

# Chapter 12

## Cancer Systems Biology

**Dana Faratian, James L. Bown, V. Anne Smith, Simon P. Langdon,  
and David J. Harrison**

### Abstract

Cancer is a complex and heterogeneous disease, not only at a genetic and biochemical level, but also at a tissue, organism, and population level. Multiple data streams, from reductionist biochemistry in vitro to high-throughput “-omics” from clinical material, have been generated with the hope that they encode useful information about phenotype and, ultimately, tumour behaviour in response to drugs. While these data stand alone in terms of the biology they represent, there is the enticing prospect that if incorporated into systems biology models, they can help understand complex systems behaviour and provide a predictive framework as an additional tool in understanding how tumours change and respond to treatment over time. Since these biological data are heterogeneous and frequently qualitative rather than quantitative, at the present time a single systems biology approach is unlikely to be effective; instead, different computational and mathematical approaches should be tailored to different types of data, and to each other, in order to test and re-test hypotheses. In time, these models might converge and result in usable tractable models which accurately represent human cancer. Likewise, biologists and clinicians need to understand what the requirements of systems biology are so that compatible data are produced for computational modelling. In this review, we describe some theoretical approaches (data-driven and process-driven) and experimental methodologies which are being used in cancer research and the clinical context where they might be applied.

**Key words:** Systems biology, Cancer, S-systems, Bayesian networks, Targeted therapeutics, Oncology

---

## 1. Introduction

### **1.1. What Is Cancer, and Why Do We Need Cancer Systems Biology?**

Cancer is the term applied to over 200 different diseases in which cells acquire a set of characteristic biological properties, namely autonomous growth, evasion of death, and the ability to invade and spread to distant sites (metastasise) (1). The underlying cause of cancer is genetic, with either inherited or acquired abnormalities of genes or the control of genes giving rise to the cancerous phenotype (2).

While cancer can arise from any cell in the body, the commonest cancers in man (such as breast and lung, which together account for over 83,000 new cases in the UK per annum) arise from the epithelial cells, which line the cavities, ducts and surfaces of the body, and are called carcinomas (to distinguish them from sarcomas, which are rarer cancers arising from mesenchymal cells such as muscle cells or vascular endothelial cells). The ready isolation and growth properties of some human carcinoma epithelial cells *in vitro* have made them excellent experimental models to study cancer for several decades, and in some cases represent patterns of genomic aberration in human disease (3), but it is also recognised that cell line models are imperfect representations of complex phenotypes *in vivo*. This is partly because cancer cell lines may carry more complex genetic abnormalities than those seen *in vivo* as part of their acquired ability to survive *in vitro*. Also, simple 2D cultures lack the cells which normally support cancerous epithelium, such as stroma and blood vessels, which are intrinsic to the tumour *in vivo*, and these cellular models therefore fail to represent the complex interplay between epithelium and stroma which can influence both how cancers form and how they respond to treatment (1). Although this additional complexity leaves theoreticians and experimentalists with the uncomfortable prospect of not only trying to understand epithelial biology but also complex tissue biology (and indeed interactions with the organism as a whole), cancer cells do not exist in isolation and the growth, death and invasive phenotypes seem exquisitely sensitive to the spatial context. Therefore, tumours do not grow without new blood vessels, and invasion does not occur without the degradation of the extracellular matrix. Furthermore, metastases in distant sites (such as lymph nodes or bone marrow) may have different sensitivity to therapy than primary tumour. A further problem is that the cancer and its environment continually evolve and change over time, particularly in the face of therapeutic insult. These spatial and temporal challenges need to be considered in advancing our approach to understanding cancer and responses to therapy.

In clinical practice, the diagnosis of cancer is rarely a problem, with histopathology still the gold standard by the microscopic examination of stained sections of tissue. Traditionally, the objective of histopathology has been to categorise and classify disease by grade (how closely the tumour cells resemble their normal counterparts), or stage (how far the tumour has spread), because this has prognostic value (4). More recently, however, there has been a move to stratify patients for optimal therapy on the basis of molecular biomarkers (5), so that the appropriate drug can be given to the patient. Thus, those who are not likely to respond should be spared ineffective therapy. The commonest clinical setting where this applies is in breast cancer, where estrogen receptor (ER) and HER2 protein or DNA copy number are measured

by immunohistochemistry or Fluorescence In Situ Hybridisation (FISH), to identify patients who should be given endocrine therapy (tamoxifen or aromatase inhibitors) and trastuzumab (Herceptin), respectively (6). Nevertheless, these markers, while very good at excluding patients who will not respond to therapy (high negative predictive value), are poor at identifying patients who will respond to therapy (low positive predictive value), either because those tumours are intrinsically resistant to that therapy or they develop resistance over time. The empirical approach to overcome this problem has been to measure more biological variables (usually using gene expression microarrays) and calculate statistical associations with disease outcome, but to date the clinical usefulness has been disappointing. We have already discussed the reasons for this situation (5, 7), but essentially the current approach to translational research by the analysis of candidate biomarkers, even within large trials, requires re-evaluation. The emerging evidence indicates that a failure to recognise the dynamic properties of signalling can result in costly mistakes. For example, a loss of feedback inhibition in tumours treated with the mammalian target of rapamycin (mTOR) inhibitors results in the induction of AKT signalling, and may be responsible for the disappointing efficacy of mTOR antagonists in the clinic (8). Negative feedback signalling mechanisms are likely to contribute to the poor efficacy of agents when studied in phase II and phase III cancer trials and to the high rate of attrition of drugs (approximately 30% due to efficacy), which is both time-consuming and expensive (9). Empirical testing of every possible agent or combination of agents in the preclinical or clinical setting becomes prohibitively expensive and impractical.

No single experimental or theoretical methodology can be used in isolation in order to de-convolute complex biological behaviour. On the contrary, the current state is that different methodologies have different strengths and weaknesses. For instance, while experimentally it is at present relatively difficult to measure variables over time (either in real-time or at sufficiently high density to be meaningful), kinetic modelling permits the simulation of biological behaviour in real time. On the other hand, it is relatively easy to quantify complex spatial data experimentally but significantly harder to do so mathematically. Likewise, given the heterogeneity of experimental data available to theoreticians, such as high density and high volume but static data gene expression microarray data from clinical trials or very low density, highly quantitative data from reductionist biology, a “one size fits all approach” to mathematical modelling is unrealistic.

The ultimate goal of these investigations is to translate knowledge from both diagnostic and biomarker data into an individualised treatment protocol, informed by a predicted outcome. This requires a predictive framework that is able to absorb experimental

results, reflect the dynamical states of signaling pathways in aberrant cells and represent the impact of treatment agents on those pathways. Increasingly, it is becoming clear that multi-target treatment regimens are necessary to overcome the inherent robustness of the cell cycle, and it is likely that combinatorial approaches to therapy will similarly be required to overcome the robustness intrinsic in most druggable signaling pathways (such as PI3K or MAPK pathways) (10). Importantly, to support the identification of such regimens in a predictive framework requires a representation of the dynamical state of the essential pathway network and their interconnections (10). However, the construction of such a systems-level model attracts several challenges. First, there is wide variation in the level of detail known for different parts of this network and it is necessary to best utilise the descriptions available. Second, the scale of a dynamical system-level model means that it is not possible to fully describe all parts and it is difficult to interpret any predictions. Third, experimentation is not able to provide a complete description of the system over space and time. To address these challenges, we turn to the interoperability of experimentation and modelling.

For the first challenge, we consider *process-driven models* that exploit areas of the network about which at least some of the molecular species interactions are known and show that these models are able to make mechanistic predictions, test assumptions about unknowns, and determine target areas to measure experimentally. We also consider the role of *data-driven schemes* to derive hitherto unknown associations among measurables, even when these comprise multiple data types. To address the second challenge, we exploit network robustness analysis for reducing model complexity, the need for a standardised representation of network subcomponents to support integration, and optimisation approaches to fill gaps in knowledge by mixing data-driven and process-driven approaches. For the third challenge, we relate state-of-the-art *experimental methods* for profiling the spatio-temporal dynamics of cancer. In the following sections, we will therefore discuss a number of modelling approaches, which types of data generated by cancer biologists and clinicians they are best suited to, and then describe advances in cancer experimental biology which may aid systems biology approaches.

---

## 2. Approaches to Modelling Cancer

### 2.1. Linking Experiment and Theory

Process-based models afford a mechanistic representation of the underlying cell dynamics and may be parameterised directly by experimental data. These models are formulated in terms of

ordinary differential equations that describe the kinetics of the concentrations of molecular species within the network over time. Here, we present two illustrative examples of the approach and the opportunity to inform subsequent experiments.

In Faratian et al. (11), we employed a process-driven approach in order to study resistance factors to receptor tyrosine kinase (RTK)-inhibitors such as trastuzumab and pertuzumab. Trastuzumab (Herceptin) is widely used as breast cancer therapy in patients who overexpress the HER2 oncogene. Unfortunately, HER2 protein expression or gene amplification status is a poor predictor of response with a very low positive predictive value (12, 13). The documented actual benefit of adjuvant trastuzumab combined with chemotherapy vs chemotherapy alone in terms of overall survival is only modest (96% vs. 95% respectively at 1 year) (12) and 91% vs. 87% respectively at 4 years (13). A large proportion of patients therefore unnecessarily receive ineffective and expensive treatments with toxic side-effects, and there is a need to identify markers which predict therapeutic response. Since the reported resistance mechanisms to trastuzumab seem to relate to aberrant MAPK/PI3K signalling (*PIK3CA* mutations and inactivation of the tumour suppressor gene *PTEN* (14, 15), we reasoned that a systems analysis of these pathways, which are the best studied process-driven models to date, would be a useful application of systems biology to a clinical problem in oncology. These canonical pathways have only been modelled in order to explain and predict physiological phenomena, such as the binding of ligand to growth factor receptors (e.g. EGF to EGFR) (16–22), but have not been so helpful for understanding therapeutic interventions, since they frequently fail to include important oncogene and tumour suppressor nodes, which are known to be fundamental to carcinogenesis and proven resistance proteins (such as HER2, PTEN, and SRC in PI3K and MAPK signalling models). A new model of MAPK/PI3K was developed to describe HER2-inhibitor antibody/receptor binding, HER2/HER3 dimerisation and inhibition, AKT/MAPK crosstalk, and the kinetic and regulatory properties of PTEN, and was based on modelling studies of the HER signalling network (19, 23–25). The inclusion of the tumour suppressor protein PTEN was deemed particularly important since it is a key negative regulator of the PI3K signalling pathway. We successfully demonstrated that resistance to RTK-inhibitors was governed by the PTEN:activated PI3K ratio (integrated resistance factor  $\gamma$ ), and that PTEN, appropriately measured in the clinical setting, could stratify patients for HER2-inhibitor or combinatorial therapy, particularly an RTK-inhibitor and PI3K-inhibitor in cancers with low  $\gamma$ . This is one of few “success stories” of how a systems biology approach can generate hypotheses that can be tested experimentally in preclinical models and which can then be applied to clinical evaluation.

Further examples of applied systems biology are required so that it might gain credibility and be accepted within the clinical community.

Clyde et al. (26), used a process-based model parameterised by experimental data to generate a hypothesis for a new and important mechanism in the ATM intracellular pathway. ATM contributes to the co-ordination of the DNA damage response pathways that protect cells from potentially harmful mutations. In doing this, ATM also has the capacity to initiate the repair of treatment-driven damage, for example, radio-therapy, and so limits the impact of some treatments. A better understanding of the mechanisms of ATM regulation is therefore important both in the prevention and treatment of disease. Clyde et al. (26) investigated the behaviour of the damage response signaling pathway by treating cells with a DNA damaging agent and measuring ATM expression levels, and demonstrated that ATM gene expression is unaffected by the damaging agent. However, following the application of a specific ATM-inhibitor, a significant increase in ATM and ATR transcription was observed. Importantly, these results cannot be explained in terms of known cellular processes. Using a process-based model of the interaction network for all of the protein species considered in the experimental data together with the impact of the inhibitor and damaging agents used in the experiment, a novel feedback process was identified which was able to explain the anomalies in the data. Model predictions are consistent both with these in vitro experiments and with in vivo studies by another group (27). The model predictions point to a possible new target for ATM inhibition that overcomes the restorative potential of the proposed feedback.

Such process-based approaches are highly dependent on the assumptions made in model formulation. Crucial assumptions relate to the architecture of the network and the strength of the interdependencies among the measurables in the system. Where existing biological knowledge is limited, statistical data-driven approaches can make a valuable contribution to determining those associations. Here we describe two such data-driven approaches, Bayesian networks and S-systems modelling.

A Bayesian network (BN) is a graphical representation of statistical dependencies among a number of variables (28–30). Variables are drawn as nodes linked by arrows, forming a network. In BN parlance, the variable at the root of an arrow is known as the *parent* and the variable at the point is the *child*: the arrow indicates a direct statistical dependence of the child on the parent. While BNs were initially developed as “expert systems” (29) – a process-driven model where domain experts were consulted to form a network that could make predictions – BNs are now commonly used as a data-driven method, where data is used to learn the structure of a BN (30, 31). A feature of learned BNs

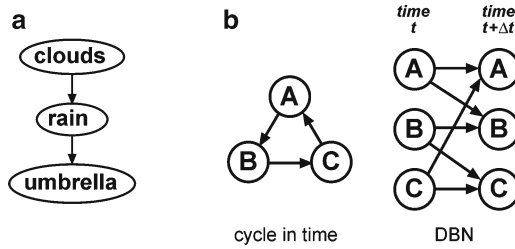


Fig. 1. Bayesian networks. **(a)** An example Bayesian network. The BN is able to distinguish that the presence of rain is a more direct predictor of people carrying umbrellas than are the presence of clouds. This prediction is in an informative, not a causal, sense, however; the *arrows* could conceivably all be reversed, for example, using the presence of umbrellas seen from a high office window to conclude that it must be raining. **(b)** A cyclic causal interaction in time represented by a dynamic Bayesian network. When time series are available, DBNs can infer causal interactions by making informative predictions about the future: a set of causal relationships (*left*) can be represented by a DBN (*right*) that represents all variables at two points in time, where  $\Delta t$  represents the time between data samples. Every variable is predictive of itself in the future, and thus identical variables are linked across time slices; the informative predictions of one variable (e.g., A) of another in the future (e.g., B) can be interpreted to be causal (if A, then, later, B).

is that they find a minimal set of direct dependencies necessary to explain statistical structure in the data; thus, they are well-suited to distinguish direct from indirect relationships among (potentially correlated) measured variables (Fig. 1a) (28). As such, BNs can be useful for identifying direct relationships to be used in further fine-grained modelling with process-driven methods. Caution must be taken, however, with relationships identified in a BN, as they are not necessarily causal (28, 29). The statistical dependence indicated by an arrow in the BN is most usefully conceptualised as “is useful for predicting”. For example, rain does not cause clouds, but the presence of rain is a useful predictor for the presence of clouds; similarly, the value of parent in a BN would be a useful predictor for the value of its child (Fig. 1a). Causal relationships can be discovered when time-series data are available: here, a dynamic Bayesian network (DBN) represents variables across time, and arrows from a parent to a child indicate that the past value of the parent is useful for predicting the future value of the child (Fig. 1b). As time series data becomes more common in biology, more biological data is likely to be modelled with DBNs (31); however, the continued need for analysis of non-time series data, such as clinical samples, means that static BNs will continue to play an important role.

Bayesian networks can be applied to a variety of different data types. Commonly, continuous data, such as from gene or protein expression, is discretised, and relationships found among variables with states such as low/medium/high. Such discrete data

enables the discovery of multiple types of relationships, including linear, non-linear, and non-monotonic (e.g., U-shaped or combinatoric (28)). Continuous data can also be analysed directly with BNs, although such models are often limited to additive (i.e., non-combinatoric) relationships (32). The most common biological application of BNs has been to microarray gene expression data, to discover networks that represent transcriptional regulation interactions (30, 31). Metabolic flux, protein expression, and phosphoprotein expression have also been used to discover networks representing signalling pathways (33–35). BNs can also incorporate multiple data types into a single network. For example, clinical data can be included as variables in a network alongside gene expression (36). This flexibility in the data types BNs can model means it is possible to determine statistical relationships between very different variables: for example, identifying which genes are most directly dependent on an experimental condition (37). However, this same flexibility means that care must be taken in interpreting networks. A discovered relationship between two gene expression values may be interpreted as translational regulation; a relationship between a metabolite and a protein may be interpreted as enzymatic activity. But the networks represent only statistical dependence without any suggestion of mechanism; interpretations are based on the user's biological knowledge and thus are only as good as our understanding of the system. Bayesian networks are thus most useful for (1) identifying direct relationships among a number of associated variables, and (2) identifying biological species that are most relevant to broader variables such as experimental condition or clinical outcome. Identified relationships can then be followed up with more detailed mechanistic modelling methods.

An alternative data-driven approach to Bayesian networks is one based on general power-law formalisms and Biochemical Systems Theory, (38) and is particularly suited to interaction networks within the cell. The Biochemical Systems Theory is based on an underlying S-system, which is a mathematical representation of non-linear systems, based on power-laws, an approach that explicitly represents the dynamics of the network in terms of differential equations that describe the rate of change of variables such as protein concentrations or gene-expression levels (38). The equations characterise the rate of change of the variables in terms of the interaction between components of the system as products of power-laws of the concentrations (or expression levels) of these components. This allows component interactions to be described in terms of rates of production and degradation of concentrations, and it can be shown that any kind of interaction can be approximated by this form, at least locally (39).

An S-system approach allows the construction of an interaction network for a given set of measured variables by fitting the

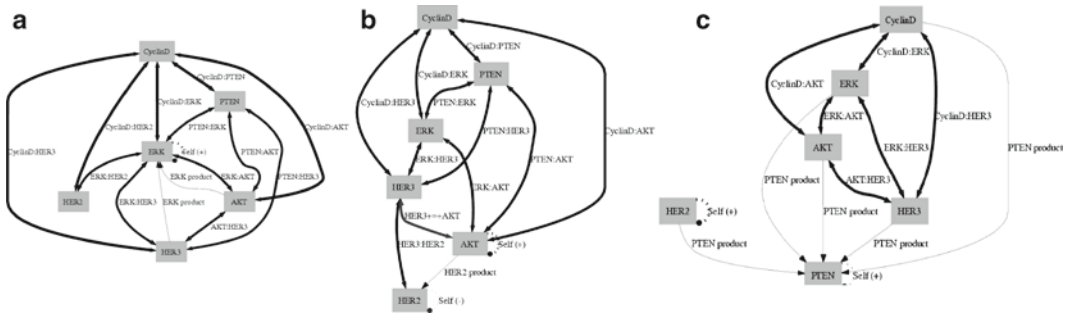


Fig. 2. Figures are visualisations of results of S-systems analyses. The protein expression data relates to a breast cancer cell line (BT474), treated with heregulin in the presence of the HER2 dimerisation inhibitor, pertuzumab. In the diagrams, *lines* represent complexes (either formation or degradation), positive feedback interactions, or production influence; *line thickness* indicates the strength of interaction. Figures (a)–(c) represent analyses on timecourse data and show the evolution of the network in response to treatment. The approach requires three at least time points to derive a network. Network (a) represents the first three time points (0, 1 and 2 min); (b) represents the next three (*overlapping*) time points, that is, 1, 2 and 5 min; (c) represents 2, 5 and 10 min. Among other results, the diagrams show the dissociation of HER2 from the interaction network in response to treatment with a receptor dimerisation inhibitor, which would be expected from the known biology.

equations to time series data in terms of a set of power-law exponents. S-systems are flexible in terms of type of data, so long as the kinetics are described by making a series of measurements over time. Where this fit returns an exponent value of zero, one can infer that the variable in question is independent of that component. Positive associations indicate activating influences; negative associations indicate inhibition. A network of interactions can be defined from this fitting by considering only the non-zero exponents for each variable in the system. This network can be visualised to indicate those links that most sensitively affect the rate of change of connected components (Fig. 2). In this sense, the derived network can be used to identify components as potential targets for changing the dynamical state, such as a new drug target or tumour-suppressor nodes. The prediction is valid provided the system is not changed too much from the dynamical state in which the original measurements were made or where combination therapies are being considered. The S-systems approach is beginning to be used in relevant areas of biology such as the analysis of experimentally-derived time courses of cDNA gene expression array data, albeit in yeast (40). This is promising since it is proof of principle that the technique may be used in gene expression data generated from cancer specimens, such as those taken at different time points in a neoadjuvant trial (see below).

## 2.2. Systems-Level Modelling

The cell is a complex interaction network that must maintain its functioning while being subjected to continuous ecological and evolutionary pressures. This inherent robustness in behaviour is a

systems-level property and is pervasive across a wide range of biological systems (41). Effective treatment of cancer cells is dependent on changing existing cell functioning, and so any treatment regime must overcome this robustness. However, it is likely that interventions aimed at affecting the cell behaviour must target multiple pathways (42, 43), as the dynamics may be extremely robust to uncoordinated changes in individual pathways (41). It is clear that any model that can contribute to our understanding of how to impact on cell behaviour by overcoming a systems-level property must be constructed at a systems-level, and so represents a sufficient level of complexity and detail. However, the formulation of such a model presents challenges in terms of constructing a model that integrates knowledge across the signalling pathway network and is able to operate in spite of the requirement of a large number of parameters.

The integration of knowledge arising from the research output of multiple, disparate groups require a common currency of exchange. Often, diagrams are used to describe the current understanding of pathway interactions but these diagrams, while heavily annotated, are informal and ambiguous (44). In recognition of both the utility of annotated diagrams for explaining pathway interactions and the benefits of formalising those diagrams, effort has been invested in developing a standardised language for reporting results. One increasingly pervasive scheme for this is SBGN, for example, as used in Calzone et al. (45), the Systems Biology Graphical Notation (46). SBGN supports the syntactic representation of biological entities and their interactions, where these interactions are semantically and visually unambiguous. This provides to different research communities a homogeneous reporting platform that enables knowledge synthesis. Because the entities in the diagrams are defined in terms of their syntax and semantics, it becomes possible to interconnect, for example, a pair of networks where the same, unambiguously defined node exists in each. In principle, this is all that is required to support knowledge integration. However, SBGN offers further support to bridge the gap between experiment and theory. SBGN offers tooling support for creating and verifying the diagrams used to describe pathway architecture and interactions. SBGN also offers tooling to translate these diagrams into formal models that aid computational model construction. Given the large number of groups working on different regions of the cell and the need for systems-level modelling, the value of such an integrative framework is clear.

Systems-level modelling can lead to very large-scale models with many parameters describing the kinetics of the system. Experimental data may only support the identification of some of these parameters and so, in principle, the sensitivity of the model to uncertainty in parameter values may be explored using sensitivity

analysis (47). Sensitivity analysis offers a highly structured approach to identify the subset of parameters that the model is most sensitive to changes in value. In its simplest form, this is achieved on a parameter-by-parameter basis: for each parameter, fix all other parameter values and systematically vary that selected parameter, measuring the extent to which the system-scale behaviour changes in response to value changes. This knowledge may be used to inform experiment and theory. It may direct experimentation, as it provides the set of parameters that are most important to system-level dynamics. It may inform model formulation as it offers the potential for model simplification. This may be in the form of determining regions of the network where the model is robust to large parameter changes, and these regions can then be simplified into an abstracted form or even removed from the network thus reducing the complexity and need for experimental parameterisation. In extreme cases, it may be used to completely reformulate the model in terms of only those sensitive parameters, as in Pachepsky et al. (48).

This form of sensitivity analysis ignores interactions among parameters, and so a more holistic approach to this form of analysis is required here to account for the interconnectivity among pathways. Saltelli et al. (49) describe global sensitivity analysis, where multiple parameters may be varied simultaneously and the impact of this measured again in terms of the degree to which parameter value modifications drive changes in system-level behaviour. The approach has been used in the optimisation of (synthetic) genetic circuits to inform experimental designs (50) and, importantly, the authors highlight the applicability of the approach to other biological networks.

In addition to sensitivity analysis, the fact that the model describes the whole system means that knowledge of that system-level behaviour, that is, model output, can be used to constrain the parameters, that is, model input. Techniques from artificial intelligence, such as Genetic Algorithms (51) can be used to reverse engineer missing parameter values that are consistent with known system-scale behaviour when combined with other, known input parameters. This means that not all variables in the system require explicit measurement; a subset of variables can be derived from simulation. The reverse engineering approach operates by comparing system-scale model predictions with observed system-scale behaviour, and refining through iteration the unknown input parameters to reduce the difference between predicted and observed behaviour. Clearly, the success of the scheme is dependent on defining an appropriate measure of difference, termed a fitness function. The definition of this fitness function is challenging but also the key to combining the strengths of process-based and data-driven approaches to address gaps in knowledge.

When considering the output from the process-based model, the fitness function can be defined to account for, for example, particular concentration values and/or particular rates of change in those values. However, there are many possible parameter configurations that could match the predicted behaviour to that which is observed. To reduce the parameter space further and determine more accurate values for unknown parameters, it is possible to integrate the results of data-driven models into the evaluation of specific process-based parameter configurations. This integration may be effected through the fitness function. The data-driven models described above provide a mechanism to infer the network structure from measurables without specifying causal links in the network. This provides a meta-level description of the experimental system that is not dependent on any assumptions relating to cell structure or architecture. By deriving the same data-driven meta-level description of the predictions of the process-based model, based on the equivalent output of those same (simulated) measurables, it is possible to further constrain the set of parameters that may be reverse-engineered to be consistent with the observed system-scale behaviour. Thus, data-driven models offer the possibility of substantially reducing the potentially large parameter search space in a rigorous way and without making a priori assumptions about parameter ranges and/or interaction network architecture.

---

### **3. Experimental Methods for Data Generation in Systems Biology**

#### ***3.1. Spatial Resolution***

We have already highlighted that cancer is more than just an epithelial disease. Therefore, in tissues, at least two spatial levels must be resolved; tissue compartments (i.e., epithelial, stromal, inflammatory component, vascular and interstitial) and cellular compartments (e.g., immediate extracellular environment, membrane, cytoplasmic, nuclear and organelles). Since the modelling approaches described above, particularly DBNs, require sufficient density of data, high-throughput experimental approaches are required. We use multiple strategies to combine compartment-specific analysis with high-throughput molecular analysis; including tissue microdissection, in situ protein quantification and reverse-phase protein arrays (RPPA; see below). Microdissection is becoming a standard strategy in gene-expression microarray (52) and genomic hybridization protocols (53) in order to enrich for epithelial-cell populations, either to overcome the inherent limitations of sensitivity of the assay (as in array CGH) or to infer compartment-specific biology (as in the case of gene expression) (52, 53). Microdissection often means a relatively crude dissection of the

epithelial area of a tumour by hand under a dissecting microscope, or alternatively the use of laser capture/catapult microdissection techniques in order to obtain populations of pure cells (54). The latter approach may be more suited for systems biology because the methodology can also be extended to separate any tissue compartment (e.g., blood vessels, stroma) and also morphologically heterogeneous elements within the epithelial areas of tumours, which have differential gene and protein expression signatures and therefore potentially different responses to therapy. A second approach to compartment-specific analysis is in situ protein quantification on automated image analysis platforms for total and phosphorylated (usually active) states of proteins within signalling pathways of interest (Fig. 3a and (55)). These methods multiplex antibodies against particular compartments (usually epithelial, although any compartment may be discriminated) with one or more targets of interest, so that compartment-specific protein expression can be quantified. The advantage of this technique over microdissection methods is the ability to discriminate protein expression levels at the compartmental and subcellular levels. However, disadvantages include the limited number of targets that can be measured from a single section (which is governed by the number of filters (usually up to five) on the fluorescence microscope), availability of high-quality specific antibodies, tissue

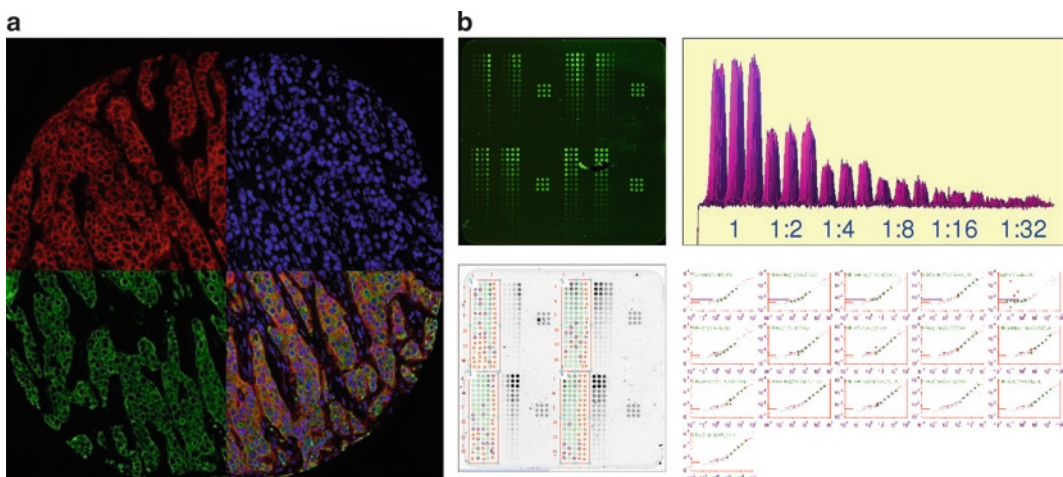


Fig. 3. Examples of quantitative methodologies for systems biology research in vivo and in vitro. (a) HistoRx AQUA fluorescence image analysis. The expression of protein per unit area of a molecularly defined compartment (tissue and subcellular) is measured using this technique. In this case, membrane expression of the HER2 oncoprotein (*top left quadrant*) in invasive cancer (*bottom left quadrant*; cytokeratin mask) has been quantified. Expression of proteins in the nuclear compartment may be measured by quantifying signal intensity in the DAPI compartment (*top right quadrant*). (b) Reverse phase protein arrays (RPPA). Cell or tissue lysates are spotted onto nitrocellulose-coated glass slides in replicates and as a dilution series, and probed for a protein of interest using specific antibodies and a fluorescent secondary (*top*). Each spot intensity is measured and plotted as a dilution curve, and the expression of protein derived from a regression analysis of all the data generated for a single sample (*bottom*). In this way, the expression of protein is guaranteed to be derived from the linear range of detection (i.e., not saturated) and the technical variance of the data is measured.

auto-fluorescence and difficulties with absolute quantification. While most in situ methods are limited to protein techniques, all macromolecules may be extracted from microdissected tissues, including RNA, microRNA, DNA and protein. Each macromolecule is amenable to high-throughput array-based technologies and, in the case of protein, the tantalizing prospect of reliable clinical mass-spectrometry proteomics (56) can offer quantitative advantages or assessment of functional post-translational modifications (e.g., phosphorylation or acetylation).

### **3.2. Temporal Resolution**

Although increasing temporal resolution seems straightforward for in vitro-based experiments, it is more difficult within the clinical setting. There are a range of biological models available for analyzing complex biological systems. Nonetheless, even the simplest in vitro models require reconsideration of design in order to generate data of sufficient quality to populate mathematical models, particularly data-hungry methodologies such as DBNs. While it is relatively trivial to increase data points in an experiment examining pathway responses to targeted therapy in vitro, it is only recently that downstream assays are sufficiently “high-throughput” to generate robust, quantitative data of sufficient quality and quantity to be used to populate mathematical models. Although robotics can help meet the demands of high-performance throughput, new techniques need to be considered to meet the data-rich demands of systems approaches.

Since many of the components assessed within mathematical models are proteins and their activated forms, newer protein assays may be used in order to address the problems outlined above. Reverse-phase protein arrays (RPPA) are high-throughput, high-density protein arrays in which protein lysates from in vitro or in vivo biological samples are immobilised as spots on nitrocellulose-coated glass slides and protein targets are detected with specific antibodies, similar to immunohistochemistry or immunofluorescence (Fig. 3b and (57)). In this way, hundreds to thousands of lysates (including technical and biological replicates) can be assayed under the same conditions on a single slide. As only picoliter quantities are spotted on each spot, tens to hundreds of identical slides may be produced for multiple target analysis from a single cellular or tissue lysate, particularly if robotic spotting is used. In addition, since each lysate is spotted as a dilution series, the signal detected can always be guaranteed to be in the linear range of detection (that is while signal intensity is still unsaturated), which is rarely the case for western blotting, and if recombinant peptide or protein controls are spotted on the same slide, accurate quantification is possible. As well as facilitating high throughput protein analysis in vitro, the small amount of lysate required and the high number of assays that can be

performed make this a useful technique for assessing clinical material. This means that large models, requiring tens to hundreds of biological measurements (e.g., the ordinary differential equation-based approach) may be populated with ease.

A second challenge to achieving sufficient temporal resolution for systems approaches is the selection of appropriate *in vitro* or *in vivo* models. While transgenic mice offer the attraction of stable genetic perturbations that can be applied to computational models, the long generation times and high numbers of time points required do not lend themselves to the iterative nature of systems biology, that is the need to refine the model on the basis of re-experimentation in order to improve it. Nevertheless, if coupled with live imaging techniques, such as the relatively new technique of intra-vital microscopy, which has been used to image tumour cell invasion in real-time (58), these animal models may become attractive models for detailed pharmacodynamic studies. We have used a combination of 3-dimensional *in vitro* primary and human adenocarcinoma cell line cultures on contracted collagen matrices (59) and cell line xenografts to model the pharmacodynamics of targeted therapies. Three-dimensional culture simplifies the tumour context but offers a flexible analysis of epithelial and stromal compartments, where both compartments may be genetically manipulated and subjected to both destructive and non-destructive temporal analysis, such as by reverse-phase protein arrays and immunofluorescence. Primary and cell-line xenografts capture the complexity of whole tissues but what makes these models ideal for high-throughput, spatially resolved analyses is their ability to assess tumours from multiple time points, and the availability of abundant tissue for fresh frozen and formalin-fixed paraffin-embedded analysis. In addition, cell lines are readily manipulated *in vitro* for the knockdown or overexpression of specific targets with small hairpin RNA and stable transfection of genetic constructs, respectively, and are then ready for re-implantation and re-testing of the system with specific perturbations. These models are an important intermediate step in validating computational models before they are sufficiently reliable to be used in clinical decision making.

---

#### **4. Clinical Considerations for Data Generation**

The real challenge lies in achieving sufficient temporal resolution using real human disease as the model. Nevertheless, there is now extensive experience in gathering tissue and biological samples from three time points in the neoadjuvant setting (that is,

patients treated with drugs or radiotherapy before surgery), in individual patients with breast cancer, such that limited pharmacodynamic studies may be performed. In this model, patients are given endocrine or chemotherapy for 3 months prior to definitive resection and samples taken at diagnosis, 2 weeks and 3 months at the time of resection (60). If basic pathological endpoints such as proliferation (immunohistochemistry analysis of Ki67 expression levels) are measured, then the proliferation index at 2 weeks (but not at diagnosis) is predictive of long-term survival in response to aromatase inhibitor therapy (61, 62). Breast cancer is amenable to this type of temporal intervention, since there is the added benefit that tumour shrinkage in the neoadjuvant setting can result in the use of breast-conserving surgery rather than mastectomy. Nevertheless, other cancers may also be amenable to multiple sampling, such as ovarian tumours treated with intraperitoneal chemotherapy (63), or colorectal tumours treated with pre-operative radiotherapy (64), which may be achieved with minimal discomfort or inconvenience to the patient. By exploiting carefully selected human models, we can begin to determine the true nature of the dynamics of tumour responses and move away from inferred biology based on static biomarker analysis. In the short term, gathering high quality, temporal data from real clinical material is essential to populate and validate computational models. In time, such mathematical models are likely to become applicable and may avoid the need for multiple biological measurements.

---

## 5. Conclusions

In this review, we have discussed a number of possible general approaches to applying systems biology to understanding therapeutic responses in cancer. It should be apparent that no one computational, mathematical or experimental methodology can be used in isolation to de-convolute the complexity of cancer. We have discussed some approaches which are being used in this infant field. On their own, some of these approaches are starting to produce results, but by using different approaches, such as process- and data-driven models in tandem in order to refine and validate models, the hope is that clinically relevant and useful models may become a reality sooner. In order for this to occur, data from clinical trials will also have to be incorporated, which will also require coordinated multidisciplinary efforts from the clinical and basic science communities.

## References

- Hanahan D, Weinberg RA (2000) The hallmarks of cancer. *Cell* 100:57–70
- Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ, Shen D, Boca SM, Barber T, Ptak J, Silliman N, Szabo S, Dezso Z, Ustyanksky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kaminker JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Willson JK, Sukumar S, Polyak K, Park BH, Pethiyagoda CL, Pant PV, Ballinger DG, Sparks AB, Hartigan J, Smith DR, Suh E, Papadopoulos N, Buckhaults P, Markowitz SD, Parmigiani G, Kinzler KW, Velculescu VE, Vogelstein B (2007) The genomic landscapes of human breast and colorectal cancers. *Science* 318:1108–1113
- 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* 10:515–527
- Sobin LH, Wittekind CH (2002) UICC: TNM classification of malignant tumors. Wiley-Liss, New York
- Faratian D, Bartlett J (2008) Predictive markers in breast cancer—the future. *Histopathology* 52:91–98
- Payne SJ, Bowen RL, Jones JL, Wells CA (2008) Predictive markers in breast cancer – the present. *Histopathology* 52:82–90
- Faratian D, Moodie SL, Harrison DJ, Goryanin I (2007) Dynamic computational modeling in the search for better breast cancer drug therapy. *Pharmacogenomics* 8:1757–1761
- O'Reilly KE, Rojo F, She QB, Solit D, Mills GB, Smith D, Lane H, Hofmann F, Hicklin DJ, Ludwig DL, Baselga J, Rosen N (2006) mTOR inhibition induces upstream receptor tyrosine kinase signaling and activates Akt. *Cancer Res* 66:1500–1508
- Kola I, Landis J (2004) Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov* 3:711–715
- Kholodenko BN (2006) Cell-signalling dynamics in time and space. *Nat Rev Mol Cell Biol* 7:165–176
- Faratian D, Goltsov A, Lebedeva G, Sorokin A, Mullen P, Kay C, Um I, Langdon SP, Goryanin I, Harrison DJ (2009) Systems biology reveals new strategies for personalising cancer medicine and confirms PTEN's role in resistance to trastuzumab. *Cancer Res* 69:6713–6720
- 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, Greaux V, Ward C, Strachle C, McFadden E, Dolci MS, Gelber RD (2005) Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N Engl J Med* 353:1659–1672
- 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
- Berns K, Horlings HM, Hennessy BT, Madiredjo M, Hijmans EM, Beelen K, Linn SC, Gonzalez-Angulo AM, Stemke-Hale K, Hauptmann M, Beijersbergen RL, Mills GB, van de Vijver MJ, Bernards R (2007) A functional genetic approach identifies the PI3K pathway as a major determinant of trastuzumab resistance in breast cancer. *Cancer Cell* 12:395–402
- Nagata Y, Lan KH, Zhou X, Tan M, Esteva FJ, Sahin AA, Klos KS, Li P, Monia BP, Nguyen NT, Hortobagyi GN, Hung MC, Yu D (2004) PTEN activation contributes to tumor inhibition by trastuzumab, and loss of PTEN predicts trastuzumab resistance in patients. *Cancer Cell* 6:117–127
- Fuss H, Dubitzky W, Downes CS, Kurth MJ (2005) Mathematical models of cell cycle regulation. *Brief Bioinform* 6:163–177
- Hatakeyama M, Kimura S, Naka T, Kawasaki T, Yumoto N, Ichikawa M, Kim JH, Saito K, Saeki M, Shirouzu M, Yokoyama S, Konagaya A (2003) A computational model on the modulation of mitogen-activated protein kinase (MAPK) and Akt pathways in heregulin-induced ErbB signalling. *Biochem J* 373:451–463
- Hendriks BS, Cook J, Burke JM, Beusmans JM, Lauffenburger DA, de Graaf D (2006) Computational modelling of ErbB family phosphorylation dynamics in response to

- transforming growth factor alpha and heregulin indicates spatial compartmentation of phosphatase activity. *Syst Biol* (Stevenage) 153: 22–33
19. 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
  20. Markevich NI, Hoek JB, Kholodenko BN (2004) Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. *J Cell Biol* 164: 353–359
  21. Shankaran H, Wiley HS, Resat H (2006) Modeling the effects of HER/ErbB1-3 coexpression on receptor dimerization and biological response. *Biophys J* 90:3993–4009
  22. Steuer R (2007) Computational approaches to the topology, stability and dynamics of metabolic networks. *Phytochemistry* 68: 2139–2151
  23. Birtwistle MR, Hatakeyama M, Yumoto N, Ogunnaike BA, Hoek JB, Kholodenko BN (2007) Ligand-dependent responses of the ErbB signaling network: experimental and modeling analyses. *Mol Syst Biol* 3:144
  24. Moehren G, Markevich N, Demin O, Kiyatkin A, Goryanin I, Hoek JB, Kholodenko BN (2002) Temperature dependence of the epidermal growth factor receptor signaling network can be accounted for by a kinetic model. *Biochemistry* 41:306–320
  25. 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
  26. Clyde RG, Craig AL, de Breed L, Bown JL, Forrester L, Vojtesek B, Smith G, Hupp T, Crawford J (2009) A novel ataxia-telangiectasia mutated autoregulatory feedback mechanism in murine embryonic stem cells. *J R Soc Interface* 6:1167–1177
  27. Gueven N, Fukao T, Luff J, Paterson C, Kay G, Kondo N, Lavin MF (2006) Regulation of the Atm promoter in vivo. *Genes Chromosomes Cancer* 45:61–71
  28. Heckerman D, Geiger D, Chickering D (1995) Learning Bayesian networks: the combination of knowledge and statistical data. *Mach Learn* 20:197–243
  29. Pearl J (1988) Probabilistic reasoning in intelligent systems. Morgan Kaufmann, San Francisco
  30. Friedman N (2004) Inferring cellular networks using probabilistic graphical models. *Science* 303:799–805
  31. Kim SY, Imoto S, Miyano S (2003) Inferring gene networks from time series microarray data using dynamic Bayesian networks. *Brief Bioinform* 4:228–235
  32. Imoto S, Kim S, Goto T, Miyano S, Aburatani S, Tashiro K, Kuhara S (2003) Bayesian network and nonparametric heteroscedastic regression for nonlinear modeling of genetic network. *J Bioinform Comput Biol* 1:231–252
  33. Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP (2005) Causal protein-signaling networks derived from multiparameter single-cell data. *Science* 308:523–529
  34. Guha U, Chaerkady R, Marimuthu A, Patterson AS, Kashyap MK, Harsha HC, Sato M, Bader JS, Lash AE, Minna JD, Pandey A, Varmus HE (2008) Comparisons of tyrosine phosphorylated proteins in cells expressing lung cancer-specific alleles of EGFR and KRAS. *Proc Natl Acad Sci USA* 105:14112–14117
  35. Li Z, Chan C (2004) Inferring pathways and networks with a Bayesian framework. *FASEB J* 18:746–748
  36. Gevaert O, De Smet F, Timmerman D, Moreau Y, De Moor B (2006) Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks. *Bioinformatics* 22:e184–e190
  37. Matthäus F, Smith VA, Fogtman A, Sommer WH, Leonardi-Essmann F, Lourdasamy A, Reimers MA, Spanagel R, Gebicke-Haerter PJ (2009) Interactive molecular networks obtained by computer-aided conversion of microarray data from brains of alcohol-drinking rats. *Pharmacopsychiatry* 42:S118–S128
  38. Sorribas A, Savageau MA (1989) A comparison of variant theories of intact biochemical systems. I. Enzyme-enzyme interactions and biochemical systems theory. *Math Biosci* 94:161–193
  39. Savageau MA, Voit EO (2008) Power-law approach to modeling biological systems. I. Theory, 60th edn. pp 519–544
  40. Voit EO (2002) Models-of-data and models-of-processes in the post-genomic era. *Math Biosci* 180:263–274
  41. Kitano H (2004) Biological robustness. *Nat Rev Genet* 5:826–837
  42. Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D, Joshi MB, Harpole D, Lancaster JM, Berchuck A, Olson JA Jr, Marks JR, Dressman HK, West M, Nevins JR (2006) Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 439:353–357

43. Wulfskuhle J, Espina V, Liotta L, Petricoin E (2004) Genomic and proteomic technologies for individualisation and improvement of cancer treatment. *Eur J Cancer* 40: 2623–2632
44. Moodie SL, Sorokin A, Goryanin I, Ghazal P (2009) Graphical notation to describe the logical interactions of biological pathways. *J Integr Bioinform* 3:36
45. 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
46. Le Novère N, Moodie SL, Sorokin A, Hucka M, Schreiber F, Demir E, Mi H, Matsuoka Y, Wegner K, Kitano H (2008) Systems biology graphical notation: process diagram level 1. *Nature Precedings*
47. Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Saisana M, Tarantola S (2008) *Global sensitivity analysis: the primer*. Wiley, Chichester
48. Pachepsky E, Crawford JW, Bown JL, Squire G (2001) Towards a general theory of biodiversity. *Nature* 410:923–926
49. Saltelli A, Tarantola S, Chan K (1999) Quantitative model-independent method for sensitivity analysis of model output. *Technometrics* 41:39–56
50. Feng XJ, Hooshangi S, Chen D, Li G, Weiss R, Rabitz H (2004) Optimizing genetic circuits by global sensitivity analysis. *Biophys J* 87:2195–2202
51. Goldberg D (1989) *Genetic algorithms in search, optimization, and machine learning*. Addison-Wesley, Reading, MA
52. Kirby J, Heath PR, Shaw PJ, Hamdy FC (2007) Gene expression assays. *Adv Clin Chem* 44:247–292
53. Kennett JY, Watson SK, Saprunoff H, Heryet C, Lam WL (2008) Technical demonstration of whole genome array comparative genomic hybridization. *J Vis Exp*, 870
54. Edwards RA (2007) Laser capture microdissection of mammalian tissue. *J Vis Exp*, 309
55. Camp RL, Chung GG, Rimm DL (2002) Automated subcellular localization and quantification of protein expression in tissue microarrays. *Nat Med* 8:1323–1327
56. Pan S, Aebersold R, Chen R, Rush J, Goodlett DR, McIntosh MW, Zhang J, Brentnall TA (2008) Mass spectrometry based targeted protein quantification: methods and applications. *J Proteome Res* 8(2):787–797
57. Tibes R, Qiu Y, Lu Y, Hennessy B, Andreeff M, Mills GB, Kornblau SM (2006) Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells. *Mol Cancer Ther* 5:2512–2521
58. Kedrin D, Gligorijevic B, Wyckoff J, Verkhusha VV, Condeelis J, Segall JE, van Rheenen J (2008) Intravital imaging of metastatic behavior through a mammary imaging window. *Nat Methods* 5:1019–1021
59. Edward M (2001) Melanoma cell-derived factors stimulate glycosaminoglycan synthesis by fibroblasts cultured as monolayers and within contracted collagen lattices. *Br J Dermatol* 144:465–470
60. Dixon JM (2004) The scientific value of pre-operative studies and how they can be used. *Breast Cancer Res Treat* 87(Suppl 1):S19–S26
61. Dowsett M, Smith IE, Ebbs SR, Dixon JM, Skene A, Griffith C, Boeddinghaus I, Salter J, Detre S, Hills M, Ashley S, Francis S, Walsh G, A'Hern R (2006) Proliferation and apoptosis as markers of benefit in neoadjuvant endocrine therapy of breast cancer. *Clin Cancer Res* 12:1024s–1030s
62. Dowsett M, Smith IE, Ebbs SR, Dixon JM, Skene A, A'Hern R, Salter J, Detre S, Hills M, Walsh G (2007) Prognostic value of Ki67 expression after short-term presurgical endocrine therapy for primary breast cancer. *J Natl Cancer Inst* 99:167–170
63. Alberts DS, Markman M, Armstrong D, Rothenberg ML, Muggia F, Howell SB (2002) Intraperitoneal therapy for stage III ovarian cancer: a therapy whose time has come! *J Clin Oncol* 20:3944–3946
64. Nagtegaal ID, Gaspar CG, Peltenburg LT, Marijnen CA, Kapiteijn E, van de Velde CJ, Fodde R, van Krieken JH (2005) Radiation induces different changes in expression profiles of normal rectal tissue compared with rectal carcinoma. *Virchows Arch* 446:127–135