1 A Path to Knowledge: from Data to Complex Systems Models of Cancer

Sylvia Nagl

> 'The definitive property of individuality at the organismal level lies in the effective suppression of the differential propagation of subparts as a necessary strategy for maintaining functional integrity . . . This suppression has been so effective, while the consequences of failure remain so devastating, that human organisms have coined a word for the cell lineage's major category of escape from this constraint, a name with power to terrify stable human organisms beyond any other threat to integrity and persistence – cancer.' (Gould, 2002, p. 695)

This chapter will chart a path of knowledge discovery, bringing together cutting edge experimental and computational methods in order to advance our understanding of the structure and dynamic function of biological systems underpinning cancer phenotypes. The aim is to provide a comprehensive overview of a very large area of current research and to highlight key developments and challenges (it is not intended as a detailed review of any of the specialist areas discussed and the reader is referred to the many excellent reviews and the primary literature for in-depth study). The past decade has seen the ascendance of high-throughput methods for measuring the global expression of different biological components – genomics, transcriptomics, proteomics, glycomics, metabolomics. Cancer researchers were among the first to extensively deploy these 'omic' technologies, and the wealth and breadth of available data (see Table 1.1 for on-line access to genomic and transcriptomic data) and technologies now make the

Cancer Bioinformatics: From therapy design to treatment Edited by Sylvia Nagl

Table 1.1 A selection of genome-focused data resources for cancer bioinformatics and systems biology

Focus	Data source	Data types	URL
Genome	Cancer Genome Project (Sanger Centre)	Cancer Gene Census, COSMIC (somatic mutations), LOH mapping, deletion mapping, small intragenic somatic mutations	www.sanger.ac.uk/genetics/CGP
	Human Genome Resources (NCBI)	Integrated information resource for human genome data	www.ncbi.nlm.nih.gov/genome/guide/human
	Genome Browser (UC Santa Cruz)	Visualization and query tools	genome.ucsc.edu
Karyotype	Cancer Chromosomes (NCBI)	SKY/M-FISH and CGH, Mitelman database, NCI Recurrent Aberrations in Cancer	www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=cancerchromosomes
	Progenetix (University of Florida)	CGH data for different cancer types	www.progenetix.net
SNPs	dbSNP (NCBI)	Single nucleotide polymorphisms	www.ncbi.nlm.nih.gov projects/SNP
	SNP500cancer (NCBI)	SNPs with relevance to epidemiology studies in cancer	snp500cancer.nci.nih.gov
Gene expression	Gene Expression Omnibus (NCBI)	A curated resource for gene expression data browsing, query and retrieval	www.ncbi.nlm.nih.gov/geo
	Oncomine (University of Michigan)	Tools to locate, query and visualize cancer microarray data for a given gene or cancer type	141.214.6.50/oncomine/main/index.jsp
	Cancer Genome Anatomy Project (CGAP)	Integrated resource for genes, chromosomal aberrations, SNP500cancer, tissues, pathways, SAGE expression data (normal, precancer and cancer cells)	cgap.nci.nih.gov
Clinical genomics	Cancer Molecular Analysis Project (CMAP)	Molecular profiles, targets, targeted agents, trials	cmap.nci.nih.gov

study of cancer from a *systems* perspective a paradigmatic arena in which to develop systems science for biology and medicine.

Systems biology aims to understand how complex molecular interactions give rise to dynamic processes in biological systems and, because it is very difficult to directly observe and measure dynamic processes in complex systems, the research relies on

data generated by 'omic' technologies and an integration of experimental and computational methods. Systems biology is already fundamentally changing the practice of cancer biology and directly addresses pressing challenges in the development of new anti-cancer therapies, particularly the lack of efficacy or toxicity due to poor understanding of the biological system they attempt to affect. It provides an integrative methodology for identifying and characterizing pathways that are critical to cancer, discovering new targets within the context of biological networks and assessing both on- and off-target effects of therapeutics. It can confidently be expected to play an increasingly central role in pharmacogenomics by helping to uncover sources of inter-individual variability in treatment response, thereby supporting the promise of individualized therapy intended to maximize effectiveness and minimize risk. (Bogdanovic and Langlands, 2004; Birney *et al*., 2005; Khalil and Hill, 2005).

Knowledge discovery needs to cut across biological levels (genome, transcriptome, proteome, metabolome, cell; and beyond to tissue, organ and patient) and is of necessity a multidisciplinary endeavour requiring an unprecedented level of collaboration between clinicians and scientists from diverse disciplines (see Chapter 3). Toyoda and Wada (2004) have coined the term 'omic space' – denoting a hierarchical conceptual model linking different 'omic' planes – and showed that this concept helps to assimilate biological findings comprehensively into hypotheses or models, combining higher order phenomena and lower order mechanisms, by demonstrating that a comprehensive ranking of correspondences among interactions in the space can be used effectively. It also offers a convenient framework for database integration (see also omicspace.riken.jp/gps and www.gsc.riken.go.jp/eng/gsc/project/genomenet.html).

Furthermore, systems-based discovery has both experimental and computational components and ideally involves an iterative cycle that integrates both 'wet' and 'dry' methods. Computational systems biology is developing a rapidly expanding methodological scope to integrate and make sense of 'omic' data, by relating it to higher level physiological data and by using it to analyse and simulate pathways, cells, tissues, organs and disease mechanisms (see Chapters 4–7). There is a diverse range of both established and newly emerging computational methods (Ideker and Lauffenburger, 2003), and it is clear that research aimed at a systems-level understanding of cancer requires advanced statistical analysis and mining of the large amounts of data obtained through 'omic' technologies to be integrated with mathematical modelling of systems dynamics. Computational data management, data mining and mathematical modelling offer research tools commensurate with powerful laboratory techniques provided that they are used appropriately (Murray, 2002; Swanson, True and Murray, 2003).

Success will depend not only on the deployment of appropriate computational methods but also, equally vitally, on the standardization of experimental data capture protocols, data quality assurance and validation procedures, and data integration and sharing standards. As discussed in the Guidance for Industry on Pharmacogenomic Data Submissions published by the Food and Drug Administration in March 2005 (www.fda.gov/cber/gdlns/pharmdtasub.pdf), substantial hurdles exist with regard to: laboratory techniques and test procedures not being well validated and not generalizable across different platforms; the scientific framework for interpreting the physiological,

toxicological, pharmacological, or clinical significance of certain experimental results not yet being well understood; and the standards for transmission, processing and storage of the large amounts of highly dimensional data generated from 'omic' technology not being well defined or widely tested. Standard development initiatives, such as caCORE of the National Cancer Institute (NCI) in the USA and the Cancer Informatics Initiative of the National Cancer Research Institute (NCRI) and Cancergrid in the UK, therefore constitute a prerequisite for further advances in cancer research.

In summary, wet–dry knowledge discovery cycles can be considered to serve as fundamental frameworks for cancer research in the 21st century (Figure 1.1) whose essential components comprise:

- An integrative 'complex systems' approach (see Sections 1.1 and 1.2 and Chapters 2 and 3).
- Experimental science and technological advances (outside the scope of this book).
- Appropriate *in vivo* model systems (Chapters 8 and 9).
- Standards for experimental design and the generation of data suitable for systems-based discovery (Chapter 3).
- Mathematical modelling (see Section 1.3 and Chapters 4–7).
- Bioinformatics and large-scale data mining (see Section 1.3 and Chapter 3).
- Data/model standardization and integration (see Section 1.4 and Chapters 3, 4, 10 and 11).
- Software design and data sharing ethics (Chapters 3 and 12–14).

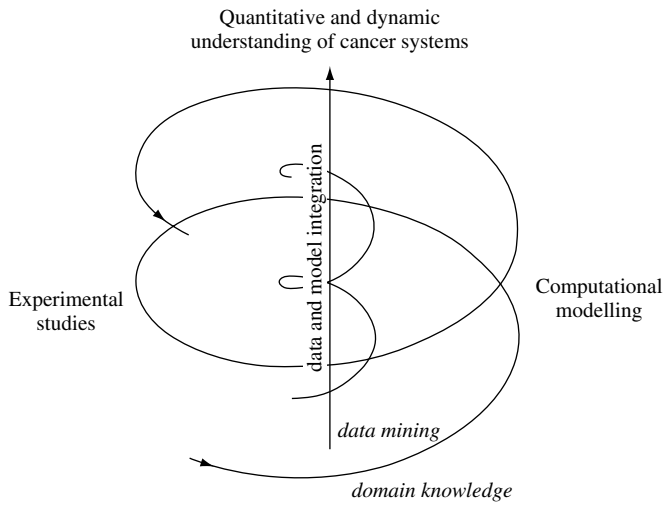


Figure 1.1 The iterative knowledge discovery cycle

1.1 Conceptual foundations: biological complexity

Systems biology seeks to address the complexity of human cancer by drawing on a conceptual framework based on the current understanding of the characteristics of complex adaptive systems in general, regardless of whether they are physical, biological or social in nature, e.g. ranging from cellular networks to social communities, ecological systems and the Internet. Complex systems are composed of a huge number of components that can interact simultaneously in a sufficiently rich number of parallel ways so that the system shows spontaneous self-organization and produces global, emergent structures (Holland, 1995; Depew and Weber, 1996). Self-organization concerns the emergence of higher level order from the local interactions of system components in the absence of external forces or a pre-programmed plan embedded in any individual component (Holland, 1995, 1998; Mitchell, 2003). The mechanisms of self-organization are amenable to analysis in terms of positive and negative feedback (amplification and damping). Importantly, complex systems are 'robust, yet fragile' – they can often be disabled catastrophically by even small perturbations to certain components (Csete and Doyle, 2002).

Cancer cells maintain their survival and proliferative potential against a wide range of anti-cancer therapies and immunological responses of the patient. Robustness is seen as an emergent property arising through abnormal feedback control, redundancy and heterogeneity. These constituent characteristics result from the interplay of genomic instability and selective pressure driven by host–tumour dynamics (see Chapter 2). The challenge then is to identify the vulnerabilities in the system through an understanding of its organization and dynamic behaviour and to systematically control the cell dynamics rather than its molecular components.

In contrast to the systems-based framework outline above, conceptual models of the dependency of human cancer upon one genetic abnormality or a very small number of abnormalitic have been extremely influential in guiding single-target strategies in therapy design. These models postulate that correction of any one key oncogenic defect, or oncogene/pathway 'addiction', would be sufficient to 'precipitate the collapse' of the tumour (Workman, 2003). Primarily, selection of single targets is based on criteria such as frequency of genetic or epigenetic deregulation of the target or pathway in cancer, demonstration in a model system that the target contributes to the malignant phenotype and evidence of at least partial reversal of the cancer phenotype by target inhibition.

However, there is strong evidence that several genetic abnormalities are causally involved in most human cancers and, very significantly, there may be dozens of genes that are aberrant in copy number or structure (due to aneuploidy) and hundreds or even thousands of genes that are abnormally expressed. The *pathobiology* of cancer is driven by mutation in oncogenes, tumour suppressors and stability genes needed for DNA repair and chromosomal integrity (e.g. BRCA1, BLM, ATR). Only mutations in oncogenes and suppressors can directly affect net cell growth. Stability genes keep genetic alterations to a minimum, and inactivation of both alleles therefore can result in an increased mutation rate in the genome that potentially can affect any other genes in a more or less random

manner. These 'bystander mutations' can have profound effects on the *cancer phenotype*, notably also including treatment resistance (Figure 1.2). Furthermore, epigenetic changes (covalent modifications of DNA or chromatin that are preserved during cell division) in expression patterns can affect hundreds or even thousands of genes as a consequence of the primary mutations and lead to a reconfiguration of the cancer cell's biology. At the moment when treatment is commonly given, most tumour cells will have acquired an abnormal phenotype that embodies complex combinations of these different types of molecular abnormality.

Novel treatment strategies need to take into consideration the high level of complexity of cancer cell phenotypes. The details of the scope of deregulated wiring of signal transduction pathways in cancer, and their interdependent effects on the cell and tumour level, are not adequately understood to make a 'rational' selection of treatment targets. What is more, the complex nature of underlying genome deregulation can be expected to make a rational approach impossible in the traditional sense. It is here where cancer systems biology seeks to make an essential contribution through application of sophisticated computational data analysis (data integration, bioinformatics, data mining) and mathematical modelling. Equally importantly, the well-orchestrated generation of high-quality matched data sets, gathered at different 'omic' levels and including frequently sampled time-series to measure response to perturbation (e.g. cytotoxic drug exposure) in appropriate models, ought to be placed high on the research agenda as a prerequisite for 'systems understanding'. Owing to the heterogeneity of cancer, this is an immense undertaking and will require a concerted international effort, not unlike the large-scale programmes associated with genome projects, and will depend on shared protocols and data standards (see Chapter 3). Validated

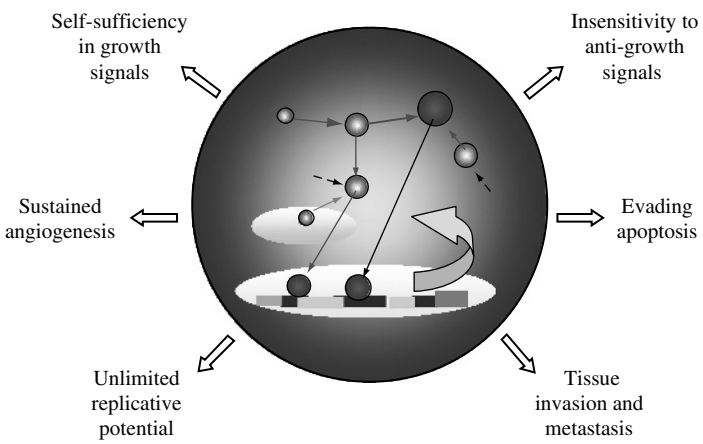


Figure 1.2 Emergence of cancer cell phenotypes. Extensively altered circuits in signal transduction networks arise through the interplay of genomic instability and selective pressure driven by host–tumour dynamics. Altered signal transduction both causes and sustains cancer cell phenotypes (together with other cell processes) (A colour reproduction of this figure can be seen in the colour section.)

quantitative and multiscale data so obtained can then be integrated and exploited through data mining and mathematical modelling.

1.2 A taxonomy of cancer complexity

The challenges posed by the complex systems properties of cancer are several-fold and can be thought about in terms of a 'taxonomy of complexity' put forward by (Mitchell, 2003, p. 4) (Figure 1.3):

- *Constitutive complexity* – organisms display complexity in structure and the whole is made up of numerous parts in non-random organization.
- *Dynamic complexity* – organisms are complex in their functional processes.
- *Evolved complexity* – alternative evolutionary solutions to adaptive problems, historically contingent.

Constitutive complexity

A central insight of systems biology is that no individual component is likely to be uniquely responsible for governing a cellular response (Prudhomme *et al*., 2004). The collective effects of mutations that lead to tumour development arise in the context of complex genetic and signal transduction networks. In cancer, extensively altered network circuits often give rise to non-intuitive cellular phenotypic outcomes because of feedback loops and cross-talk between pathways. Dependence on biological context and dynamic interconnectedness is at the core of biological function. Critically, in order to advance treatment strategies through the identification of more effective targets, analysis must be aimed at the discovery of functional links between (multiple) cell components and processes at different levels of organization (Hanash, 2004); cellular

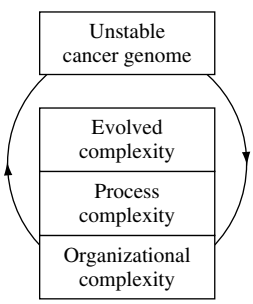


Figure 1.3 Taxonomy of cancer complexity. The inter-relationship between genome instability and the three types of complexity within cellular evolution in cancer is highlighted

networks and functional systems must be studied in multivariate mode (Prudhomme *et al*., 2004). A predictive understanding of cancer cells and their response to treatment requires a framework that can relate underlying genome structure and molecular circuitry to time-varying expression profiles and cellular phenotypes in a mathematically rigorous manner (Figure 1.4) (see Begley and Samson, 2004; Christopher *et al*., 2004; Eungdamrong and Iyengar, 2004; Khalil and Hill, 2005).

The role of biomolecular networks in cancer systems biology

Metabolic and signal transduction networks are located midway between the genome and the phenotype, and can be conceptualized as an 'extended genotype' or 'elementary phenotype' (Huang, 2004). Thus, these networks provide a stepping stone for the integrative study of gene function in complex living systems and are a major focus of systems biology.

Network biology, a distinct research area within systems biology, addresses the aspect of topology (or 'wiring') and seeks to identify organizational rules underlying large-scale topologies of cell networks that can provide insights into pathway and network function. For example, protein networks contain highly connected hub proteins that have been shown to correlate with evolutionarily conserved proteins, and in yeast with proteins encoded by essential genes (Jeong *et al*., 2001). Another challenge is to understand how representations of signalling networks can be expanded to include

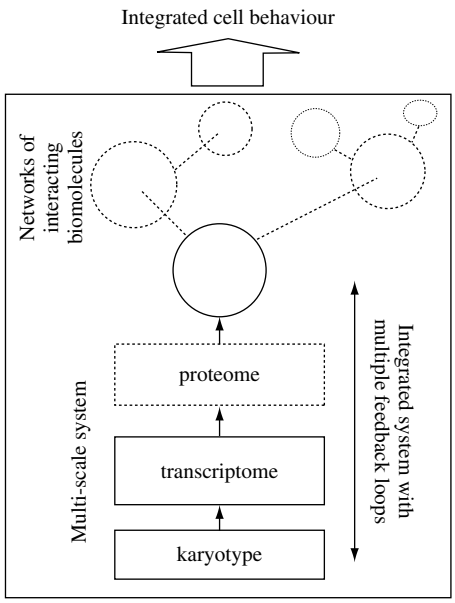


Figure 1.4 Vertically integrated cell framework

other regulatory networks, e.g. metabolic, gene expression and cytoskeletal networks, and how cell signalling networks can be integrated into the larger networks of interacting cells, tissues and physiological systems. Systems biology then aims to formalize dependencies between network topology and dynamic behaviour, with the goal of ultimately linking dynamic network behaviour to cell function. Research in this area is growing rapidly and the reader is referred to the body of literature. A good starting point is the *FEBS Letters* special issue 'Systems Biology Understanding the Biological Mosaic', 21 March 2005 (Vol. 579, Issue 8).

Shared characteristics exhibited by networks of interacting agents ranging from cellular networks, ecological systems to the Internet suggest a common logic in their function, in terms of their connectivity and dynamics. As already mentioned, robust systems are able to maintain their function in the presence of certain perturbations (such as those frequently encountered), but are often vulnerable to other types of perturbations (such as those they are rarely exposed to). In general, cells are highly robust to uncertainty in their environments and the failure of component parts, yet can be disabled catastrophically by even small perturbations to certain genes (mutation, dosage change), trace amounts of toxins (drugs) that disrupt the structural elements or regulatory control networks or inactivation of essential network components. Cancer cells reconfigure normal cellular networks to establish a pathological kind of robustness, including evasion of apoptosis and treatment resistance in response to selection pressure through anti-cancer drugs or radiation therapy (Albert, Jeong and Barabási, 2000; Barabási and Oltvai, 2004; Cork and Purugganan, 2004; Galitski, 2004; Kitano, 2004; Papin and Subramaniam, 2004; Papin *et al*., 2005).

Dynamic complexity

Biological complexity has become associated more recently with non-linear mathematical functions representing processes in space and time. Process complexity is linked to a range of dynamic characteristics such as sensitivity to initial conditions, discontinuous change (bifurcation), self-organization and negative and positive feedback control. Striking generalities in the models of complex dynamic processes found in chemical and physical systems have led to their increasing application to biological systems (von Bertalanffy, 1968; Holland, 1995; Mitchell, 2003).

To study the emergent properties of cell behaviour in relation to the function of genes it is necessary to: interpret gene expression at the level of the transcriptome and the proteome within the topology of gene regulatory and protein interaction networks; and go beyond network topology and address the global dynamics of networks that will reveal the collective behaviour of the interacting gene products (Huang, 2004). Linking gene expression to pathway dynamics is critical as, for example, the concentrations of signalling proteins can have a very significant quantitative influence on the outcome of signal transduction (see Section 1.3 and Chapter 4). There is ample evidence that the extent of cell surface receptor expression can determine whether a cell enters

the cell cycle, arrests growth or undergoes apoptosis, and the concentration of members of signal pathways downstream of receptors can also have profound effects. Overexpression of MAPKK or MAPK beyond a certain optimal level can lead to signal inhibition rather than signal enhancement (Levchenko, 2003). A number of dynamic models have been developed already for well-characterized pathways such as the epidermal growth factor (EGF) and the MAP kinase pathways (Bhalla and Iyengar, 1999, 2001; Asthagiri and Lauffenburger, 2001; Schoeberl *et al.*, 2002; Resat *et al.*, 2003) (see Chapter 4 and also www.cellml.org/examples/repository/index.html for further models in CellML format and the extensive primary literature). In addition, Table 1.2 lists various collaborative projects of interest for cancer systems biology. Dynamic pathway models may represent theorized or validated pathways and need to have kinetic data attached to every connection – this enables one to simulate the change in concentrations of the components of the pathway over time when given the initial parameters. Using standard principles of biochemical kinetics, a complex regulatory network can be cast as a set of non-linear differential equations according to the network topology and the types of protein–protein interactions present. Using a basal parameter set, the equations are then solved numerically. However, for many pathways that are highly relevant to cancer the available data are far too incomplete for modelling, which again highlights the necessity for systematic generation of comprehensive data sets (including interaction and activation kinetics).

Evolved complexity

New insights also may be gained by approaching the subject of cancer within an evolutionary framework. In complex adaptive systems, the regularities of experience are encapsulated in highly compressed form as a model or schema (Holland, 1995). An agent (cancer cell in the present context) must create internal models by selecting patterns in the input it receives and then convert these patterns into changes in its internal structure. Schemata can change to produce variants that can compete with each other and selection will act on the agents' internal schemata. Changes can be either gradual or sudden, and success is measured by survival.

Table 1.2 A selection of international systems biology initiatives with relevance to cancer

Initiative	URL
Alliance for Cellular Signaling	www.signaling-gateway.org
E-Cell	www.e-cell.org, ecell.sourceforge.net
Institute for Systems Biology (Seattle)	www.systemsbiology.org
Systems Biology Institute (Tokyo)	www.systems-biology.org/index.html
Computational and Systems Biology (MIT)	csbi.mit.edu
TUMATHER	calvino.polito.it/~mcrtn

Cells are linked to their environments through feedback loops that enable adaptive modification and reorganization. Selection acts on cancer cells and selects for altered internal schemata (genome mutations and changes in cellular network structure and dynamics) that form the basis of altered signal processing by intracellular networks and the abnormal cancer phenotype (Hanahan and Weinberg, 2000) (Figures 1.2 and 1.3). Change in cell function emerges from gradual accumulation of small alterations (multiple mutations over extended time) or simultaneous large-scale change (aneuploidy).

Progression from normal tissue to malignancy is associated with the evolution of neoplastic cell lineages with multiple genetic lesions that are not present in the normal tissues from which the cancers arose. *Cellular* evolution, at a vastly accelerated rate and guided by natural selection, transforms normal cells into malignant cells. Multiple neoplastic clones may coexist and compete with each other for resources and space during the progression to malignancy. In this evolutionary process neoplastic cells develop genome-wide instability and variants are selected, leading to the emergence of clonal populations with multiple genomic abnormalities and selective proliferative advantages, including, for example, the evasion of cell death and anti-cancer treatment resistance. This can be exacerbated by exerting selective pressure through exposure to therapeutic agents (Nowell, 1976; Novak, Michor and Iwasa, 2003; Maley *et al*., 2004).

Genomes are dynamic entities at evolutionary and developmental time-scales. In cancer, dynamic structural rearrangements occur at dramatically increased frequency – an unstable genome is a distinguishing characteristic of most types of cancer (Nygren and Larsson, 2003; Vogelstein and Kinzler, 2004). In addition to mutations in individual oncogenes and tumour suppressors, extensive gross chromosomal change (aneuploidy, which is quantitatively measurable through cytogenetic analysis, including new high-throughput chip-based methods) is observed in liquid and nearly all solid tumours. The most common mutation class among the known cancer genes is chromosomal. Copy-number changes, such as gene amplification and deletion, can affect several megabases of DNA and include many genes. These large-scale changes in genome content can be advantageous to the cancer cell by simultaneous activation of oncogenes, elimination of tumour suppressors and the production of variants that can rapidly evolve resistance to drug exposure.

Given the irreversible nature of evolutionary processes, the randomness of mutations relative to those processes and the modularity by which complex wholes are composed of simpler parts, there exists in nature *a multitude of ways to 'solve' the problems of survival and reproduction* (Mitchell, 2003, p. 7). Because each patient's cancer cells evolve through an independent set of mutations and selective environments, the resulting cell population in each patient will be heterogeneous and will exhibit certain unique features. The fact that the population of cells includes significant heterogeneity means that they will be unlikely to respond to therapy in a uniform manner and that most treatments will not eradicate all the cells. Furthermore, this also implies that we are unlikely to find general treatments that will work for all or even most patients.

These challenges, which in significant part arise from the processes of cellular evolution decoupled from controls normally operating in multicellular organisms (Buss, 1987; quote in Gould, 2002, p. 696), have given rise to the new field of

'pharmacogenomics', which has as its ultimate aim the design of individualized treatments based on a patient's (and his or her tumour's) molecular characteristics. Motivated by an evolutionary perspective on the complexity of cancer, the methods of systems biology can be applied to address three fundamental questions underlying pharmacogenomics. Firstly, can we discover key features of the 'evolutionary logic' of cancer cell and tumour *systems* emerging from the interplay between the unstable cancer genome, higher level cellular systems and the tumour microenvironment (including exposure to drugs)? A systems-based approach to finding answers to this question extends biomarker identification and molecular profiling as presently practised, because its aim would be not only to show statistical dependence relationships between a small number of markers and high-level physiological phenomena, but to provide explanatory power in terms of biological process. One of the challenges involved is to develop methods for integrative analysis encompassing different levels of 'omic space' within cells (Toyoda and Wada, 2004) and selection dynamics within the tumour microenvironment. This is a tall order and progress also will involve innovative application of established methods, such as multivariate techniques, Bayesian networks, cellular automata and agent-based modelling for example, and integration of models representing different aspects of cell and tumour biology (see Section 1.3 regarding the requirement for prediction and modelling from vertically integrated data sets). The second, and of course related, question concerns a formalized methodology for the discovery of system *vulnerabilities* from investigations of this kind. Here, general systems theory and control systems engineering are already finding useful cross-disciplinary application (Ogunnaike and Ray, 1995). The third question, which also requires extensive multidisciplinary attention, relates to the major scientific, medical and social changes that will be precipitated by the integration of systems-based pharmacogenomics in preclinical therapy development, clinical trials and clinical practice (see also Chapters 3 and 14).

1.3 Modelling and simulation of cancer systems

Increasing use of mathematics is inevitable as biology becomes more complex and more quantitative, as has been stated very eloquently by Murray *et al*. (1998):

> 'We suggest that mathematics, rather theoretical modeling, must be used if we ever hope to genuinely and realistically convert an understanding of the underlying mechanisms into a predictive science. Mathematics is required to bridge the gap between the level on which most of our knowledge is accumulating (…cellular and below) and the macroscopic level of the patterns we see. In wound healing and scar formation, for example, a mathematical approach lets us explore the logic of the repair process. Even if the mechanisms were well understood – and they certainly are far from it at this stage – mathematics would be required to explore the consequences of manipulating the various parameters associated with any particular scenario. In the case of such

> things as wound healing – and now in angiogenesis with its relation to possible cancer therapy – the number of options that are fast becoming available to wound and cancer managers will become overwhelming unless we can find a way to simulate particular treatment protocols before applying them in practice…. The very process of constructing a mathematical model can be useful in its own right. Not only must we commit to a particular mechanism, but we are also forced to consider what is truly essential to the process, the central players (variables) and mechanisms by which they evolve. *We are thus involved in constructing frameworks on which we can hang our understanding. The model equations, the mathematical analysis and the numerical simulations that follow serve to reveal quantitatively as well as qualitatively the consequences of that logical structure'* [italics added].

The translation of highly detailed knowledge of the molecular changes in cancer into new treatments requires a synthesis of knowledge and data only attainable through computational methods. Eventually, the predictive power of mature models of cancer systems may greatly enhance target identification, therapy development, diagnostics and treatment by focusing attention on particular molecules and pathways, while avoiding unnecessary tests and procedures.

Mathematical modelling provides a formal language for the expression of complex biological knowledge, assumptions and hypotheses in a form amenable to logical analysis and quantitative testing. This is increasingly necessary as the scope and depth of information and knowledge, with the accompanying uncertainty, surpass the analytical capabilities of the unaided human mind (Swanson *et al*., 2003; Rao, Lauffenburger and Wittrup, 2005). Computational models, by their nature, serve as repositories of the current knowledge, both established and hypothetical (Figure 1.1).

Within the knowledge discovery cycle, mathematical modelling can make a major contribution to *hypothesis-driven* research (Swanson, True and Murray, 2003) (Figure 1.1): isolation of key steps in the process under study (drawing on prior experimental results and domain knowledge); formulation of a model mechanism (equations) that reflects these key elements and involves actual biological quantities; mathematical investigation of the theoretical model and generation of solutions with biologically realistic boundary and initial conditions; *and*, iteratively, in the light of the theoretical results, return to the biology with predictions and suggestions for illuminating experiments that will help to elucidate the underlying mechanisms. Models can be especially useful if they are designed to represent competing mechanisms proposed by different sources, so that a set of criteria allowing one to distinguish between different hypotheses can be formulated based on the underlying computational predictions (Levchenko, 2003). Alternatively, *data-driven* approaches include the application of data mining technologies to large-scale 'omic' data sets in order to identify key molecular features and correlations between system components, and subsequently 'reverse-engineer' models from the observed data (Figure 1.1).

Vertical genomics: data mining and systems modelling in tandem

Mining of the large amounts of data obtained through omic technologies, already an essential methodology for contemporary target discovery, will become even more critical for systems-based discovery. Data mining seeks new knowledge via an iterative execution of several knowledge discovery steps. Each step focuses on a specific discovery task that is accomplished through the application of a suitable discovery technique. Neural networks, decision trees, Bayesian techniques, hierarchical and fuzzy clustering and classical statistics are commonly applied (Brenner and Duggan, 2004; Prendergast, 2004) (the reader is also referred to the very large literature on data mining, e.g. for DNA microarray data). Systems-based discovery faces an urgent challenge because available techniques will need to be tested rigorously and, if necessary, extended for application to increasingly more complex, particularly *multiscale*, data sets generated by systems biology. A particular challenge is posed by the need for software tools that can effectively visualize, analyse and model both the functional and dynamic relationships between genome structure, expression and dynamic cell processes. Integrative *in silico* environments are needed that can jointly deploy data mining tools and mathematical modelling of pathway, cell and, eventually, tumour dynamics.

This vision lies at the heart of the Systems Complexity Interface for *path*ways (SCI*path*) project, which delivers an object-oriented framework acting as an integrative hub together with data mining, modelling and visualization tools and Systems Biology Markup Language (SBML)-enabled software connectivity (Table 1.3). The SCI*path* project is specifically designed to facilitate the exploitation of data sets that are vertically matched across 'omic space' (Toyoda and Wada, 2004) and may include karyotype, transcriptome, proteome and cell physiology data (Figure 1.5a). Several object-oriented analysis and visualization tools for vertically integrated analysis of cell signalling have been built already and new *java* tools tailored to user needs can be integrated easily with existing features. Currently implemented tools include (Figures 5b–5e):

- Custom-designed pathway mapping, automated layout and pathway merging.
- Easy upload of SBML-compliant pathways (see also Section 1.4).
- Pathway sharing with other SBML-compliant applications (e.g. Virtual Cell, E-Cell, Gepasi).
- Links to external databases facilitating bioinformatics analysis of pathway nodes.
- Data normalization and statistical testing for gene expression microarrays.
- Analysis of gene expression on pathways (customized for Affymetrix data and also applicable to dual-channel technology).
- Interactive visualization.
- Visualizations can be overlaid with other data types, e.g. gene copy number and proteomics data.

Table 1.3 A selection of open access tools for cancer systems biology[a]

Software	Platform	GO	SBML	Other standards	Linkage to SBW-powered simulator modules	Main application areas	URL
Systems Biology Workbench (SBW)	Windows, FreeBSD and Linux (with MacOS X planned)	N/A	Yes	N/A	Linkage hub	Modular, broker-based, message-passing framework for simplified communication between applications	sbw.sourceforge.net
Over 80 SBML-compliant software systems	Various	Check specific software	Yes	Check specific software	Check specific software	Dynamic pathway and cell modelling	sbml.org/index.psp
SCI*path*	Platform-independent *java* application	Yes	Yes	MAGE-ML in version 2.0 (under development)	Yes	See main text, Section 1.3	www.scipath.org.uk
Cell Designer (also integrated in PANTHER, Applied Biosystems)	Platform-independent *java* application		Yes		Yes	Structured diagram editor for drawing networks	www.celldesigner.org
Bioconductor (run with 'R')	Unix, Linux, MacOS X, Windows	Yes		MAGE-ML; XML annotation for genomic data		Powerful statistical and graphical methods for the analysis of genomic data; integration of biological metadata; analysis of graphical structures in computational biology (graph, Rgraphviz and RBGL packages)	www.bioconductor.org

[a] See text for further details.

Sources: Funahashi, A., Tanimura, N., Morohashi, M. and Kitano, H. 2003. CellDesigner: a process diagram editor for gene-regulatory and biochemical networks, *BIOSILICO* **1**: 159–162; Carey, V.J., Gentry, J., Whalen, E. and Gentleman, R. 2005. Network structures and algorithms in Bioconductor. *Bioinformatics* **21**: 135–136.

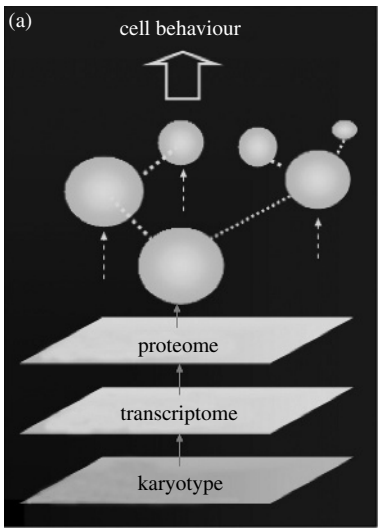
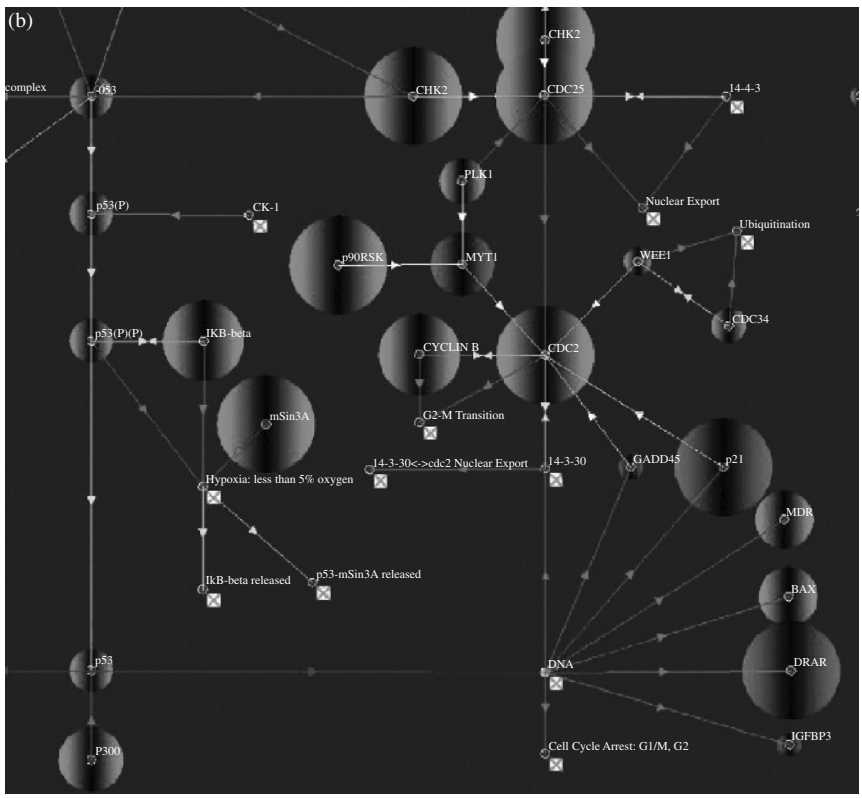
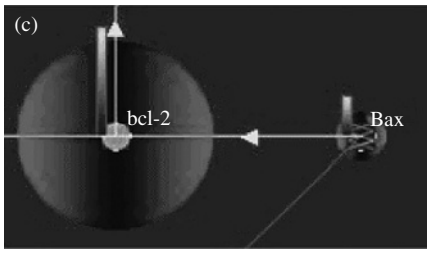


Figure 1.5 The SCI*path* project. (a) Software is specifically designed to facilitate the exploitation of vertically integrated data sets. (b) Differential gene expression ratios (relative up- or down-regulation, *relative size of turquoise and purple circles*) based on microarray data can be mapped to user-defined pathways





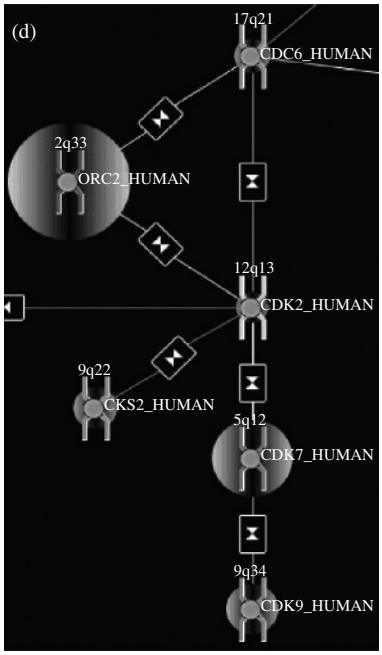
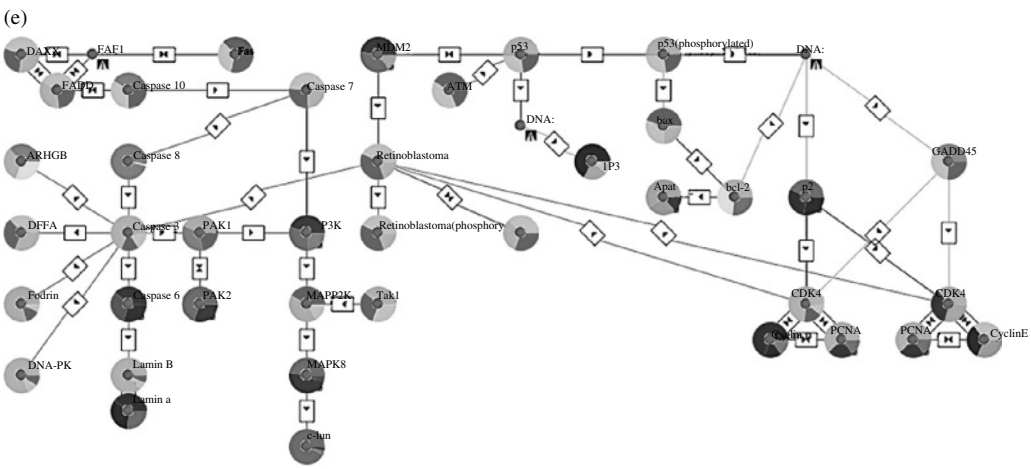


Figure 1.5 (*Continued*) (c) Visualizations can be overlaid with other data types, e.g. proteomic data (*orange bars*). (d) Genome scanning data (e.g. from array Comparative Genome Hybridization experiments, aCGH) can be mapped to pathways. The chromosomal location (single band resolution) of each node's gene locus is shown by *colour-coded stylized chromosomes* and copy number changes of associated genomic regions can be visualized (here, *size of yellow circles* represents relative increase in copy number). (e) Fuzzy *k*-means clustering can reveal complex co-expression relationships between pathway nodes dependent on biological context. The *colour-coded 'pie chart'* mapped to each node represents membership scores related to a node's three top scoring fuzzy clusters. Shared context-dependent cluster membership between nodes can be identified easily by segments of the same colour (A colour reproduction of this figure can be seen in the colour section.)



- Data mining tools written in *java* can be plugged into the SCI*path* framework to take advantage of data sharing functionality and visualization tools aiding complex visual reasoning.
- Linkage to SBW (Systems Biology Workbench)-powered simulator modules for dynamic pathway and cell system modelling.
- Fuzzy *k*-means clustering: identification of complex co-expression patterns.
- Data mining based on the Gene Ontology (GO) hierarchies.

The SBW (Sauro *et al*., 2003; Table 1.3) is one of the foremost efforts to bring data and visualization integration to bioinformatics. It provides support for a variety of different programming languages on the most popular platforms and is therefore the most powerful open-source integration package for bioinformatics to date. The SBW architecture rests on a broker service that, through providership, offers application services to other SBW-enabled modules. The user can therefore, quite seamlessly, borrow the functionalities of multiple modules without having to open up a multitude of new applications manually to get the desired result. As far as programming SBW compatibility goes, the designers have provided a simple interface-writing approach for *java* developers and ample documentation at their website. This simple yet robust approach presents the opportunity to make effective use of a wide range of otherwise quite specialized applications. Many third-party, SBW-enabled modules already exist and new programs are in development (for an up-to-date list, please see sbw.sourceforge.net/sbw/software/index.shtml).

1.4 Data standards and integration

We are currently not in a position to make maximal use of the existing or future data sets for computational analysis and mathematical modelling, because data have not yet been standardized in terms of experimental and clinical data capture (protocols, data reproducibility and quality) and computational data management (data formats, vocabularies, ontologies, metadata, exchange standards, database interoperability) (Figure 1.6). Integration of different data types, spanning the range from molecular to clinical and epidemiological data, poses another challenge.

Data integration initiatives

In order for the potential of cancer bioinformatics and *in silico* systems analysis to be fulfilled, the basic requirements are the generation of validated high-quality data sets and the existence of the various data sources in a form that is intelligible to computational analysis. This has been well recognized, as is amply demonstrated by the aims

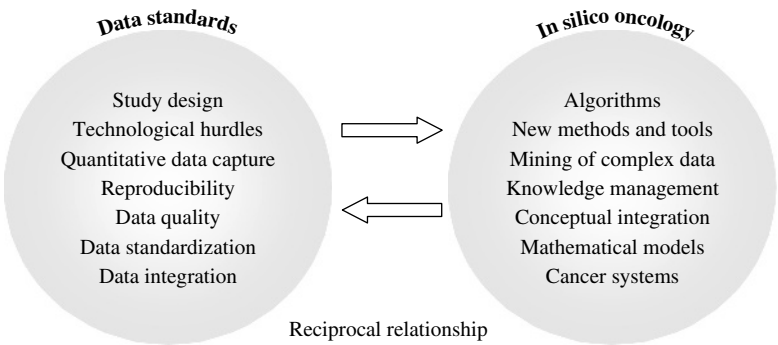


Figure 1.6 Reciprocal relationship between standard development and cancer systems biology and bioinformatics

and activities of a collaborative network of several large initiatives for data integration within the cancer domain that work towards shared aims in a coordinated fashion (the initiatives mentioned below are meant to serve as example projects and do not represent the sum total of these efforts on an international scale).

The National Cancer Institute Center for Bioinformatics (NCICB) in the USA has developed caCORE, which provides an open-source suite of common resources for cancer vocabulary, metadata and data management needs (biological and clinical), and the latest release (Version 3.0) achieves semantic interoperability across disparate biomedical information systems. It uses concepts from description logic thesauri to build up the data classes and attributes in Unified Modelling Language (UML) information models. The models are registered in a metadata registry and then turned into model-driven data management software. The caCORE Software Development Kit gives any developer the tools needed to create systems that are consistent and interoperable with caCORE (for detailed information and access to the caCORE components, see ncicb.nci.nih.gov/core). The caCORE infrastructure plays an essential integrative role for the Mouse Models of Human Cancers Consortium (see Chapter 9) and the cancer Biomedical Informatics Grid (caBIG), a voluntary network connecting individuals and institutions to enable the sharing of data and tools, creating a 'World Wide Web of cancer research' whose goal is to speed up the delivery of innovative approaches for the prevention and treatment of cancer (cabig.nci.nih.gov).

In the UK, the National Cancer Research Institute (NCRI) is developing the NCRI Strategic Framework for the Development of Cancer Research Informatics in the UK (www.cancerinformatics.org.uk/index.html; see also Chapter 3). The ultimate aim is the creation of an internationally compatible informatics platform that would facilitate data access and analysis. The NCRI Statement of Intent projects that 'enabling this sharing of knowledge across disciplines, from genomics through to clinical trials, will benefit patients and researchers by channelling the development of novel therapeutics and diagnostics in a more effective way' (published in *Nature*

and the *British Medical Journal* in March 2004). CancerGRID develops open standards and information management systems (XML, ontologies and data objects, web services, GRID technology) for clinical cancer informatics, clinical trials, multi-site development and distributed computing, integration of molecular profiles with clinical data and effective translation of clinical trials data to bioinformatics and genomics research (www.cancergrid.org). The Clinical E-Science Framework (CLEF) aims to implement a high-quality, safe and interoperable information repository derived from operational electronic patient records to enable ethical and user-friendly access to the information to support clinical care and biomedical research, and is also designing complementary information capture and language tools (www.clef-user.com).

Semantic web technologies

Using simple page layout information, the current web represents information using natural language, numerical data, graphics, multimedia, etc. in a way that often requires humans to process this information by deducing facts from partial information, creating mental associations and integrating various types of sensory information. In addition, data that a user wishes to integrate are often presented in incompatible formats and undefined nomenclature at distributed sites. In spite of these difficulties, humans can combine data reasonably easily even if different terminologies and presentation formats are used.

However, to make a global cancer data grid a reality, data need to be accessed, integrated and processed automatically by computers. Therefore, web service technology and high-bandwidth data grids need to comply with standards for the 'Semantic Web', which can be defined as a metadata-based infrastructure for reasoning (www.w3.org/2001/sw). The Semantic Web provides a common framework that allows data to be shared and reused across application, institution and community boundaries and is based on the Resource Description Framework (RDF), which integrates a variety of applications using XML for syntax.

Within this framework, the data resource provides information about itself, i.e. metadata, in a machine-processable format, and an agent accessing the resource should be able to reason about the (meta)data. To make metadata machine-processable, a common data model for expressing metadata (i.e. RDF) and defined metadata vocabularies and concept relationships are needed.

Ontologies for translational cancer research

Ontologies (formal representations of vocabularies and concept relationships) and common data elements based on these definitions are prerequisites for successful data integration and interoperability of distributed data sources. Various standard vocabularies and object models have been developed already for genomics, molecular profiles,

certain molecular targeted agents, mouse models of human cancer, clinical trials and oncology-relevant medical terms and concepts (SNOMED-RT/CT, ICD-O-3, MeSH, CDISC, NCI Health Thesaurus, caCORE, HUGO). There are also existing ontologies describing histopathology (standards and minimum data sets for reporting cancers, Royal College of Pathologists; caIMAGE, National Cancer Institute). The European Bioinformatics Institute (EBI) is developing standards for the representation of molecular function (Gene Ontology) and the Microarray Gene Expression Data (MGED) Society is developing MIAME, MAGE and the MAGE ontology, a suite of standards for microarray users and developers including an object model, document exchange format, toolkit and an ontology. However, significant gaps still exist and eventually all cancer-relevant data types (see the NCRI Planning Matrix, www.cancerinformatics.org.uk/planning_matrix.htm) will need to be formalized in ontologies. These efforts are ongoing and pursued by a large community of researchers (see above and ftp1.nci.nih.gov/pub/cacore/ExternalStds for further details on available standards).

Protégé-2000 Protégé is a freely available tool that allows users to construct domain ontologies, customize data entry forms and enter data. It is also a platform that can be extended easily to include graphs and tables, media such as sound, images and video, and various storage formats such as OWL, RDF, XML and HTML (protege.stanford.edu/index.html). Protégé is a mature technology and it is especially appropriate for knowledge acquisition from domain experts and the design of sharable ontologies because of its emphasis on flexibility and extensibility. Protégé supports the development of knowledge bases in a fashion that facilitates the reuse of encoded knowledge for a variety of purposes.

The OWL format unifies frame and description logics into one language. Its encoding to RDF schema makes it a semantic metadata language for the web and it supports the goals of the Semantic Web initiative for languages, expressing information in a machine-processable form (www.w3.org/TR/owl-features). By offering these capabilities, OWL is establishing itself as the current state-of-the-art ontology exchange language. It facilitates greater machine interpretability of web content than that supported by XML, RDF and RDF Schema (RDF-S) by providing vocabulary along with a formal semantics.

XML exchange standards for pathways and models

An increasing number of model building tools include integrated databases of genomic, proteomic and/or other information, or provide close links to such data, and these need to be standardized for input into models; XML exchange standards are being developed in areas such as transcriptomics (e.g. MAGE-ML) and proteomics (e.g. PSI, PEDRO, BioPAX), which will enable increased efficiency and automation of data use.

Information standards are also needed if the models themselves are to be shared, evaluated and developed cooperatively. A uniform Systems Biology Markup

Language (SBML) has therefore been developed to facilitate data and model exchange, and closely allied initiatives are also underway (Table 1.3). SBML is a computer-readable format for representing models of biochemical reaction networks, and is applicable to metabolic networks, cell-signalling pathways, regulatory networks and many others (sbml.org/index.psp). It is currently supported by over 80 software systems and its widespread adoption enables the use of multiple tools without rewriting network models for each tool, supports network model sharing between different software environments and ensures the survival of models beyond the lifetime of the software used to create them. The purpose of CellML is to store and exchange computer-based mathematical models and it includes information about model structure (how the parts of a model are organizationally related to one another), mathematics (equations describing the underlying processes) and metadata (additional information about the model that allows scientists to search for specific models or model components in a database or other repository) (http://www.cellml.org/public/about/what_is_cellml.html). CellML includes mathematics and metadata by leveraging existing languages, including MathML (http://www.w3.org/Math/) and RDF. AnatML is aimed at exchanging information at the organ level, and FieldML is appropriate for storing geometry information inside AnatML, the spatial distribution of parameters inside compartments in CellML or the spatial distribution of cellular model parameters across an entire organ (http://www.cellml.org/public/about/what_is_cellml.html).

1.5 Concluding remarks

Cancer systems biology seeks to elucidate complex cell and tumour behaviour through the integration of many different types of information. Enhanced understanding of how genome instability and complex interactions within cells and tissues give rise to cancer, and its confounding heterogeneity, through a hierarchy of biochemical and physiological systems is expected to improve prevention, diagnosis and treatment. Advanced experimental technologies and computational methods need to be applied together in mutually complementary fashion to address the challenges ahead. The classical techniques of statistics and bioinformatics for analysis of the genome, biological sequences, large-scale ‘omic’ data sets and protein three-dimensional structure will continue to form an indispensable backbone for computational cancer research, whereas new systems-based approaches will extend our knowledge of the organization and dynamic functioning of the implicated biological systems. Cancer systems biology is already addressing pressing challenges in the development of new anti-cancer therapies and is poised to take an even more leading role in our quest for deeper insights into the biological complexity of cancer. Complementing the methods of systems biology, new data management technologies to enable the integration and sharing of data and models are also a prerequisite for advancement.

Acknowledgements

The author gratefully acknowledges funding support from the National Translational Cancer Research Network (NTRAC), the Medical Research Council and Cancer Research UK.

References

Albert, R., Jeong, H. and Barabási, A. L. 2000. Error and attack tolerance in complex networks. *Nature* **406**: 378–382.

Asthagiri, A. R. and Lauffenburger, D. A. 2001. A computational study of feedback effects on signal dynamics in a mitogen-activated protein kinase (MAPK) pathway model. *Biotechnol Prog*. **17**: 227–239.

Barabási, A. L. and Oltvai, Z. N. 2004. Network biology: understanding the cell's functional organization. *Nature Rev. Genet*. **5**: 101–114.

Begley, T. J. and Samson, L. D. 2004. Network responses to DNA damaging agents. *DNA Repair* **3**: 1123–1132.

Bhalla, U. S. and Iyengar, R. 1999. Emergent properties of networks of biological signaling pathways. *Science* **283**: 381–387.

Bhalla, U. S. and Iyengar, R. 2001. Functional modules in biological signalling networks. *Novartis Found. Symp*. **239**: 4–13.

Birney, E., Ciliberto, A., Colding-JØrgensen, M., Goldbeter, A., Hohmann, S., Kuiper, M., *et al*. 2005. *EU Projects Workshop Report on Systems Biology*. Available at http://www.sbi.uni-rostock.de/dokumente/csbworkshop 2005 01 en.pdf.

Bogdanovic, S. and Langlands, B. 2004. *Systems Biology: the future for Integrated Drug Discovery*. Available at http://www.pibpubs.com/scrip reports/systems biology.

Brenner, C. and Duggan, D. 2004. *Oncogenomics: Molecular Approaches to Cancer*. John Wiley: Hoboken, NJ.

Buss, L. W. 1987. *The Evolution of Individuality*. Princeton University Press: Princeton, NJ.

Christopher, R., Dhiman, A., fox, J., Gendelman, R., Haberitcher, T., Kagle, D., *et al*., 2004. Data-driven computer simulation of human cancer cell. *Ann. NY Acad. Sci*. **1020**: 132–153.

Cork, J. M. and Purugganan, M. D. 2004. The evolution of molecular genetic pathways and networks. *BioEssays* **26**: 479–484.

Csete, M. E. and Doyle, J. C. 2002. Reverse engineering of biological complexity. *Science* **295**: 1664–1669.

Depew, D. J. and Weber, B. H. 1996. *Darwinism Evolving: Systems Dynamics and the Genealogy of Natural Selection*. MIT Press: Cambridge, MA.

Eungdamrong, N. J. and Iyengar, R. 2004. Modeling cell signaling networks. *Biol. Cell* **96**: 355–362.

Galitski, T. 2004. Molecular networks in model systems. *Annu. Rev. Genom. Hum. Genet*. **5**: 177–187.

Gould, S. J. 2002. *The Structure of Evolutionary Theory*. Belknap Press: Cambridge, MA.

Hanahan, D. and Weinberg, R. A. 2000. The hallmarks of cancer. *Cell* **100**: 57–70.

Hanash, S. 2004. Integrated global profiling of cancer. *Nat. Rev. Cancer* **4**: 638–644.

Holland, J. H. 1995. *Hidden Order: How Adaptation Builds Complexity*. Helix Books: New York.

Holland, J. H. 1998. *Emergence: From Chaos to Order*. Addison-Wesley: Redwood City, CA.

Huang, S. 2004. Back to the biology in systems biology: what can we learn from biomolecular networks? *Briefi. Funct. Genom. Proteom.* **2**: 279–297.

Ideker, T. and Lauffenburger, D. 2003. Building with a scaffold: emerging strategies for high- to low-level cellular modeling. *Trends Biotechnol*. **21**: 255–262.

Jeong, H., Mason, S. P., Barabási, A. L. and Ottrai, Z. N. 2001. Lethality and centrality in protein networks. *Nature* **411**: 41–42.

Khalil, I. G. and Hill, C. 2005. Systems biology for cancer. *Curr. Opin. Oncol*. 17: 44–48.

Kitano, H. 2004. Biological robustness. *Nat. Rev. Genet*. **5**: 826–837.

Levchenko, A. 2003. Dynamic and integrative cell signaling: challenges for the new biology. *Biotechnol. Bioeng*. **84**: 773–782.

Maley, C. C., Galipeau, P. C., Li, X., Sanchez, C.A., Paulson, T. G., Blount, P. L. and Reid, B. J. 2004. The combination of genetic instability and clonal expansion predicts progression to esophageal adenocarcinoma. *Cancer Res*. **64**: 7629–7633.

Mitchell, S. D. 2003. *Biological Complexity and Integrative Pluralism*. Cambrige University Press: Cambridge.

Murray, J. D. 2002. *Mathematical Biology*. Springer Verlag: New York.

Murray, J. D., Cook, J., Tyson, R. and Lubkin, S.R. 1998. Spatial pattern formation in biology, I: dermal wound healing; II: bacterial patterns. *J. Franklin Inst*. **335B**: 303–332.

Novak, M. A., Michor, F. and Iwasa, 2003. The linear process of somatic evolution. *Proc. Natl. Acad. Sci*. **100**: 14966–14969.

Nowell, P. C. 1976. The clonal evolution of tumor cell populations. *Science* **194**: 23–28.

Nygren, P. and Larsson, R. 2003. Overview of the clinical efficacy of investigational anti-cancer drugs. *J. Intern. Med*. **253**: 46–75.

Ogunnaike, B. and Ray, W. H. 1995. *Process Dynamics, Modeling, and Control*. Oxford University Press: Oxford.

Papin, J. and Subramaniam, S. 2004. Bioinformatics and cellular signaling. *Curr. Opin. Biotechnol*. **15**: 78–81.

Papin, J. A., Hunter, T., Palsson, B. O. and Subramaniam, S. 2005. Reconstruction of cellular signalling networks and analysis of their properties. *Nat. Rev. Mol. Cell Biol*. **6**: 99–111.

Prendergast, G. C. 2004. *Molecular Cancer Therapeutics*. John Wiley: Hoboken, NJ.

Prudhomme, W., Daley, G. Q., Zandstra P. and Lauffenburger, D. A. 2004. Multivariate proteomic analysis of murine embryonic stem cell self-renewal versus differentiation signaling. *Proc. Natl. Acad. Sci*. **101**: 2900–2905.

Rao, B. M., Lauffenburger, D. A. and Wittrup, K. D. 2005. Integrating cell-level kinetic modeling into the design of engineered protein therapeutics. *Nature Biotechnol*. **23**: 191–194.

Resat, H., Ewald, J. A., Dixon, D. A. and Wiley, H. S. 2003. An integrated model of epidermal growth factor receptor trafficking and signal transduction. *Biophys. J*. **85**: 730–743.

Sauro, H. M., Hucka, M., Finney, A., Wellock, C., Bolouri, H., Doyle, J. and Kitano, H. 2003. Next generation simulation tools: the Systems Biology Workbench and BioSPICE integration. *OMICS* **7**: 355–372.

Schoeberl, B., Eichler-Jonsson, C., Gilles, E. D. and Muller, G. 2002. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nature Biotechnol*. **20**: 370–375.

Swanson, K. R., True, L. D. and Murray, J. D. 2003. On the use of quantitative modeling to help understand prostate-specific antigen dynamics and other medical problems. *Am. J. Clin. Pathol*. **119**: 1 (editorial).

Swanson, K. R., Bridge, C., Murray, J. D. and Alvord, E. C. 2003. Virtual and real brain tumors: using mathematical modeling to quantify glioma growth and invasion. *J. Neurol. Sci*. **216**: 1–10.

Toyoda, T. and Wada, A. 2004. 'Omic space': coordinate-based integration and analysis of genomic phenomic interactions. *Bioinformatics* **20**: 1759–1765.

Vogelstein, B. and Kinzler, K. W. 2004. Cancer genes and the pathways they control. *Nat. Med*. **10**: 789–799.

von Bertalanffy, L. 1968. *General Systems Theory*. George Braziller: New York.

Workman, P. 2003. Strategies for treating cancers caused by multiple genome abnormalities: from concepts to cures? *Curr. Opin. Invest. Drugs* **4**: 1410–1415.