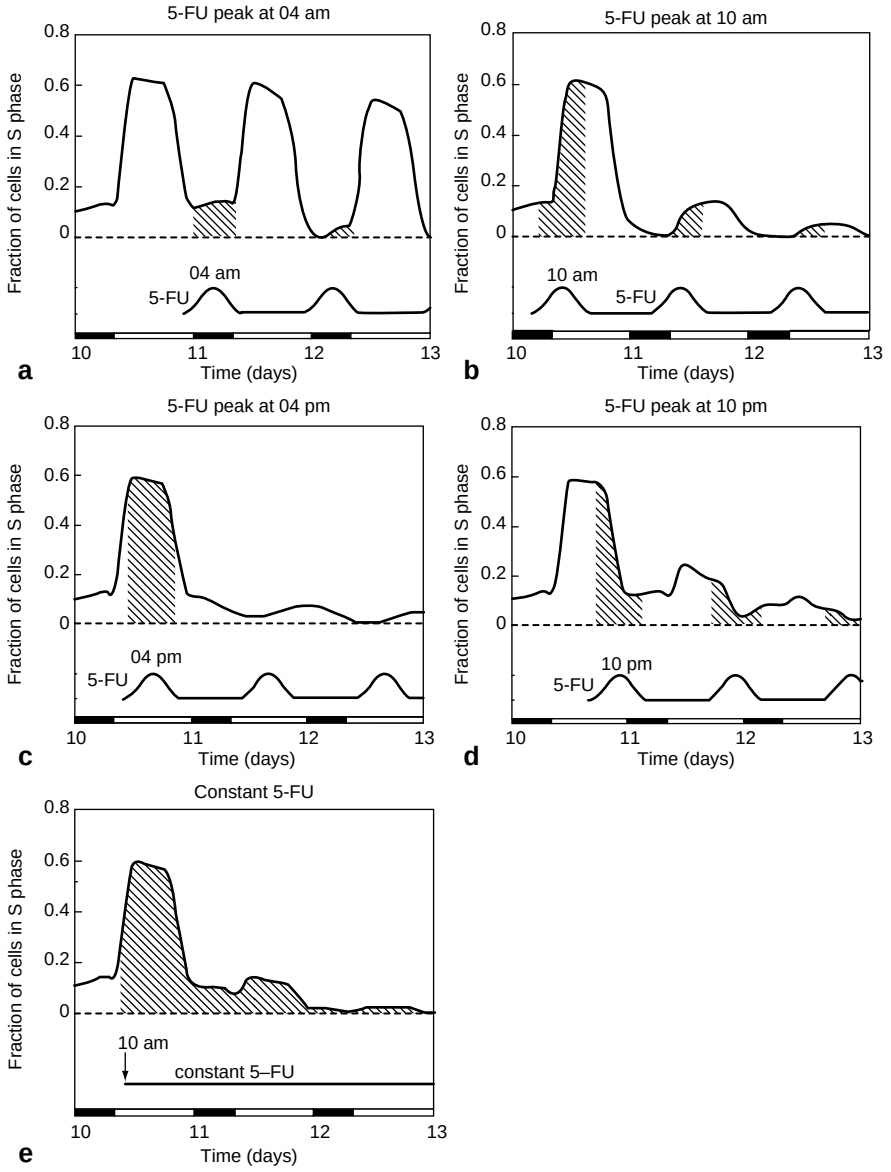


Fig. 15.4 Differential cytotoxicity of circadian administration of 5-FU peaking at 4 a.m. (a), 10 a.m. (b), 4 p.m. (c) and 10 p.m. (d), compared to constant administration (e) as predicted by the cell cycle automaton model. The variability of the duration of cell cycle phases is equal to 15%. Data is obtained for a cell cycle duration of 22 h, in the presence of entrainment by the circadian clock. The hatched area shows the fraction of cells in S phase exposed to 5-FU and thus probably marked to exit the cell cycle at the next G2/M transition. The curves showing the cumulated number of cells killed (in units of 10^4 cells) indicate that the schedule with peak delivery at 4 a.m. is the one that causes minimal damage to the cells, because the peak in 5-FU then coincides with the trough of the oscillations of S-phase cells. (from Altinok et al. 2007b)

15.2.3.4 Chronotolerance and Chronoefficacy: Insights from a Computational Approach

In searching for optimal patterns of anticancer drug delivery, two distinct goals are pursued. Healthy tissues should be protected as much as possible from drug toxicity, while at the same time maximum damage is attempted to be caused to the tumour. This issue was addressed for oxaliplatin, using other modelling approaches based on optimal theory (Clairambault 2007). In the clock-gated cell cycle automaton, drug toxicity was determined in a single-cell population. We focused on conditions corresponding to minimum cytotoxicity. The question arises as to how a temporal pattern ensuring maximum protection to healthy tissue could at the same time co-exist with enhanced toxicity toward tumour cells (see above, Sect. 15.2.1.2). To address this issue, let us consider two populations of cells differing by cell cycle duration, variability in duration of the cell cycle phases, and/or entrainment by the circadian clock. Besides these three factors, it is likely that additional differences exist between the normal and tumour cell populations.

Shown in Fig. 15.5a is the cytotoxicity over five consecutive days determined for the circadian pattern of 5-FU delivery peaking at 4 a.m., which is the least toxic pattern for a cell cycle duration of 22 h (see Fig. 15.4). The three curves show the evolution for a cell population entrained (E) by the circadian clock and characterized by a low variability ($V = 5\%$), a second population entrained but characterized by a higher variability (V = 15%), and a third population with the same variability, but which is not entrained (NE) by the circadian clock. We observe a marked difference in cytotoxicity as a function of either factor (arrows marked (a) and (b) in Fig. 15.5a). Both factors are additive, so that the differential effect becomes larger when the two populations differ by both cell cycle variability and circadian entrainment. Such differential effects only occur for circadian administrations of 5-FU and are not observed for constant 5-FU delivery (Fig. 15.5b), a schedule which is as cytotoxic as the circadian pattern peaking at 4 p.m. (see Fig. 15.4; Altinok et al. 2007). The differential cytotoxic effects predicted by the model for circadian administration of 5-FU as a function of circadian entrainment and variability are not observed for l-OHP, whether its delivery pattern is circadian or constant (Altinok et al. 2009).

One further difference between the two cell populations pertains to the duration of the cell cycle prior to entrainment. Thus, if the normal cell population has a cell cycle duration of 22 h prior to entrainment and if the tumour cell population has a different cell cycle duration, then the circadian pattern of drug delivery peaking at 4 a.m., which is least cytotoxic to the normal population, might be more cytotoxic toward the second population (Altinok et al. 2009). This is illustrated in Fig. 15.6, where panels A-C show the time evolution of a population of proliferating cells until 5-FU is administered in a circadian manner with a peak at 4 a.m., under the same conditions as in Fig. 15.5, i.e. for a cell cycle duration of 22 h. The curves show the evolution of the total number of cells in the three conditions depicted in Fig. 15.5. The data of Fig. 15.6 indicates that cytotoxicity is minimal when the variability V is low and cells are entrained by the circadian clock (a), but progressively increases when the variability is larger, with (b) or without (c) entrainment. Interestingly, the

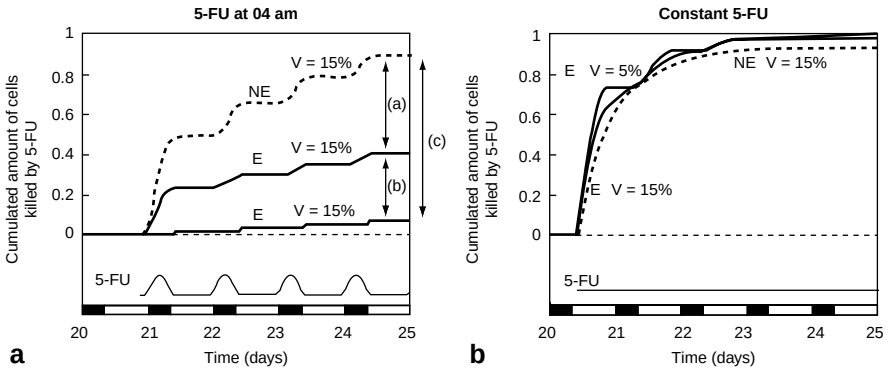


Fig. 15.5 Differential effects of circadian or continuous administration of 5-FU in cell populations differing by variability in cell cycle durations or/and entrainment by the circadian clock. **a** Comparison of cumulative cell kill (in units of 10^4 cells) upon circadian delivery of 5-FU peaking at 4 a.m. for two cell populations differing by variability *V* or/and by circadian entrainment. (a) The variability between the two populations is 15%, but one is entrained (E) and the other not (NE). (b) Both populations are entrained by the circadian clock but the variabilities differ. (c) The first population is not entrained and the variability is 15%, whereas the second one is entrained and the variability is 5%. **b** All the differences observed in (a) disappear when 5-FU is administered in a constant manner. The total quantity of drug delivered over 24 h is the same as in (a). (from Altinok et al. 2009)

simulations indicate (Fig. 15.6d) that cytotoxicity is even larger for a cell population of similarly large variability in the absence of entrainment, when the cell cycle length is 18 h rather than 22 h. As mentioned above, a similar, though weaker, effect is predicted in the case of l-OHP.

These results show that chronoefficacy may coincide with chronotolerance when the same temporal pattern of drug delivery peaking at 4 a.m. is applied to two populations differing by variability in duration of cell cycle phases; capacity of being entrained by the circadian clock; and/or the duration of the cell cycle. As demonstrated in Fig. 15.6, maximum differential effects are obtained when the two populations differ in all three characteristics (Altinok et al. 2009).

15.3 From Standard to Personalized Cancer Chronotherapeutics

15.3.1 Experimental and Clinical Data

15.3.1.1 Gender as a Determinant of Optimal Treatment Schedule

A large international randomized trial compared the chronomodulated administration of the same three drugs over four days (chronoFLO), with a two-day conven-

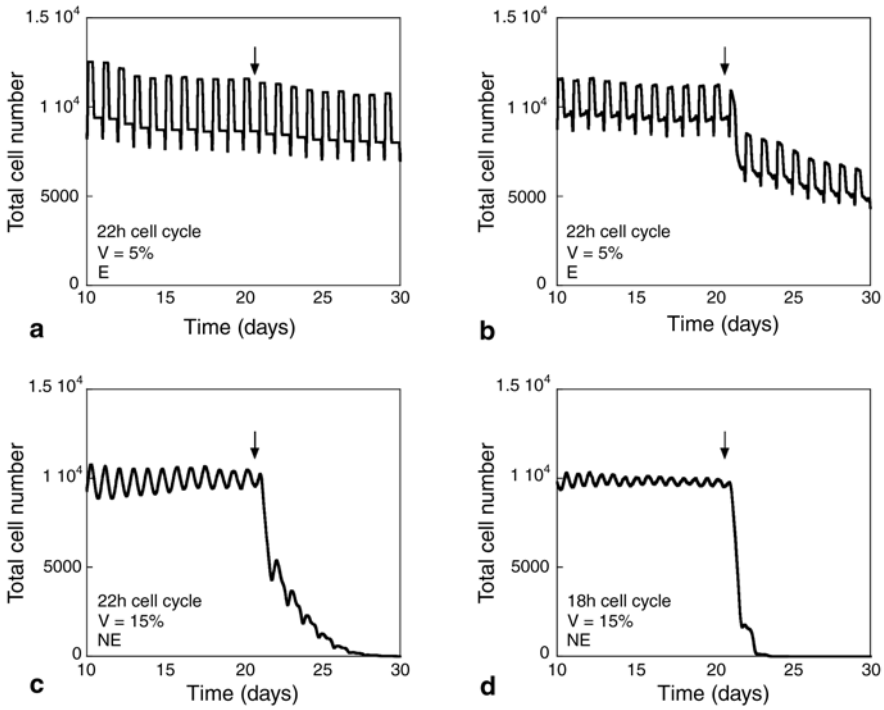


Fig. 15.6 The same circadian pattern of 5-FU administration can simultaneously display minimum toxicity for a first cell population (chronotolerance), and significant toxicity for a second cell population (chronoefficacy). **a** Cell population with cell cycle duration of 22 h and variability $V = 5\%$, entrained (E) by the circadian clock. **b** Same as (a) with variability $V = 15\%$. **c** Same as (b) without entrainment (NE) by the circadian clock. **d** Same as (c) with a cell cycle duration of 18 h. Comparison of cases (a) and (d) indicates strong cytotoxic effects of the same temporal pattern of 5-FU delivery, where the two cell populations differ by properties such as entrainment by the circadian clock, variability in duration of cell cycle phases, and cell cycle duration. (from Altinok et al. 2009)

tional administration schedule without any timing stipulation (FOLFOX2), in 564 previously untreated patients with metastatic colorectal cancer (Giacchetti et al. 2006). The trial was intended to treat each patient at the near-maximum tolerated dose. Overall survival, the main endpoint in this large study, did not differ as a function of treatment schedule. However, the relative risk of an earlier death on chronoFLO4 significantly increased by 38% in women, and significantly decreased by 25% in men, compared with conventional delivery (Giacchetti et al. 2006). A recent meta-analysis of the three randomized trials comparing chronoFLO to a conventional infusion modality of the same drugs in 842 patients with metastatic colorectal cancer confirms that for women, administration by chronoFLO achieves similar or worse efficacy, compared with conventional delivery. In men, however, the same chronoFLO treatment significantly increases tumour response and survival compared with conventional delivery, independently of other prognostic factors. The

most likely explanation for such gender dependency of optimal delivery schedule is the occurrence of excessive toxicity in the female patients causing circadian disruption, and thus the impairment of chronotherapeutic mechanisms. This hypothesis is supported by the occurrence of 20–50% more toxicities in the women patients (Giacchetti et al. 2006; Lévi et al. 2007). Experimental and clinical studies have shown that cancer treatments can disrupt circadian biomarkers in experimental models and in cancer patients, and that this disruption is associated with systemic toxicities (Lévi et al. 2010).

15.3.1.2 Circadian Biomarkers and Clock Genes

Individual cancer patients are characterized by circadian alteration or disruption of plasma cortisol and melatonin, blood cell count, liver enzymes, or renal tests. As a result, healthy tissues in these patients display disrupted circadian coordination of cell division cycle. Minimally invasive techniques, such as rest-activity monitoring or iterative salivary cortisol determinations, show that nearly a third of patients with metastatic cancer display poor circadian biomarker rhythms at baseline (Lévi et al. 2010; Innominato et al. 2010). Moreover, circadian disruption, which hinders both chronotherapeutic mechanisms and host control of malignant processes, appears as an independent prognostic factor of survival (Innominato et al. 2009, 2010).

Human clock genes are highly polymorphic, as documented by large population-based studies. Constitutive polymorphisms of the clock gene *Npas2*, a *Clock* homolog that predominates in specific tissues, are associated with a decreased risk of non-Hodgkin's lymphoma and breast and prostate cancers. Conversely, *Cry2* polymorphisms are associated with an increased risk of non-Hodgkin's lymphoma and prostate cancer (Innominato et al. 2010). Polymorphisms in tissue-specific clock-controlled genes can also account for inter-individual differences in relevant circadian rhythms. In human tumours, the mRNA or protein expression of the clock genes *Per1*, *Per2*, or *Per3* as well as *Npas2* or *Dec1* is markedly decreased on average, or deregulated, in comparison with reference tissues. This is the case for cancers of the breast, lung, colon, endometrium, ovary, pancreas, and bone marrow, a finding supporting the frequent presence of circadian disruption in human malignancies (Gery et al. 2006; Innominato et al. 2010). The altered expression of clock genes in human tumours can influence the efficacy of cancer chronotherapeutics, possibly through the deregulation of cell cycle, and/or drug pharmacodynamics in cancer cells (Innominato et al. 2010).

15.3.1.3 Experimental Models for the Personalization of Cancer Chronotherapeutics

The pharmacological effects of anticancer drugs differ widely according to cell lines and the species, strain, or sex of experimental animal models. The heterogeneity of cancer cells and cancer tissues is an additional cause of variability of anticancer

drug effects. For instance, strain-specific differences in glucuronidation reactions and irinotecan detoxification explain strain- and sex-dependent toxicity and efficacy of this drug. The overall tolerability of therubicin and irinotecan is about threefold better in female than in male B6D2F1 mice. Moreover, the optimal circadian timing for tolerability occurs nearly 4 h earlier in males as compared to females for both drugs (Lévi et al. 2010). No consistent relationship is found between circadian pharmacokinetics and circadian toxicity pattern. Ongoing studies are mapping and modelling the underlying molecular and physiological circadian determinants that account for sex and genotype specificities in chronotolerance patterns and properties. Such models could subsequently aid in guiding the clinical development of personalized cancer chronotherapeutics.

15.3.2 Insights from a Modelling Approach

The issue of individual variability and its impact on cancer treatment, is still some distance from resolution through a modelling approach. However, some results reported above point in this direction. We have seen, for example, that cytotoxicity of 5-FU is markedly affected by the degree of variability V of duration of the cell cycle phases, by entrainability to the circadian clock, and by cell cycle length (see Fig. 15.6). All these factors may vary between tissues and also between individuals, thereby affecting the efficacy of anticancer drugs.

15.3.2.1 Variability of Cell Cycle Phase Durations

For the circadian pattern peaking at 4 a.m., which is least cytotoxic when cell cycle duration is of 22 h, the data indicates that cytotoxicity rises with an increase in the degree of variability (Altinok et al. 2007a, b). When variability increases, the cells in fact desynchronize more rapidly, so that at any moment the fraction of cells in S phase, and thus sensitive to 5-FU, is larger than in the case where cells are better synchronized, at relatively smaller values of variability (Lévi et al. 2008).

15.3.2.2 Cell Cycle Length

The marked dependence of cytotoxicity on the circadian pattern of 5-FU delivery shown in Fig. 15.4 was observed when the cell cycle length was 22 h. This dependence altered with the duration of the cell cycle. Indeed numerical simulations of the CCA model at different values of the cell cycle length show that a minimum in 5-FU cytotoxicity only occurs when the cell cycle duration ranges from 18 to 26 h. Simulations by the model indicate that the minimum progressively shifts from 12 p.m. to 1 a.m., 3 a.m., 5 a.m., and 12 a.m. when the duration of the cell cycle increases from 18 to 26 h. The depth of the trough, which corresponds to reduced

cytotoxicity, is most significant when the duration of the cell cycle ranges from 20 to 24 h. When the cell cycle length is 16 h, no minimum 5-FU toxicity is apparent, while for a cell cycle duration of 26 h, the minimum as a function of peak time in circadian delivery becomes very shallow (for an explanation of these results, see Altinok et al. 2009).

15.3.2.3 Cell Cycle Entrainment by the Circadian Clock

Yet another factor that influences the cytotoxicity of 5-FU is entrainment by the circadian clock. The CCA model shows, in fact, that cytotoxicity is markedly different in the presence or absence of circadian entrainment. This is because entrainment by the circadian clock results in cells becoming more synchronized than in the absence of entrainment. Hence, the effect of administration time of the anticancer drug is enhanced in conditions of entrainment.

15.3.2.4 A detailed Computational Model for the Mammalian Cell Cycle

The automaton model for the cell cycle provides a useful tool for assessing the cytotoxic effect of various temporal patterns of anticancer drug delivery. Although the model is relatively simple and does not explicitly take into account the detailed molecular machinery controlling cell proliferation, it nevertheless shows how a population of cells can progressively desynchronize due to the stochastic nature of the transitions between the successive phases of the cell cycle, and the variability that characterizes their duration. The results obtained by numerical simulations of the cell cycle automaton indicate that the least cytotoxic patterns of 5-FU and l-OHP circadian administration match those used clinically. The model therefore confirms the utility of those patterns that were initially selected on the basis of experimental studies in animals, and subsequently tested in humans. It further shows that continuous administration of 5-FU and l-OHP has the same effect as the most cytotoxic circadian pattern of drug delivery. Additionally, the model helps us identify factors that may contribute to explain a long-standing puzzle, namely, why temporal patterns corresponding to minimum cytotoxicity for a population of healthy cells could at the same time prove more cytotoxic toward a population of tumour cells.

At the opposite end of the modelling spectrum, a detailed model for the cyclin/Cdk network driving the mammalian cell cycle is now available (Gérard and Goldbeter 2009). The model incorporates the main cyclin-dependent kinases (Cdk) that are important for an ordered progression along the phases G1, S, G2 and M of the cell cycle (Morgan 2006). The activity of the cyclin/Cdk complexes can be regulated by phosphorylation/dephosphorylation, association with Cdk protein inhibitors, and modulation of protein synthesis or degradation. The model shows how the presence of sufficient amounts of growth factor triggers the occurrence of self-sustained oscillations in the Cdk network. These oscillations correspond to the successive transient activation along the cell cycle phase of the complexes cyclin D/Cdk4-6 in

phase G1; cyclin E/Cdk2 at the G1/S transition; cyclin A/Cdk2 in phase S and at the S/G2 transition; and cyclin B/Cdk1 at the G2/M transition. The following properties of the mammalian cell cycle are accounted for by this modelling approach: (1) repetitive cell cycling in presence of supra-threshold amounts of growth factor; (2) control of cell cycle progression by the balance between the tumour suppressor pRB and the transcription factor E2F; (3) existence of a restriction point located in G1, beyond which completion of the cell cycle becomes independent of growth factor. By incorporating the DNA replication checkpoint mediated by the kinases ATR and Chk1, this approach further shows how checkpoints can modulate the oscillatory dynamics of the cell cycle.

The detailed computational model for the mammalian cell cycle thus provides an integrated framework showing how the regulatory interaction between four Cdk modules results in their repetitive sequential activation, which brings about an orderly progression along cell cycle phases. It also shows how the cell can switch from a non-oscillatory stable steady state (corresponding to quiescence), to a self-sustained oscillatory regime (corresponding to cell proliferation). The model has direct relevance to cancer. It shows how the over-expression of an oncogene product, such as the phosphates Cdc25, can trigger the transition to cell proliferation even in the absence of growth factor (Gérard and Goldbeter 2009).

The cell cycle and circadian clocks are coupled through the effect of the circadian complex CLOCK-BMAL1 (see Sect. 15.1.1.2), which induces the expression of the *Wee1* gene. The gene codes for a kinase that inhibits Cdk1. Other components of the cell cycle machinery, e.g. the expression of cyclins, are induced by the circadian clock. By coupling detailed models for the cell cycle to that of the circadian clock via CLOCK-BMAL1, circadian entrainment of the cell cycle can readily be observed (Gérard and Goldbeter 2009). The model for the Cdk network should prove useful not only for probing entrainment by the circadian clock, but also for searching for optimal patterns of circadian administration of anticancer drugs.

15.4 Conclusions and Perspectives

A large body of knowledge has been gathered during the past decade regarding the molecular circadian clock and the clock-controlled pathways, several of which are involved in cancer processes and/or anticancer treatment activities. Detailed mathematical models of the circadian clock enable *in silico* testing of clock properties and downstream dynamic disturbances on cellular proliferation and metabolism. Both experimental and clinical data demonstrates that circadian disruption favours cancer processes, while circadian reinforcement or induction can impair cancer growth. The clinical data also document the relevance of circadian timing for designing safer and more effective anticancer treatment modalities. Mathematical models integrate circadian clock control of cell cycle and pharmacology processes, and allow for *in silico* assessment of drug effects according to circadian scheduling. Such a theoretical approach proposes mechanisms that explain why chronotoler-

ance and chronoefficacy of anticancer drugs usually coincide. However, optimal circadian delivery profiles of anticancer drugs require further fine-tuning, so as to personalize cancer chronotherapeutics according to the characteristics of individual patients or patient subgroups. Mathematical models reveal that inter-patient differences in response to cancer chronotherapeutics could stem from differences in circadian entrainment, a finding verified experimentally and clinically, and/or in cell cycle length. A systems biology approach to the interactions between circadian clocks, cell cycle and pharmacological pathways, from cell to tissue to whole organism, may prove indispensible for a proper adjustment of cancer chronotherapeutics according to the status of the circadian timing system of the patient, tumour dynamics, and the choice of the therapeutic strategy.

Acknowledgments This work was supported by the European Commission (EC) through the Network of Excellence BioSim (contract No. LSHB-CT-2004-005137). Work by FL was further supported by the EC through the Scientific Targeted Research Project TEMPO (contract LSHG-ct-2006-037543), by the Association for Research on Cancer (ARC, Villejuif, France), and by the Association pour la Recherche sur le Temps Biologique et la Chronothérapie (ARTBC International, hospital Paul Brousse, Villejuif, France). Work by AG was further supported by grant n° 3.4607.99 from the *Fonds de la Recherche Scientifique Médicale* (F.R.S.M., Belgium), and by the Belgian Federal Science Policy Office (IAP P6/25 "BioMaGNet": "Bioinformatics and Modelling—From Genomes to Networks").

References

Adam R, Wicherts DA, Haas RJ de, Ciacio O, Lévi F, Paule B, Ducreux M, Azoulay D, Bismuth H, Castaing D (2009) Patients with initially unresectable colorectal liver metastases: is there a possibility of cure? J Clin Oncol 27:1829–1835

Altinok A, Lévi F, Goldbeter A (2007a) A cell cycle automaton model for probing circadian patterns of anticancer drug delivery. Adv Drug Deliv Rev 59:1036–1053

Altinok A, Lévi F, Goldbeter A (2007b) Optimizing temporal patterns of anticancer drug delivery by simulations of a cell cycle automaton. In: Bertau M, Mosekilde E, Westerhoff HV (eds) Biosimulation in drug development. Wiley-VCH, Weinheim, pp 275–297

Altinok A, Lévi F, Goldbeter A (2009) Identifying mechanisms of chronotolerance and chronoefficacy for the anticancer drugs 5-fluorouracil and oxaliplatin by computational modelling. Eur J Pharmaceut Sci 36:20–38

Bjarnason GA, Jordan R (2000) Circadian variation of cell proliferation and cell cycle protein expression in man: clinical implications. Prog Cell Cycle Res 4:193–206

Clairambault J (2007) Modelling oxaliplatin drug delivery to circadian rhythms in drug metabolism and host tolerance. Adv Drug Deliv Rev 59:1054–1068

Coudert B, Focan C, Genet D, Giacchetti S, Cvickovic F et al (2008) A randomized multicenter study of optimal circadian time of vinorelbine combined with chronomodulated 5-fluorouracil in pretreated metastatic breast cancer patients: EORTC trial 05971. Chronobiol Int 25:680–696

Csikasz-Nagy A, Battogtokh D, Chen KC, Novak B, Tyson JJ (2006) Analysis of a generic model of eukaryotic cell-cycle regulation. Biophys J 90:4361–4379

Dibner C, Schibler U, Albrecht U (2010) The mammalian circadian timing system: organization and coordination of central and peripheral clocks. Annu Rev Physiol 72:517–549

Filipski E, Innominato PF, Wu MW et al (2005) Effects of light and food schedules on liver and tumour molecular clocks in mice. J Natl Cancer Inst 97:507–517

Forger DB, Peskin CS (2003) A detailed predictive model of the mammalian circadian clock. Proc Natl Acad Sci U S A 100:14806–14811

Fu L, Lee CC (2003) The circadian clock: pacemaker and tumour suppressor. Nat Rev Cancer 3:350–361

Gachon F, Olela FF, Schaad O, Descombes P, Schibler U (2006) The circadian PAR-domain basic leucine zipper transcription factors DBP, TEF, and HLF modulate basal and inducible xenobiotic detoxification. Cell Metab 4:25–436

Gallego M, Eide EJ, Woolf MF, Virshup DM, Forger DB (2006) An opposite role for tau in circadian rhythms revealed by mathematical modelling. Proc Natl Acad Sci U S A 103:10618–10623

Gérard C, Goldbeter A (2009) Temporal self-organization of the cyclin/Cdk network driving the mammalian cell cycle. Proc Natl Acad Sci U S A 106:21643–21648

Gery S, Komatsu N, Baldjyan L, Yu A, Koo D, Koeffler HP (2006) The circadian gene per1 plays an important role in cell growth and DNA damage control in human cancer cells. Mol Cell 22:375–382

Giacchetti S, Itzhaki M, Gruia G, Adam R, Zidani R, Kunstlinger F, Brienza S, Alafaci E, Bertheault-Cvitkovic F, Jasmin C, Reynes M, Bismuth H, Misset JL, Lévi F (1999) Long-term survival of patients with unresectable colorectal cancer liver metastases following infusional chemotherapy with 5-fluorouracil, leucovorin, oxaliplatin and surgery. Ann Oncol 10:663–669

Giacchetti S, Bjarnason G, Garufi C, Genet D, Iacobelli S et al (2006) Phase III trial comparing 4-day chronomodulated therapy versus 2-day conventional delivery of fluorouracil, leucovorin, and oxaliplatin as first-line chemotherapy of metastatic colorectal cancer: the European Organisation for Research and Treatment of Cancer Chronotherapy Group. J Clin Oncol 24:3562–3569

Goldbeter A, Claude D (2002) Time-patterned drug administration: insights from a modelling approach. Chronobiol Int 19:157–175

Gorbacheva VY, Kondratov RV, Zhang R, Cherukuri S, Gudkov AV, Takahashi JS, Antoch MP (2005) Circadian sensitivity to the chemotherapeutic agent cyclophosphamide depends on the functional status of the CLOCK/BMAL1 transactivation complex. Proc Natl Acad Sci U S A 102:3407–3412

Granda TG, Liu XH, Smaaland R, Cermakian N, Filipski E, Sassone-Corsi P, Lévi F (2005) Circadian regulation of cell cycle and apoptosis proteins in mouse bone marrow and tumour. FASEB J 19:304–306

Gréchez-Cassiau A, Rayet B, Guillaumond F, Teboul M, Delaunay F (2008) The circadian clock component BMAL1 is a critical regulator of p21WAF1/CIP1 expression and hepatocyte proliferation. J Biol Chem 283:4535–4542

Innominato PF, Focan C, Gorlia T et al (2009) Circadian rhythm in rest and activity: a biological correlate of quality of life and a predictor of survival in patients with metastatic colorectal cancer. Cancer Res 69:4700–4707

Innominato PF, Lévi F, Bjarnason GA (2010) Chronotherapy and the molecular clock: clinical implications in oncology. Adv Drug Delivery Rev 62:979–1001

Iurisci I, Filipski E, Reinhardt J, Bach S, Gianella-Borradori A et al (2006) Improved tumour control through circadian clock induction by Seliciclib, a cyclin-dependent kinase inhibitor. Cancer Res 66:10720–10728

Iurisci I, Filipski E, Sallam H, Harper F, Guettier C et al (2009) Liver circadian clock, a pharmacological target of cycline-dependent kinase inhibitor seliciclib. Chronobiol Int 26:1169–1188

Leloup JC, Goldbeter A (2003) Toward a detailed computational model for the mammalian circadian clock. Proc Natl Acad Sci U S A 100:7051–7056

Leloup JC, Goldbeter A (2004) Modelling the mammalian circadian clock: sensitivity analysis and multiplicity of oscillatory mechanisms. J Theor Biol 230:541–562

Leloup JC, Goldbeter A (2008) Modelling the circadian clock: from molecular mechanism to physiological disorders. Bioessays 30:590–600

Leloup JC, Goldbeter A (2010) Modelling the dual role of PER phosphorylation and its effect on the period and phase of the mammalian circadian clock. IET Syst Biol 5:44

Lévi F, Zidani R, Vannetzel JM, Perpoint B, Focan C, Faggiuolo R, Chollet P, Garufi C, Itzhaki M, Dogliotti L (1994) Chronomodulated versus fixed-infusion-rate delivery of ambulatory chemotherapy with oxaliplatin, fluorouracil, and folinic acid (leucovorin) in patients with colorectal cancer metastases: a randomized multi-institutional trial. J Natl Cancer Inst 86:1608–1617

Lévi F, Zidani R, Misset JL (1997) Randomised multicentre trial of chronotherapy with oxaliplatin, fluorouracil, and folinic acid in metastatic colorectal cancer. International Organization for Cancer Chronotherapy. Lancet 350:681–686

Lévi F, Focan C, Karaboué A, la Valette V de, Focan-Henrard D, Baron B, Kreutz M, Giacchetti S (2007) Implications of circadian clocks for the rhythmic delivery of cancer therapeutics. Adv Drug Deliv Rev 59:1015–1035

Lévi F, Schibler U (2007) Circadian rhythms: mechanisms and therapeutic implications. Annu Rev Pharmacol Toxicol 47:593–628

Lévi F, Altinok A, Clairambault J, Goldbeter A (2008) Implications of circadian clocks for the rhythmic delivery of cancer therapeutics. Philos Transact A Math Phys Eng Sci 366:3575–3598

Lévi F, Okyar A, Dulong S, Innominato PF, Clairambault J (2010) Circadian timing of cancer treatments. Annu Rev Pharm Toxicol 50:377–421

Li XM, Delaunay F, Dulong S, Claustrat B, Zampera S, Fujii Y, Teboul M, Beau J, Lévi F (2010) Cancer inhibition through circadian reprogramming of tumour transcriptome with meal timing. Cancer Res 70:3351–3360

Matsuo T, Yamaguchi S, Mitsui S, Emi A, Shimoda F, Okamura H (2003) Control mechanism of the circadian clock for timing of cell division in vivo. Science 302:255–259

Mirsky HP, Liu AC, Welsh DK, Kay SA, Doyle FJ 3rd (2009) A model of the cell-autonomous mammalian circadian clock. Proc Natl Acad Sci U S A 106:11107–11112

Morgan DO (2006) The cell cycle: principles of control. Oxford Univ Press, Oxford

Novak B, Tyson JJ (2004) A model for restriction point control of the mammalian cell cycle. J Theor Biol 230:563–579

Ohdo S, Koyanagi S, Suyama H, Higuchi S, Aramaki H (2001) Changing the dosing schedule minimizes the disruptive effects of interferon on clock function. Nat Med 7:356–360

Oklejewicz M, Destici E, Tamanini F, Hut RA, Janssens R, Van Der Horst GT (2008) Phase resetting of the mammalian circadian clock by DNA damage. Curr Biol 18:286–291

Okyar A, Lévi F (2008) Circadian control of cell cycle pathways: relevance of cancer chronotherapeutics. In: Yoshida K (ed) Trends in cell cycle research. Research Signpost, Kerala, pp 293–317

Qu Z, Weiss JN, MacLellan WR (2003) Regulation of the mammalian cell cycle: a model of the G1-to-S transition. Am J Physiol Cell Physiol 284:349–364

Swat M, Kel A, Herzel H (2004) Bifurcation analysis of the regulatory modules of the mammalian G1/S transition. Bioinformatics 20:1506–1511

Takahashi JS, Hong HK, Ko CH, McDearmon EL (2008) The genetics of mammalian circadian order and disorder: implications for physiology and disease. Nat Rev Genet 9:764–775

Toh KL, Jones CR, He Y, Eide EJ, Hinz WA, Virshup DM, Ptácek LJ, Fu YH (2001) An hPer2 phosphorylation site mutation in familial advanced sleep phase syndrome. Science 291:1040–1043

Vanselow K, Vanselow JT, Westermark PO, Reischl S, Maier B, Korte T, Herrmann A, Herzel H, Schlosser A, Kramer A (2006) Differential effects of PER2 phosphorylation: molecular basis for the human familial advanced sleep phase syndrome (FASPS). Genes Dev 20:2660–2672

Xu Y, Toh KL, Jones CR, Shin JY, Fu YH, Ptácek LJ (2007) Modelling of a human circadian mutation yields insights into clock regulation by PER2. Cell 128:59–70

Chapter 16
Clinical Applications of Systems Biology Approaches

Sergio Iadevaia, Adel B. Tabchy, Prahlad T. Ram and Gordon B. Mills

Abstract Completion of the Human Genome Project, rapid progress in The Cancer Genome Atlas (TCGA) and genome-wide association studies, and the ensuing development of high-throughput technologies for analysing patient samples at the DNA, RNA, protein, and metabolic levels, have resulted in a rapid accumulation of data capable of providing an atlas or 'tool kit' that describes the events that may occur during tumour initiation and progression. This body of data is challenging because of the incredible heterogeneity among types of tumours, and the number of genetic aberrations found in epithelial cancer cells. The complexity of these aberrations makes identification of the emerging properties of tumours overwhelmingly difficult using current technologies. Thus, converting the abundant data into useful information about tumour initiation and progression, and, more importantly, obtaining the knowledge required to improve patient outcomes, will require new integrated approaches. Systems approaches that integrate data collected using multiple platforms and modalities with mathematical models of functional and phenotypic tumour outcomes will be necessary for the discovery of principles that can accelerate progress in translating preclinical studies into improved patient outcomes. This chapter focuses on the current and future clinical applications of systems biology approaches related to cancer, which include identification of diagnostic, prognostic, and therapeutic biomarkers, selection of suitable intervention strategies tailored to individual patients using combinatorial targeted therapy, and design of clinical trials.

16.1 Chapter Introduction

The daily life of human beings often presents us with surprises and unexpected events. However, daily life may also (and more usually) appear repetitive. We see the same people, drive to work on the same streets, eat at the same restaurants, and perform the same activities at home, integrating what we most like to do into recursive patterns that simplify our lives. Although the mechanism of this repetition may be perceived as limiting and boring, it is one of the fundamentals of life. Every

S. Iadevaia (✉)
Departments of Systems Biology, The University of Texas M. D. Anderson Cancer Center, 7435 Fannin St, Unit 950, P.O. Box 301429, Houston, TX, USA
e-mail: siadevai@mdanderson.org

A. Cesario, F. B. Marcus (eds.), *Cancer Systems Biology, Bioinformatics and Medicine,*
DOI 10.1007/978-94-007-1567-7_16,

day the human body repeats with astonishing accuracy a large number of biological processes to assure its proper function as an integrated system. These biological events are tightly regulated by robust, redundant (back-up) regulatory machineries that have evolved for the optimal integration, synchronization, and coordination of functions produced by different molecules, cells, and organs experiencing complex environmental cues at different times. Moreover, the regulation of cellular physiology is complemented by quality control systems that can detect molecular abnormalities and correct errors. When this corrective process fails, control systems can eliminate aberrant cells to minimize malfunctions at the molecular, cellular, organ, and organism levels. In a multicellular organism, the preferred default decision process of a 'damaged' cell is to enter a death pathway or irreversible senescence. Indeed, the loss of an aberrant cell has no consequence on the survival of a multicellular organism, whereas persistence of aberrant cells can lead to diseases, and cancer in particular.

Despite the remarkable reliability of the control systems that have co-evolved with the organism, regulation of biological functions in the human body may fail and result in a myriad of disease states. Given the number and complexity of the integrated processes required for survival of an organism, failure of functional regulation is not surprising. The control systems appear to be resistant to single stresses or aberrations but may be susceptible to coordinated abnormalities targeting multiple molecules or functions. When several molecular aberrations accumulate in a single cell, quality control systems may no longer be able to adequately correct genetic aberrations, or to trigger the decision of the cell to enter a death pathway. Thus, concurrent accumulation of several aberrations can result in acquisition of a previously non-existent functional ability to evade quality control, to survive cell-fate decisions that should lead to death, and to redirect the regulatory control systems into sustaining deregulated cellular functions. These events can lead to the initiation of cancer.

Cancer is a multistage heterogeneous disease in which groups of cells evade limits on division (growth), avoid death (apoptosis), and acquire the ability to invade adjacent tissues (invasion), and to migrate to other organs via lymph nodes and the bloodstream (metastasis). Staging of cancer (see Chap. 2) depends on the size of the primary tumour, the status of regional lymph-nodes, and the presence of metastases, and describes the extent to which the disease has spread in the body at the time of diagnosis (NCI 2009b). Although most tumours appear to be clonal in origin, the degree of intratumoral heterogeneity that can arise from acquisition of additional mutations after the initiation of the original clone varies widely (Bertucci and Birnbaum 2008). Differentiation of the intrinsic gene expression patterns of the tumour stem-cell lineage, the environmental context of the cells, and underlying germline susceptibility or resistant loci in the host, may contribute to determining the types of aberrations that can transform a particular cell or, alternatively, the coordinated set of aberrations that can be tolerated (i.e., that do not lead to death or senescence in cells of a particular lineage). Cancers that form in the same organ can thus consist of tumour subclasses which display differences in cellular proliferation, the ability to invade and produce metastases, drug sensitivity, and importantly, patient outcomes.

Accumulation of genetic aberrations results in the heritable changes that drive tumour initiation and progression. These abnormalities are either acquired through genetic (NIH 2009a) and epigenetic (Esteller 2008) mutations, hormone imbalance (Zanetta et al. 2000) and immune conditions (Mellemkjaer et al. 2002); or they are the result of toxic environmental insults to cellular DNA that occur in everyday life, such as inflammation (Mellemkjaer et al. 2002), radiation exposure (English et al. 1997), smoking (Kuper et al. 2002), and hazardous chemical exposure (O'Reilly et al. 2007).

Standard approaches to cancer treatment include local therapy (surgery and radiation therapy) and systemic therapy (hormone therapy, chemotherapy, and targeted therapy), either alone or, more often, in combination (ACS 2009). One of the precepts of combined therapy is that the heterogeneity of most epithelial tumours protects them against the effects of monotherapy, requiring a coordinated inhibition of several critical processes. Systemic therapy is administered to patients having no detectable cancer after surgery (adjuvant), to eliminate both occult metastases and any cells that exit primary tumours and spread via the bloodstream or lymphatics during surgery. Systemic therapy also may be given to patients before surgery (neoadjuvant) to shrink tumours and facilitate resection (see Chap. 2). Additionally, neoadjuvant therapy can provide important information on tumour sensitivity or resistance to therapeutic agents. Extended analyses can be done on the bioptic material obtained for histological diagnosis before the administration of the neoadjuvant therapy, thereby providing biological information about the tumour at the very beginning of the treatment path, so that therapy can be tailored. Hormonal therapy aims to block the effect of hormones on stimulating the growth and survival of cells derived from hormonally responsive tissue lineages. Unfortunately, because cytotoxic drugs target all rapidly dividing cells, chemotherapy is not specific to cancer cells and can harm healthy cells that have high replacement rates. Whereas chemotherapy affects both normal and cancer cells and thus can have high morbidity rates, targeted therapy interferes with specific molecular targets responsible for tumour formation and progression. The so-called 'oncogene addiction' concept, that tumour cells with behaviour driven by particular genomic aberrations may be more dependent on aberrant pathways than normal cells may account for the remarkable efficacy of some targeted therapies, and their attendant low collateral toxicity levels.

Ironically, cellular aberrations that lead to cancer initiation and progression frequently target the very regulatory systems that enable organisms to develop and undergo normal cellular functions. For effective diagnosis, treatment, and prevention of cancer, we need to be able to identify the molecular abnormalities that are most likely to be the drivers of carcinogenesis. Targeting these aberrations with effective therapeutic agents will aid in regaining control of the deregulated system, and in many cases reinstitute the decision of cells to undergo apoptosis or other forms of death. The launch of the Human Genome Project in 1990 (NIH 2009b) paved the way for a new era of personalized medicine, in which patients could receive individualized treatment with minimally toxic, highly selective drugs capable of compensating for cellular deregulations driven by specific genetic aberrations. Following the one-gene/one-drug approach pioneered by Ehrlich (Kaufmann 2008),

investigators have used several small molecules for targeted therapy (see Chap. 2): imatinib (Gleevec; Lassila et al. 2004); gefitinib (Iressa; Pao et al. 2004); erlotinib (Tarceva; Katzel et al. 2009); bortezomib (Velcade; Adams and Kauffman 2004); and tamoxifen (Jordan 2006) and the monoclonal antibodies rituximab (Maloney et al. 1997); trastuzumab (Herceptin) (Hudis 2007); cetuximab (Van Cutsem et al. 2009); and bevacizumab (Rini 2007). Several of these have demonstrated remarkable efficacy, indicating the potential use of the one-gene/one-drug approach. However, this approach is often effective in only very limited patient populations.

A number of challenges being encountered in designing effective targeted therapy for treating cancer in individual patients will be discussed in the following paragraphs. The first challenge is to distinguish the genetic aberrations that drive tumour behaviour and constitute optimal therapeutic targets or biomarkers, from those aberrations that develop because of genomic instability or that contribute to tumour initiation, but which do not drive tumour behaviour and thus do not predict outcomes. Moreover, genetic aberrations of transformed cells may not necessarily correlate with abnormal production of proteins and metabolites, so that biomarkers identified as drivers of tumorigenesis may not accurately predict patient outcome.

The second challenge is to determine the mechanisms that limit the efficacy of targeted therapies. Regulation of homeostasis (Olovnikov et al. 2009; Semenza 2009), redundancy (Tortora et al. 2004), bypass mechanisms (van Amerongen and Berns 2008; Vasudevan et al. 2009; Wetherill et al. 2006), and feedback compensation (Mirzoeva et al. 2009; Tremblay and Marette 2001) may result in the interaction networks that control cellular functions becoming self-regulatory, and thus insensitive to targeted therapy (see Chap. 17). Therapeutic interventions themselves may possibly favour the development of resistance mechanisms arising out of adaptation of tumour cells to the effects of targeted therapy (Dexter and Leith 1986; Geisler et al. 2002).

The third challenge is to identify biomarkers that specify patient populations likely to benefit from targeted therapies with minimal ‘off-side’ toxicity.

The pilot phase of The Cancer Genome Atlas (TCGA) project is now complete and entering the full project phase, the aim of which is to provide an atlas of the genomic, mutational, methylation, and transcriptional changes occurring in 500 tumours, each of 23 different tumour lineages. The initial glioma data has provided particularly exciting new conceptual insights into the networks and pathways that contribute to development of this tumour. The TCGA project, and other related programmes such as the International Cancer Genomics Consortium (ICGC), will provide an atlas of the aberrations that drive tumour behaviour in the transformed cells. The most convincing piece of early evidence from TCGA studies is that heterogeneity among the different types of cancers, and the diversity of the identified aberrations, demand novel approaches to understanding the complexity of tumour initiation and progression. Systems biology approaches integrating high-throughput genomic, transcriptomic, proteomic, and metabolic data with predictive mathematical models will facilitate the identification of those genes, proteins, and metabolites that are activated or inactivated in specific disease processes, in order to rationally identify prognostic and predictive molecular targets detected in individual patients

for personalized targeted therapy. This chapter focuses on the current and future clinical applications of systems biology approaches related to cancer, which include identification of diagnostic, prognostic, and therapeutic biomarkers; selection of intervention strategies tailored to individual patients, using combinatorial targeted therapy; and design of clinical trials.

16.2 Systems Biology Approaches to Identifying Diagnostic, Prognostic, and Therapeutic Biomarkers for Cancer

16.2.1 Genomic, Transcriptomic, Proteomic, and Metabolic (Omics) Analysis of Human Tumours

Identifying the molecular abnormalities that underlie cancer is crucial for early diagnosis, prognosis, and determination of optimal therapeutic approaches. Researchers have extensively employed genomic and transcriptional profiling of patient biopsy samples to identify alterations of gene-copy numbers and gene (mRNA)-expression levels that could direct the search for biomarkers for all tumour types (Apolo et al. 2009; Clark-Langone et al. 2007; Dhanasekaran et al. 2001; Haass and Smalley 2009; Harry et al. 2009; Kosari et al. 2008; Phillips et al. 2006; Ross 2009; Scott and Salgia 2008; Walther et al. 2009) (See also Chap. 3 and 13). The recently developed next-generation sequencing platform and the so-called 'thousand dollar genome' efforts have the potential to rapidly supplant current approaches to genomic and transcriptional profiling. Identifying genetic aberrations in cancer cells is equally fundamental to guiding clinical decisions about treatment. For example, the Onco*type* DX test is used to analyse the expression profiles of 21 genes in patients with node-negative breast tumours (Paik et al. 2004). The result of this test is reported as the recurrence score, which is employed to correlate the gene activity with the likelihood of distant recurrence of cancer, and to guide clinical decision-making regarding whether chemotherapy is warranted after surgery (Cronin et al. 2007). The recurrence score is designed to integrate a series of functions important to tumour behaviour (e.g. proliferation and survival function), rather than to quantify the levels of expression of specific transcripts. Thus, the Onco*type* DX test integrates crucial information about breast tumours in a systems biology approach to predicting the natural history of breast cancer, which is unobtainable using standard clinico-pathological procedures such as staging and histological grading (see Chap. 2).

Genomic analysis identifies molecular biomarkers that may be correlated with survival (tumour suppressor genes) or drivers of pathogenesis, invasion, and metastasis (oncogenes). However, the functional changes at the proteomic level cannot be predicted from analyses at the DNA or mRNA level. An integrative approach, combining data obtained from DNA sequencing with mRNA and protein data obtained from patient tumours, will be necessary for optimal understanding of the content of the tumours (and their microenvironments), in order to increase the ac-

curacy of identification of prognostic biomarkers and predictive targets. Tian and colleagues (2004) used DNA microarrays and quantitative proteomic analysis to determine the correlation between the expression levels of mRNAs, and proteins in the ELM and MPRO cell lines and livers of mice. The analysis revealed that differential expression of the mRNAs could capture up to 40% of the variation in protein expression, and emphasized the importance of post-transcriptional modifications in mammalian cells, which cannot currently be predicted by DNA and RNA analysis alone. In 2005, Varambally and co-workers combined transcriptomic analysis with high-throughput immunoblotting to determine genetic and proteomic abnormalities in tissues specimens of benign prostates, clinically localized, and metastatic prostate cancer from different patients (Varambally et al. 2005). Using this approach, the authors identified 64 proteins that were deregulated in clinically localized prostate tumours compared to benign prostate tumours, and 156 proteins that were altered in metastatic prostate tissues compared to clinically localized prostate tumours. They also observed 48–65% concordance between genomic and proteomic aberrations in the different tumours. Thus, differential proteomic alteration between clinically localized and metastatic prostate cancers that mapped concordantly with genetic abnormalities served as better predictors of patient outcome, as compared with predictors obtained using only DNA and RNA analyses. In 2008, Stemke-Hale and colleagues analysed the subtype specificity and signalling effects of *PIK3CA*, *AKT*, and *PTEN* mutations in 547 human breast tumours and 41 cell lines, using mass spectroscopy-based sequencing in combination with reverse-phase protein arrays (Stemke-Hale et al. 2008). This integrative analysis revealed that aberrations in the phosphoinositide 3-kinase (PI3K) pathway are likely to play distinct roles in the pathogenesis of different breast cancer subtypes, which may have implications for the selection of PI3K-targeted therapies for hormone receptor-positive breast cancer. Further, the effects of the aberrations on cell function and protein levels were different from those predicted on the basis of prior findings, demonstrating the complexity of human tumours and the need for approaches to integrate information on patient tumours, including the tumour itself, its stroma, and its microenvironment.

Although investigators have extensively profiled gene and protein expression in human tumours, little is known about the metabolic aberrations that characterize cancer development and progression. Sreekumar et al. (2009) used high-throughput gas-liquid chromatography-based mass spectrometry to profile more than 1126 metabolites in 262 clinical samples of benign prostates, clinically localized prostate tumours, and metastatic prostate tumours, along with samples of urine and plasma obtained from the same patients. They found that sarcosine levels were increased during prostate tumour progression to metastasis, and higher in invasive prostate cancer cell lines than in benign prostate epithelial cells. Therefore, sarcosine was identified as a potentially important metabolic intermediary for cancer cell invasion and aggressiveness, a finding not predicted from studies of DNA or RNA.

In the near future, integrative analysis of genomic, transcriptomic, proteomic, and metabolic (omics) profiles will generate more comprehensive data sets that could reveal the molecular mechanisms of cancer progression, thereby guiding the identification of biomarkers for accurate diagnosis of, prognosis for, and treatment

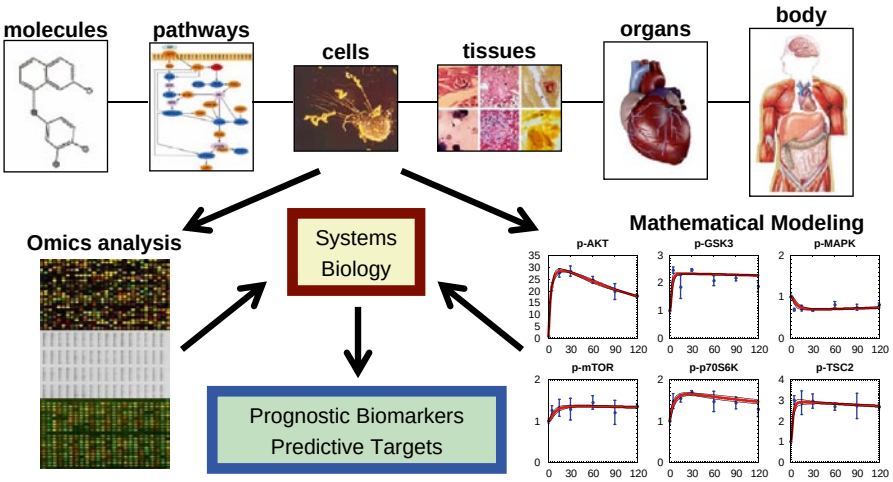


Fig. 16.1 Systems biology approaches to identifying prognostic biomarkers and predictive targets. Omics analysis generates sets of high-throughput data reporting the activity of genes, mRNAs, proteins, and metabolites measured in tumour samples at the levels of molecule, pathways, cells, tissue, organ, and body. Mathematical modelling attempts to derive detailed models that describe the mechanisms of cancer formation and progression, to predict the effect of targeted therapy on tumour samples. Systems biology approaches are focused on integrating omics data with mathematical modelling to form a comprehensive picture of cancer that can be studied as a unified system, so as to identify biomarkers and targets for pharmacological treatment and prevention of the disease

of human tumours in individual patients. However, large sets of experimental data are often noisy and may contain hidden information that cannot be easily extracted via inspection. Hence, translating candidate biomarkers into preclinical and clinical trials requires high-quality selection of hallmarks across multiple tissues, organs, and disease states. For that reason, 'omics' analyses of human tumours are often complemented by the use of computational tools to improve the selection of candidate biomarkers (Fig. 16.1). However, the ability to integrate omics data into robust, visualizable predictive models that can be mined to develop testable hypotheses remains a progress-limiting factor.

16.2.2 Computational Mining of Omics Data

Computer-aided data mining of omics profiles is being introduced and routinely applied to refining the selection of clinically relevant molecular hallmarks in large lists of potential candidate biomarkers of human tumours. Researchers have performed data-mining using expressed sequence tags (Aouacheria et al. 2006; Campagne and Skrabanek 2006); serial analyses of gene expression (Hustinx et al. 2004); massively parallel signature sequencing (Chen et al. 2005); and microarray expression

databases (Shen et al. 2005), to guide the selection of mRNA transcripts that are differentially expressed in diseased and normal tumours. Ma et al. (2007) used a combination of random forests and linear discriminant analysis to investigate the expression profiles of 7650 genes in 99 patients with node-negative or positive breast cancer, and identified new gene signatures that can be used to predict breast cancer recurrence and metastasis, tumour grade, and nodal status. In 2008, Ou and colleagues employed two-dimensional electrophoresis and mass spectrometry to generate proteomic maps of breast cancer (MCF-7 and HCC-38) and control (CCD-1059Sk) cell lines. They mapped the differentially expressed proteins therein onto mRNA databases of cancer cell lines and primary tumours, to identify candidate biomarkers that were concordantly expressed at the gene level (Ou et al. 2008). Using a support vector machine algorithm, they further ranked the concordantly expressed biomarkers on the basis of their individual contributions to normal versus tumour classification accuracy. Of the top nine ranked candidate biomarkers, they successfully validated four, using immunohistochemistry with breast tissue microarrays. In 2004, Weinstein developed a number of bioinformatic software programs (CIMminer, MedMiner, MatchMiner, and GoMiner) to biologically interpret the profiles of the NCI-60 cancer cell line panel, measured using genomic, transcriptomic, and proteomic experimental platforms (Weinstein 2004). Using an integromic approach, this study identified tumour-subtype biomarkers and cancer therapy targets. The developers of TCGA, through the Genome Data Analysis Center programme, have taken on the development and implementation of technologies able to analyse the mass of omics data obtained using high-throughput analysis of tumour samples. The Integrative Cancer Biology Program (NIH 2009c), in turn, will partially support the challenge of integrating this data into understandable systems biology principles.

The design of new bioinformatic software programs and quantitative omics technologies will enable the identification of novel biomarkers with a high probability of being drivers of tumour formation and progression in individual patients. This will improve cancer management, because of more accurate diagnosis, assessment of prognosis, and selection of personalized targeted therapy. However, identifying a therapeutic target is not the same as knowing the effect of inhibiting that target. Moreover, inhibition of single targets in deregulated networks may be ineffective because of redundancy of biological pathways; the presence of self-regulatory loops; the heterogeneity of human tumours; and mechanism-based resistance and toxicity to targeted therapy. Therefore, testing and validating the efficacy of combinatorial targeted therapy in preclinical trials is a crucial prerequisite before designing personalized intervention strategies. However, preclinical data cannot predict the effect of targeted therapy on a patient's tumour without the development of usable predictive models, which in turn will need to be refined using iterative hypothesis development and testing. In this regard, 'omics' profiles of human tumours must be mapped onto omics profiles of animals and cell lines, to develop *in vivo* and *in vitro* models of cancer capable of recapitulating the mechanisms of cancer formation and progression observed in patients. Animal and cell line models thus provide the experimental platforms necessary for validation of those hypotheses

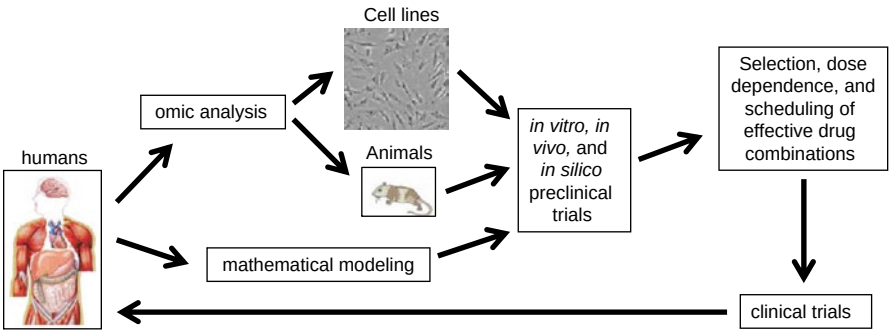


Fig. 16.2 Systems biology approaches to designing preclinical and clinical trials. Mapping omics data of human tumours onto omics data of cell lines and animal models, and formulating mathematical models that recapitulate the mechanisms of cancer formation and progression observed in human tumours, provides the integrated experimental-theoretical platform for preclinical identification of candidate strategies for cancer therapy. The integrative use of *in vitro*, *in vivo* and *in silico* models may aid the selection, dose dependence, and scheduling of effective drug combinations, and guide the design of clinical trials aimed at improving personalized patient management

on the preclinical treatment of cancer that arise from the use of predictive models, either experimental or mathematical. Mathematical modelling can be employed to reconstruct the mechanisms of cancer formation and progression in networks of molecular interactions, and to perform *in silico* experiments that predict the effect of combinatorial targeted inhibition of relevant molecules within these networks. Therefore, combining mathematical models with omics analyses of animal and cell line models, mapped onto omics profiles of human tumours, is the basis for systems biology efforts to test the efficacy of candidate drug combinations and to design combinatorial targeted therapy for cancer in preclinical trials (Fig. 16.2).

16.3 Systems Biology Approaches to the Design of Combinatorial Targeted Therapy for Cancer

16.3.1 Animal and Cell Line Models

Animal and cell line models of cancer can recapitulate the *in vivo* and *in vitro* patterns, respectively, of tumour formation and progression in patients. Di Cristofano et al. (2001) used a mouse model to study simultaneous inactivation of the tumour suppressor gene PTEN and the Cdknlb gene (encoding $p27^{KIPI}$), which are commonly detected in tumour cells and in most advanced human prostate cancer cases. They showed that concomitant inactivation of one Pten allele and one or both Cdkn1b alleles resulted in accelerated carcinogenesis, and an increased incidence of tumours of various histological origins. All of the mutant mice displayed a higher rate of cell proliferation, but not cell survival, than did wild-type mice, and developed prostate carci-

noma at complete penetrance within 3 months after birth. However, mice deficient in Cdkn1b did not develop prostate cancer, whereas $Pten^{+/-}$ mice died within 8 months after birth. Hence, this study revealed the relevance of combinatorial alteration of Pten and p27 tumour-suppressive activity in a single tumour cell, and its ultimate effect on prostate tumour development. It further supported the contention that well-designed animal models have the potential to reflect the natural history and pathological features of human prostate cancer, and to provide useful predictive information.

The NCI-60 cell line panel is a collection of 60 tumour-cell lines derived from human cancers (NCI 2009a). Included in the panel are nine categories of tumour cells, specifically, leukaemia, melanoma, breast, ovarian, prostate, colon, lung, renal, and central nervous system cancer cells. Researchers have profiled the NCI-60 panel for mRNA and protein expression, mutational status, chromosomal aberrations, and DNA copy number, and the National Cancer Institute has used it to screen more than 100,000 anticancer chemical compounds since 1990. Covell et al. (2007) compared the performance of anticancer compounds used in various clinical trials with that of their structural analogues screened in the NCI-60 panel, and found that the NCI-60 screen could be used to identify appropriate anticancer compounds with selective sensitivity to leukaemia, colon, central nervous system, melanoma and ovarian cancer cell lines, but not to renal, prostate, or breast cancer cell lines. In 2006, Neve and colleagues integrated comparative genomic hybridization analysis with immunoblotting, to establish a panel of 51 breast cancer cell lines with genomic and transcriptomic features that recapitulated those of 145 primary breast tumours (Neve et al. 2006). They found that the cell lines display heterogeneity in gene copy numbers, gene expression, and protein profiles similar to those in the primary tumours, although they reported some differences between human tumours and cancer cell lines. To further assess the ability of these cell lines to predict responses of breast cancer to targeted therapy, the authors treated nine Her2-amplified cell lines and two control cell lines with Herceptin, and found that the frequency of responses to the treatment was similar to that reported in clinical evaluations of Herceptin-based monotherapy. Therefore, the Her2-amplified cell line panel can be used as a model system mirroring the functional contribution of the genes involved in human breast carcinogenesis, and serves to predict outcomes of targeted therapy for cancer. Importantly, these studies emphasized the need to study large sets of cancer cell lines under physiologically relevant conditions, to determine the heterogeneity of human cancers. Indeed, the failure of early studies of cell lines and animal models to predict the efficacy of therapeutic interventions in human cancers probably resulted from the failure of these models to reflect the heterogeneity of, and aberrations present in, the tissue samples studied. One example of this is the widespread use of the MDA-MB-231 breast cancer cell line as a platform for the development of therapies for this cancer. This line has coordinate mutations of RAS and RAF that are rarely present in human breast cancers ($<0.1\%$).

Appropriate cautious use of animal and cell line models, with an understanding of both their strengths and weaknesses, remains fundamental to the discovery and implementation of novel anticancer therapeutics and biomarkers; quantification of the pharmacology and toxicity of existing drugs; performance of preclinical drug

studies; and validation of hypotheses regarding the most effective approaches to treating cancer in humans. Designing clinical trials, however, cannot depend solely on preclinical analyses of a few pairwise drug combinations tested experimentally using a limited number of tumour platforms. In this regard, mathematical modelling is extremely useful, because it enables computational exploration of the effect of both individual therapeutics and combinations of them at much lower costs and more rapidly than in experimental studies. Mathematical modelling thus provides a theoretical platform for generating hypotheses regarding effective treatment of human tumours. However, such hypotheses must be tested experimentally prior to implementation of human trials.

16.3.2 Pharmacodynamic Modelling

Pharmacodynamic modelling is used to predict the physiological effects of drugs on the human body and to direct the choice of targeted therapy. Researchers have formulated several mathematical models to describe the dynamics of processes that may be deregulated in human cancers, such as cell signalling networks (Bhalla and Iyengar 1999; Brightman and Fell 2000; Borisov et al. 2009; Park et al. 2003; Hoffmann et al. 2002; Soebiyanto et al. 2007); the cell cycle (Tyson 1991; Chassagnole et al. 2006); and apoptosis (Fussenegger et al. 2000; Legewie et al. 2006). Although these models offer only a partial view of the complex mechanisms that regulate tumour formation and progression, they can be rapidly implemented to study the effect of drug combinations on rewiring biological pathways deregulated in patients with cancer. To this end (see also Chap. 14), Physiomics (2009) developed pharmacodynamic models of the mammalian cell cycle, apoptosis, and growth factors for studying available drugs, such as cyclin-dependent kinase inhibitors, gemcitabine, Aurora kinase inhibitors, Herceptin, Iressa, Tarceva, and the Raf inhibitor ZM3363732. In addition, Charusanti et al. (2004) modelled the effects of Gleevec on BCR-ABL autophosphorylation and signalling in the Crkl pathway in chronic myeloid leukaemia cases, and predicted the minimum concentration for drug effectiveness. Their modelling results also indicated that the mechanisms of drug clearance may reduce the efficacy of Gleevec in blast crisis cells (a phase of chronic myelogenous leukemia), and may be present from the onset of disease. In 2004, Christopher and colleagues formulated a detailed mathematical model that combined gene expression networks with signalling transduction pathways to describe the processes of receptor activation, mitogenic signalling, cell cycle initiation, passage of checkpoints, and apoptosis in colon cancer cell lines (Christopher et al. 2004). They employed sensitivity analysis and optimization methods to identify unknown regulatory interactions and kinetic parameters from time-course experiments measuring mRNA abundance and protein activity. They also identified previously unknown pathways and crosstalk between known pathways by analysing fluorescence-activated cell sorter, RNA-knockdown, cell-growth, and apoptosis data. Using this mathematical framework, the authors computationally tested the

efficacy of various drugs and experimentally validated their modelling predictions. This theoretical platform can potentially be used to assess the efficacy of therapeutics in specific patient populations, using patient-specific omics data. However, the use of this theoretical platform remains limited by the lack of quantitative high-throughput omics data. Although such platforms are rapidly being implemented, their application for human tumours remains to be realized.

16.3.3 Pharmacokinetic Modelling

Pharmacokinetic modelling facilitates prediction of how the underlying genomics of an individual, and how the physiological functions of the human body, affect the drugs used for treating diseases in patients. It also may help to guide dosing and scheduling of treatment to concurrently minimize toxicity and optimize efficacy. Kuh et al. (2000) created a pharmacokinetic model to determine the time courses for intracellular paclitaxel concentrations at different numbers and binding affinities of intracellular binding sites, free binding site fractions, and concentrations of the drug in extracellular fluid. They experimentally validated the results predicted by the model using the MCF-7 breast cancer cell line. Thus, this model is a useful tool for computationally quantifying the dose dependence of paclitaxel in breast cancer cells, and for scheduling paclitaxel administration. Furthermore, Chen et al. (2007) combined pharmacokinetic and toxicodynamic models to study the relationship between the time courses of topotecan disposition and topotecan-induced toxic effects and to develop a computational tool for an inverse-targeting strategy designed to enhance topotecan safety and efficacy. Also, Physiomics (2009) formulated a detailed pharmacokinetic model that simulates the time courses of drug concentrations in different organs. A unique feature of this model is its capacity for simultaneous prediction of drug absorption and distribution in the entire body, together with specific adaptation for use in individual preclinical experiments. Most of these models, however, remain to be experimentally tested and implemented in cancer management.

16.3.4 Combined Pharmacodynamic-Pharmacokinetic Modelling

Combined pharmacodynamic and pharmacokinetic modelling can direct the selection of drug combinations, determine the drug dose effect, and guide drug scheduling for effective treatment of human cancers. Panetta et al. (2008) developed a mathematical model that integrated topotecan pharmacokinetics, tumour growth, and neutrophil and platelet dynamics to analyse different topotecan-based treatment strategies for neuroblastoma, and to compare them with systemic exposure to, and scheduling of, topotecan. They used this integrative approach to optimize topotecan schedules and improve outcomes while also reducing drug toxicity. Additionally, Yamazaki et al. (2008) derived a mathematical model to characterize the phar-

macodynamic/pharmacokinetic relationship between the concentration of a cMet inhibitor and the level of phosphorylation of the cMet receptor tyrosine kinase, and to correlate cMet phosphorylation with antitumoral efficacy. Using xenograft mouse models of human gastric carcinoma and glioblastoma as an experimental platform, the authors correlated the time delay observed between the inhibitor concentrations and the cMet phosphorylation response with a rate-limiting distribution from plasma into tumour. They also established that inhibition of tumour growth demands near-complete inhibition of cMet phosphorylation. Thus, this integrative approach can guide the dosing and dose escalation to achieve effective systemic exposure. Physiomics (2009) developed a complete pharmacodynamic/pharmacokinetic framework that predicts the behaviour of individual cells in growing tumours (see Chap. 14). This model combines pharmacodynamic simulations of the cell cycle and apoptosis for each cell in the population, with pharmacokinetic models that generate concentration profiles of various drugs, and therefore can be used to simulate experiments involving cell cultures or actual tumours in order to optimize clinical trial design.

16.3.5 Combined Therapy Modelling

Combined therapy modelling serves as a computational tool to predict the effect of drug combinations on lowering the doses of drugs with similar functional targets, in order to achieve efficacy while minimizing toxicity, reducing the frequency of drug resistance, sensitizing the cells to the action of a drug by another drug, and enhancing drug potency via the addition of another drug. Araujo et al. (2005) devised a mathematical model to demonstrate that combinatorial inhibition of multiple kinases in the EGFR signalling network may induce signal attenuation with lower doses of selective inhibitors than those used for targeting of individual nodes in the network. Their modelling results also indicated that weakening of signals is enhanced in pathways downstream from serially connected target points, and that inhibition of multiple upstream processes may synergistically attenuate signal transduction. Komarova and Wodarz (2005) studied the resistance of cancers to treatment with targeted small-molecule inhibitors, using stochastic modelling based on the turnover rate of tumour cells and the rate at which resistant mutants arose. The model predicted that resistance develops mainly before the start of treatment, and that drug combinations are unlikely to be more effective than monotherapy for cancers with high turnover rates. Michor et al. (2006) investigated the probability of resistance of cancer to targeted therapy, using a mathematical framework based on multitype branching processes. By applying this mathematical model to studying the evolution of resistance of cells to targeted therapy, the authors estimated the escape dynamics for arbitrary mutation networks necessary to confer resistance, and determined the probability of success and failure of treatment regimens. Also, Fitzgerald et al. (2006) used a mechanistic model to compute the dose-response curves for signalling proteins inhibited by treatment with individual or combina-

tions of drugs, and to calculate curves corresponding to Loewe additivity and Bliss independence. Using this theoretical platform, the authors predicted the potency of drugs that inhibited single enzymes, signalling pathways, and inter-regulated signalling cascades.

Successful generation of accurate hypotheses regarding combinatorial targeted therapy for preclinical studies depends on the ability of mathematical models to adequately recapitulate biological functions of normal and cancer cells. Therefore, a systems understanding of the mechanisms that control tumour formation and progression is critical for developing computational tools that can predict cancer states at the level of cell, tissue, organ, and, ultimately, the entire human body. Although such a sophisticated version of mathematical modelling of cancer is as yet far from accomplished, researchers will be able to formulate more detailed computational frameworks as more comprehensive omics data becomes available. These models will incorporate key features of cancers, such as cell morphology, post-translational modification of signalling proteins, and cell heterogeneity, and will be tested using quantitative high-throughput experimental data for validation. Thus, integrative use of detailed mathematical modelling with cell-line and animal models will lead to the design of preclinical *in vitro* and *in vivo* trials that should aid in the selection, dose dependence, and scheduling of effective drug combinations, as well as in the design of clinical trials aimed at improving personalized patient management (Fig. 16.2).

16.3.6 Biopsy and Virtual Biopsy Approaches to Measuring Tumours and Assessing Treatment Activity

The characteristics of a tumour can be defined by direct sampling (biopsy) or determining its phenotype using indirect peripheral sampling or imaging approaches (virtual biopsy). One rate-limiting factor in advancing knowledge of tumour biology at the molecular level is the paucity of readily available tumour samples obtained at different time points during treatment. Patients with metastatic cancers are rarely biopsied more than once during clinical trials; even more rarely are such patients biopsied at recurrence and during therapy. Furthermore, it is likely that each metastasis has its own unique characteristics based on molecular evolution, selection, and microenvironment. This may contribute to mixed responses, in which one metastasis resolves completely, whereas another progresses unabatedly after therapy. A change in the tumour's molecular phenotype and interaction with its unique microenvironment underlies any change in tumoral behaviour. Thus, identification of baseline condition, and the change in molecular characteristics of tumours and their microenvironment, are of paramount importance in increasing our understanding of the determinants of tumours' responses and resistance to treatment. Therefore, direct biopsy and rebiopsy of tumours will be instrumental to increasing our knowledge of tumour response and resistance mechanisms. Also, emerging sensitive technologies that can detect tumour fragments peripherally and noninvasively, such as detection of tumour fragments using a regular blood test, are becoming ripe for clinical

use. Antigenic tumour DNA, microRNA, and protein fragments enter the bloodstream and can be measured reliably in peripheral blood samples, with minimal trauma to the patient. In addition, measurement enumeration of circulating tumour cells (cancer cells originating in a tumour that detach and enter the bloodstream) is now performed reliably in the clinic. The least invasive technique used to analyse baseline tumour characteristics and response to treatment is *in situ* molecular imaging. One commercially available *in situ* molecular imaging technique is positron emission tomography (PET), in which labelled fluorodeoxyglucose, a glucose analogue, is taken up by active tumour cells and appears on an image as an area of increased activity corresponding to a tumour. Radioactive or fluorescently labelled small molecules and antibodies that specifically or preferentially target tumour cells can be detected by measuring their activity at the tumour site using external imaging detectors, thereby defining molecular characteristics of the tumour, changes in these characteristics over time, and the tumour volume and viability in response to treatment. When combined with sensitive imaging technologies, molecular imaging can reveal minimal tumour burdens at the single-cell level, and thus identify occult metastases and minimal residual disease. Combinations of the various diagnostic modalities including direct tissue biopsy analysis, peripheral sampling of tumour fragments, and *in situ* molecular imaging, are slowly making their way into the clinic, and often provide information essential for understanding the mechanisms that drive tumour evolution and for guiding the design of targeted treatments.

16.4 The Future of Clinical Trials: Applying Systems Approaches to Clinical Trial Design

In the not to distant future, clinical trials could be very different from the trials performed in the present day. The vast majority of current clinical trials suffer from two significant obstacles. First, they require very large sample sizes to detect small but statistically significant differences in outcomes between experimental and standard therapy arms. This is because in most studies, the experimental therapy is only marginally more effective than the standard therapy, and current trial designs are directed primarily at the ‘average’ patient rather than patient populations most likely to benefit from the therapy. Secondly, most current clinical trials evaluate the efficacy of a single agent, or rarely, a combination of two agents. The development of rational combinatorial therapies designed to overcome intrinsic and adaptive resistance to targeted therapy is currently in its infancy, and represents one of the most promising opportunities for systems biology approaches. We predict that in the foreseeable future, systems biology approaches will facilitate the use of rational combinations of *multiple* agents that target the driving molecular lesions in individual patients and thus overcome the adaptive mechanisms of drug resistance driven by therapeutic modalities. Hopefully, these combinations will be sufficiently effective at producing statistically significant results, with much smaller sample sizes and lower costs than those used in current clinical trials, to permit an increase

in the rate of translation of new effective therapies to the clinic. Knowledge gained from systems approaches that employ integrated preclinical *in silico*, cell line, and animal models will guide the rational selection of combinations of targeted agents designed to inhibit the aberrant molecular pathways that drive cancer in individual patients. The simultaneous use of more than two agents will evade bypass of these pathways, activation of interconnected pathways through feedback loops, and emergence of mechanisms that render human tumours resistant to targeted therapy (see Chap. 15 for chronotherapeutics).

Neoadjuvant therapy trials examining the delivery of potentially efficacious therapy for cancer prior to definitive surgery, and window-of-opportunity trials assessing short-term manipulation of targets, provide unique opportunities for identifying useful biomarkers and selecting patients, and are rapidly becoming the gold standard for validation of biomarkers and drugs in patients with cancer. This is because at the time of surgery one can assess the tumour response and determine the accuracy of biomarkers in predicting responses to specific treatments (i.e. predictive biomarkers), as well as the effectiveness of a drug in eliciting a response. Clinical trials presently carried out on patients with metastatic cancer rarely require either initial biopsy analysis or re-biopsy analysis at progression. The limitations of these trials thus result in missed opportunities for obtaining information that is critical to rational drug delivery, and for understanding treatment-resistance mechanisms (i.e., identifying the changes in the molecular composition of a tumour that occur between its initial response to treatment and later progression). Future clinical trials should enrol patients with locally advanced and metastatic cancer; test them for the presence of specific genetic lesions in specific pathways; measure the activity of identified biomarkers, using biopsy analysis or imaging approaches; and deliver rationally selected agents that target multiple genetic lesions in each pathway. Properly designed trials using a systems biology approach should require many fewer patients, and much less time, to elucidate answers to critical questions. Patients will undergo rebiopsy analysis and reimaging, when their cancers recur or progress, to identify mechanisms of resistance to treatment and target new combinations of pathway aberrations and emerging properties of therapy-induced resistance. Failure to implement such integrative approaches will limit the ability of oncologists to produce a rapid improvement in outcomes of this devastating disease.

References

ACS (2009) American cancer society. General types of treatment. http://www.cancer.org/docroot/cri/content/cri_2_4_4x_local_vs_systemic_therapy_5.asp. Accessed 12 Oct 2009

Adams J, Kauffman M (2004) Development of the Proteasome Inhibitor Velcade (Bortezomib). Cancer Invest 22(2):304–311

Aouacheria A, Navratil V, Barthelaix A, Mouchiroud D, Gautier C (2006) Bioinformatic screening of human ESTs for differentially expressed genes in normal and tumour tissues. BMC Genomics 7:94

Apolo AB, Milowsky M, Bajorin DF (2009) Clinical states model for biomarkers in bladder cancer. Future Oncol 5(7):977–992

Araujo RP, Petricoin EF, Liotta LA (2005) A mathematical model of combination therapy using the EGFR signalling network. Biosystems 80(1):57–69
Bertucci F, Birnbaum D (2008) Reasons for breast cancer heterogeneity. J Biol 7:6
Bhalla US, Iyengar R (1999) Emergent properties of networks of biological signalling pathways. Science 283(5400):381–387
Brightman FA, Fell DA (2000) Differential feedback regulation of the MAPK cascade underlies the quantitative differences in EGF and NGF signalling in PC12 cells. FEBS Lett 482(3):169–174
Borisov N, Aksamitiene E, Kiyatkin A, Legewie S, Berkhout J, Maiwald T, Kaimachnikov NP, Timmer J, Hoek JB, Kholodenko BN (2009) Systems-level interactions between insulin-EGF networks amplify mitogenic signalling. Mol Syst Biol 5:256
Campagne F, Skrabanek L (2006) Mining expressed sequence tags identifies cancer markers of clinical interest. BMC Bioinformatics 7:481
Charusanti P, Hu X, Chen L, Neuhauser D, DiStefano JJ III (2004) A mathematical model of BCR-ABL autophosphorylation signalling through the Crkl pathway, and Gleevec dynamics in chronic myeloid leukaemia. Disc Cont Dyn Systs Ser B 4(1):99–114
Chassagnole C, Jackson RC, Hussain N, Bashir L, Derow C, Savin J, Fell DA (2006) Using a mammalian cell cycle simulation to interpret differential kinase inhibition in anti-tumour pharmaceutical development. Bio Systems 83(2–3):91–97
Chen YT, Scanlan MJ, Venditti CA, Chua R, Theiler G, Stevenson BJ, Iseli C, Gure AO, Vasicek T, Strausberg RL, Jongeneel CV, Old LJ, Simpson AJ (2005) Identification of cancer/testis-antigen genes by massively parallel signature sequencing. Proc Natl Acad Sci U S A 102(22):7940–7945
Chen J, Lu Q, Balthasar JP (2007) Mathematical modelling of topotecan pharmacokinetics and toxicodynamics in mice. J Pharmacokinet Pharmacodyn 34(6):829–847
Christopher R, Dhiman A, Fox J, Gendelman R, Haberitcher T, Kagle D, Spizz G, Khalil IG, Hill C (2004) Data-driven computer simulation of human cancer cell. Ann N Y Acad Sci 1020:132–153
Clark-Langone KM, Wu JY, Sangli C, Chen A, Snable JL, Nguyen A, Hackett JR, Baker J, Yothers G, Kim C, Cronin MT (2007) Biomarker discovery for colon cancer using a 761 gene RT-PCR assay. BMC Genomics 8:279
Covell DG, Huang R, Wallqvist A (2007) Anticancer medicines in development: assessment of bioactivity profiles within the National Cancer Institute anticancer screening data. Mol Cancer Ther 6(8):2261–2270
Cronin M, Sangli C, Liu ML, Pho M, Dutta D, Nguyen A, Jeong J, Wu J, Langone KC, Watson D (2007) Analytical validation of the Oncotype DX genomic diagnostic test for recurrence prognosis and therapeutic response prediction in node-negative, estrogen receptor-positive breast cancer. Clin Chem 53(6):1084–1091
Dexter DL, Leith JT (1986) Tumour heterogeneity and drug resistance. J Clin Oncol 4(2):244–257
Dhanasekaran SM, Barrette TR, Ghosh D, Shah R, Varambally S, Kurachi K, Pienta KJ, Rubin MA, Chinnaiyan AM (2001) Delineation of prognostic biomarkers in prostate cancer. Nature 412(6849):822–826
Di Cristofano A, De Acetis M, Koff A, Cordon-Cardo C, Pandolfi PP (2001) Pten and p27KIP1 cooperate in prostate cancer tumour suppression in the mouse. Nat Genet 27(2):222–224
English DR, Armstrong BK, Kricker A, Fleming C (1997) Sunlight and cancer. Cancer Causes Control 8(3):271–283
Esteller M (2008) Epigenetics in cancer. N Engl J Med 358(11):1148–1159
Fitzgerald JB, Schoeberl B, Nielsen UB, Sorger PK (2006) Systems biology and combination therapy in the quest for clinical efficacy. Nat Chem Biol 2(9):458–466
Fussenegger M, Bailey JE, Varner J (2000) A mathematical model of caspase function in apoptosis. Nat Biotechnol 18(7):768–774
Geisler JP, Rose SL, Geisler HE, Miller GA, Wiemann MC (2002) Drug resistance and tumour heterogeneity. CME J Gynecol Oncol 7:25–28
Haass NK, Smalley KS (2009) Melanoma biomarkers: current status and utility in diagnosis, prognosis, and response to therapy. Mol Diagn Ther 13(5):283–296

Harry VN, Gilbert FJ, Parkin DE (2009) Predicting the response of advanced cervical and ovarian tumours to therapy. Obstet Gynecol Surv 64(8):548–560
Hoffmann A, Levchenko A, Scott ML, Baltimore D (2002) The IkappaB-NF-kappaB signalling module: temporal control and selective gene activation. Science 298(5596):1241–1245
Hudis CA (2007) Trastuzumab—mechanism of action and use in clinical practice. N Engl J Med 357(1):39–51
Hustinx SR, Cao D, Maitra A, Sato N, Martin ST, Sudhir D, Iacobuzio-Donahue C, Cameron JL, Yeo CJ, Kern SE, Goggins M, Mollenhauer J, Pandey A, Hruban RH (2004) Differentially expressed genes in pancreatic ductal adenocarcinomas identified through serial analysis of gene expression. Cancer Biol Ther 3(12):1254–1261
Jordan VC (2006) Tamoxifen (ICI46,474) as a targeted therapy to treat and prevent breast cancer. Br J Pharmacol 147(Suppl 1):269–276
Katzel JA, Fanucchi MP, Li Z (2009) Recent advances of novel targeted therapy in non-small cell lung cancer. J Hematol Oncol 2:2
Kaufmann SHE (2008) Elie Metchnikoff's and Paul Ehrlich's impact on infection biology. Microbes Infect 10(14–15):1417–1419
Komarova NL, Wodarz D (2005) Drug resistance in cancer: principles of emergence and prevention. Proc Natl Acad Sci U S A 102(27):9714–9719
Kosari F, Munz JM, Savci-Heijink CD, Spiro C, Klee EW, Kube DM, Tillmans L, Slezak J, Karnes RJ, Cheville JC, Vasmatzis G (2008) Identification of prognostic biomarkers for prostate cancer. Clin Cancer Res 14(6):1734–1743
Kuh HJ, Jang SH, Wientjes MG, Au JLS (2000) Computational model of intracellular pharmacokinetics of paclitaxel. J Pharmacol Exp Ther 293(3):761–770
Kuper H, Boffetta P, Adami HO (2002) Tobacco use and cancer causation: association by tumour type. J Intern Med 252(3):206–224
Lassila M, Allen TJ, Cao Z, Thallas V, Jandeleit-Dahm KA, Candido R, Cooper ME (2004) Imatinib attenuates diabetes-associated atherosclerosis. Arterioscler Thromb Vasc Biol 24(5):935–942
Legewie S, Blüthgen N, Herzel H (2006) Mathematical modelling identifies inhibitors of apoptosis as mediators of positive feedback and bistability. PLoS Comput Biol 2(9)
Ma Y, Qian Y, Wei L, Abraham J, Shi X, Castranova V, Harner EJ, Flynn DC, Guo L (2007) Population-based molecular prognosis of breast cancer by transcriptional profiling. Clin Cancer Res 13(7):2014–2022
Maloney DG, Grillo-López AJ, White CA, Bodkin D, Schilder RJ, Neidhart JA, Janakiraman N, Foon KA, Liles TM, Dallaire BK, Wey K, Royston I, Davis T, Levy R (1997) IDEC-C2B8 (Rituximab) anti-CD20 monoclonal antibody therapy in patients with relapsed low-grade non-Hodgkin's lymphoma. Blood 90(6):2188–2195
Mellemkjaer L, Hammarstrom L, Andersen V, Yuen J, Heilmann C, Barington T, Bjorkander J, Olsen JH (2002) Cancer risk among patients with IgA deficiency or common variable immunodeficiency and their relatives: a combined Danish and Swedish study. Clin Exp Immunol 130(3):495–500
Michor F, Nowak MA, Iwasa Y (2006) Evolution of resistance to cancer therapy. Curr Pharm Design 12(3):261–271
Mirzoeva OK, Das D, Heiser LM, Bhattacharya S, Siwak D, Gendelman R, Bayani N, Wang NJ, Neve RM, Guan Y, Hu Z, Knight Z, Feiler HS, Gascard P, Parvin B, Spellman PT, Shokat KM, Wyrobek AJ, Bissell MJ, McCormick F, Kuo WL, Mills GB, Gray JW, Korn WM (2009) Basal subtype and MAPK/ERK kinase (MEK)-phosphoinositide 3-kinase feedback signalling determine susceptibility of breast cancer cells to MEK inhibition. Cancer Res 69(2):565–572
NCI (2009a) National cancer institute. DTP human tumour cell line screen. http://dtp.nci.nih.gov/docs/cancer/cancer_data.html. Accessed 12 Oct
NCI (2009b) National cancer institute. Staging fact sheet. http://www.cancer.gov/cancertopics/factsheet/Detection/staging. Accessed 12 Oct
Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L, Bayani N, Coppe JP, Tong F, Speed T, Spellman PT, DeVries S, Lapuk A, Wang NJ, Kuo WL, Stilwell JL, Pinkel D, Albertson DG, Waldman FM, McCormick F, Dickson RB, Johnson MD, Lippman M, Ethier S,

Gazdar A, Gray JW (2006) A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell 10(6):515–527
NIH (2009a) National institute of health. Handbook illustrations. http://ghr.nlm.nih.gov/handbook/illustrations/cancer. Accessed 12 Oct 2009
NIH (2009b) National institute of health. Human genome project fact sheet. http://www.nih.gov/about/researchresultsforthepublic/HumanGenomeProject.pdf. Accessed 12 Oct
NIH (2009c) National institute of health. The integrative cancer biology program. http://icbp.nci.nih.gov. Accessed 12 Oct
Olovnikov IA, Kravchenko JE, Chumakov PM (2009) Homeostatic functions of the p53 tumour suppressor: regulation of energy metabolism and antioxidant defense. Semin Cancer Biol 19(1):32–41
O'Reilly KM, Mclaughlin AM, Beckett WS, Sime PJ (2007) Asbestos-related lung disease. Am Fam Physician 75(5):683–688
Ou K, Yu K, Kesuma D, Hooi M, Huang N, Chen W, Lee SY, Goh XP, Tan LK, Liu J, Soon SY, Bin Abdul Rashid S, Putti TC, Jikuya H, Ichikawa T, Nishimura O, Salto-Tellez M, Tan P(2008) Novel breast cancer biomarkers identified by integrative proteomic and gene expression mapping. J Proteome Res 7(4):1518–1528
Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, Hiller W, Fisher ER, Wickerham DL, Bryant J, Wolmark N (2004) A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 351(27):2817–2826
Panetta JC, Schaiquevich P, Santana VM, Stewart CF (2008) Using pharmacokinetic and pharmacodynamic modelling and simulation to evaluate importance of schedule in topotecan therapy for pediatric neuroblastoma. Clin Cancer Res 14(1):318–325
Pao W, Miller V, Zakowski M, Doherty J, Politi K, Sarkaria I, Singh B, Heelan R, Rusch V, Fulton L, Mardis E, Kupfer D, Wilson R, Kris M, Varmus H (2004) EGF receptor gene mutations are common in lung cancers from 'never smokers' and are associated with sensitivity of tumours to gefitinib and erlotinib. Proc Natl Acad Sci U S A 101(36):13306–13311
Park CS, Schneider IC, Haugh JM (2003) Kinetic analysis of platelet-derived growth factor receptor/phosphoinositide 3-kinase/Akt signalling in fibroblasts. J Biol Chem 278(39):37064–37072
Phillips HS, Kharbanda S, Chen R, Forrest WF, Soriano RH, Wu TD, Misra A, Nigro JM, Colman H, Soroceanu L, Williams PM, Modrusan Z, Feuerstein BG, Aldape K (2006) Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. Cancer Cell 9(3):157–173
Physiomics (2009) Pharmacodynamic modelling. http://www.physiomics-plc.com/pd_modelling.htm. Accessed 12 Oct
Rini BI (2007) Vascular endothelial growth factor-targeted therapy in renal cell carcinoma: current status and future directions. Clin Cancer Res 13(4):1098–1106
Ross JS (2009) Multigene classifiers, prognostic factors, and predictors of breast cancer clinical outcome. Adv Anat Pathol 16(4):204–215
Scott A, Salgia R (2008) Biomarkers in lung cancer: from early detection to novel therapeutics and decision making. Biomark Med 2(6):577–586
Semenza GL (2009) Regulation of oxygen homeostasis by hypoxia-inducible factor 1. Physiology 24:97–106
Shen D, He J, Chang HR (2005) In silico identification of breast cancer genes by combined multiple high throughput analyses. Int J Mol Med 15(2):205–212
Soebiyanto RP, Sreenath SN, Qu CK, Loparo KA, Bunting KD (2007) Complex systems biology approach to understanding coordination of JAK-STAT signalling. Biosystems 90(3):830–842
Sreekumar A, Poisson LM, Rajendiran TM, Khan AP, Cao Q, Yu J, Laxman B, Mehra R, Lonigro RJ, Li Y, Nyati MK, Ahsan A, Kalyana-Sundaram S, Han B, Cao X, Byun J, Omenn GS, Ghosh D, Pennathur S, Alexander DC, Berger A, Shuster JR, Wei JT, Varambally S, Beecher C, Chinnaiyan AM (2009) Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. Nature 457(7231):910–914
Stemke-Hale K, Gonzalez-Angulo AM, Lluch A, Neve RM, Kuo WL, Davies M, Carey M, Hu Z, Guan Y, Sahin A, Symmans WF, Pusztai L, Nolden LK, Horlings H, Berns K, Hung MC,

Vijver MJ van de, Valero V, Gray JW, Bernards R, Mills GB, Hennessy BT (2008) An integrative genomic and proteomic analysis of PIK3CA, PTEN, and AKT mutations in breast cancer. Cancer Res 68(15):6084–6091

Tian Q, Stepaniants SB, Mao M, Weng L, Feetham MC, Doyle MJ, Yi EC, Dai H, Thorsson V, Eng J, Goodlett D, Berger JP, Gunter B, Linseley PS, Stoughton RB, Aebersold R, Collins SJ, Hanlon WA, Hood LE (2004) Integrated genomic and proteomic analyses of gene expression in Mammalian cells. Mol Cell Proteomics 3(10):960–969

Tortora G, Bianco R, Daniele G (2004) Strategies for multiple signalling inhibition. J Chemother 16(4):41–43

Tremblay F, Marette A (2001) Amino acid and insulin signalling via the mTOR/p70 S6 kinase pathway. A negative feedback mechanism leading to insulin resistance in skeletal muscle cells. J Biol Chem 276(41):38052–38060

Tyson JJ (1991) Modelling the cell division cycle: cdc2 and cyclin interactions. Proc Natl Acad Sci U S A 8(16):7328–7332

van Amerongen R, Berns A (2008) Targeted anticancer therapies: mouse models help uncover the mechanisms of tumour escape. Cancer Cell 13(1):5–7

Van Cutsem E, Köhne CH, Hitre E, Zaluski J, Chang Chien CR, Makhson A, D'Haens G, Pintér T, Lim R, Bodoky G, Roh JK, Folprecht G, Ruff P, Stroh C, Tejpar S, Schlichting M, Nippgen J, Rougier P (2009) Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer. N Engl J Med 360(14):1408–1417

Varambally S, Yu J, Laxman B, Rhodes DR, Mehra R, Tomlins SA, Shah RB, Chandran U, Monzon FA, Becich MJ, Wei JT, Pienta KJ, Ghosh D, Rubin MA, Chinnaiyan AM (2005) Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression. Cancer Cell 8(5):393–406

Vasudevan KM, Barbie DA, Davies MA, Rabinovsky R, McNear CJ, Kim JJ, Hennessy BT, Tseng H, Pochanard P, Kim SY, Dunn IF, Schinzel AC, Sandy P, Hoersch S, Sheng Q, Gupta PB, Boehm JS, Reiling JH, Silver S, Lu Y, Stemke-Hale K, Dutta B, Joy C, Sahin AA, Gonzalez-Angulo AM, Lluch A, Rameh LE, Jacks T, Root DE, Lander ES, Mills GB, Hahn WC, Sellers WR, Garraway LA (2009) AKT-independent signalling downstream of oncogenic PIK3CA mutations in human cancer. Cancer Cell 16(1):21–32

Walther A, Johnstone E, Swanton C, Midgley R, Tomlinson I, Kerr D (2009) Genetic prognostic and predictive markers in colorectal cancer. Nat Rev Cancer 9(7):489–499

Weinstein JN (2004) Integromic analysis of the NCI-60 cancer cell lines. Breast Dis 19:11–22

Wetherill YB, Hess-Wilson JK, Comstock CE, Shah SA, Buncher CR, Sallans L, Limbach PA, Schwemberger S, Babcock GF, Knudsen KE (2006) Bisphenol A facilitates bypass of androgen ablation therapy in prostate cancer. Mol Cancer Ther 5(12):3181–3190

Yamazaki S, Skaptason J, Romero D, Lee JH, Zou HY, Christensen JG, Koup JR, Smith BJ, Koudriakova T (2008) Pharmacokinetic-pharmacodynamic modelling of biomarker response and tumour growth inhibition to an orally available cMet kinase inhibitor in human tumour xenograft mouse models. Drug Metab Dispos 36(7):1267–1274

Zanetta GM, Webb MJ, Li H, Keeney GL (2000) Hyperestrogenism: a relevant risk factor for the development of cancer from endometriosis. Gynecol Oncol 79(1):18–22

Chapter 17
Cancer Robustness and Therapy Strategies

Hiroaki Kitano

Abstract Cancer is robust, in the sense that many forms of the disease are resistant to therapeutic intervention and continue to proliferate. Understanding the fundamental principles of robustness is critical for developing novel and more effective therapies, such as better targeted drugs and improved methods of administration. Robustness is a fundamental and inherent property of complex evolved biological systems, optimized for certain functions and environments. No system, however, can maintain robustness against every possible perturbation. Systems which are robust against some perturbations have been shown to be extremely fragile in the face of others. Cancer is not an exception in this respect. Therefore, an enhanced understanding of cancer robustness and its limitations will assist in uncovering fragilities that off the potential of becoming novel targets for therapeutic interventions in cancer.

17.1 Introduction

17.1.1 Cancer as a Robust System

Cancer is a heterogeneous and highly robust disease that represents the worst-case scenario of entire system failure: a fail-on fault where malfunctioning components are protected by hijacking those mechanisms that support robustness in the normal physiology of the host (Kitano 2003, 2004b). The capability for survival and proliferation of tumour cells is robustly maintained against a range of therapies by such mechanisms as intra-tumoral genetic diversity, feedback loops for multi-drug resistance, tumour-host interactions, etc. This chapter examines why cancer is robust against therapeutic interventions, and tries to identify possible options transcending and controlling that robustness.

H. Kitano (✉)
Sony Computer Science Laboratories, Inc., Tokyo, Japan

The Systems Biology Institute, Tokyo, Japan

Department of Cancer Systems Biology, The Cancer Institute, Tokyo, Japan

Okinawa Institute of Science and Technology, Japan
e-mail: kitano@sbi.jp

A. Cesario, F. B. Marcus (eds.), *Cancer Systems Biology, Bioinformatics and Medicine,*
DOI 10.1007/978-94-007-1567-7_17,

17.1.2 What is Robustness?

Robustness is a biological property of the system that preserves a certain degree of functionality despite external and internal perturbations (Kitano 2004a). It is peculiar to the whole system, and cannot be observed by looking only at individual components. Robustness is observed ubiquitously within and among biological systems, from bacteria to human beings. For example, bacteria robustly maintain chemotaxis against a broad range of perturbations such as external chemical changes and internal fluctuations of enzyme dosages (Alon et al. 1999). Robustness also applies to engineering systems, see Sect. 17.2.4. Specific aspects of the system, the functions to be maintained, and the types of perturbations against which the system is robust, must be well defined in order to make solid deductions.

Robustness is not identical to homeostasis or stability. the term 'homeostasis' was coined by Walter Cannon, from the Greek *homoios*, of the same nature or similar, and *stasis*, state or condition. In 'Wisdom of the body', he describes it as follows (Cannon 1932):

> The everlasting state maintained in the body may be called an equilibrium state. But, this word has a fairly precise meaning now since it has been used for a comparatively simple physicochemical state in which known forces are balanced, namely used for closed systems. The interrelated physiologic action that maintains the main portions of stable states in a living body, which contains the brain, the nerves, the heart, the lungs, the kidneys and the spleen to fulfil their functions collaboratively, is so complex and unique that I have been proposing to use a special word of 'homeostasis' for such a state. This word does not indicate something that is fixed and does not move, or a stasis. It means a certain state that may change but is relatively constant.

Stability is a concept similar to homeostasis, because it is basically judged by the degree of maintenance of a state. Robustness, on the other hand, is judged on the criterion of the maintenance of functions, not of a state. That is to say, if a state changes substantially in order to maintain the functions, it can be considered as a kind of robustness, but not as homeostasis or stability.

With mathematical abstraction, the state of the system can be expressed in N-dimensional phase space. Figure 17.1 represent a simplified two-dimensional phase space. When perturbations are imposed, the state of the system drifts in phase space. If the degree of perturbation is small and the system's state tends to be stable, then its phase space trajectory orbits around the 'basin' (local minimum) of the 'attractor' (a region towards which a dynamical system evolves over time), and gradually returns to the initial state. The machinery of bacterial chemotaxis is an example of such a case.

However, there are cases where perturbations trigger a transition of the state from one basin of the attractor to another. Systems may be robust in this case if they are able to switch to a new state which maintains functions, rather than returning to the original state which may now be unstable. Nature provides a satisfactory example in the tardigrade (Crowe and Crowe 2000). This miniscule water-dwelling animal in its normal state crawls slowly around, and when it enters a high-salt or very dry environment, it becomes dehydrated, stops metabolic activity, and enters

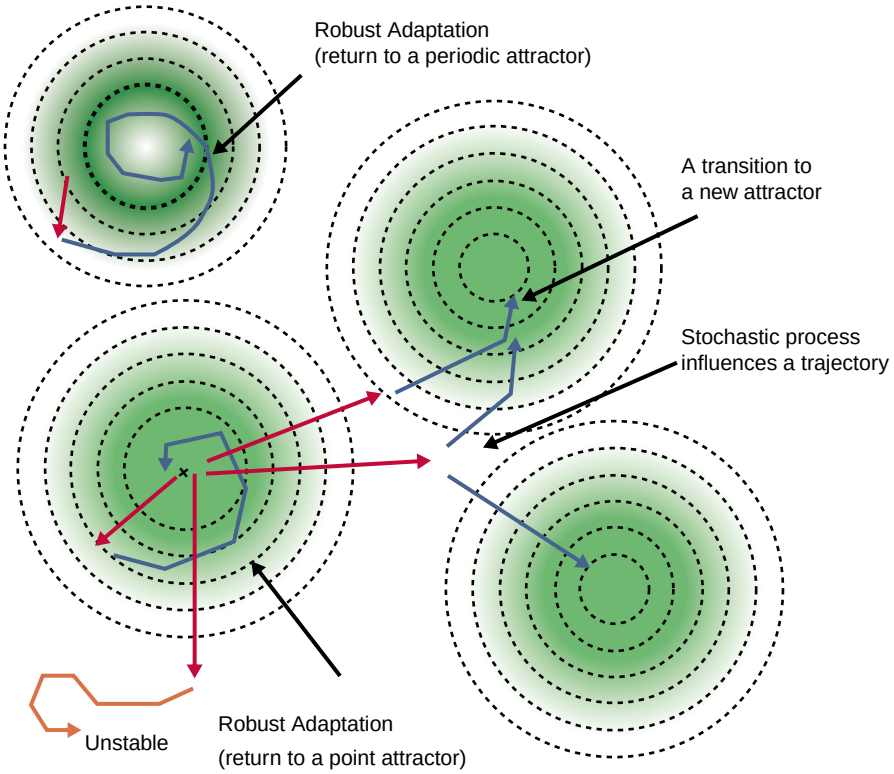


Fig. 17.1 Robustness and stability

a kind of freeze-dry state. It survives in this state for years, and can start crawling again when water is available. In this case, it gave up the homeostasis of metabolism, and moved into the state of waterless survival, to cope with a major external perturbation. This creature is robust against dehydration stress, but that robustness is generated by sacrifice of metabolic homeostasis.

17.1.3 Robustness and Homeostasis

Attention should be paid to the essential difference between robustness and homeostasis. There is a major impact on evolutionary selection from the functions associated with higher probabilities of survival and reproduction, not directly from the state. However, the functions related to maintaining the state of an individual are subject to selection only when linked to survival and reproduction. The system's behaviour of increasing the survival probability by moving into a completely different state to cope with perturbations, just like the response of the tardigrade against dehydration, may be considered as robustness, but not homeostasis. When the continuation of a state

permits the functions of a system to be maintained, robustness and homeostasis (or stability) are equivalent, whereas when the transition of the system into a new state is an effective response to a disturbance, a robust system will abandon homeostasis.

In addition, there are some cases in which robustness is maintained because a system is unstable. HIV has a very high mutation rate, with part of its gene sequence changing frequently. Because of this property, HIV eludes immune system control. In the case of cancer, the chromosomes become more unstable during advanced stages, and produce a variety of abnormalities which coexist even in a solid cancer. Therefore, even if part of a tumour cluster is eliminated by an anticancer drug, subsets of tumour cells with genetic properties tolerant to anticancer drugs will survive and continue to proliferate, as a consequence of chromosomal instability.

17.2 Mechanisms for Robustness

Herein we investigate the basic mechanisms which make systems robust. There are at least four such: systems control, fault-tolerance, modularity, and decoupling.

17.2.1 System Control

Biological systems employ extensive mechanisms of control. Negative and positive feedback and feed-forward are used to make a system dynamically stable around a specific state, or to form bi-stable and multi-stable switches that drive transitions of the state to different attractors. Bacterial chemotaxis is an example of how negative feedback enables robust control against a wide range of fluctuations in chemical concentration (Alon et al. 1999; Barkai and Leibler 1997; Yi et al. 2000). Using integral feedback, bacteria can sense changes in chemo-attractant and chemo-repellent environments, independently of absolute ligand concentrations, so that proper chemotaxis behaviour is maintained over a wide range. In addition, the same mechanism is insensitive to changes in rate constants in the circuit. This is an example of how systems control is used to maintain a state even within the environment of an attractor so that proper functioning can continue.

Positive feedbacks are often used to create bi-stability in signal transduction and the cell cycle, so that the system is tolerant against minor perturbations from stimuli (Chen et al. 2004; Ferrell 2002; Tyson et al. 2001).

17.2.2 Fault-tolerance

Fault-tolerant mechanisms increase resistance to component failure and environmental changes, by providing alternative components or methods aiming ultimately

maintain a function of the system. In the phenomenon referred to as diversity, there exist multiple system components that are similar to each other and are redundant. In other cases, multiple components or circuits with overlapping functions compensate for any component insufficiency. The difference between redundancy and diversity may be elucidated by the example of modern methods of communication. Multiple phone lines are 'redundant', but alternative access to internet, phone, FAX, and other means of communication is 'diversity'. Redundancy and diversity are often considered as opposites, but it is more reasonable to view them as different ways of providing alternative fail-tolerant mechanisms.

17.2.3 Modularity

Modularity is a mechanism providing isolation of a perturbation from the rest of the system. The cell is the most significant example: where one dies, the multi-cellular organism which contains it survives. More subtle examples are modules of biochemical and gene-regulatory networks. Modules that buffer perturbations also play an important role during developmental processes, so that proper pattern formation can be accomplished (Eldar et al. 2002; Meir et al. 2002; von Dassow et al. 2000). There is still some controversy over the definition of what constitutes a module and over methods for module detection, but the general consensus is that modules do exist and they play an important role (Schlosser and Wagner 2004).

17.2.4 Decoupling

Decoupling isolates low-level noise and fluctuations from functional structures and dynamics. One example is genetic buffering by Hsp90, in which misfolding of proteins due to environmental stresses is repaired. In this way, circuit functions are protected from the effects of such perturbations. The decoupling mechanism also applies to genetic variations, where genetic changes in coding regions that may affect protein structures are masked by the fixing of protein folding by Hsp90, unless such masking is removed by extreme stress (Queitsch et al. 2002; Rutherford 2003; Rutherford and Lindquist 1998). Emergent behaviour of complex networks exhibits similar buffering properties (Siegal and Bergman 2002). Waddington suggests that these effects may constitute canalization (Waddington 1957).

The aeroplane, already mentioned as an example of a sophisticated engineering system, illustrates how decoupling mechanisms work within a whole system: see Fig. 17.2. An aeroplane is supposed to maintain the pilot's desired flight path against atmospheric perturbations and various internal perturbations, including changes in the centre of gravity due to fuel consumption and movement of passengers, as well as mechanical inaccuracies. This function is carried out by an automatic flight control system (AFCS) using movable surfaces (rudder, flaps, elevators, etc.) and a

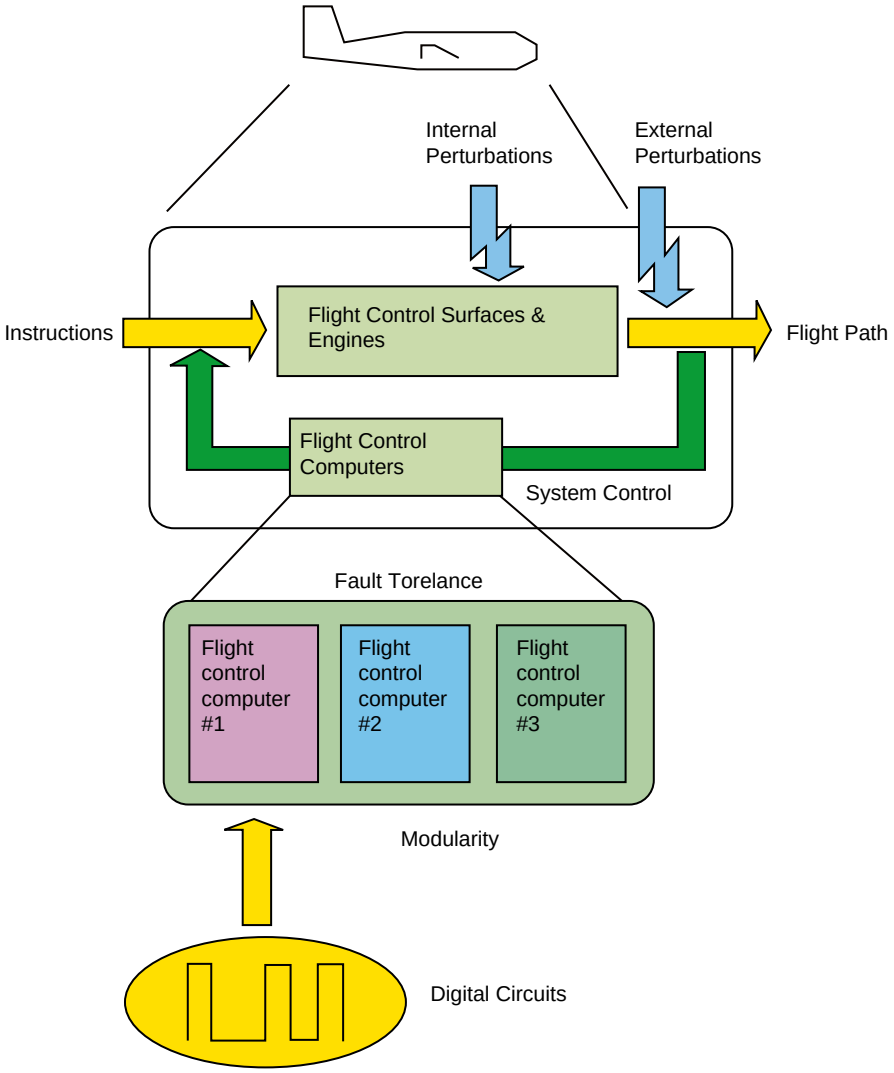


Fig. 17.2 Robustness mechanisms in an aeroplane

propulsion system (engines). Extensive negative feedback control is used to correct deviations from the desired flight path. Precisely because the reliability of the AFCS is crucial for stable flight, it is composed of three independently implemented modules (a triple redundancy system) all of which meet the same functional specifications. Most parts of the AFCS are digitalized, so that low-level noise from voltage fluctuations is effectively decoupled from digital signals that define the function of the system. Thanks to these mechanisms, modern aeroplanes are highly robust against various perturbations.

17.3 Mechanisms for Cancer Robustness

The robustness of cancer against various therapeutic interventions is facilitated by three groups of mechanisms: intratumoral heterogeneity; intracellular feedback loops; and host-tumour entrainment.

Intratumoral genetic heterogeneity is a major source of robustness in cancer. Chromosomal instability enables the generation of intratumoral genetic heterogeneity through gene amplification, chromosomal translocation, point mutations, aneuploidy, etc. (Lengauer et al. 1998; Li et al. 2000; Rasnick 2002; Tischfield and Shao 2003). This heterogeneity is one of the most important features of cancer. Fault-tolerance for tumour survival and growth, even after the application of various therapies, depends on some tumour cells having genetic profiles with therapeutic resistance.

While there are only a few studies on inter-tumour genetic heterogeneity, available observations in certain types of solid tumours indicate that within one tumour cluster there are multiple sub-clusters of tumour cells, each exhibiting different chromosomal aberrations (Baisse et al. 2001; Frigyesi et al. 2003; Fujii et al. 2000; Gorunova et al. 1998, 2001). This implies that each sub-cluster is developed as a clonal expansion of a single mutant cell. Creation of a new sub-cluster is initiated by the emergence of a new mutant that is viable for clonal expansion. A computational study has demonstrated that the spatial distribution within a tumour cluster enables coexistence of multiple sub-clusters (Gonzalez-Garcia et al. 2002). Heterogeneity is also a feature of cancer stem cells at the genetic and epigenetic levels.

Multi-drug resistance is a cellular-level mechanism that provides robustness of viable tumour cells against toxic anti-cancer drugs. In general, this mechanism involves over-expression of genes such as MDR1 that encode an ATP-dependent efflux pump, P-glycoprotein (P-gp), which ejects a broad range of cytotoxins (Juliano and Ling 1976; Nooter and Herweijer 1991). Trials to counteract the function of P-gp, by using verapamil and its cyclosporine derivative PSC833, have been disappointing (Tsuruo et al. 1981).

Host-tumour entrainment (alignment of two systems to each other) involves various tumour-host interactions as well as remodelling and control of the tumour micro-environment. Tumour-host interactions play major roles in tumour growth and metastasis (Bissell and Radisky 2001). When the rate of tumour growth is not matched by vascular growth, a hypoxic condition emerges in tumour clusters (Harris 2002). This triggers HIF-1 up-regulation which induces a series of reactions that maintain normal physiological conditions (Sharp and Bernaudin 2004). Up-regulation of HIF-1 induces up-regulation of VEGF, which facilitates angiogenesis, and of uPAR and other genes that enhance cell motility (Harris 2002). These responses solve the problem of hypoxia for tumour cells, either by providing oxygen to the tumour cluster or by moving tumour cells to a new environment, resulting in further tumour growth or metastasis.

Interestingly, macrophages are found to move via chemotaxis into a tumour cluster. Such a macrophage is called a Tumour Associated Macrophage (TAM), and is

found to over-express HIF-1 (Bingle et al. 2002). This means that a macrophage that is supposed to remove tumour cells may instead be incorporated into feedback loops to facilitate tumour growth and metastasis. Cancer cells can also carry out cell fusion with macrophages to change their character to support metastasis (Sole 2003; Dropulic et al. 1996; Weinberger et al. 2003). Furthermore, macrophages that accumulate around a cancer may over-express genes that contribute to the growth of a tumour.

In addition, tumour cells may evolve through self-extending symbiosis (Kitano and Oda 2006). In this case, tumour cells can enhance their robustness against various perturbations through horizontal gene transfer and up-take of chromosomes, leading to symbiosis with other cells in the form of cell fusion, as well as to the formation of symbiotic relationships with surrounding environments (Ogle et al. 2005; Pawelek et al. 2006; Pawelek 2005; Vignery 2005). This implies that tumour cells may be considered as a group of cells that become somewhat detached from the host system and begin evolving independently, so that a wide range of phenomena, such as self-extending symbiosis, occur in tumour cells, thereby enhancing their robustness against perturbations. Patients suffering from an immune deficit due to AIDS present with a high percentage of Kaposi sarcomas, whereas the incidence rate in those patients of other cancers, such as breast cancer and lung cancer, is substantially lower than usual. This seems to indicate that the innate immune systems are hijacked by cancer cells so as to contribute to their proliferation and metastasis. If these observations are correct, this particular cancer seems to evolve to convert the surrounding host systems into extended cancer support systems. So far, such phenomena have only been reported independently, not as a unified overview. Re-organizing these findings with a view to a more coherent understanding of the aspect of cancer robustness will provide us with guidelines for further research.

17.4 Robustness Trade-offs

Systems that acquire robustness against certain perturbations through design or evolution have intrinsic trade-offs between the competing demands of robustness, fragility, performance, and resource demands. Carlson and Doyle argued, using simple examples from physics and forest fires, that systems which are optimized for specific perturbations are extremely fragile against unusual perturbations (Carlson and Doyle 1999, 2002). Csete and Doyle further argued that robustness is a conserved quantity (Csete and Doyle 2002), in that when enhanced against a range of perturbations, it is compensated by fragility elsewhere, accompanied by compromised performance and increased resource demands (Fig. 17.3).

Robust yet fragile trade-offs can be understood intuitively using the aeroplane example. When comparing modern commercial aeroplanes with the Wright flyer, modern aircraft are far more robust against atmospheric perturbations, a fact attributable to a sophisticated flight control system. However, the unlikely situation of total power failure means that the aeroplane cannot be controlled at all. On the

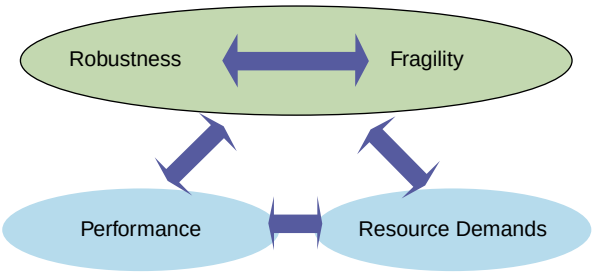


Fig. 17.3 Robustness trade-offs

other hand, the Wright flyer did not rely on an electrical control system. This extreme example illustrates that systems which are optimized for certain perturbations could be extremely fragile against unusual disturbances. Trade-offs are expected to exist not only between robustness and fragility, but also between performance and resource demands.

17.5 Theoretically-motivated Therapeutic Strategies

Because of the highly complex control systems and genetic heterogeneity of tumours, random trials of potential targets are not as satisfactory as one would wish. There is a need for theoretically-motivated approaches to guide us to therapies likely to counter the disease more effectively. The general principle behind the concept of robustness is that there are specific patterns of behaviour, each accompanied by weaknesses, in robust systems, as well as a rational way of controlling and fixing systems. The same general principle also applies to cancer, and can be harnessed in order to develop novel approaches to the prevention and treatment of cancer.

Strategies for cancer therapy may depend upon the level of robustness of tumours in specific patients. When both robustness and genetic heterogeneity are low, then there is a good chance that the use of drugs with specific molecular targets may be effective by causing a common mode failure, in which all redundant subsystems fail for the same reason. The example of Chronic Myeloid Leukaemia (CML) therapy by Imatinib Metylate may provide us with some insights (Hochhaus 2003; Hochhaus et al. 2001). Although this is speculative, the dramatic effect of Imatinib Metylate on early stage CML may be due to a common mode failure, while resistance in advanced stages may be due to high heterogeneity. For this drug to be more efficacious, there is a requirement for an appropriate method of diagnosing the degree of intratumoral genetic variations. Also, the most effective choice of a targeted therapeutic molecule needs to be determined according to processes of identification and optimization.

For patients with advanced-stage cancer, intratumoral genetic heterogeneity may already be high, and various feedback controls may be significantly up-regulated. Drugs that are effective for early stage cancer may not work as well in later stages, due to the heterogeneous response of tumour cells, and feedback responses which

compensate for perturbations. In these cases, therapy and drug design need a drastic shift from a single target-oriented approach to one that is systems-oriented. The question then becomes which particular approach should be taken for targeting the system. There seem to be only a few theoretically-motivated countermeasures, of which we discuss six:

1. The robustness/fragility trade-off implies that cancer cells that have gained increased robustness against various therapies may nevertheless have a point of extreme fragility—an Achilles heel—which, when targeted, may provide dramatic results in combating the disease. The major challenge is to find such a point of fragility. When a cancer is resistant to a specific drug, but vulnerable to another, one therapeutic strategy might consist of a series of regimens that continue to attack the point of fragility. For example, cancer cells that survived because somatic evolution resulted in increased robustness to a first drug, may have fragility to another drug. Eventually, cancer cell clusters establishing robustness to the second drug will appear, for which a third drug should be prepared. Consecutive treatment changes is therefore one promising method for exploring the fragility of cancer. A further extension of the above approach could involve the development of a method for artificial control and creation (or induction) of fragility.
   Some readers may wonder if such a convenient set of circumstances is in fact feasible. It is not, however, a new idea. The notion whereby a cancer has achieved tolerance to a certain type of anticancer drug becomes fragile to other anticancer drugs was already known in the 1970s as ‘Collateral Sensitivity’ (Skipper et al. 1972). Skipper and his colleagues investigated the circumstances in which a cancer becomes tolerant to one of 30 drugs, and attempted to identify other drugs to which the cancer becomes sensitive or fragile. Figure 17.4 shows the tolerance to 10 anticancer drugs out of a total of 30. This research was not pursued afterwards, but it demonstrates the trade-offs of robustness, and hence deserves to be studied anew.
2. There are possibilities for deploying more ‘ecological’ approaches. Where cancer cells with multiple kinds of genetic diversity coexist, they may be in competitive and symbiotic relationships. There is an interesting example in the treatment of prostate cancers, where androgen-dependent cancer cells proliferate to grow a tumour. Blocking androgen to destroy those cells is the main purpose of treatment. Unfortunately, after the shrinkage of a tumour primarily consisting of androgen-dependent cells, cancer cells independent of adrogen begin to proliferate. The androgen-independent cancer cells are higher in malignancy and harder to cure. In the situation where both androgen-dependent and -independent cells coexist, it has been observed that there is a possibility of slight reductions in androgen-independent cancer cell clusters. In this case, it should be possible to establish a treatment to control proliferation and reduce the tumour by intermittent blocking of androgen. Actual clinical results show the effectiveness of just such a method (Bruchovsky et al. 2000; Gleave et al. 1997, 1998). Aihara and his team at the University of Tokyo conducted mathematical analyses of these

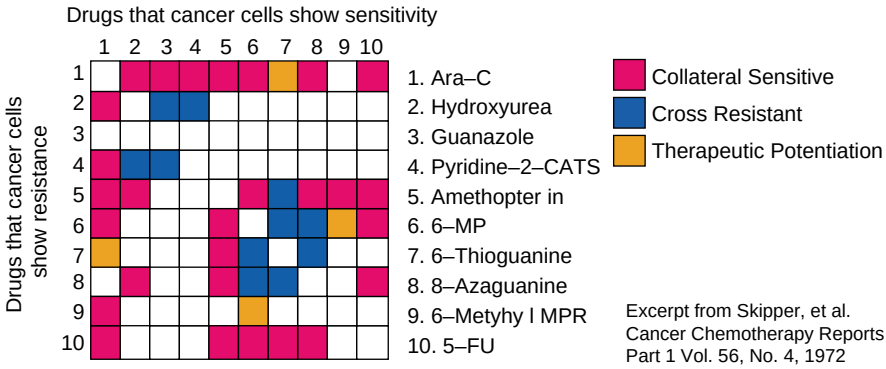


Fig. 17.4 Collateral sensitivity

dynamics to derive an optimum treatment strategy (Shimada and Aihara 2008). Intermittent androgen blocking offers a way of controlling the ecology of genetically diverse cancer cell clusters. The further generalization of the ecology-controlling method, and the systematic use of multiple cancer treatments, should provide an important therapeutic resource in the future.

3. Approaches that avoid increase in robustness are another possibility. Since genetic heterogeneity is enhanced, at least in part, by somatic recombination, selectively inducing cell-cycle arrest in tumour cells can be an effective means of controlling the robustness. There is a theoretical possibility that such subtle control can be achieved by careful combination of multiple drugs that specifically perturb biochemical interactions. A computational study indicates that removal or attenuation of specific feedback loops involved in the cell cycle reduces robustness of the cell cycle against changes in rate constants (Morohashi et al. 2002). The challenge is to find appropriate combinations of drugs capable of inducing cell-cycle arrest in tumour cells only, not in other cells. While this approach uses a combination of multiple drugs, the hope is to find a set of drugs that can be administered at minimum dosage and toxicity. This approach results in dormancy of tumours, which has already been proposed (Takahashi and Nishioka 1995; Uhr et al. 1997). Several reports indicate that induced dormancy has been observed in the mouse (Holmgren et al. 1995; Murray 1995). These studies also report cases where tumour cell proliferation is offset by increases in apoptosis. However, this type of dormancy, which may be termed ‘pseudo dormancy,’ does not prevent an increase of cell proliferation and resulting heterogeneity; hence, robustness is not controlled. Genuine dormancy requires the induction of cell cycle arrest to be as selective as possible.
4. An approach to actively reducing intratumoral genetic heterogeneity, followed by therapy with molecular targeted drugs, may be a viable option. If we can design an initial therapy to impose a specific selection pressure on tumours, whereby only cells with specific genetic variations survive the therapy, then a reduction of genetic heterogeneity may be achieved. If a tumour cell population

became sufficiently homogeneous, a drug that specifically targets a certain molecule might have significant impact on the remaining population. An important point here is that the drugs used should not enhance mutation and chromosomal instability; otherwise, particularly after the initial therapy, heterogeneity may quickly increase, so that second-line therapy is ineffective. The drawback of this approach is that it does not eliminate the fundamental chromosomal instability that continues to generate tumour cells with diverse genetic backgrounds.

Alternatively, a method of enhancing chromosomal instability selectively in cells that already have unstable chromosomes could be employed. The question is whether or not such effects can be achieved with sufficient selectivity. A non-selective approach to increasing chromosomal instability has been proposed (Sole 2003), but it may promote instability in cells that are relatively stable, potentially leading to malignancy.

5. Another possible desideratum might be to retake control of those feedback loops that give rise to robustness in disease states. Since the robustness of a tumour is often caused by host-tumour feedback controls, robustness can be effectively countered by interrupting the feedback loops. One possible approach is to introduce a decoy virus that disrupts feedback control and hence the invasive mechanisms of the disease state. Such an approach is proposed in AIDS therapy, so that a conditionally replicating HIV-1 (crHIV-1) vector, which has only a *cis* region but not *trans*, is introduced (Dropulic et al. 1996; Weinberger et al. 2003). This decoy dominates the replication machinery, so that the HIV-1 virus is pushed into latency, instead of being eradicated. In the case of solid tumours, an interesting idea has been proposed of using a tumour-associated macrophage (TAM) as a delivery vehicle of the vector (Bingle et al. 2002; Owen et al. 2004). The TAM migrates into a solid tumour cluster and up-regulates HIF-1, which facilitates angiogenesis and metastasis. If TAM can be used to retake control, robustness may be well controlled, and self-extending symbiosis in cancer evolution may be aborted.
6. It is critically important that therapeutic interventions specifically target tumour cells, without damaging normal cells. A simple identification of disease-causing genes is not enough, since the same genes perform. Simply identifying disease causative genes is not enough, since the same genes perform important functions for the operation of normal cells. The differences between target and off-target cells have to be identified. The author's team created a novel biological assay method called genetic Tug-of-War (gTOW), which enabled us to measure the quantitative upper bound of genes that can be over-expressed without perturbing cellular functions such as proliferation (Moriya et al. 2006). This method, when extended to mammalian cell systems, may help us discover differences in susceptibility to perturbation of robustness between target and off-target cells. The same perturbation should ideally hit the point of fragility in target cells, but be tolerable for normal cells. As these differences between cells may arise out of the properties of networks, focusing on such differences may provide us with a broader margin of therapeutic windows than current dosage-based methods. At the same time, the use of multiple components to explore differences of robust-

ness has been proposed. Spread-spectrum control and the Long-tail approach (Sect. 17.7) have been proposed as mathematical constructs that may provide us the means to design drugs with large numbers of components (Kitano 2007).

17.6 An Appropriate Index of Treatment Efficacy

It is important to recognize that, in the light of cancer robustness theory, tumour mass reduction is not necessarily an appropriate index for judging therapeutic and drug efficacy. As discussed already, reduction of tumour mass does not mean that the proliferation potential of a tumour has been decreased. It merely means that a sub-population of tumour cells that respond to the therapy were eradicated, or significantly reduced. The problem is that the remaining tumour cells may be more malignant and aggressive, so that therapies for the relapsed tumour could be quite ineffective. This is particularly the case when drugs used to reduce tumour mass are toxic and potentially promote mutations and chromosomal instability in non-specific ways. They may even enhance malignancy by imposing selective pressures favouring resistant phenotypes and enhancing genetic diversity, as well as providing niches for growth by eradicating a more susceptible sub-population of tumour cells.

A proper index should rather be based on control of robustness: either to minimize an increase in robustness or to reduce it. This can be achieved by inducing dormancy, actively imposing selective pressure to reduce heterogeneity; or exposing fragility that can be the target of therapies to follow, thereby retaking control of feedback regulation. The outcome of controlling the robustness may vary from moderating the growth of the tumour, to dormancy where there is no growth in tumour mass, or even significant reduction, a possibility not excluded by robustness control. If we can target a point of fragility within the tumour, it may trigger a common mode failure and result in significant tumour mass reduction. However, that would be a consequence of controlling robustness, and should not be confused with therapies aimed at tumour mass reduction, because robustness has to be controlled first, in order to actively exploit a point of fragility.

However, this criterion poses a problem for drug design, because the current efficacy index by which anti-tumour drugs are measured is based on tumour mass reduction. Drugs that induce dormancy will not satisfy this efficacy criterion. Whether such an approach can be taken will depend on a perception change among practitioners, drug industries, and regulatory authorities.

17.7 Long-tail Drugs

One of the points highlighted in the above discussions is that the living body is a robust and evolving system, control of which may be ineffective if only one factor is identified and suppressed; indeed, in this way resulting in undesirable conse-

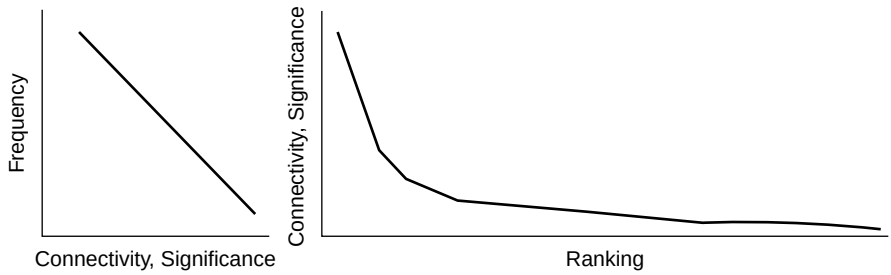


Fig. 17.5 Long-tail distribution, double and single logarithmic

quences. Actually, the genes involved in the initiation and maintenance of cancers are quite often the hubs of networks (Goh et al. 2007; Yao and Rzhetsky 2008). This means that, when the causative gene is set as a target of a drug, there is an inherent risk of severe side-effects. Statistical studies indicate that the average number of interactions involving targeted molecules of anti-cancer drugs is significantly higher than those of non-cancer drugs (Hase et al. 2009). This difference seems to reflect the higher degree of side-effects of anti-cancer medications compared to other drugs. To minimize these side-effects, it is necessary to adopt logical and strategic approaches during cancer treatment to avoid perturbation of major hubs in molecular interaction networks.

The number of interactions involving each molecule seems to exhibit a power law distribution. That is to say, their distribution indicates that only a few proteins have a large number of interacting partners, while a majority of proteins interact with only small numbers of other molecules. In Fig. 17.5, on a double logarithmic graph, the appearance frequency is plotted on the vertical axis and the number of interactions on the horizontal axis; whereas with a single logarithm, the number of interactions is plotted on the vertical axis and the appearance frequency on the horizontal axis. The shape of the distribution is a straight declining line on the double logarithmic chart; while on the single logarithmic chart (vertical axis: logarithm), it is a curved line with a very long tail in the right-ward direction, hence referred to as the 'long-tail distribution'.

In an analogy to illustrate implications for cancer treatment, we observe that this long-tail distribution is well-known in the world of Internet business, having the theoretical background of the profit structure of online shops. In a study conducted by a team from the MIT business school (Brynjolfsson et al. 2003) of Amazon.com, which sells more than 2.5 million book titles online, the book sales amount and ranking form a Long-tail distribution containing a few best sellers and a great many bad sellers. It turns out that half of the total sales came from the 40,000 best-selling book titles, and the other half from the remaining millions. In other words, the small-volume sellers have a big impact on total sales. Therefore, a concentration on best-sellers may result in missing the business opportunity provided by the thin but long tail.

What does this imply for the intracellular interactions in cancer that show the same statistical distribution pattern? Instead of using a powerful drug to suppress

a small number of key genes and proteins, might it be possible to apply weak suppression more broadly to the relatively large number of less important genes and proteins? This suggests that an effective synergy could be obtained by targeting multiple non-hub genes and proteins and intervening in them weakly rather than strongly using a medication we might call a '*long-tail drug*'. If its performance proved equal to its promise, such a therapy could be applied without great disturbances to the hub molecules, thus producing fewer side effects.

A group of Hungarian researchers (Agoston et al. 2005; Csermely et al. 2005) calculated, on the basis of the metabolism model of budding yeasts and bacteria, how much the metabolic rate decreases when the largest hub is eliminated. They then computed how many interactions are necessary for achieving a similarly decreased metabolic rate, using a method of decreasing the number of interactions by half. Their conclusion was that 10–50 interactions need to be suppressed. This demonstrates the theoretical possibility of producing a big impact on cells by combining weak interactions.

These results suggest the importance of the option of analysing network structures, and intervening in multiple targets with multiple chemical substances, for the development of new drugs, as well as for exploring combination therapy using existing drugs. The ultimate goal is a long-tail drug combination that generates the synergy of weak interactions with the simultaneous use of very many substances. The important targets, the so-called hubs, need to be avoided as far as possible, with the synergy effect being applied to targets that are not thought to be important, but exist abundantly in the long tail of the distribution of the number of interactions of proteins and genes (Kitano 2007).

Since there are enormous numbers of possibilities in the selection of target molecules and the combinations of drugs to be used in long-tail drug therapy, it is necessary to establish a technology for efficient selection of effective combinations producing the fewest side effects. The author considers that the approach of measuring differences of robustness against the altered expression of candidate genes and the altered activity of candidate proteins, between the target cells and the off-target cells, will become a workable option. Right now, the approach of using the budding yeast system is undergoing verification, and in the future, it will be possible to define an experimental system for mammalian cells, with ultimate application to humans.

17.8 Conclusion

In this chapter, we have discussed the basic ideas behind biological robustness, and its implications for cancer research and treatment. Robustness is one of the essential features of living systems, and is tightly coupled with evolution. It may also shape the basic architectural feature of biological systems, which are robust and evolving. We believe we have provided a useful overview, and presented a framework that may lead to more solid theories of biological robustness. One of the major neces-

sities for the development of treatment strategy is the identification of trade-offs between robustness, fragility, resource demands, and performance. Fragility is particularly relevant to diseases. At the same time, cancer establishes its own robustness as it evolves in a patient. Its success may be a result of hijacking the robustness intrinsic to the host system. Understanding this feature of the complex nature of biological systems may have profound implications for biomedical research in the future, inspiring useful and novel theoretically motivated strategies. Some of these ideas may not be practical, but if even one of them were realized, it could result in ground-breaking progress in cancer therapeutics.

Acknowledgements This research is, in part, supported by ERATO-SORST Program (Japan Science and Technology Agency: JST), the Genome Network Project (Ministry of Education, Culture, Sports, Science, and Technology), BBSRC-JST Strategic Collaboration Program of JST to the Systems Biology Institute.

References

Agoston V, Csermely P, Pongor S (2005) Multiple weak hits confuse complex systems: a transcriptional regulatory network as an example. Phys Rev E Stat Nonlin Soft Matter Phys 71:051909

Alon U, Surette MG, Barkai N, Leibler S (1999) Robustness in bacterial chemotaxis. Nature 397:168–171

Baisse B, Bouzourene H, Saraga EP, Bosman FT, Benhattar J (2001) Intratumour genetic heterogeneity in advanced human colorectal adenocarcinoma. Int J Cancer 93:346–352

Barkai N, Leibler S (1997) Robustness in simple biochemical networks. Nature 387:913–917

Bingle L, Brown NJ, Lewis CE (2002) The role of tumour-associated macrophages in tumour progression: implications for new anticancer therapies. J Pathol 196:254–265

Bissell MJ, Radisky D (2001) Putting tumours in context. Nat Rev Cancer 1:46–54

Bruchovsky N, Klotz LH, Sadar M, Crook JM, Hoffart D, Godwin L, Warkentin M, Gleave ME, Goldenberg SL (2000) Intermittent androgen suppression for prostate cancer: Canadian Prospective Trial and related observations. Mol Urol 4:191–199 discussion 201.

Brynjolfsson E, Smith MD, Hu Y (2003) Consumer surplus in the digital economy: estimating the value of increased product variety at online booksellers. MIT Sloan School of Management, Cambridge

Cannon W (1932) The wisdom of the body. Norton, New York

Carlson JM, Doyle J (1999) Highly optimized tolerance: a mechanism for power laws in designed systems. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics 60:1412–1427

Carlson JM, Doyle J (2002) Complexity and robustness. Proc Natl Acad Sci U S A 99(Suppl 1):2538–2545

Chen KC, Calzone L, Csikasz-Nagy A, Cross FR, Novak B, Tyson JJ (2004) Integrative analysis of cell cycle control in budding yeast. Mol Biol Cell 15:3841–3862

Crowe JH, Crowe LM (2000) Preservation of mammalian cells-learning nature's tricks. Nat Biotechnol 18:145–146

Csermely P, Agoston V, Pongor S (2005) The efficiency of multi-target drugs: the network approach might help drug design. Trends Pharmacol Sci 26:178–182

Csete ME, Doyle JC (2002) Reverse engineering of biological complexity. Science 295:1664–1669

Dropulic B, Hermankova M, Pitha PM (1996) A conditionally replicating HIV-1 vector interferes with wild-type HIV-1 replication and spread. Proc Natl Acad Sci U S A 93:11103–11108

Eldar A, Dorfman R, Weiss D, Ashe H, Shilo BZ, Barkai N (2002) Robustness of the BMP morphogen gradient in Drosophila embryonic patterning. Nature 419:304–308

Ferrell JE Jr (2002) Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. Curr Opin Cell Biol 14:140–148
Frigyesi A, Gisselsson D, Mitelman F, Hoglund M (2003) Power law distribution of chromosome aberrations in cancer. Cancer Res 63:7094–7097
Fujii H, Yoshida M, Gong ZX, Matsumoto T, Hamano Y, Fukunaga M, Hruban RH, Gabrielson E, Shirai T (2000) Frequent genetic heterogeneity in the clonal evolution of gynecological carcinosarcoma and its influence on phenotypic diversity. Cancer Res 60:114–120
Gleave ME, Sato N, Goldenberg SL, Stothers L, Bruchovsky N, Sullivan LD (1997) Neoadjuvant androgen withdrawal therapy decreases local recurrence rates following tumour excision in the Shionogi tumour model. J Urol 157:1727–1730
Gleave M, Bruchovsky N, Goldenberg SL, Rennie P (1998) Intermittent androgen suppression for prostate cancer: rationale and clinical experience. Eur Urol 34(Suppl 3):37–41
Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL (2007) The human disease network. Proc Natl Acad Sci U S A 104:8685–8690
Gonzalez-Garcia I, Sole RV, Costa J (2002) Metapopulation dynamics and spatial heterogeneity in cancer. Proc Natl Acad Sci U S A 99:13085–13089
Gorunova L, Hoglund M, Andren-Sandberg A, Dawiskiba S, Jin Y, Mitelman F, Johansson B (1998) Cytogenetic analysis of pancreatic carcinomas: intratumour heterogeneity and nonrandom pattern of chromosome aberrations. Genes Chromosomes Cancer 23:81–99
Gorunova L, Dawiskiba S, Andren-Sandberg A, Hoglund M, Johansson B (2001) Extensive cytogenetic heterogeneity in a benign retroperitoneal schwannoma. Cancer Genet Cytogenet 127:148–154
Harris AL (2002) Hypoxia—a key regulatory factor in tumour growth. Nat Rev Cancer 2:38–47
Hase T, Tanaka H, Suzuki Y, Nakagawa S, Kitano H (2009) Structure of protein interaction networks and their implications on drug design. PLoS Comput Biol 5:e1000550
Hochhaus A (2003) Cytogenetic and molecular mechanisms of resistance to imatinib. Semin Hematol 40:69–79
Hochhaus A, Kreil S, Corbin A, La Rosee P, Lahaye T, Berger U, Cross NC, Linkesch W, Druker BJ, Hehlmann R et al (2001) Roots of clinical resistance to STI-571 cancer therapy. Science 293:2163
Holmgren L, O'Reilly MS, Folkman J (1995) Dormancy of micrometastases: balanced proliferation and apoptosis in the presence of angiogenesis suppression. Nat Med 1:149–153
Juliano RL, Ling V (1976) A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants. Biochim Biophys Acta 455:152–162
Kitano H (2003) Cancer robustness: tumour tactics. Nature 426:125
Kitano H (2004a) Biological robustness. Nat Rev Genet 5:826–837
Kitano H (2004b) Cancer as a robust system: implications for anticancer therapy. Nat Rev Cancer 4:227–235
Kitano H (2007) A robustness-based approach to systems-oriented drug design. Nat Rev Drug Discov 6:202–210
Kitano H, Oda K (2006) Self-extending symbiosis: a mechanism for increasing robustness through evolution. Biol Theory 1:61–66
Lengauer C, Kinzler KW, Vogelstein B (1998) Genetic instabilities in human cancers. Nature 396:643–649
Li R, Sonik A, Stindl R, Rasnick D, Duesberg P (2000) Aneuploidy vs. gene mutation hypothesis of cancer: recent study claims mutation but is found to support aneuploidy. Proc Natl Acad Sci U S A 97:3236–3241
Meir E, Dassow G von, Munro E, Odell GM (2002) Robustness, flexibility, and the role of lateral inhibition in the neurogenic network. Curr Biol 12:778–786
Moriya H, Shimizu-Yoshida Y, Kitano H (2006) In vivo robustness analysis of cell division cycle genes in Saccharomyces cerevisiae. PLoS Genet 2:e111
Morohashi M, Winn AE, Borisuk MT, Bolouri H, Doyle J, Kitano H (2002) Robustness as a measure of plausibility in models of biochemical networks. J Theor Biol 216:19–30
Murray C (1995) Tumour dormancy: not so sleepy after all. Nat Med 1:117–118

Nooter K, Herweijer H (1991) Multidrug resistance (mdr) genes in human cancer. Br J Cancer 63:663–669
Ogle BM, Cascalho M, Platt JL (2005) Biological implications of cell fusion. Nat Rev Mol Cell Biol 6:567–575
Owen MR, Byrne HM, Lewis CE (2004) Mathematical modelling of the use of macrophages as vehicles for drug delivery to hypoxic tumour sites. J Theor Biol 226:377–391
Pawelek JM (2005) Tumour-cell fusion as a source of myeloid traits in cancer. Lancet Oncol 6:988–993
Pawelek J, Chakraborty A, Lazova R, Yilmaz Y, Cooper D, Brash D, Handerson T (2006) Co-opting macrophage traits in cancer progression: a consequence of tumour cell fusion? Contrib Microbiol 13:138–155
Queitsch C, Sangster TA, Lindquist S (2002) Hsp90 as a capacitor of phenotypic variation. Nature 417:618–624
Rasnick D (2002) Aneuploidy theory explains tumour formation, the absence of immune surveillance, and the failure of chemotherapy. Cancer Genet Cytogenet 136:66–72
Rutherford SL (2003) Between genotype and phenotype: protein chaperones and evolvability. Nat Rev Genet 4:263–274
Rutherford SL, Lindquist S (1998) Hsp90 as a capacitor for morphological evolution. Nature 396:336–342
Schlosser G, Wagner G (eds) (2004) Modularity in development and evolution. The University of Chicago Press, Chicago
Sharp FR, Bernaudin M (2004) HIF1 and oxygen sensing in the brain. Nat Rev Neurosci 5:437–448
Shimada T, Aihara K (2008) A nonlinear model with competition between prostate tumour cells and its application to intermittent androgen suppression therapy of prostate cancer. Math Biosci 214:134–139
Siegal ML, Bergman A (2002) Waddington's canalization revisited: developmental stability and evolution. Proc Natl Acad Sci U S A 99:10528–10532
Skipper HE, Hutchison DJ, Schabel FM Jr, Schmidt LH, Goldin A, Brockman RW, Venditti JM, Wodinsky I (1972) A quick reference chart on cross resistance between anticancer patients. Cancer Chemother Rep 56:493–498
Sole RV (2003) Phase transitions in unstable cancer cell populations. Eur Phys J B 117–123
Takahashi Y, Nishioka K (1995) Survival without tumour shrinkage: re-evaluation of survival gain by cytostatic effect of chemotherapy. J Natl Cancer Inst 87:1262–1263
Tischfield JA, Shao C (2003) Somatic recombination redux. Nat Genet 33:5–6
Tsuruo T, Iida H, Tsukagoshi S, Sakurai Y (1981) Overcoming of vincristine resistance in P388 leukemia in vivo and in vitro through enhanced cytotoxicity of vincristine and vinblastine by verapamil. Cancer Res 41:1967–1972
Tyson JJ, Chen K, Novak B (2001) Network dynamics and cell physiology. Nat Rev Mol Cell Biol 2:908–916
Uhr JW, Scheuermann RH, Street NE, Vitetta ES (1997) Cancer dormancy: opportunities for new therapeutic approaches. Nat Med 3:505–509
Vignery A (2005) Macrophage fusion: are somatic and cancer cells possible partners? Trends Cell Biol 15:188–193
von Dassow G, Meir E, Munro EM, Odell GM (2000) The segment polarity network is a robust developmental module. Nature 406:188–192
Waddington CH (1957) The strategy of the genes: a discussion of some aspects of theoretical biology. Macmillan, New York
Weinberger LS, Schaffer DV, Arkin AP (2003) Theoretical design of a gene therapy to prevent AIDS but not human immunodeficiency virus type 1 infection. J Virol 77:10028–10036
Yao L, Rzhetsky A (2008) Quantitative systems-level determinants of human genes targeted by successful drugs. Genome Res 18:206–213
Yi TM, Huang Y, Simon MI, Doyle J (2000) Robust perfect adaptation in bacterial chemotaxis through integral feedback control. Proc Natl Acad Sci U S A 97:4649–4653

Part V
Perspectives and Conclusions

Chapter 18
Synthetic Biology and Perspectives

Toru Yao and Frederick B. Marcus

Abstract This chapter describes synthetic biology and associated technologies in the context of establishing perspectives for future systems approaches to cancer. We focus on long term advances made possible by new technologies that operate at the molecular, cellular, and microorganism levels. World trends in synthetic biology are described, plus examples of a highly integrated program in Japan. Detailed possibilities for application to cancer research and clinical applications are explored. These promising technological and research developments establish a basis for translation to clinical applications, allowing us to summarize and evaluate future perspectives and their infrastructure requirements elaborated in workshops and related publications, including those organized by US/NCI-EU-Germany-Japan, BBSRC/UK, and JST/Japan collaborations.

18.1 Introduction

Perspectives provide a view of the current state of the field and an outlook for the future. In this chapter, feasibility in the near and medium term, mostly the next 5 years, is presented by emphasizing prerequisites for progress. Advances do not happen automatically, they require preparation, infrastructure and supporting research, and an integrated vision of cancer systems approaches involving new research paradigms.

In particular, advances involving systems computer model development and verification rely heavily on new methods of gathering data and manipulating organisms to test new theories. The emerging discipline of synthetic biology provides many such tools, especially when joined by advances in other disciplines such as nanotechnology, developmental biology and immunology. These tools enable the exploration of systems level areas such as applied evolution, robustness, and chronotherapy. For example, the integration of the circadian clock into algorithms for cancer therapies (Altinok et al. 2007a, b, 2009; see chapter 15) may prove to be crucial for optimal chronotherapeutic delivery of anti-cancer drugs. The inclusion into models of relevant clinical factors such as circadian biomarkers and gender

T. Yao (✉)
Genomic Sciences Center, RIKEN, Yokohama (Major), Japan
e-mail: yao@riken.jp

A. Cesario, F. B. Marcus (eds.), *Cancer Systems Biology, Bioinformatics and Medicine,*
DOI 10.1007/978-94-007-1567-7_18,

constitutes an innovative and promising approach for personalized cancer medicine. Lévi and Schibler (2007) further discuss the therapeutic implications of taking account of circadian rhythms, providing an excellent demonstration of the potential advantages of a systems approach which integrates organism-level activities.

Developing the foundations for advances in systems biology and medicine depends on a multi-faceted approach, involving a mixture of steady technological advances, construction of appropriate research infrastructures, choice of model organisms and patient data, and careful development and continuous validation of computer models. In addition to this steady progress on a broad front, paradigm-changing areas such as systems-based synthetic biology are already offering new methods of inducing perturbations in model systems, as well as altering the system itself. We unify this picture (Sect. 18.4) by describing and evaluating a series of workshops, their resulting publications, and 'Grand Challenges' that develop roadmaps for future progress.

18.2 Synthetic Biology for Cancer Research and Applications

18.2.1 Introduction to Synthetic Biology

The relatively new research domain of synthetic biology combines science and engineering to design and construct biological functions and systems not found in nature, mainly achieved by manipulations at the molecular level. It is an emerging and rapidly expanding field applying powerful tools and technologies to various areas of biology, including basic science, healthcare, food, materials, energy, and environment (NEST 2005; Serrano 2007; Parliamentary Office 2008; Purnick and Weiss 2009).

Synthetic biology has a wide-ranging scope of applications, covering molecular, cellular, tissue, organ and organism levels. Initial molecular-level protein engineering research (Ulmer 1983) subsequently led to cell factory projects being set up in several countries (Cell Factory 2009). Synthetic biology now plays a major role in systems approaches to biology and medicine, since tools enabling the modification and rational design of molecules and cells provide a means of model verification by applying precise system perturbations. Molecular manipulation and engineering offers great promise for the development of new drugs and treatment methods.

Applications of synthetic biology to cancer are especially effective, since the disease itself functions in rather similar ways. Cancer, like science, manipulates the genetic code; modifies existing bodily systems such as vascularisation for its own purposes; provokes accelerated evolution; and creates important modifications to cellular properties. Hence, synthetic biology techniques are highly appropriate in four key areas applicable to cancer: fundamental biology; inspection and diagnostics; drug development; and treatment and therapies.

Synthetic biology can especially enhance understanding of biological systems when integrated with systems biology numerical modelling approaches. Throughout the world, intensive activities are under way for the development of synthetic biology methodologies and applications to various fields. These methodologies will strongly influence the evolution of cancer diagnostics, therapies and drug discoveries during the next decade.

There are many other current activities within synthetic biology not specifically targeted at cancer applications; some of these, however, will certainly contribute. Synthetic biologists with an interest in cancer include Frances Arnold, Jonathan Weissman, Elowitz, David Tirrell, Wendall Lim, Jim Collins, and Bonnie Bassler. The difficult aspects of Bioethics and Law are under active discussion (Alta Char, Jonathan Moreno, and Laurie Zoloth).

Every year several conferences or symposiums on synthetic biology are organized. In addition, there is significant funding activity promoting the field of synthetic biology in the USA, Europe, and Japan, among other countries. Synthetic biology is undergoing rapid growth, and holds huge promise for future contributions to the area of cancer (Parliamentary Office 2008; Haseloff and Ajioka 2009; Purnick and Weiss 2009).

18.2.2 Manipulation at the Molecular Level

Pioneers in synthetic biology recognized that an infrastructure of standard parts would be required for industrial-scale engineering of biological systems. Tom Knight at MIT proposed such a biological infrastructure for the design of artificial systems (Knight 2003). In 2005, Drew Endy, Kim Knight and others organized BioBrick, a collection of biologically functional DNA sequence parts, and established the BioBrick Foundation for the registry of Standard Biological Parts (http://bbf.openwetware.org), later extended to BioBrick vectors (Shetty et al. 2008). The BioBrick collection comprises more than 3000 biological parts, and is continually increasing. The same researchers established in 2004 an annual competition in synthetic biology, the International Genetically Engineered Machine (iGEM); the number of participants grows with every year. About 100 teams from around the world gathered at MIT in 2009 (http://2009.igem.org). An interesting result from the Stanford team concerned the design of Immuni-T. Coli, offering a probiotic approach to diagnosing and treating inflammatory bowel disease (http://2009.igem.org). The J. Peccoud group at Virginia Bioinformatics Institute, like Biobrick, is collecting biological parts for targeted purposes, as well as developing other approaches to genetic design (Peccoud et al. 2008; Goler et al. 2008).

Major advances have recently been made in genetic engineering. The Serrano group at the Center for Gene Regulation in Barcelona has developed new synthetic biology approaches involving the engineering of integrated genetic, network, and developmental systems (Serrano 2007; http://serrano.crg.es). G. Church's group at Harvard University has produced a high-throughput method for genome engineer-

ing (Wang et al. 2009); a synthetic gene network (Friedland et al. 2009); and the integration of RNA and protein synthesis (Jewett and Church 2011). R. Weiss's group at Princeton University is exploring a method for synthetic gene networks (Batt et al. 2007). Saito and Inoue (2007, 2009) investigated the role of RNA in synthetic biology applications.

Protein engineering has a long history dating back more than 20 years and has led to major progress, especially recently after its reinforcement with the new discipline of synthetic biology. Papapostolou and Howorka (2009) have described engineering and exploiting protein assemblies using synthetic biology methods. Cao and his colleagues (2009) studied the evolution, complex structures and function of septin proteins. Semenza (2008) investigated the hypoxia-inducible factor 1 (HIF-1), a transcriptional activator that mediates adaptive responses to hypoxia. Bloom and Arnold (2009) employed directed evolution to investigate pathways of adaptive protein evolution, and Prabhaker's team (2006) investigated accelerated evolution of conserved non-coding sequences in humans.

The related fields of nanotechnology (Heath et al. 2009) and nanomedicine are currently opening up fresh avenues for cancer detection and treatment. Technologies of miniaturization are offering new possibilities for detecting extremely low biomarker levels and conducting analyses at the cellular or sub-cellular level.

18.2.3 Applications in Cells

Several approaches are being developed for changing or modifying existing cells, leading to exciting new discoveries. A particularly interesting result was a method for making induced pluripotent stem cells (iPS) by introducing transcriptional genes into somatic cells such as skin cells (Takahashi and Yamanaka 2006). This work was subsequently followed up in studies by the same and other researchers (Takahashi et al. 2007; Yu *et al.* 2007; Zhao and Daley 2008; Maherali and Hochedilinger 2008; Nirmalanandhan and Sittampalam 2009). However, the mechanisms of these modifications need to be more clearly elucidated, and systems approaches will be required for these advances. C. Venter's group (http://www.jcvi.org) developed a genome engineering method by chemical synthesis of genome sequences (Gibson et al. 2008; Lartigue et al. 2009), and established a company for the applications of that technology to areas such as biofuels (http://www.syntheticgenomics.com). They have recently engineered and inserted an entire artificial chromosome (similar to the original) into a bacterium, over-publicized as creation of 'artificial life'.

Synthetic biology at the microbial level is being used to develop a whole range of technologies and drug development applications (Schwille and Diez 2009; Weber and Fussenegger 2009; Prather and Martin 2008; Brenner et al. 2008; Babu 2008; Williams and Haque 1997; Anderson et al. 2006, 2007; Wang et al. 2009). J. Keasling's group at Berkeley succeeded in modifying *E. coli* for efficient production of ar-

temisinin by engineering a mevalonate pathway (Martin et al. 2003). They followed this up by studying applications to biofuels (Keasling and Chou 2008), and established a new company, Amyris Biotechnologies (http://www.amyrisbiotech.com).

18.2.4 Synthetic Biology in Japan

The history of synthetic biology applications in Japan is an excellent example of an integrated development of the field and its applications. In the mid 1980s, protein engineering was very promising for the design of new artificial proteins. Japan established the Protein Engineering Research Institute (PERI) in 1986 as a large national project, with T. Yao as the research director of the computational analysis group. The S. Yokoyama group developed an efficient method of producing membrane proteins *in vitro* (Shimono et al. 2009). In the 1990s, evolutionary engineering became a popular method for generating functional biomolecules or microorganisms by accelerated evolution *in vitro*. The New Energy and Industrial Technology Development Organization (NEDO) in Japan initiated a project for this purpose in 1995 for a period of 7 years.

A new stage of synthetic biology was entered with genome-wide experimental and computational approaches. Several research groups have associated themselves with these developments. M. Itaya presented the concept of 'genome engineering' (Itaya 1995), going on to develop a unique method of connecting parts from different genomes to construct a new genome with several altered functions (Itaya et al. 2008). In November 2007, the forum 'Making Artificial Cells' was initiated by D. Kiga, H. Ueda, H. Iwasaki and others (http://jscsr.org). Kiga is pursuing the generation of artificial cells through a new concept involving a DNA computer (Ayukawa and Kiga 2007). K. Shiba's team at the Japanese Foundation for Cancer Research in Tokyo has produced an artificial protein for cancer cells using Mol-Craft, a novel type of *in vitro* protein evolution system also developed by the group (Saito et al. 2007). T. Fujio's group at Biofrontier Laboratories in Tokyo developed an *E. coli* minimum genome factory in a NEDO project (Mizoguchi et al. 2007).

One of the greatest discoveries to have been made in cellular level engineering was the work of the Yamanaka Group, which produced iPS cells from somatic cells. This method is applicable to tumour-initiating cells, which are among the most important targets for anticancer drug discovery (Zhou et al. 2009). Major advances depend on future research being aimed at gaining a detailed understanding of the mechanisms of iPS cells.

In order to fully exploit the wealth of possible applications of synthetic biology approaches, a very fundamental understanding of living systems is undoubtedly needed before major advances can be secured. Professor Akiyoshi Wada in 2003 organized a challenging workshop in Tokyo around his concept of 'Strategy of Life' (Yao 2003), in which it was emphasized that synthetic biology applications should take account of two main aspects, natural and man-made design (Wada

2009). The latter term is self-evident; the term 'natural design' refers to the fact that all existing organisms are the result of an evolutionary process involving natural mutations and selection over billions of years. A better understanding of evolution, a central concept in cancer progression, can be achieved by mimicking it by means of accelerated evolutionary engineering (AEE). The AEE approach imparts a certain environmental constraint to the targeted proteins or micro-organisms, and makes it possible for the adaptation generation cycles to be observed occurring thousands or million times faster than in natural cycles. The success story that is AEE has resulted in the generation of many useful proteins, cells and microorganisms (Wang et al. 2009).

18.3 Synthetic Biology Applications to Cancer

The four key areas of application to cancer, all amenable to techniques of synthetic biology and associated nanotechnologies, are biology, diagnostics, drug development, and therapy, with the latter divided into gene/protein- and immunology-related therapy.

18.3.1 Cancer Biology

Synthetic biology constitutes a serviceable tool for the investigation of fundamental cancerous processes. Various designed and synthetically manufactured proteins have proved useful in enhancing our understanding of cancer biology, diagnosis, therapy, and drug development. Saito's team (2008) employed a motif-programmed artificial protein to induce apoptosis in cancer cells by disrupting mitochondria. Engineered proteins were also used by the same researchers to investigate apoptosis in cancer cells (Saito et al. 2008) and pathway changes (Saito et al. 2007; Bloom and Arnold 2009). A particularly exciting recent development is the technology of cell-free synthesis of membrane proteins (Shimono et al. 2009). This method is applicable to the production of GPCR proteins for cancer biology, among other fields. In the majority of human cancers, HIF-1 activity is increased as a result of genetic alterations and intratumoral hypoxia. Inhibition of HIF-1 activity may therefore turn out to have therapeutic effects, especially in combination with other anticancer drugs.

Various methods have been discussed for changing or engineering cancer or cancer-related cells (Anderson et al. 2006; Chan and Giaccia 2008; Mizauri et al. 2008; Papait et al. 2009). Other researchers have described a variety of cell-related technologies and approaches (Zhang et al. 2009; Webb et al. 2009; Heng et al. 2009; Papait et al. 2009; Jorgensen 2009; Mizuari et al. 2008; Smailus et al. 2007; Le Meur and Gentleman 2008; Saito et al. 2007; Robins et al. 2008).

18.3.2 Diagnostics

Very small colonies of cancer cells will hopefully soon be detectable, even at a very early stage, by their secretions of proteins and other substances into the urine or blood, these constituting early-stage biomarkers (see Chap. 13). Systems approaches will become essential for predicting the appearance and interpretation of biomarker concentrations, given that false positives are a major problem (Harris 2005).

Wu and Olafsen (2008) discuss the use of antibodies for molecular imaging of cancer. Antibodies have attained a central role as targeted therapeutics; several significant drugs are already on the market, with many more in clinical development for oncological applications. Expansion of the role of antibodies in cancer imaging has been accelerated by a number of factors, including the recognition that antibodies can provide a powerful class of molecular imaging probes for interrogating cell surfaces *in vivo*. Identification of relevant cell-surface biomarkers as imaging targets, coupled with advances in antibody technology, together facilitate the generation of antibodies optimized for non-invasive imaging. Developments in imaging instrumentation and radionuclide availability have paved the way for broader evaluation and implementation of radioimmunoscintigraphy and immunoPET.

18.3.3 Drug Development

Advances in biotechnology have now created a capacity to produce therapeutically active proteins on a commercial scale, opening the potential for their application in an array of disease conditions. Delivery technologies are an important new area in drug development. Stolnik and Shakesheff (2009) reviewed formulations for delivery of therapeutic proteins. Yang's (2009) group examined ways of formulating protein drugs into particulates feasible for practical pharmaceutical dosage, allowing inhalation and sustained-release delivery. Shoyele (2008) studied the engineering of protein particles for pulmonary drug delivery. Advanced techniques such as spray-drying and spray freeze-drying, and supercritical fluid technology, have been developed to produce particles in a size range and morphology suitable for deep long deposition, without altering the native conformation of these biomolecules.

Quaglia (2008) exploits tissue engineering methodologies, which hold great promise for protein delivery. Other applications include incorporating proteins into tissue-engineering scaffolds and medical devices, and targeting protein therapeutics in an *in vivo* environment. The concept of developing a tissue either *in vitro* or *in vivo* by analogy with natural physiological events has prompted the integration of molecular signals such as growth factors (GFs), with the aim of guiding cell proliferation, differentiation and migration; this is clearly of high relevance to cancer applications. Moss's team (2009) have discussed the rational design and protein engineering of growth factors for regenerative medicine and tissue engineering, while

Schmidt's group (2009) have studied protein engineering of carboxyl esterases by rational design and directed evolution.

Platis and Labrou (2008) described chemical and genetic engineering strategies for improving the potency of pharmaceutical proteins and enzymes. Even though proteins are important therapeutic agents, their use presents a number of disadvantages in comparison with small-molecule drugs: immunogenicity and antigenicity; poor efficacy and oral bioavailability; and, in many cases, short serum half-lives. Genetic engineering and site-specific chemical synthesis/modification techniques offer the best promise for improved protein therapeutics. The potency of protein drugs is reinforced by employing modern recombinant DNA technologies and novel chemical synthesis techniques. Ioerger and Sacchettini (2007) examined structural genomics approaches to drug discovery.

Designed ankyrin repeat proteins (DARPins) are a novel class of binding molecules with the potential of overcoming the limitations of monoclonal antibodies, thereby facilitating novel therapeutic approaches (Stumpp et al. 2008). DARPins are small, single domain proteins (14 kDa) which can be selected to bind any given target protein with high affinity and specificity. These characteristics make them ideal agonistic, antagonistic or inhibitory drug candidates; hence, they are a prominent member of the next generation of protein therapeutics with the potential to surpass the performance of existing antibody drugs.

18.3.4 Gene/Protein Therapy

Vázquez and colleagues (2009) employed modular protein engineering in innovative cancer therapies based on targeted nanoconjugates, designed to be specifically directed against target cells. These constructs, although suitable as vectors of conventional chemical drugs, are especially appropriate for delivering novel biopharmaceuticals such as expressible or antisense DNA molecules, silencing RNAs, or functional proteins. Natural or modified proteins or short peptides offer appropriate tools for functionalizing vehicles for targeted drug delivery. Procedures consist of identifying, obtaining and engineering functional and multifunctional polypeptides for target drug delivery and for potential applications of such constructs in emerging cancer therapies. Certain molecular traits in the biology of cancer are critical for the identification and selection of suitable targets for protein-based drug delivery. Engineered affinity proteins offer a wide range of applications for tumour targeting (Friedman and Ståhl 2009).

Jones et al. (2008) and colleagues reviewed recent advances and accomplishments in the use of rational and combinatorial protein engineering approaches to developing ligands and receptors as agonists and antagonists against clinically important targets. Ligand-receptor interactions govern myriad cell signalling pathways that regulate homeostasis and ensure that cells respond properly to stimuli. Growth factors, cytokines and other regulatory elements use these interactions to mediate the cell responses of proliferation, migration, angiogenesis, immune reactions, and

cell death. Proteins that inhibit these processes bear great potential as therapeutics for cancer and autoimmune disorders.

Vucic and Fairbrother (2007) discussed the role of inhibition of apoptosis (IAP) proteins in cancer, as well as options for targeting IAP proteins for therapeutic intervention. IAP proteins are expressed in the majority of human malignancies at elevated levels, and play an active role in promoting tumour maintenance, both by inhibiting cellular death and by participating in signalling pathways associated with malignancies.

18.3.5 Immunotherapy

Major advances have been made in enlisting the aid of the immune system to destroy cancerous cells and tumours, and in developing a possible cancer 'vaccination' to fight existing tumours (De Duve Institute 2010). Work is in progress on systems approaches to T-cell activation (Sybilla 2010), a complex process relying on multiple layers of tightly controlled intracellular signalling molecules forming an intricate and dynamic network. Defects in this network can cause autoimmune responses that destroy normal body cells. In order to understand and predict the behaviour of this network, it is crucial to study it as a complete system and not as isolated parts. The Sybilla project aims at a systems-level understanding of how T-cells discriminate foreign- from self-peptides by activating quantitatively distinct signalling pathways, and by developing new analytical and mathematical tools to generate and integrate high-density quantitative data. Proteomics, transcriptomics, imaging and biochemical techniques are being applied to obtain holistic maps of the T-cell signalling network and to achieve a quantitative and dynamic understanding of signalling networks and their regulation in response to different signal inputs. Applications include the identification of new drug targets and the discovery of new biomarkers to refine prognosis of autoimmune diseases.

Michielin (2007) studied the application of molecular modelling to new therapeutic cancer approaches. Recent progress in the experimental determination of protein structures aids in understanding molecular recognition mechanisms. This level of understanding makes it possible to design rational therapeutic approaches, in which effectors molecules are adapted or created *de novo* to perform a given function. An example is small inhibitory molecules, which are designed using *in silico* simulations and tested *in vitro*. Killer T lymphocyte receptors are being made more efficient against melanoma cells.

Friedman and Ståhl (2009) examine the use of engineered affinity proteins for tumour-targeting applications. Targeting of tumour-associated antigens is an expanding treatment modality in clinical oncology as an alternative to, or in combination with, conventional treatments, such as chemotherapy, external-radiation therapy and surgery. Targeting of antigens that are unique or more highly expressed in tumours than in normal tissues can be used to increase the specificity and reduce the cytotoxic effect on normal tissues. Affinity proteins in tumour-targeted therapy

can affect tumour progression by altering signal transduction or by delivering a payload of toxins, drugs or radionuclides.

Advances in genetic engineering and *in vitro* selection technology have enabled high-throughput generation of monoclonal antibodies (mAbs), i.e. antibody derivatives. In recent decades, monoclonal antibodieshave emerged as therapeutics. Chames and Baty (2009) review bi-specific antibodies for cancer therapy. Nine mAbs have been approved for cancer therapy. However, the efficiency of mAbs is far from optimal, and antibody engineering is actively used to improve the molecules. Because of their ability to simultaneously bind two different antigens, bi-specific antibodies are unique, and their wide potential as targeting reagents has been demonstrated over the years. Presta (2008) studied molecular engineering and design of therapeutic antibodies. Monoclonal antibody therapeutics can be engineered to improve their efficacy for enhanced effector functions; control of half-life; tumour and tissue accessibility; augmented biophysical characteristics, such as stability; and more efficient (and less costly) production. Significant progress has been made in designing antibodies with improved pharmacokinetic properties, using modified interaction with the neonatal Fc receptor (FcR). The ability to alter the communication of a therapeutic antibody with the immune system has been enhanced, by manipulation of the immunoglobulin protein sequence and its glycosylation.

In conclusion, the capabilities provided by synthetic biology approaches, especially when integrated with systems approaches, will continue to provide important advances in cancer understanding, diagnostics and therapies even more rapidly in the coming decade.

18.4 Review Articles and Workshops—Integrated Perspectives

With the technological and research perspectives provided by synthetic biology and the other areas covered in this book, integrated visions developed in workshops of key desiderata for systems biology and medicine may now be assessed.

Previously, cancer research was partly driven by ‘grand challenges’ of ‘eliminating cancer’ and ‘curing cancer’, in the spirit of the successful eradication of smallpox. The challenge presented by cancer is however more difficult. Smallpox, as an external pathogen transmitted from person to person was susceptible to successful elimination by vaccine development, whereas cancer and cancerous processes are intimately linked to our fundamental nature as living, growing and ageing multicellular organisms. DNA damage and repair mechanisms at the genetic level, apoptosis at the cellular level, and developmental processes at the multi-cellular level, tend to favour the development of cancer. While ‘eliminating cancer’ may not be a realistic goal, major advances in understanding, classifying, detecting, treating, and reducing the burden of the disease may nevertheless be achieved through systems-based approaches.

The conference on The Future Challenges of Systems Biology (FCSB 2008) produced the Tokyo Declaration: 'Recent advances in Systems Biology indicate that the time is now ripe to initiate a grand challenge project to create over the next 30 years a comprehensive, molecules-based, multi-scale, computational model of the human ("the virtual human"), capable of simulating and predicting, with a reasonable degree of accuracy, the consequences of most of the perturbations that are relevant to healthcare'. The Japanese JST and British BBSRC sponsors proposed a collaborative programme to drive this project forward.

Whatever the final outcome of such a highly ambitious challenge, near-term advances in much more limited sub-systems modelling will require continuous model verification, to be provided by close interaction with experimentation and clinical results. A near-term 'grand challenge' for cancer might be to achieve a steady development of verified models at appropriate physiological levels, with successful application to patient outcomes.

An important precondition for implementation of such a challenge is to persuade cancer researchers and clinicians that systems approaches are indeed relevant. Procedures (excluding modelling) for cancer research and clinical applications have already received extensive attention. Abeloff's Clinical Oncology (Abeloff et al. 2008) describes basic science, pathology, diagnosis, management, outcomes, rehabilitation, and prevention; Weinberg's Biology of Cancer (Weinberg 2007) covers the subject with a highly systematic approach; and papers in research areas such as the p53 'cell death' gene amount to over 56,000 publications.

Weinberg (2007) writes: '… Successes in these efforts, involving the new discipline of 'systems biology,' will surely benefit cancer research…. For now, at least, we need to wrestle with the grim realities of drug development, the inadequate animal models, our ignorance of the behaviour of cellular regulatory circuitry, and the confounding biological complexities of human cancer.' Acceptance of new paradigms requires proof of fresh understanding based on modelling, analysis results and their translation to the clinic. Our Part IV highlights the progress that has been made and what can be expected.

18.4.1 How Systems Biology Can Advance Cancer Research

A series of workshops were conducted to explore systems approaches to cancer. Integrated perspectives were developed at the first European Union (EU) and American National Cancer Institute (NCI) workshop on cancer systems biology (Aebersold et al. 2009). The main conclusion of the workshop participants was that 'systems biology approaches can indeed advance cancer research, having already proved successful in a very wide variety of cancer-related areas, and are likely to prove superior to many current research strategies.' They summarize the foundations for establishing these perspectives and state the following strategies to follow:

- Systems biology and computational approaches can make important contributions to research and development in key clinical aspects of cancer and of cancer

treatment, and should be developed for understanding and application to diagnosis, biomarkers, cancer progression, drug development and treatment strategies.
- Development of new measurement technologies is central to successful systems approaches, and should be strongly encouraged. The systems view of disease combined with these new technologies and novel computational tools will, over the next 5–20 years, lead to medicine that is predictive, personalized, preventive and participatory (P4 medicine).
- Major initiatives are in progress to gather extremely wide ranges of data for both somatic and germ-line genetic variations, as well as gene, transcript, protein and metabolite expression profiles that are cancer-relevant. Electronic databases and repositories play a central role to store and analyse this data. These resources need to be developed and sustained.
- Understanding cellular pathways is crucial in cancer research, and these pathways need to be considered in the context of the progression of cancer at various stages. At all stages of cancer progression, major areas require modelling via systems and developmental biology methods including immune system reactions, angiogenesis and tumour progression.
- A number of mathematical models of an analytical or computational nature have been developed that can give detailed insights into the dynamics of cancer-relevant systems. These models should be further integrated across multiple levels of biological organization in conjunction with analysis of laboratory and clinical data.
- Biomarkers represent major tools in determining the presence of cancer, its progression and the responses to treatments. There is a need for sets of high-quality annotated clinical samples, enabling comparisons across different diseases and the quantitative simulation of major pathways leading to biomarker development and analysis of drug effects.
- Education is recognized as a key component in the success of any systems biology programme, especially for applications to cancer research. It is recognized that a balance needs to be found between the need to be interdisciplinary and the necessity of having extensive specialist knowledge in particular areas.
- One or more types of cancer should be investigated over the full scale of its progression, for example glioblastoma or colon cancer, requiring extensive experimental and computational tools for generating and analysing quantitative data from several biological levels, in order to understand, detect and treat cancerous processes.

18.4.2 Cancer Systems Biology—2nd Workshop

The second workshop (Wolkenhauer et al. 2009) further concludes that ‘advancing biomedical applications through systems biology approaches requires the development of new theoretical methodologies, such as novel techniques for databases, system identification, concepts for the design of experiments, good methods for

hypothesis testing, frameworks to couple processes occurring at and across different spatial and temporal scales, and effective algorithms to solve problems of computational and non-linear complexity.' The participants expand the conclusions of the first workshop to note that other cancers, such as prostate, breast and lung, would also benefit from an integrated systems approach. Despite the complexity of the various stages of these cancers, there are a number of common areas with experimentally developed and validated mathematical models, including of cellular and multiscale pathways involved in tumorigenesis (e.g. ERK-, Wnt-, TGF-β signalling) and tumour progression. The report also summarizes further foundation areas requiring development:

- Experimental systems: Basic cellular processes still need further study and quantification, often via model organisms. In many cases, it is preferable to have high quality and consistent data from artificial cell lines. However, cell lines being developed from patients and differentiated iPS cells are perhaps more realistic, but much more variable. Genetically modified animals provide important additional possibilities.
- Linking basic research with clinical research: In developing knowledge and testing procedures in model systems, the driver of research should be the clinical situation/question, while the animal and cellular models are required for validation and verification. In particular, host/tumour relationships need to be taken into account. It is possible to have access to early stages of carcinogenesis, involving blood sampling for metabolomics, proteomics, and transcriptomics studies and access to tissues from biopsies.
- Transcriptomics and epigenomics: A new trend in the analysis in cancer susceptibility and progress of cancer, especially as related to ageing, is linked to epigenetic and transcription effects, in particular in relation to histones and methylation. New technologies in this area will result in key analyses.
- Proteomics: For biomarker proteomics and systems biology, however, absolute quantification rather than relative quantification is required, and a number of new technologies are being developed and/or expanded: Western blotting, Xmap technology, QconCAT, nanostring.
- Metabolomics and fluxomics: The identification of metabolite biomarkers is essential to research in cancer systems biology. Fairly advanced technologies, including gas chromatography, mass spectroscopy, nuclear magnetic resonance and combinations thereof, allow for extremely sensitive analysis of metabolites. Metabolomic data complement genetic and proteome information, provide valuable biomarkers, and as a link between genotype and phenotype, help to refine models.

New modelling perspectives therefore require the integration of:

- Specific systems (cell cycle, MAP kinases, etc)
- Technologies: the nano-world and single-molecule biochemistry/biophysics
- Multiscale modelling from genes to cell activity
- Individualized medicine

- Identifiability and distinguishability for parameter and model estimation
- Links between different cell/tissue/organ level systems.

In comparing the propositions and perspectives developed in these workshops, with the progress discussed in this book, it becomes evident that a large amount of the technology and research is in place, and progress towards the goals expressed is realistic.

18.4.3 Systems Medicine: The Future of Medical Genomics and Healthcare

Auffray et al. (2009) and colleagues have described the combined advances in genetics, systems biology, modelling, and application to medical problems. Together with extensive progress in genetics (ICGC 2009) and genomics, the increased efficiency of DNA sequencing opens up the possibility of analysing a large number of individual genomes and transcriptomes. The provision of complete reference proteomes and metabolomes is within reach, using powerful analytical techniques based on chromatography, mass spectrometry and nuclear magnetic resonance. Computational and mathematical tools have enabled the development of systems approaches for deciphering the functional and regulatory networks underlying the behaviour of complex biological systems. The science of medical genomics has attempted to overcome the initial limitations of genome-wide association studies and has identified a limited number of susceptibility loci for many complex and common diseases. Iterative systems approaches are starting to provide deeper insights into the mechanisms of human diseases, and to facilitate the development of better diagnostic and prognostic biomarkers for cancer and many other diseases. Systems approaches will transform the way drugs are developed through academic-industry partnerships, as for example those implemented in the European Commission's Innovative Medicine Initiative (IMI—http://www.imi.europe.eu), drugs designed to target multiple components of networks and pathways perturbed in diseases. Through systems-based approaches, medicine will be enabled to become predictive, personalized, preventive and participatory (the '4 Ps' paradigm). The Sysbiomed (2007) project, for example, has identified a number of key perspectives for cancer through its workshop and summary reports (Wolkenhauer et al. 2009) as well as a useful strategy paper (Makarow et al. 2008).

The second TACB (2007) conference (TACB CONF 2007) expanded on these themes and associated challenges therein. Advances in cancer drug development will require:

- Identification of critical targets within oncological/signalling pathways for therapeutic intervention.
- Analysis and prediction of cellular responses to anti-tumour agents, providing clues to mechanisms behind drug efficiency in experimental and cell culture models.

- Understanding the evolution of individual cancers from initial cellular degeneration to the aggressive metastasis, leading to the discovery of novel targets and the insights needed for improving therapeutic regimes.
- Improved prevention lowering the burden of the disease.

Environmental factors are often involved in cancer onset, as discussed in the EC-US Workshop on Virtual Tissues (Kolar and Francis 2009). Two Environmental Protection Agency projects are described: the Virtual Embryo and the Virtual Liver, which are both targeted towards understanding toxicity mechanisms. The Virtual Embryo examines disruption of morphogenesis and embryonal differentiation, and the Virtual Liver investigates pre-cancerous changes as effects of carcinogens. In both cases, *in vitro* data were integrated into an *in vivo* context, taking into account the kinetic dimension of changes.

As an example of an integrated approach to cancer systems medicine, The Virtual Cancer Patient model (Cancer Research UK 2006) provides a computational approach to predicting disease progression and optimizing drug combinations and schedules, by simulating the dynamics of key processes underlying drug-patient interactions. It allows drug developers to perform numerous rapid virtual clinical trials and to forecast optimal drug treatments. (See also Chap. 14.) The groundwork is already in place for systems medicine to contribute important advances over the next few years.

18.5 Resources Needed to Support Systems Approaches to Cancer Research and Diagnosis

The topic of the resources needed to support systems approaches to cancer research and diagnosis has been extensively discussed elsewhere in the present book. The following section aims to supplement that discussion by focusing on declarations and conclusions stemming from workshops devoted to the requirements necessitated by such approaches.

18.5.1 Infrastructure Requirements for Systems Biology

Cassman and Brunak (2007) explored the broad range of infrastructure needed to advance systems biology approaches, much of which is generally applicable to cancer: biobanks, related databases, protocols and standards, bioinformatic and medical informatic tools, laboratory high-throughput and high precision measurement platforms, clinical information, biomarkers, drugs and treatment protocols, and systems models. Workshops on cancer systems biology (Aebersold et al. 2009; Wolkenhauer 2009) furthered discussion of the particular infrastructures needed for cancer applications.

Resources often need additional organization specifically adapted to cancer systems approaches. Biological samples require an adequate data and database structure to make information accessible to appropriate analytical tools and to facilitate connections to clinical data and applications. There is a particular need for careful and strategical advance planning when large amounts of data are to be assembled for the purpose of systems-based analysis.

18.5.2 Clinical Resources

Integrated clinical research resources are crucial for the success of systems medicine. The Stockholm Declaration (Ringborg 2008) signed by 18 cancer research centres proposed a European initiative to work together to provide coordinated access to clinical and research infrastructures and resources. The 18 centres behind the Declaration, together with 10 more participants, very recently founded the European Platform for Translational Cancer Research (EUROCANPLATFORM) project, funded as a European Commission 'Network of Excellence', with the aim of providing the field with an appropriate infrastructure (see Chap. 2). By sharing patients, biological materials, technological resources and competence, an acceptable critical mass for development of personalized cancer medicine will be obtained. New research strategies will be implemented based on disrupting the molecular pathways that drive tumour initiation and progression. The continuous molecular alterations during tumour progression will be followed up as they evolve, from risk-factor exposure to manifest clinical disease. An important focus for future research is to link prevention and therapeutics by early detection of premalignant lesions. Development of personalized cancer medicine will be based on using biomarkers to predict the metastatic phenotype, as well as treatment effects and side-effects of anticancer agents and radiation therapy. Advanced biomarker research involving bioinformatics and systems biology approaches is being planned. The Eurocanplatform will also be valuable for expanding late translation cancer research, i.e. implementation and evaluation of new technologies in the clinical care of patients.

18.5.3 Data Resources, Analysis and Cancer Modelling Tools

It is recognized that major and dedicated data resources and analysis tools for cancer systems research will be required for the future. The Cancer bioinformatics grid caBIG (2009) forms an important infrastructure providing an integrated range of databases and bioinformatics tools in a form that lends itself to systems analysis. The GEN2PHEN (2009) collaborative project aims to supply grid-enabled data linkages between the areas of genotype, phenotype, clinical phenotype, and disease. A useful body of cancer-relevant information will be generated on gene expression, transcription, regulatory RNAs, protein interactions, time and space dependent

data, kinetics and non-equilibrium thermodynamics, 'omics' databases, and immunology. The ICGC (2009) project is yielding exceptionally detailed information on cancer genetic somatic variations from tumour samples.

Ambitious projects for the more distant future include model development at the levels of tissue and tumour, to be linked with cellular, network and pathway subsystems. The formulation of these models will require large amounts of complex quantitative data. Obtaining this data will be challenging; processing it will push the limits of database linkage to molecular biology models; and analysing it will require the full panoply of computational cell biology, integrated multiscale and multilevel modelling, virtual tumour models (CVIT 2009), and virtual cancer patient models. Fully comprehensive models are beyond any near-term possibilities, but useful preliminary models are already in existence and providing important answers to key questions.

18.6 Conclusions

Systems-based approaches are already beginning to offer effective applications for cancer research and treatment, thanks to the concurrent availability of new tools and infrastructures. Synthetic biology and nanotechnology are major new enabling tools in areas such as accelerated evolution studies and clinical oncology. Developments such as inexpensive and high-throughput genome-sequencing and nanotechnology tools, designed for protein and other biomarker detection, will enable major advances to be made towards the ultimate goal of personalized medicine. A wide range of modelling capabilities will greatly aid in the interpretation of the data generated by these tools, resulting in advances in the areas of biomarker and drug development and therapeutic applications, specifically optimized drug combinations and applied chronotherapy. The new tools and research infrastructures additionally permit the testing of novel systems-level concepts, such as robustness and evolution in cancer progression. There is a real hope that enhanced understanding of the interaction of various bodily systems, such as the harnessing of natural immunological processes, will lead to new treatment options. Major examples already exist of success in such areas as synthetic biology, nanotechnology, and chronotherapeutics; these advanced technologies, when coupled with information from living systems obtained through systems-based modelling, offer great promise for progress in the fight against cancer.

References

Abeloff et al. (2008) Abeloff's clinical oncology, 4th edn. Elsevier, Churchill Livingston, London

Aebersold R et al (2009) How systems biology can advance cancer research. Mol Oncol 3(1):9–17

Altinok A, Lévi F, Goldbeter A (2007a) A cell cycle automaton model for probing circadian patterns of anticancer drug delivery. Adv Drug Deliv Rev 59:1036–1053

Altinok A, Lévi F, Goldbeter A (2007b) Optimizing temporal patterns of anticancer drug delivery by simulations of a cell cycle automaton. In: Bertau M, Mosekilde E, Westerhoff HV (eds) Biosimulation in drug development. Wiley, Weinheim, pp 275–297

Altinok A, Lévi F, Goldbeter A (2009) Identifying mechanisms of chronotolerance and chronoefficacy for the anticancer drugs 5-fluorouracil and oxaliplatin by computational modeling. Eur J Pharmaceut Sci 36:20–38

Anderson JC, Clarke EJ, Arkin AP, Voigt CA (2006) Environmentally controlled invasion of cancer cells by engineered bacteria. J Mol Biol 355(4):619–627

Anderson JC, Voigt CA, Arkin AP (2007) Environmental signal integration by a modular AND gate. Mol Syst Biol 3:133

APO-SYS (2009) Apoptosis systems biology applied to cancer and AIDS. http://www.apo-sys.eu. Accessed 11 Mar 2009

Auffray C, Chen Z, Hood L (2009) Systems medicine: the future of medical genomics and healthcare. Genome Med 1:2

Ayukawa S, Kiga D et al (2007) SYANAC: SYnthetic biological Automaton for Noughts And Crosses. IET Synthetic Biol 1(1–2):64–67

Babu MM (2008) Computational approaches to study transcriptional regulation. Biochem Soc Trans 36(Pt 4):758–765 Review

Batt G, Yordanov B, Weiss R, Belta C (2007) Robustness analysis and tuning of synthetic gene networks. Bioinformatics 23:2415–2422

Bloom JD, Arnold FH (2009) In the light of directed evolution: pathways of adaptive protein evolution. Proc Natl Acad Sci U S A 106(Suppl 1):9995–10000 Epub 2009

Brenner K, You L, Arnold FH (2008) Engineering microbial consortia: a new frontier in synthetic biology. Trends Biotechnol 26(9):483–489. Epub 2008 Jul 31. Review

Cao L, Yu W, Wu Y, Yu L (2009) The evolution, complex structures and function of septin proteins. Biochem Soc Trans 37(Pt 4):717–721

caBIG (2009) Cancer Biomedical Information Grid (caBIG™) of the NCI (2009) https://cabig.nci.nih.gov. Accessed 12 Jan 2009

Cancer Research UK (2006) "Virtual cancer patient" predicts how breast cancer patients respond to treatment. http://info.cancerresearchuk.org/news/archive/pressreleases/2006/october/230118

Cassman M, Brunak S (2007) The US-EC Workshop on Infrastructure needs for Systems Biology. http://bnmc.caltech.edu/doku.php?id=us-ec-workshop and http://ec.europa.eu/research/biotechnology/ec-us/index_en.html. Accessed 1 Dec 2007

Cell Factory (2009) Cell factory fifth framework programme project resutls. http://ec.europa.eu/research/quality-of-life/cell-factory/volume2/index_en.html

Chames P, Baty D (2009) Bispecific antibodies for cancer therapy. Curr Opin Drug Discov Devel 12(2):276–283. Review

Chan DA, Giaccia AJ (2008) Targeting cancer cells by synthetic lethality: autophagy and VHL in cancer therapeutics. Cell Cycle 7(19):2987–2990. Epub 2008 Oct 12. Review

CVIT (2009) The Center for the Development of a Virtual Tumour. https://www.cvit.org. Accessed 12 Jan 2009

De Duve Institute (2010) http://www.deduveinstitute.be/cancer_immunology.php. Accessed 20 July 2010

FCSB (2008) FCSB first future challenge for systems biology. http://systems-biology.org/conference/report/2008-calendar-1/000005.html. Accessed

Friedman M, Ståhl S (2009) Engineered affinity proteins for tumour-targeting applications. Biotechnol Appl Biochem 53(Pt 1):1–29. Review

Friedland AE, Lu TK, Wang X, Shi D, Church GM, Collins J (2009) Synthetic gene networks that count. Science 324(5931):1199–1202

Gatenby R (2009) A change of strategy in the war on cancer. Nature, Vol. 459, pp. 508–9. 28 May 2009

GEN2PHEN (2009) Genotype to phenotype databases. http://www.gen2phen.org

Gibson D, Venter C, Hutchison C, Smith H et al (2008) Complete chemical synthesis, assembly and cloning of a mycoplasma genitalium genome. Science 319:1215

Goler J, Peccoud J et al (2008) Genetic design: rising above the sequence. Trends Biotechnol 26:538–544

Harris A (2005) REporting recommendations for tumour MARKer prognostic studies (REMARK) Editorial. Br J Cancer 93:385–386. doi:10.1038/sj.bjc.6602730 http://www.bjcancer.com. Published online 16 Aug 2005
Haseloff J, Ajioka J (2009) Synthetic biology: history, challenges and prospects. J R Soc Interface 6(Suppl 4):S389–S391. Epub 2009 Jun 3. PubMed PMID:19493895
Heath JR, Davis ME, Hood L (2009) Nanomedicine targets cancer. Sci Am 300(2):44–51
Heng HH, Bremer SW, Stevens JB, Ye KJ, Liu G, Ye CJ (2009) Genetic and epigenetic heterogeneity in cancer: a genome-centric perspective. J Cell Physiol 220(3):538–547. Review
ICGC (2009) International cancer genome consortium. http://www.icgc.org. Accessed
Ioerger TR, Sacchettini JC (2007) Structural genomics approach to drug discovery for Mycobacterium tuberculosis. Curr Opin Microbiol 12(3):318–325. Epub 2009 May 28. Review
Itaya M (1995) Toward a bacterial genome technology: integration of the Escherichia coli prophage lambda genome into the Bacillus subtilis 168 chromosome. Mol Gen Genet 248:9–16
Itaya M, Fujita K, Kuroki A, Tsuge K (2008) Bottom-up genome assembly using the Bacillus subtilis genome vector. Nat Methods 5(1):41–43
Jewett MC, Church GM (2011) In vitro integration of ribosomal RNA synthesis, ribosome self-assembly and protein synthesis. Nature for publication
Jones DS, Silverman AP, Cochran JR (2008) Developing therapeutic proteins by engineering ligand-receptor interactions. Trends Biotechnol 26(9):498–505. Epub 2008 Jul 31. Review
Jorgensen TJ (2009) Enhancing radiosensitivity: targeting the DNA repair pathways. Cancer Biol Ther 8(8):665–670. Epub 2009 Apr 27. Review
Keasling J, Chou H (2008) Metabolic engineering delivers next-generation biofuels. Nat Biotechnol 26:298–299. doi:10.1038/nbt0308–298
Kitano H (2004) Opinion: cancer as a robust system: implications for anticancer therapy. Nat Rev Cancer 4:227–235. doi:10.1038/nrc1300
Knight TF (2003) Idempotent vector design for standard assembly of BioBricks. Tech. rep, MIT
Kolar P, Francis E (2009) Virtual tissues. Report on the EC-US Workshop on virtual tissues. http://ec.europa.eu/research/biotechnology/ec-us/pdf/19th-meeting/francis_kolar_virtual_tissues_en.pdf. Accessed 20 Dec 2009
Lartigue C, Hutchison C, Smith H, Venter C et al (2009) Creating bacterial strains from genomes that have been cloned and engineered in yeast. Science 325:1693–1696
Le Meur N, Gentleman R (2008) Modeling synthetic lethality. Genome Biol 9(9):R135. Epub 2008 Sept 12
Lévi F, Schibler U (2007) Circadian rhythms: mechanisms and therapeutic implications. Annu Rev Pharmacol Toxicol 47:593–628
Maherali N, Hochedlinger K (2008) Induced pluripotency of mouse and human somatic cells. Cold Spring Harb Symp Quant Biol 73:157–162. Epub 2008 Nov 6. Review
Makarow M et al (2008) Advancing systems biology for medical applications. http://www.esf.org. Accessed
Marcus FB (2008) Bioinformatics and systems biology: collaborative research and resources. Springer, Berlin
Martin V, Keasling J et al (2003) Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. Nat Biotechnol 21:796–802. Published online: 1 June 2003. doi:10.1038/nbt833
Michielin O (2007) Application of molecular modelling to new therapeutic cancer approaches. Bull Cancer 94(9):763–768. Review
Mizoguchi H, Mori H, Fujio T (2007) Escherichia coli minimum genome factory. Biotechnol Appl Biochem 46:157–167
Mizuarai S, Irie H, Schmatz DM, Kotani H (2008) Integrated genomic and pharmacological approaches to identify synthetic lethal genes as cancer therapeutic targets. Curr Mol Med 8(8):774–783. Review
Moss AJ, Sharma S, Brindle NP (2009) Rational design and protein engineering of growth factors for regenerative medicine and tissue engineering. Biochem Soc Trans. 2009 Aug 37 Pt4 717–21

NEST (2005) NEST high-level expert group—synthetic biology—applying engineering to biology. European Commission Report
Nirmalanandhan VS, Sittampalam GS (2009) Stem cells in drug discovery, tissue engineering, and regenerative medicine: emerging opportunities and challenges. J Biomol Screen. 2009 Aug;14(7):755–68. Epub 2009 Aug 12
Nishikawa S, Goldstein RA, Nierras CR (2008) The promise of human induced pluripotent stem cells for research and therapy. Nat Rev Mol Cell Biol 9(9):725–729. Review
Papait R, Monti E, Bonapace IM (2009) Novel approaches on epigenetics. Curr Opin Drug Discov Devel 12(2):264–275. Review
Papapostolou D, Howorka S (2009) Engineering and exploiting protein assemblies in synthetic biology. Mol Biosyst 5(7):723–732. Epub 2009 May 7
Parliamentary Office (2008) Synthetic biology: postnote by parliamentary office of science and technology. Number 298. http://www.parliament.uk/documents/upload/postpn298.pdf
Peccoud J et al (2008) Targeted development of registries of biological parts. PLoS One 3(7):e2671
Platis D, Labrou NE (2008) Chemical and genetic engineering strategies to improve the potency of pharmaceutical proteins and enzymes. Curr Med Chem 15(19):1940–1955. Review
Prabhakar S, Noonan JP, Paabo S, Rubin EM (2006) Accelerated evolution of conserved noncoding sequences in humans. Science 314:786
Prather KL, Martin CH (2008) De novo biosynthetic pathways: rational design of microbial chemical factories. Curr Opin Biotechnol 19(5):468–474. Epub 2008 Sep 5. Review
Presta LG (2008) Molecular engineering and design of therapeutic antibodies. Curr Opin Immunol 20(4):460–470. Review
Purnick P, Weiss R (2009) The second wave of synthetic biology: from modules to systems. Nat Rev Mol Cell Biol 10:410–422
Quaglia F (2008) Bioinspired tissue engineering: the great promise of protein delivery technologies. Int J Pharm 364(2):281–297. Epub 2008 Apr 26
Ringborg U (2008) The stockholm declaration. Mol Oncol 2:10–11
Robins H, Krasnitz M, Levine AJ (2008) The computational detection of functional nucleotide sequence motifs in the coding regions of organisms. Exp Biol Med (Maywood) 233(6):665–673. Epub 2008 Apr 11. Review
Rowe A (2009) Experimental drug makes the immune system revolt against cancer. http://www.wired.com/wiredscience/tag/synthetic-biology
Saito H, Inoue T (2007) RNA and RNP as new molecular parts in synthetic biology. J Biotechnol 132:1–7
Saito H, Inoue T (2009) Synthetic biology with RNA motifs. Int J Biochem Cell Biol 41:398–404
Saito H, Kashida S, Inoue T, Shiba K (2007) The role of peptide motifs in the evolution of a protein network. Nucleic Acids Res 35:6357–6366
Saito H, Minamisawa T, Yamori T, Shiba K (2008) A motif-programmed artificial protein induces apoptosis in several cancer cells by disrupting Mitochondria. Cancer Sci 99:398–406
Schmidt M, Böttcher D, Bornscheuer UT (2009) Protein engineering of carboxyl esterases by rational design and directed evolution. Protein Pept Lett. 2009;16(10):1162–71
Schwille P, Diez S (2009) Synthetic biology of minimal systems. Crit Rev Biochem Mol Biol 44(4):223–242
Semenza GL (2008) Hypoxia-inducible factor 1 and cancer pathogenesis. IUBMB Life 60(9):591–597. Review
Serrano L (2007) Editorial; synthetic biology: promises and challenges. Mol Syst Biol 3(158):1–5
Shetty R, Endy D, Knight T (2008) Engineering BioBrick vectors from BioBrick parts. J Biol Eng 2:5. doi:10.1186/1754-1611-2-5
Shimono K, Yokoyama S et al (2009) Production of functional bacteriorhodopsin by an E. Coli cell-free synthesis system. Protein Sci 18:2160–2171
Shoyele SA (2008) Engineering protein particles for pulmonary drug delivery. Methods Mol Biol 437:149–160. Review
Smailus DE, Warren RL, Holt RA (2007) Constructing large DNA segments by iterative clone recombination. Syst Synth Biol 1(3):139–144. Epub 2008 Jan 24

Stolnik S, Shakesheff K (2009) Formulations for delivery of therapeutic proteins. Biotechnol Lett 31(1):1–11. Epub 2008 Sep 11. Review
Stumpp MT, Binz HK, Amstutz P (2008) DARPins: a new generation of protein therapeutics. Drug Discov Today 13(15–16):695–701. Epub 2008 Jul 11. Review
Sybilla (2010) Sybilla project. http://www.sybilla-t-cell.de. Accessed 20 July 2010
Synthetic Biology Working Group Technical Reports (2003) http://hdl.handle.net/1721.1/21168
SYSBIOMED (2007) Systems biology for medical applications. http://www.sysbiomed.org. Accessed 1 Dec 2007
Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126:663–676
Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S (2007) Induction of pluripotent stem cells from adult human Fibroblasts by defined factors. Cell 131:861–872
Yao T (2003) Meeting report –symposium on the elucidation of "Strategy of Life" Kagaku. Science 73:925–927
Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R, Slukvin II, Thomson JA (2007) Induced pluripotent stem cell lines derived from human somatic cells. Science 318:1917–1920
Ulmer KM (1983) Protein engineering. Science 219(4585):666–671
Vázquez E, Ferrer-Miralles N, Mangues R, Corchero JL, Schwartz S Jr, Villaverde A (2009) Modular protein engineering in emerging cancer therapies. Curr Pharm Des 15(8):893–916. Review
Vucic D, Fairbrother WJ (2007) The inhibitor of apoptosis proteins as therapeutic targets in cancer. Clin Cancer Res 13(20):5995–6000. Review
Wada A (2009) Private communication. http://www.gsc.riken.jp. Message by AW
Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM (2009) Programming cells by multiplex genome engineering and accelerated evolution. Nature 460(7257):894–898. Epub 2009 Jul 26
Webb TJ, Bieler JG, Schneck JP, Oelke M (2009) Ex vivo induction and expansion of natural killer T cells by CD1d1-Ig coated artificial antigen presenting cells. J Immunol Methods 346(1–2):38–44. [Epub May 14 2009]
Weber W, Fussenegger M (2009) The impact of synthetic biology on drug discovery. Drug Discov Today. 2009 Oct;14(19-20):956–63. Epub 2009 Jul 4
Weinberg RA (2007) The biology of cancer. Garland Science, New York
Williams BR, Haque SJ (1997) Interacting pathways of interferon signaling. Semin Oncol 24(3 Suppl 9):S9-70–S9-77. Review
Wolkenhauer O (2009) Cancer systems biology workshop. Warnemunde, Germany
Wolkenhauer O et al (2009) Sysbiomed report: advancing systems biology for medical applications. IET Syst Biol 3(3):131–136
Wu AM, Olafsen T (2008) Antibodies for molecular imaging of cancer. Cancer J 14(3):191–197. Review
Yang S, Yuan W, Jin T (2009) Formulating protein therapeutics into particulate forms. Expert Opin Drug Deliv. 2009 Oct; 6(10):1123–33
Zhang J, Huang S, Zhang H, Wang H, Guo H, Qian G, Fan X, Lu J, Hoffman AR, Hu JF, Ge S (2009) Targeted knockdown of Bcl2 in tumor cells using a synthetic TRAIL 3'-UTR microRNA. Int J Cancer. 2010 May 1;126(9):2229–39
Zhao R, Daley GQ (2008) From fibroblasts to iPS cells: induced pluripotency by defined factors. J Cell Biochem 105(4):949–955. Review
Zhou B, Zhang H, Damelin M, Geles K, Grindley J, Dirks P (2009) Tumour-initiating cells: challenges and opportunities foranticancer drug discovery. Nat Rev Drug Discov 8:806–823

Chapter 19
Conclusions

Alfredo Cesario and Frederick B. Marcus

Abstract Systems-based approaches in biology, bioinformatics, and medicine have a major role to play in cancer research and treatment. The key points from each chapter are summarized, and overall conclusions about the field are then discussed, based on themes running throughout the book. Important progress has already been demonstrated in many areas, such as the modelling of gene regulation networks, cell growth, biomarker identification, and drug treatment optimization. The near future is full of promise for scientific and clinically important advances based upon systems approaches.

19.1 Key Points

19.1.1 Introduction and Background

Cancer is a massively diverse set of diseases, yet with a number of common properties that allow systems-based methodologies to be established for basic understanding, potentially leading to the development of improved clinical applications and treatment. In view of the immense suffering and economic cost inflicted by cancer, there is an urgent need for novel therapeutic advances.

Diagnosis and treatment could be greatly enhanced by better understanding of the various interrelated aspects of cancer at the molecular, cellular, tumoral, histological and clinical levels. The daunting diversity of cancer types, grades, stages, and evolution is to some extent offset by the unifying features of malignancies. Systems-based analysis offers the possibility of integrating data from cytological, histological, imaging, and genetic and expression profiling, within comprehensive models. Signalling pathways analysis needs to be extended from traditional biological investigation to the inclusion of grading and clinical staging, as part of a process which will aid in the discovery of novel targets and biomarkers for diagnosis and prognosis. While major challenges and limitations remain in implementing a comprehensive systems-

A. Cesario (✉)
Deputy Scientific Director, IRCCS San Raffaele Pisana, Via di Valcannuta 247, 00166 Rome, Italy and Assistant Professor of Thoracic Surgery, Catholic University,
Largo Agostino Gemelli 1, 00166, Rome, Italy
e-mail: alfredo.cesario@sanraffaele.it; alfcesario@rm.unicatt.it

A. Cesario, F. B. Marcus (eds.), *Cancer Systems Biology, Bioinformatics and Medicine,*
DOI 10.1007/978-94-007-1567-7_19,

based approach to cancer, there is nevertheless much to be gained from these fresh methodologies and technologies, which offer the hope of accelerating the translation of new discoveries into prevention and clinical applications.

19.1.2 Laboratory, Clinical, Data and Educational Resources for Cancer Research

The successful development of systems-based approaches requires an interlinking set of resources: laboratory tools for measurement, cell and tissue samples, model organisms, standardized databases, educational facilities, and relevant infrastructure.

Measurement technologies applicable to the fields of molecular and cellular biology are currently being developed with amazing rapidity, most apparently in gene sequencing, where capacity and throughput have increased by several orders of magnitude. Systems approaches are further enabled by technologies facilitating genome-wide readout of DNA, mRNA, proteins and metabolomes. Array-based assays are being replaced with second-generation digital sequencing technologies that analyse gene expression, genotype, single nucleotide polymorphisms, and methylation patterns. Preparatory enrichment techniques are being developed to reduce complexity of samples. Proteomics and functional assays have benefitted from recent progress in mass spectrometry, automatic microscopy and image analysis, though the routine measurement of whole proteomes is still unattainable, owing to the problems of multiple transcripts and post-translational modifications. As a result of the breakthrough in human genome sequencing and its associated technological development, the management of cancer is being transformed. Molecular substaging for many types of tumour is approaching clinical reality and is aiding the decision-making processes involved in choosing and using novel chemotherapeutic agents.

One of the major bottlenecks impeding advances in the treatment of cancer patients is the difficulty of achieving sufficiently detailed and consistent modelling of the disease to yield useful results, often because of the limitations of data from cells and organisms. Advances in targeted therapies are dependent on well-characterized cell lines, biopsies and biobanks for drug evaluation and testing, as a preclinical entry point into the pharmaceutical pipeline. All of these components are key parts of any systems-based strategy for understanding and individualizing the treatment of human cancer. The use of model organisms, especially mice, as a means of investigating cancer tumours that are living, growing and interacting with other cells, will increase in importance for testing novel therapies and calibrating computational models, given the limited utility of data obtained from cell lines.

The amount of data generated in cancer research is growing rapidly, a process accelerated by new second-generation gene-sequencing technologies. High-density array-based technologies, such as genome-wide single nucleotide polymorphism (SNP) genotyping and gene expression microarrays, are producing bodies of data that are quantitatively larger and qualitatively much more complex, and tougher

to analyse and to exploit for interpretative meta-data. The utility of germ-line and somatic genetic variation databases is further restricted by their focus on a limited range of cancer phenotyping. Huge amounts of data relating to tumour genotypes are being generated by major international initiatives such as The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC). The resulting databases provide catalogues of single disease-causing variants associated with specific phenotypes, as well as genome-wide single variants of both structures and SNPs. There are also large repositories of information for complete genome-wide association studies. New access and analysis tools are available to aid users in handling the large quantity and diversity of experimental data stored in these repositories. Bioinformatics analysis and systems modelling are capable of transforming and simplifying high-dimensionality data to a level directly useful to biologists and ultimately clinicians.

Until very recently, progress in cancer research was hampered by the slow rate of acquisition of experimental data, as well as by inconsistencies in methodology and materials available. The recent explosion in high-throughput methodologies has radically changed the situation, so that analysis and interpretation of data, rather than its acquisition is now the greatest challenge. Computational tools and services have become an essential feature of the modern molecular scientist's workbench. Researchers and clinicians must be equipped with the necessary educational and infrastructural resources before they can proceed to identify quantitative and predictive relationships within these newly available collections of molecular, genomic and clinical information. The infrastructure resources need to be highly interdisciplinary, either based in single institutions or available through collaborations.

19.1.3 Bioinformatics and Systems Biology Research Results

Cancer systems modelling requires flexible strategies which depend on trade-offs between the nature of the biochemical network under investigation, the biomedical question to be elucidated, and the quantity and quality of the experimental data available. A wide range of mathematical models, each with its own area of applicability, is employed in the analysis of signalling networks in cancer: differential equations, stochastic analysis, Boolean logic, etc.

Detailed analysis of cancer-relevant biochemical networks is facilitated by computational modelling and web-based resources such as databases, bioinformatic analysis tools, and large and tested libraries of systems biology models. For example, tools are available for modelling cancer-related areas of molecular and cellular pathway interactions. Systems-based modelling demands consistent standards for database and model library creation, many of which are already in place. A number of international conventions on minimum information and ontologies have been established with a view to standardizing description and annotation of models, methods, and data, so that different groups working in the same field can share data and exchange information in a systematic and precise way.

The six hallmarks of cancer defined ten years ago by Hanahan and Weinberg, based on decades of molecular, cellular and clinical investigations, have been further explored and refined using systems approaches. A wide variety of globally-generated data has been incorporated into mathematical and computational models of those molecular and cellular pathways and networks that become dysregulated in cancer, thereby taking into account the large-scale properties of complex biological networks. Analysis of transcription and protein interaction networks reveals cancer biomarkers. The processes of cell growth, proliferation and apoptosis are analysed by means of modelling of signalling networks, whereas multiscale pathway, cellular and tumour modelling is required for the description of sustained angiogenesis and metastasis. Enhanced understanding of the hallmarks of cancer is beginning to open up new avenues for cancer diagnosis and treatment.

A systems approach with mathematical modelling is needed to fully describe the processes involved in the dysregulation of genetically-regulated programmed cell death, with its accompanying complexity and interweaving of pathways. Apoptosis, together with proliferation and differentiation, plays an important role in the pathology and physiology of cancer and other disorders; understanding of cell death is vitally important for optimal treatment of individual cancers. The models derived from systems biology studies are crucial to developing and testing biomarkers and drugs, and for optimizing the choice of combination therapies, since extensive molecular data for each type of cancer, personalized to each patient's history and genetic makeup needs to be taken into consideration. Collaborative research is often an essential means of assembling the resources needed to tackle the complex tasks of data taking, modelling and clinical testing.

Given the complexity of carcinogenesis and tumour development, it is essential to understand their underlying organizing principles. Translating biological motifs into pathway signatures creates the possibility of identifying pathological pathways in individual tumours and the perspective of prognostically stratifying and selecting adequate drugs. Much progress has been achieved in the field of spatial-temporal modelling of tissue regeneration or three-dimensional tumour infiltration. However, the ambitious goal of fully linking tumour development to single-cell decisions created by molecular models of signalling network constellations has not yet been achieved.

The role of mathematical modelling of cancers at the tumoral level is to identify the first principles governing tumour growth, invasion, metastases, and response to therapy. Cancers are complex dynamical systems dominated by non-linear processes, where most critical parameters exhibit significant heterogeneity. This variability, while central to the ability of cancers to adapt to a wide range of environmental perturbations including therapy, tends to be lost in molecular-level measurements which provide typically 'average' values for large numbers of heterogeneous tumour cells obtained at a single time point. Biologically realistic mathematical models are necessary to transform the reductionist approach of modern cancer biology into comprehensive models of the host-cancer interactions governing the dynamics of tumour growth and therapy, and important beginnings have been made towards that goal. Evolutionary theory is a necessary paradigm for developing these new approaches.

19.1.4 Translation to Clinical Applications

The challenges and difficulties involved with the translation of fundamental discoveries into useful clinical biomarkers are many and varied. Approaches using bioinformatics, system analysis and modelling are highly relevant for effective discovery of biomarkers and their prioritization and quantification, along with elimination of false-positive indications. There is a need for large-scale efforts to achieve harmonization, standardization and integration of multilevel data into systems modelling, in order to ensure that the potential offered by the unprecedented availability of data and information is fully exploited to develop validated and useful biomarkers.

New approaches are currently being investigated in the process of drug development, in an attempt to improve upon the large failure rate seen in many clinical trials. Systems biology shows great promise for changing the way we think about disease and drug development, and over the last decade, several biotechnology companies have been established with the aim of incorporating systems approaches into the oncological drug development process. An integrative philosophy has been widely adopted by industry/academia-generated platforms. Both descriptive and predictive models are used in order to provide the presumed best approach for different situations. Predictive 'virtual tumour' models, coupled with pharmacokinetic and pharmacodynamic models, are being developed to aid in the design of optimized drug schedules and combinations.

The circadian timing system (CTS) controls cellular proliferation and drug metabolism over a twenty-four hour period through molecular clocks in each cell. The CTS down-regulates malignant growth in cancer patients, and also generates large and predictable changes, during the twenty-four hour 'day', in the toxicity and efficacy of experimental and clinical anticancer treatments, which have been validated in randomized clinical trials. Systems approaches have become essential for further advances in this complex area. Computer models have been developed for the interactions between circadian clocks, cell division cycle and pharmacology pathways, revealing why the same circadian timing jointly optimizes tolerability and efficacy of a given anticancer drug, both in experimental models and in cancer patients.

Current and future clinical applications of systems biology approaches related to cancer include the identification of diagnostic, prognostic, and therapeutic biomarkers; the design of intervention strategies tailored to individual patients, using combinatorial targeted therapy; and the setting up of clinical trials leading to personalized cancer medicine. As a result of the completion of the Human Genome Project, together with rapid progress in The Cancer Genome Atlas (TCGA), genome-wide association studies, and the ensuing development of high-throughput technologies for analyzing patient samples at the DNA, RNA, protein, and metabolic levels, there has been a rapid accumulation of data capable of providing an atlas or 'tool kit' describing the events that may occur during tumour initiation and progression. This data is challenging because of the incredible heterogeneity among tumour types and the number of genetic aberrations found in epithelial cancer cells. Systems approaches capable of integrating large bodies of data collected using multiple platforms and modalities with mathematical models of functional and phenotypic

tumour outcomes, will be necessary for the discovery of new principles that can accelerate progress in translating the pre-clinical studies into improved patient outcomes. Many useful models have already been established in these areas.

Most cancers are resistant to a variety of therapeutic interventions, continuing to proliferate both by evolving rapidly, and harnessing the body's own mechanisms. This is in large part attributable to a systems-level property known as 'robustness', a fundamental feature of complex evolved systems, including cancer, that is optimized for certain functions and environments, but which may be offset by 'fragility' to unusual stresses. The modelling and application of this property is a key area for research and treatment. Understanding the workings of both robustness and fragility (*i.e.* resistance and susceptibility to perturbation) is crucial for the development of novel and effective therapeutic strategies, meaning both new drugs and improved administration modalities. Several new treatment methodologies have been based on analyses of the phenomenon of robustness.

19.1.5 Perspectives

Long term advances in cancer research and treatment will be made possible by systems approaches coupled with new technologies that operate at the molecular, cellular, and microorganism levels. Synthetic biology and its associated technologies provide an important basis for establishing the perspectives for future systems approaches. Both at individual laboratories and integrated programmes such as that in Japan are making important advances thanks to synthetic biology, in cancer research and clinical applications. These promising technological and research developments demonstrate the utility of systems approaches in facilitating translation to the clinical context. Perspectives for future progress and infrastructure requirements have been elaborated in several collaborative workshops, notably those organized by US/NCI-EU-Germany-Japan, BBSRC/UK, and JST/Japan.

19.2 Overall Conclusions

Systems-based approaches have a major and highly positive role to play in cancer research, and indeed the first steps have been taken for the application of such approaches to treatment. The field is already well established, with important advances demonstrated in areas such as models for gene regulation networks and cell and tumour growth; diagnostic and treatment strategies for some cancers; and study of the effects of drug therapy. The future is full of promise for rapid and major advances, thanks partly to systems approaches in understanding and treating cancer.

However, except for very select cases where molecular sub-staging for some tumour types has approached clinical reality, direct intervention in standard clinical decision-making processes has not yet happened. Decisions are still being taken on

the basis of information coming from grossly morphological criteria, these being the TNM classification system, some sub-classifications, and standard histology. For the ultimate user of system-based advances, the oncologist trying to diagnose and treat a patient suffering from cancer, considerations based on 'omics' may seem rather theoretical and remote, something for the distant future.

The gigantic amount of information stemming from the boost of 'omics' technologies in biomedical translational research is certainly capable of changing the matrix of the inferential decision-making process in the clinical environment, specifically in clinical oncology (both surgical and medical). However, the data, which is frustratingly more complex and increasingly difficult to structure and interpret than initially assumed, remains currently in a fragmented state, and does not contribute optimally to understanding and applications. As a result, the failure to deliver broadly validated and useful clinical applications (i.e. improvements in diagnosis and treatment of patients) in significant terms, has given a substantial push towards the re-structuring of basic research, from a reductionist and linear methodology towards a more integrative systems-based approach.

In fact, many systems-based techniques are potentially applicable to clinical practice, and more will be developed in the near future by means of systems analysis; for example:

- Complex biomarker and biopsy readings could be interpreted in a more quantitative way in order to stratify patients in terms of diagnosis and treatment.
- Germ-line genetics and somatic profiling of biopsies could be employed for more accurate prognosis of cancer evolution.
- During drug development and clinical trials, drug dosages and their combination could be applied in more precisely measured ways to stratified patient categories, with further refinements from cancer chronotherapy modelling.
- Enhanced understanding of the evolutionary and robustness properties of a combination of cancer cells in tumours and metastases could result in better informed treatment decisions.

Successful systems approaches will be possible only once a cultural bridge has finally been established between fundamental and clinical sciences, so as to transfer and merge information on the basis of fully integrative attitudes.

However, before reaching the 'daily clinical practice' level, which translates to guidelines involving clinical validation, several years are legally required for thorough evaluation of the effectiveness of systems-based approaches. Any new approach has to go through the entire clinical testing pathway, and this process is strictly regulated, ethically controlled, and needs very significant patient numbers. As documented in Part IV of this book, systems modelling has already been involved in certain limited areas of clinical trials, such as biomarker testing, drug development, cancer chronotherapy, and patient stratification. Many more possibilities are available today based upon systems approaches, which if not immediately validated for daily clinical practice, at least allow the initiation of clinical experimentation. We conclude that systems approaches and supporting informatics are not only major research tools, but offer the hope of significant progress in cancer treatment.

Index

A. Cesario, F. B. Marcus (eds.), *Cancer Systems Biology, Bioinformatics and Medicine,*
DOI 10.1007/978-94-007-1567-7, © Springer Science+Business Media B.V. 2011