Published in final edited form as:

Biochim Biophys Acta. 2009 December ; 1796(2): 129–139. doi:10.1016/j.bbcan.2009.06.001.

A Systems Biology View of Cancer

Reinhard Laubenbacher^{1,6,*}, Valerie Hower², Abdul Jarrah¹, Suzy V. Torti^{3,4}, Vladimir Shulaev^{1,6}, Pedro Mendes^{1,5,6}, Frank M. Torti^{4,6}, and Steven Akman^{4,6}

¹ Virginia Bioinformatics Institute, Washington St. (0477), Blacksburg, VA 24061. ² School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0160. ³ Department of Biochemistry, Wake Forest University School of Medicine, Winston-Salem, NC 27157

⁴Comprehensive Cancer Center, Wake Forest University School of Medicine, Winston-Salem, NC 27157 ⁵ School of Computer Science, University of Manchester, Manchester, England. ⁶ Department of Cancer Biology, Wake Forest University School of Medicine, Winston-Salem, NC 27157.

SUMMARY

In order to understand how a cancer cell is functionally different from a normal cell it is necessary to assess the complex network of pathways involving gene regulation, signaling, and cell metabolism, and the alterations in its dynamics caused by the several different types of mutations leading to malignancy. Since the network is typically complex, with multiple connections between pathways and important feedback loops, it is crucial to represent it in the form of a computational model that can be used for a rigorous analysis. This is the approach of systems biology, made possible by new –omics data generation technologies. The goal of this review is to illustrate this approach and its utility for our understanding of cancer. After a discussion of recent progress using a network-centric approach, three case studies related to diagnostics, therapy, and drug development are presented in detail. They focus on breast cancer, B cell lymphomas, and colorectal cancer. The discussion is centered on key mathematical and computational tools common to a systems biology approach.

Keywords

systems biology; cancer; mathematical modeling

I. INTRODUCTION

The insight that cancer has a strong genetic component has focused attention on the search for oncogenes [1] and the pathways in which they participate [2]. It is well appreciated that several different types of mutations must occur in a multi-stage process before a cell becomes malignant, thereby altering different molecular pathways involving the affected genes. These pathways represent a variety of molecular mechanisms relating to gene regulation, signaling,

© 2009 Elsevier B.V. All rights reserved.

* corresponding author: FAX: 540-231-2606; reinhard@vbi.vt.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Agent-based models represent individuals, or agents, in a population explicitly, for instance, individual cells in a tumor. Each agent is equipped with a set of rules that describe how it interacts with its environment and with other agents. Such models are being used increasingly in population biology, social science, and computational biology.

and cell metabolism. In order to understand how a cancer cell is functionally different from a normal cell it is necessary to assess the overall network and its changed dynamics rather than the individual pathways. Since the network tends to be complex, with extensive cross-talk between pathways as well as important feedback loops, it is crucial to represent it in the form of a computational model that can be used for a rigorous analysis. As for most complex diseases, an additional complication arises from the fact that for many aspects of cancer it is important to integrate several different scales in addition to the molecular one. For example, for certain therapies, it might be important to also consider the cellular scale, e.g. the cell cycle, and the tissue scale, e.g. certain effects of angiogenesis. What emerges is a view of cancer as an exceedingly complex array of components that operate at different spatial and temporal scales, all working together as a highly non-linear dynamic system. Recognizing dynamic signatures of this system can help improve diagnostics, and an understanding of the mechanisms can help with the design of sophisticated perturbations of this system that disrupt cancer progression.

The case for a systems biology approach to complex diseases such as cancer or diabetes has been made eloquently in [3] and in [4,5]. As the authors point out in [4], "Reductionism pervades the medical sciences and affects the way we diagnose, treat, and prevent diseases. While it has been responsible for tremendous successes in modern medicine, there are limits to reductionism, and an alternative explanation must be sought to complement it." This alternative explanation lies in a systems-oriented view. It is only in the last two decades that technological advances in molecular data generation, imaging, and computing have come together to allow the realization of such a view in a practical way.

This review is intended to highlight the usefulness of a systems biology approach to cancer for the clinic. We begin with a discussion of why systems biology tools are an essential part of the repertoire of approaches to this complex disease, using the Akt signaling pathway as an example. We then discuss in detail three case studies related to cancer diagnostics, therapy, and drug development, providing an overview of different mathematical and computational approaches to problems in the framework of these case studies.

II. CANCER IS A SYSTEMS BIOLOGY DISEASE

Systems biology has emerged during the last decade as a powerful new paradigm for research in the life sciences. It is based on the premise that the properties of complex systems consisting of many components that interact with each other in nonlinear, nonadditive ways cannot be understood solely by focusing on the components. The system as a whole has emergent properties that are not visible at the parts level. The availability of high-throughput data generation technologies, such as DNA microarrays, has made it possible to apply this paradigm in molecular biology and biomedicine. Transcriptional data from a large part or all of the genome allows a shift in focus from individual genes in linear pathways to large-scale networks of interacting genes. In many cases, it is the properties of the aggregate network, possibly interacting with the extracellular environment, which are at the heart of pathological changes, such as cancer. Moreover, the system properties of interest often are embodied in its dynamics. For instance, changes in the structure of the network through mutations or epigenetic effects can lead to changes in network dynamics that result in different physiological properties. In many cases, in particular when it comes to cancer, phenomena at different scales are connected, such as the molecular and tissue level scales in tumor growth. The natural framework for the study of the structure and dynamics of networks is through the use of mathematical models. This is perhaps the defining characteristic of the systems biology paradigm.

As an example, we discuss the Akt signaling pathway. Akt is a serine/threonine kinase whose signaling is increased in numerous types of cancer. Akt occurs downstream of P13K (phosphatidylinositol 3-kinase) and can be altered by many different mechanisms in malignant

cells, including PTEN (phosphatase and tensin homolog) mutation[6] and activity of a number of oncogenes including Ras [7] and Her2/neu [8]. When unstimulated, the Akt protein resides in the cytosol. Akt can translocate to the plasma membrane and is subsequently phosphorylated. This process activates Akt as a kinase, and Akt can then affect many biological processes occurring within the cell. For instance, Akt promotes cell survival by inhibiting pro-apoptotic proteins such as Bad (Bcl2-antagonist of cell death) [9], caspase-9 [10], and FOXO1 (forkhead box O1) [11] and by activating anti-apoptotic proteins like Mdm2 (transformed mouse 3T3 cell double minute 2) [12]. The tumor suppressor TSC2 (tuberous sclerosis 2) is phosphorylated by Akt, which activates the mTOR kinase [13]. This causes an increase in both translation and protein synthesis while promoting cell growth. Metabolically, data suggests that Akt serves as regulator of glycolysis in cancer cells. Akt activity increases glucose consumption and causes cancer cells to rely on glucose for survival [14]. A common way to represent these interactions is in a simple schematic cartoon, as shown in Fig 1A.

However, the Akt pathway cannot be considered in isolation. There is in fact cross-talk between this pathway and several others implicated in different types of cancers. For example, NF- κ B transcription factors are induced by Akt signaling through its regulation of I κ B [15], thus linking the Akt pathway to NF- κ B signaling [16]. Akt also cross-talks with the CCR2 (chemokine (C-C motif) receptor 2) [17] and COX2 (PTGS2) pathways [18]. Recent software packages have been designed to depict the additional complexities represented by these interconnected networks, whose nodes represent molecular species. An example derived using the Ingenuity Pathways software package is shown in Figure 1B. It becomes clear that an analysis and understanding of this complex network is difficult or impossible without analytical tools.

While this graphical representation of the network provides additional information not available by considering the individual pathways separately, it is of course still a vast simplification, since it is merely a static representation of several dynamic processes happening at the same time, with several intertwined feedback loops. The only way to effectively study the effect of either mutations or therapeutic interventions is to create a quantitative model of this network which integrates the dynamics of the individual pathways and their interconnections and which can be simulated on a computer. Representing aspects of an *in silico* cell, the model can then be used to explore a variety of questions. One can use it to identify network nodes that have a particularly strong effect on network dynamics, or one can simulate the compound effect of a multi-target drug. As an example, in [19] a mathematical model was constructed of HRG-induced signaling cascades induced by HRG (histidine-rich glycoprotein) and mediated by ErbB tyrosine kinase receptors involving the Akt and MAPK pathways and their cross-talk. These receptors play an important role in cell proliferation and differentiation, and altered expression or mutation has been implicated in several human cancers. The mathematical model allows a detailed simulation of several different dynamic events, resulting in temporal concentration curves. For example, the graph shown in Fig. 1C produced by dynamic modeling reveals that the extent of Raf-Akt crosstalk can profoundly affect the magnitude and duration of the response of ErbB to receptor stimulation. Such an insight can clearly not be gleaned from the simple cartoon shown in Fig 1A or even the complex network depicted in Figure 1B.

This detailed example demonstrates how one is naturally led to consider quantitative models of complex molecular networks if one wants to comprehend the net effect of a mutation or the application of a therapeutic. This path is even more essential if one wants to link intracellular processes with events at the multi-cellular level, such as angiogenesis. A key feature of cancer development is encapsulated in the interaction between intracellular networks and the extracellular environment. Ultimately, a comprehensive understanding of cancer has to link events at the molecular level with those at the tissue-level, organism level, and even the

environment. This can only be done using multi-scale mathematical models that integrate quantitative information at different scales. This is precisely the approach of systems biology. Section IV and Case Study 3 in Section V illustrate this view by placing intracellular networks in the context of the extracellular environment. Before reviewing systems biology results specific to cancer, we provide a general introduction to the systems biology paradigm and methodological “toolbox.”

III. THE SYSTEMS BIOLOGY APPROACH

The ability to collect data for a large number of molecular species from the same sample, including data measuring gene expression and the levels of many proteins and metabolites, opens the possibility of obtaining network-level data. This type of data also makes possible a more discovery-oriented approach. Since the ultimate goal of a systems biology approach to cancer is to understand a relevant biological network with the help of a mathematical model derived from a data-mining effort on heterogeneous, system-level data, systems biology projects require a specialized experimental design, several steps for data analysis, and an iterative approach alternating between computational prediction and experimental verification. This process is schematically represented in Figure 2.

The fundamental premise of systems biology is that a complete understanding of a molecular network requires not just knowledge of its parts but also of their dynamic interactions. In order to construct meaningful system-level models, quantitative whole-cell measurements of the network's components as they change over time in response to one or more perturbations are needed. They form the basis of a computational model of the network. The model incorporates one or more hypotheses about the mechanisms that are responsible for system dynamics and provides a theoretical explanation of the observations. Model validation with further experiments tests these hypotheses. A validated model can then serve as a tool for the generation of further hypotheses. As such, the mathematical model represents a tool for hypothesis generation about a system that is typically too large and complex to understand on the basis of a static wiring diagram. Essential features of molecular networks, such as feedback loops and synergistic effects of different pathways, cannot be discovered and understood without the theoretical basis provided by the model.

Constructing dynamic mathematical models of networks is often not possible, however, due to the lack of appropriate time course data sets. In these cases it is still beneficial and often possible to construct and analyze static networks representing, for example, all possible protein-protein interactions in a cell type. Many data sets from functional genomics are available for this purpose. In Section VI we will describe two such examples.

Another important tool set in the systems biology repertoire is network inference from system-level data, typically DNA microarray data measuring gene transcription levels, 2D-gel, or large-scale mass spectrometry data. Inference methods allow the construction of phenomenological networks solely based on information available in the data, without the bias of prior biological knowledge or preconceived notions of the network structure. These phenomenological models can then be used to focus attention on certain features and provide constraints for building mechanistic models like the one in [19].

Sometimes network inference methods are referred to as “top-down” modeling and mechanistic models as “bottom-up.” As these terms have begun to enter into general use, it is worth commenting on their distinction. The availability of high-throughput data sets such as DNA microarrays or large-scale protein and metabolite profiles has opened up the possibility to create system-wide models of networks through inference of relationships between network nodes from data sets. This approach, sometimes referred to as *top-down modeling*, is in contrast to the more traditional approach of constructing mathematical or statistical models by first

identifying a list of molecular species and their interactions and then constructing a model that encodes these relationships. Model parameters are then frequently estimated by fitting the model to available data. This approach might be called *bottom-up modeling*. (An example is provided in the next case study.) The fundamental difference between the two approaches is that in the bottom-up approach one begins with a determination of which molecules are important and which interactions are to be considered. In the top-down approach the aim is to reverse-engineer this information in an unbiased way from the system-level data. Network inference has emerged as a central problem in systems biology [20]. A variety of algorithms have been developed for this purpose, whose output ranges from statistical models such as ARACNE and Bayesian networks models [21], providing correlation measures of network nodes, to discrete models, such as Boolean networks [22] and systems of ordinary differential equations[23], giving full dynamic network models. In most cases, top-down models are phenomenological rather than mechanistic, like bottom-up models. The ultimate goal is to combine the two approaches, with top-down models providing constraints for the bottom-up approach.

Systems biology approaches to the study of cancer to date have mostly focused on the construction and understanding of the molecular networks being altered by malignant transformations. But several studies have been undertaken that move the application of systems biology techniques closer to the clinic. While still far from the personalized medicine paradigm suggested by the genomics revolution it is feasible to envision some of the ways in which parameters in mathematical models might depend on characteristics of individuals and determine the nature of dynamic processes related to cancer progression or diagnosis and treatment. Relevant questions impacting clinical practice are:

- How can we use network-level analysis to improve prognostic forecasting? Case Study 1 in Section V. suggests an answer with respect to the prediction of the metastatic potential of breast tumors. Using a combination of DNA microarray data and protein-protein interaction information, the study proposes subnetworks of the protein-protein interaction network as predictors, rather than collections of individual genes.
- How can we use alterations of the molecular network in malignant cells rather than differential gene expression in order to identify oncogenes? Case Study 2 focuses on this question in the context of B cell lymphomas. Using a variety of data sets and statistical techniques, a network is established that integrates a range of molecular interactions. Oncogenes in different lymphoma phenotypes are identified by the changes in their network neighborhood rather than by differential expression.
- How can we use multi-scale dynamic models spanning the molecular and the tissue level in order to improve treatment methods? This question is addressed in Case Study 3, focused on improved protocols for radiation therapy. Using a dynamic mathematical model that integrates the molecular, cellular, and tissue level scales, one can study strategies for optimal timing of radiation treatment sessions.

Before presenting the case studies that address specific instances of these questions, we first review briefly the systems biology literature as it pertains to cancer. This review is not intended to be comprehensive, but rather illustrative of the systems biology approach.

IV. A SYSTEMS BIOLOGY VIEW OF THE HALLMARKS OF CANCER

In their much-cited review [24], Hanahan and Weinberg identify several crucial acquired traits that are necessary for a cell to become malignant. These include the ability to mimic normal growth signaling, the insensitivity to anti-growth signals, the ability to evade apoptosis, limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis.

Although these hallmarks focus on the tumor cell itself and do not encompass contributions of the extracellular environment, they have formed a useful framework for many systems-based approaches to the study of cancer. We discuss here progress made in understanding these hallmarks from a systems biology point of view. Specific studies illustrating a range of different approaches are presented in Table 1.

Independence from external growth signaling

As suggested in [24], there is ample evidence that most or all cancers display dysregulation of several signaling pathways resulting in the cells' acquired independence of external growth factors. A common alteration is the upregulation of growth factor receptors. In particular, the pathway of the epidermal growth factor receptor EGF-R is hyperactivated in approximately 30% of all cancers, including malignancies of the colon, lung, pancreas, ovary, kidney, as well as some leukemias [40]. Several detailed mathematical models of this pathway have been constructed, e.g., [38,41], and analyzed.

The point is made in [24] that to properly understand the phenomenon of becoming independent of external growth signals will require an understanding of how mutated cells interact with other cells in their microenvironment, in addition to intracellular events. During the last decade so-called agent-based or individual-based models have become increasingly popular which provide an excellent modeling paradigm for such a multi-scale approach. Several mathematical models of the cellular signaling pathways involving the family of *Her* receptor proteins have been constructed; see, e.g., [19,36,42,43].

In [37], a multiscale model of malignant glioma is constructed that represents the interior of each cell as a system of ordinary differential equations representing mass balance equations of molecular species related to the EGFR (*Her1*) interaction network. Multiple cells reside on a 2-dimensional grid on which they proliferate and move between grid points, based on a collection of rules. The rules incorporate cell-cell signaling and the sensing of environmental conditions such as nutrient sources. One interesting conclusion suggested by model simulation is that tumor dynamics cannot be explained through molecular processes inside the cells alone. Instead, cells with a higher receptor density show an early increase in the switching between proliferative and migratory traits, suggesting a more active protein-level interaction between tumor cells.

Downstream pathways that process signals from these and other receptors can also be altered in malignant cells, with the SOS-Ras-Raf-MAPK cascade playing a central role. Here too a systems biology approach has resulted in mathematical models focusing on different aspects of this network [44-46]. Altogether, these efforts are creating a more and more extensive and detailed quantitative representation of the inter- and intracellular signaling network that helps mutated cells acquire self-sufficiency in growth signals. The complexity of this network renders traditional "cartoon-level" representations of increasingly limited value.

Insensitivity to antigrowth signaling and evasion of apoptosis

Cell growth can be arrested through either a temporary or permanent interruption of the cell cycle, with cells being directed to the quiescent (G0) state or to senescence. At the extreme, cells are directed to programmed cell death, or apoptosis. For example, many antiproliferative signals involve the retinoblastoma protein pRb and its relatives. Disruption of the pRb pathway through hypophosphorylation prevents antigrowth factors such as TGF β from blocking advance through G1 phase. The cell cycle has been studied extensively from the point of view of systems biology. While there are mathematical models of the cell cycle in model organisms, e.g., yeast [47], only limited progress has been made on a systems biology description of the mammalian

cell cycle. However, a mathematical model was used in [48] to study, among other things, pRb dynamics during G1 phase in mammalian cells.

Preventing terminal cell differentiation is frequently accomplished by directly affecting the c-myc oncogene. A recent transcriptional profiling study has made significant progress toward identifying the transcriptional network that regulates c-myc expression [39], using statistical network inference algorithms. See the review [49] for further progress on understanding the function of c-myc in the epidermis.

There are several possible pathways that activate the apoptosis program in a cell. The most common mechanism involves the tumor suppressor protein p53, one of the most extensively studied mammalian proteins. Enough information and data is available to make the p53 network a good candidate for mathematical modeling. In recent years several systems biology models have appeared. For instance, the papers [50,51] study the oscillatory properties of the negative feedback loop between p53 and the oncogene Mdm2. This feedback loop was investigated in [35] as a possible tool in the design of optimal radiotherapy protocols.

Limitless replicative potential

Telomere maintenance is a feature in virtually all cancer cells and is regulated by a complex cellular network. A first step toward understanding this network and its role in carcinogenesis in mammalian cells is to elucidate the network topology, that is, the pattern of the interactions between the molecular species involved, and eventually its dynamics. Here too, mathematical and statistical methods have aided in making progress. Several authors have studied the network. Most recently, in [34] a general method was developed to infer a draft of the topology of the telomere length-control network from high-throughput phenotypical data. The method was then applied to data from gene knockout yeast strains. The nodes of the network are proteins and an edge indicates a protein-protein-interaction. (Networks are typically represented through the mathematical notion of a graph. The nodes, or vertices, of the graph represent the molecular species involved in the network. There is a directed or undirected edge between two nodes if they interact in some way. For instance, a directed edge from gene *A* to gene *B* might indicate that gene *A* (through its protein) activates gene *B*. An undirected edge between two proteins might indicate a protein-protein interaction.)

Sustained angiogenesis and metastasis

To understand the ability of solid tumors to invade neighboring tissues and to stimulate the growth of blood vessels to assure nutrient supply, it is important to understand the interplay between molecular events inside tumor cells, such as the upregulation of the vascular endothelial growth factor (VEGF) and the nature of the angiogenic switch, and events in the microenvironment surrounding the tumor, such as communication with neighboring cells, flow characteristics in the nearby vasculature, and other factors favoring or discouraging tissue invasion. These different scales represent a significant challenge to a mathematically formulated systems biology view. Fortunately, in the last decade, several modeling frameworks have been developed that allow a multi-scale modeling approach. In particular, the combination of continuous models, using systems of differential equations to capture gradient flows of nutrients or signal cascades, and discrete, so-called agent-based¹, models to represent collections of individual cells and their individual or communal actions based on their specific environment appears to be a very promising approach; see, e.g., [52-54]. Increased computational power is making it possible in the near future to simulate three-dimensional multi-scale models of tumor growth. Preliminary results are already providing intriguing insights and new hypotheses, such as the possibility that a tumor's microenvironment might have more of a selective rather than a regulatory effect on invading cancer cells [55], which departs from the conventional view [56].

V. CASE STUDIES

We now present three exemplary case studies that demonstrate how a systems-level view can provide insights of direct relevance to the problem of human cancer. They are chosen to be representative of different clinical issues, different mathematical methods, and different cancers. The first study uses a combination of gene expression data and a protein-protein-interaction network to discover network motifs that can be used for the prediction of developing distant metastases in breast cancer patients. The quantitative methodology consists of a statistical analysis. The second uses a more complex network of molecular interactions instead of just gene expression data to identify oncogenes in certain types of B cell lymphomas. In addition to statistical tools, a network inference methods based on information theory is used to construct the network. Finally, the third study constructs a detailed mechanistic mathematical model of the effects of radiation therapy on cell proliferation. It uses three different coupled models to represent a genetic network (at the lowest level) that regulates the cell cycle (at the middle level), which, in turn, affects cell proliferation in the microenvironment (top level). Furthermore, conditions in this microenvironment feed back to control the gene regulation network.

1. Network-based classification of breast cancer metastasis

Fatalities from breast cancer, the most common malignant disease in Western women, result primarily from metastases at distant sites rather than from the primary tumor. As described in [57], more than 80% of patients currently receive adjuvant chemotherapy. Only approximately 10-15% of breast cancer patients develop distant metastases within 3 years of initial diagnosis, however, it is not uncommon that metastases appear after 10 years or more from initial diagnosis. The availability of accurate prognostic markers would allow a more targeted application of adjuvant therapy, thereby avoiding many unnecessary chemotherapy treatments. However, traditional clinical markers provide an accurate prognosis for only about 30% of patients. With the availability of large-scale data collection through DNA microarrays, the question was investigated whether one could assemble a profile consisting of a relatively small number of genes that could be used to predict the likelihood for a given patient to develop distant metastases.

The study [58] used data from 295 patients to identify a profile consisting of 70 genes whose differential expression was a good predictor of disease outcome in a group of patients 52 years or younger, tumor-negative lymph nodes, and no previous history of cancer. Another study, [59], identified a profile of 76 genes with good predictive properties in patients with lymph-node-negative breast cancer. Interestingly, the overlap between the profiles from the two studies consisted of only 3 genes. In order to improve the predictive value of gene expression for this purpose it is reasonable to group genes according to the pathways they belong to and attempt to find pathway profiles instead. However, there are several obstacles to this strategy. For one, our knowledge of the many relevant molecular pathways is still very incomplete. Furthermore, most genes have not been assigned to individual pathways yet. And, as we have seen, the notion of pathway is not clearly defined in the first place, and one might better focus on networks; but these are even less well understood at this time.

In this section we focus on a study [31] that proposes as an alternative the combination of DNA microarray data and our increasingly good knowledge of the human protein-protein-interaction network. Overlaying gene expression on top of the proteins in this network one can then search for collections of sub-networks that provide better discrimination between patients with good prognosis and those with poor prognosis. Chuang et al. tested this idea on the data from the two studies [58,59], with promising results.

The mathematical model in [46] is a human protein-protein-interaction network N . Its nodes consist of a collection of 11,203 human proteins. An edge between two nodes/proteins represents evidence of a physical interaction between the two proteins. Such evidence can be obtained from the literature. In [60], for instance, a literature mining effort resulted in 31,609 interactions between 7,748 human proteins. In addition, in [61] a high-throughput yeast two-hybrid system was used to obtain approximately 2,800 interactions between human proteins. Using a variety of literature sources, the network in [31] contains 57,235 interactions.

The role of model simulation for a given set of experimental data is the computation of certain types of subnetworks of N . The input here is a gene expression data set consisting of a collection of DNA microarrays from patient samples, classified as either metastatic or non-metastatic, which is superimposed on the nodes of the subnetworks of the network N . A subnetwork M of N is simply a connected subgraph, that is, a collection of nodes and edges in N , so that any two nodes in this collection are connected by a path. We can assign a score $S(M)$ to a subnetwork M through the following steps:

- After a suitable transformation, compute the average of the expression values of the genes corresponding to all proteins in M , for each DNA microarray in the data set. We obtain a collection X of values. We can view X as a discrete random variable.
- Let Y be the discrete random variable that takes on the values 1 or 2 (defined by a probability distribution on these values), depending on whether a given DNA microarray corresponds to a patient with metastatic or non-metastatic breast cancer.
- Compute the mutual information between the random variables X and Y to obtain $S(M)$.

The mutual information between two random variables X and Y is used in information theory and measures the mutual dependence of the two variables. It is computed as follows:

$$MI(x, Y) = \sum_x \sum_y p(x, y) \log(p(x, y) / p(x) p(y)).$$

Here, the sums extend over all x in X , respectively y in Y , $p(x, y)$ is the joint probability distribution function of X and Y , and $\log$ is the logarithm function to the base 2. In our case, the mutual information measure computes essentially how much information about Y we obtain from knowing X , that is, how much we know about the classification of the samples if we know the activity score of the subnetwork with respect to the given DNA microarray data set.

The approach in [31] is to now identify a collection of subnetworks that can serve as markers to discriminate between metastatic and non-metastatic cancer. This is accomplished with a heuristic search algorithm. For the microarray data set S1 in [58] a total of 149 such networks, including a total of 618 genes, were identified as discriminative, and 243 were found for the data set S2 in [59], totaling 906 genes. For a new expression profile, the subnetworks from S1 predicted outcome with 70.1% accuracy while the subnetworks from S2 had 72.2% accuracy. This compares favorably to the rate of 62% for single gene marker classification (that is, use of differential expression of each of a collection of genes) of S1 in [58] and 63% for S2 in [59]. The subnetworks derived from S1 classified data from S2 with 48.8% accuracy and the subnetworks from S2 classified data from S1 with 55.8% accuracy. In comparison, the differential expression-based tests in [58] and [59] could classify the other data set with 45.3% and 41.5% accuracy, respectively. This shows a significant shortcoming of this and other methods, which seem to be dependent on the data set used to derive the subnetworks. Across data sets the subnetwork markers seems to perform only marginally better than single gene markers, even though there was a significantly larger overlap in subnetworks from the two data sets S1 and S2 (12.7%) than in individual genes (1.7%).

As an example of an interesting biological conclusion to be drawn from the subnetwork markers, while thioredoxin was not differentially expressed in S2, it instead connected many nodes of differentially expressed genes involved in DNA replication and cell mobility. The corresponding subnetworks were predictive of the S1 data. The Ras-related subnetwork and RAD54L-related proteasome from S1 were predictive of the S2 data. While many well-established prognostic markers for breast cancer disease outcome, such as HER-2/neu (ERBB2) or Myc, were not present in single gene marker analysis alone [62], they played a central role in the discriminative subnetworks by interconnecting many expression-responsive genes. We will observe a similar phenomenon in the next case study.

This paper represents a novel approach to integrate transcript and protein level information into networks to identify markers for disease outcome. While this approach may improve predictive performance, it also raises questions. For example, how good a substitute are protein-protein-interaction networks for more detailed realistic signaling networks involving transcripts, proteins, and a variety of small molecules? Undoubtedly, the performance of markers based on them might improve method performance. With improved data collection and improved mathematical and computational algorithms, network-based markers for disease outcome might soon become much more sophisticated, and [31] is a step in this direction.

2. Prediction of oncogenes and molecular perturbation targets in B-cell lymphomas

A network-centric approach, as in [31], can also be used to distinguish normal from malignant cells and identify targets and effectors of specific biochemical perturbations, potentially useful for the identification of drug targets. This approach is taken in [33] with a focus on B cell lymphomas. The central tool in this approach is the *B-cell interactome (BCI)*, a genome-wide cellular network that incorporates several different molecular interaction types in a human B cell, including transcriptional and signaling interactions, as well as complex formation. The BCI was assembled using information from a variety of sources, such as the literature and several databases, in addition to information obtained through the use of statistical network inference algorithms applied to DNA microarray data. It consists of 64,649 unique pairwise interactions and represents an “average” set of molecular interactions supported by the majority of B-cell samples from several stages of normal development (naïve, memory, germinal center) as well as from several tumor phenotypes.

This network is used as a background against which to test DNA microarray data from different tumor phenotypes. For a given phenotype, an edge in the BCI may disappear (loss of correlation, LoC) in this phenotype or may appear (gain of correlation, GoC), based on an information-theoretic test. One can then rank individual genes based on the statistical significance of the LoC/GoC enrichment among the interactions in which they directly participate. Note that this is quite different from a differential-expression-based ranking of genes. Here, the ranking is made on the basis of the extent to which the BCI network neighborhood of a gene changes in different phenotypes. The method was validated using three well-characterized B-cell tumor phenotypes, follicular lymphoma, Burkitt's lymphoma, and mantle cell lymphoma, as well as a set of biochemical perturbation assays.

The BCI interactome—The B-cell interactome (BCI) used in the study, based on an earlier version published in [63], integrates information from several different sources into a network of protein-protein, protein-DNA, and modulatory interactions in mammalian B-cells. This network can be represented as a graph with three types of edges. The nodes of the graph represent gene products. An undirected edge $P - Q$ between two proteins P and Q represents a protein-protein interaction, whereas a directed edge $P \rightarrow G$ from a protein P to a gene G represents a protein-DNA interaction. Finally, a directed edge from a node to an edge between nodes represents a modulatory interaction, arising from a protein, such as an activating kinase,

modulating the regulation of a target by a transcription factor. As mentioned earlier, it represents a “background” network that can be used as a control against which to evaluate different phenotypes.

Whether a specific potential interaction does in fact exist is determined by a Bayesian evidence integration model (that is, a probabilistic model that weights the reliability of different types of input data in computing likelihood), which evaluates several different “clues” about the interaction. Specifically, the *posterior odds* O_{post} is computed as the product of the *prior odds* O_{prior} and a *likelihood ratio* LR . Here, the prior odds are computed as the average probability that two random gene products interact, based on the number of interactions expected in a specific cellular context. The likelihood ratio of an interaction between gene products x and y , given a collection of clues $c_1, \dots, c_N$ is computed as

$$LR(x, y; c_1, \dots, c_N) = P(c_1 \cdots c_N | x \text{ and } y \text{ interact}) / P(c_1 \cdots c_N | x \text{ and } y \text{ do not interact}),$$

where $P(X/Y)$ denotes the conditional probability of event X happening, given the occurrence of event Y . In order to compute these probabilities one can use a *positive*, respectively *negative*, *gold standard set* of interactions, respectively non-interactions and calculate how many times a specific clue is observed in the positive and negative gold standard set. The gold standard sets should include only gene product pairs for which it is known whether they do or do not interact.

The positive gold standard set for protein-protein interactions (GSP) was constructed by extracting human protein-protein interactions from several existing data bases, which were obtained through low-throughput, high quality experiments. The resulting network includes 34,842 unique protein-protein interactions involving 7,323 genes. Since negative interactions are less well documented in the literature, the negative gold standard set (GSN) was constructed using an indirect method. Proteins were classified into four subcellular compartments, including the cell periphery and exocytic pathway, cytoplasm, mitochondria, and nucleus. For each pair of compartments, the enrichment of interactions involving proteins from the two compartments was computed, which resulted in compartment pairs that were likely to contain proteins involved in an interaction. The GSN was then taken to contain all pairs of proteins contained in cell compartment pairs not enriched for interactions, resulting in 18,359,948 negative interactions. Since many proteins are known to migrate between intracellular compartments, this approach may be problematic, however.

For protein-DNA interactions the amount of experimentally validated data is significantly smaller. Here, the approach taken in [63] was to focus on the well-studied transcription factor MYC. The GSP was taken to include a set of 1,041 B-cell specific genes known to be regulated by MYC. The GSN was taken to be its genomic complement.

The resulting four gold standard sets were then used to assign levels of confidence to evidence for interactions obtained using several other methods. For protein-protein interactions, interaction data from the model organisms *Caenorhabditis elegans*, *Drosophila melanogaster*, *Mus musculus*, and *Saccharomyces cerevisiae* were used by mapping model organism genes to human genes based on published information [64]. In addition, interactions from two yeast two-hybrid data sets were included. Secondly, interactions were obtained by applying a literature data-mining algorithm, based on a collection of keywords. These were supplemented by information from the Gene Ontology (GO) database, motivated by the observation that interacting proteins tend to share the same biological process. Finally, a large set of human gene expression profiles were used to extract interaction information based on gene co-

expression. This was done by computing the pairwise mutual information (introduced in the first case study) of over 6000 genes.

For protein-DNA interactions information was extracted from databases of interactions in the mouse, mapped to human interactions via [64]. The largest contribution to the set of protein-DNA interactions, however, came from the application of the statistical network inference algorithm ARACNE [65] to the above-mentioned collection of DNA microarray profiles. ARACNE also relies on the mutual information measure to establish correlations between genes. These different methods for establishing pairwise interactions represent the evidence, or clues, that are then used to compute the likelihood ratio needed to compute the odds that a given pair of gene products interact, as described earlier. The final assembly of this information comprises the B-cell interactome, used in [33] as the background, against which different phenotypes of B-cells are tested.

The first step is to apply the mutual information measure introduced earlier to a large collection of DNA microarray gene profiles in B-cells comprising three different cancer phenotypes, to test, for each pair of gene products in the BCI, whether its relationship (interaction/non-interaction) is consistent with the mutual information measure obtained for this pair.

Accordingly, the edge between the two is classified as Loss of Correlation (LoC) if the edge is present in the BCI but not in the new phenotype, or as Gain of Correlation (GoC) if the edge is absent in the BCI but present in the new phenotype. The next step is to rank genes according to the statistical significance of the LoC/GoC enrichment among the interactions in which they directly participate. This provides a measure of dysregulation for a given gene in a given cancer phenotype which is based on the changes in the gene's network neighborhood rather than just its individual differential expression. The approach can identify small network modules that are dysregulated, providing clues to the mechanistic processes underlying oncogenesis.

Three cancer phenotypes are represented in the DNA microarray set. Follicular lymphoma, one of the most common B-cell non-Hodgkin's lymphomas, exhibited a relatively small amount of dysregulation, with only 192 LoC/GoC interactions. Two genes, BCL2 (B-cell CLL/lymphoma 2) and SMAD1, were ranked very high in this analysis (first and sixth). BCL2 is involved in 8 of the 192 interactions, and SMAD1 has been shown to be involved in abnormal pathway activation in follicular lymphoma, mediated by tumor-transforming growth factor- β . By comparison, neither gene is ranked highly by a differential expression analysis. For Burkitt's lymphoma the transcription factor MYC is known to play an important regulatory role [18]. The network dysregulation analysis ranked MYC in place 15, whereas a differential expression analysis ranks it 32. Furthermore, a key effector of MYC, MTA1 (metastasis associated 1), was ranked third, even though a differential expression analysis does not place it among the top 1000 genes. For mantle cell lymphoma, the method also identified several relevant genes, in agreement with differential expression analysis.

In summary, the systems biology method developed in [33] identifies genes whose function within the cellular network of gene products in a given B-cell lymphoma phenotype is significantly altered. The viewpoint underlying the method is that the network neighborhood of a gene is more predictive of altered function than mere differential expression of the gene alone, and provides more information about the possible mechanisms underlying the phenotypic changes. The method has been validated in different cancer phenotypes, including Burkitt's lymphoma, follicular lymphoma, and mantle cell lymphoma, and with a validation experiment, consisting of a perturbation of the *CD40* (TNF receptor superfamily member 5) pathway, with promising results. This is despite the fact that the method requires a number of assumptions to compensate for the lack of qualitatively and quantitatively adequate data to establish the B-cell interactome and estimate likelihood ratios. It can be expected that with the improved availability of data, the statistical accuracy of methods such as this one will improve.

3. A multiscale mathematical model of cancer, and its use in designing radiation therapies

Currently, the three principal cancer treatments are surgery, chemotherapy, and radiotherapy. The latter is administered to the majority of cancer patients today, possibly in conjunction with other treatments. Treatment protocols typically include daily doses of ionizing radiation over the course of several weeks. A major limitation of the efficacy of radiotherapy is that only proliferative cells in a tumor are susceptible to DNA damage from radiation. If solid tumors outgrow their blood supply, then many of their cells become quiescent due to hypoxia. Furthermore, sensitivity of cells to ionizing radiation also depends on what phase of the cell cycle they are in. In order to study the optimal timing of a radiation dose it is therefore important to understand the cumulative effect of these factors. Since the cellular response to hypoxia is controlled at the molecular level through a complex signaling network, a full understanding of the cumulative effect of environmental conditions and molecular events on cell sensitivity to ionizing radiation must integrate dynamic events at several spatial and temporal scales. The paper [29] carries out this integration in the case of colorectal cancer through a complex multi-scale mathematical model, which can be used to test the efficacy of different radiation protocols. Application of this model is used to suggest an alternative protocol that administers radiation doses at times coinciding with a maximal proportion of proliferative cells.

The model integrates events at three different scales, including a genetic signaling network that responds to environmental conditions and controls the proliferative state of a cell, the progression of a cell through the different stages of the cell cycle, as well as tumor growth at the tissue level and environmental conditions such as hypoxia and overpopulation. The 2-dimensional computational domain of the model is an 8 cm square area of tissue which contains five small circular tumor masses, one in the center and the other four toward the corners, with two sources of oxygen to the right and left of the central tumor. The oxygen concentration in the tissue is described by a diffusion equation (that is, a partial differential equation which describes density fluctuations in a material undergoing diffusion) with a diffusion coefficient that remains constant throughout the tissue (so that the diffusion equation becomes linear).

The propagation of tumor cells through the tissue is modeled using a macroscopic continuous fluid dynamics model, which describes the flow of tumor cells in the extracellular matrix:

$$v = -k \nabla p.$$

Here, v represents the flux (discharge per unit area), k is a constant that indicates media permeability, and ∇p is the gradient vector of the pressure field p , indicating direction and magnitude of its change. Let $n_\phi(x, y, t)$ denote the density of cells at time t at position (x, y) which are in cell cycle phase ϕ , where ϕ is one of the phases G_1 , S , G_2M , G_0 (= quiescent), *Apop*. Then n_ϕ can be described by the advection equation

$$\partial n_\phi / \partial t + \nabla \cdot (v n_\phi) = P_\phi,$$

where P_ϕ is the cell density proliferation term at time t . (Here, ∂ denotes the partial derivative of the multivariate function n_ϕ with respect to variable t , and ∇ denotes the gradient vector of its argument function.) Advection equations are standard tools to describe transport of a substance in a moving fluid. Summing up the equations for the different cell cycle phases we obtain the global age-structured equation describing cell density:

$$-\nabla \cdot (k \nabla p) = \sum_{\phi} P_\phi.$$

As time passes in this dynamic model, cells at different locations (x,y) in the computational space progress through the various stages of the cell cycle, which is implemented through a discrete mathematical model at the cellular scale. The cell cycle is represented as three stages, G_1 , S, and G_2 -M, together with a stage G_0 for quiescent cells and *Apop* for cells that have been targeted for apoptosis. The different stages have different lengths and are divided into discrete time intervals. A given cell progresses through the time intervals in discrete steps at constant speed. At the end of the G_1 phase there is a checkpoint at which two tests are performed, one for DNA damage and the other for two environmental conditions, hypoxia and overpopulation. In the case of DNA damage, the cell is targeted for apoptosis (*Apop*) through activation of *p53*, if it is not mutated. Similarly, in the event of hypoxia or overpopulation in the location (x,y) , the cell is directed to the quiescent state G_0 through activation of the APC or SMAD pathway (see below). When environmental conditions change the cell can return to the proliferative state and progress through the cell cycle. If the cell passes this checkpoint it progresses through the S and G_2 -M phases of the cell cycle to produce offspring.

As mentioned, at the molecular level of each cell a network of signaling pathways is active that reacts to environmental conditions the cell finds itself in and regulates progression through the cell cycle. This network consists of four interconnected pathways, which process signals from overpopulation (APC), growth factor signals (RAS), hypoxia (SMAD), and DNA damage (*p53*). It is implemented as a Boolean network model in which a given gene/protein in the network is either ON or OFF. The determination of the state of a given network node at time $t+1$ is determined using a logical expression involving the states at time t of all the nodes that provide input to it in the network. For instance, the logical expression for c-myc is:

$$c\text{-myc}^{t+1} = \text{RAS}^t \wedge \text{B-cat}^t \wedge \neg \text{SMAD}^t.$$

This formula is to be interpreted as saying that c-myc at time $t+1$ will be ON precisely if RAS and B-cat are ON at time t and SMAD is OFF at time t . There are similar formulas for the other nodes in the network. Mutation of the genes SMAD and APC are implemented in the model by setting $\text{SMAD}^{t+1} = 0 = \text{APC}^{t+1}$. Using this Boolean network model one can compute the cell cycle signal resulting from a given set of environmental conditions. If hypoxia occurs at time t , then $\text{SMAD}^{t+1} = 1$; similarly for overpopulation.

To simulate the model, one needs to make several assumptions. For the discrete cell cycle and gene regulation models the time step is chosen to be 1 hour, and simulations are run for 320 hours. The model is initialized with a distribution of normal and malignant (i.e., at least one mutation) cells at different points of the cell cycle. First, hypoxia and overpopulation conditions are checked, using model parameters, and the genetic model is run for each cell with these inputs until it reaches a steady state. Then the cell cycle model is run for one time step (=1 hour) and new values are computed for the cell densities n at different cell cycle stages and proliferation terms P . Finally, the advection equations are solved with these new values.

Radiation therapy—As mentioned earlier, it is assumed that only proliferative cells are sensitive to radiation treatment. Furthermore, DNA damage is measured as the number of double strand breaks, given by the product of the radiosensitivity and the radiation dose. DNA damage is implemented in the model by setting a threshold for the number of double strand breaks. If the number of breaks exceeds this threshold, *p53* is activated and the cell is subsequently directed to apoptosis. The standard radiotherapy protocol used in the simulations consists of a 2 Gy dose per day, five days a week, repeated for several weeks. In the model the dose is assumed to be uniformly distributed over the spatial domain. Thus, a treatment is implemented in the simulation by computing for each cell its susceptibility and the damage it

sustains, and activating *p53* as appropriate. Then, at the next time step, the genetic model will target the cell for apoptosis.

Simulation results—The goal of the paper is to assess the efficacy of the standard radiotherapy protocol, which is done by counting the number of cells directed to apoptosis after repeated applications of a radiation dose. The simulation results show clearly that during 10 treatments on the standard schedule an increasingly large number of cells in the interior of the tumors are in a quiescent state, due to the effect of increasing hypoxia. As a result, repeated treatments on a 24-hour schedule become increasingly less effective since the majority of cells are not sensitive. In the paper, an alternative treatment schedule is explored, which takes into account the length of the cell cycle. Here, after an initial treatment, the follow-up treatments are timed to coincide with a maximum number of mitotic cells and are delivered before hypoxia sets in. The result is an increase by a factor of 20 in the number of cells marked for apoptosis.

Summary—The model described here integrates relevant phenomena at several spatial and temporal scales into a heterogeneous mathematical model of tumor growth and the effect of radiotherapy. The key characteristic of this systems biology model is that it takes into account a complex network of relevant factors whose aggregate contribution can only be understood with the help of sophisticated mathematical modeling tools. It allows an in-depth exploration of different variants of the standard treatment protocol scheduling, which can provide the basis for the design of appropriate clinical trials. It also allows the exploration of the effects of different mutations on treatment response. At the tissue level, a fluid dynamics model is used, working with cell densities rather than numbers of individual cells. Using individual-based simulation techniques to create a fully discrete model might improve its accuracy. Also, constructing a true three-dimensional model might bring to light important spatial effects. At the same time, both of these improvements would come at substantial computational cost. Ultimately, as depicted in Fig. 2, comparing simulated outcomes with experimental results will point to gaps or shortfalls that can be addressed through additional modeling.

VI. OUTLOOK AND CONCLUSION

In order to obtain a useful understanding of cancer, it is necessary to take a systemic view of the disease, and to utilize mathematical modeling techniques. We have described in detail three examples the utility of such a view. The first two examples use a graphical depiction of a large-scale intracellular network as a mathematical tool, representing interactions between molecular species. Changes in the network structure, rather than changes in individual genes, then form the basis for an analysis of the differences between healthy and malignant cells. In the first case this leads to prognostic markers for the propensity to metastasis in breast cancer patients, in the second case to a new method for identifying oncogenes. The third example represents an attempt to bridge different spatial and temporal scales to understand the effect of different environmental parameters on cell proliferation, integrating the molecular, cellular, and tissue scales. It can then be used to study a variety of different treatment approaches at each of these scales. The example here focuses on the effect of radiation therapy.

In order to use such models effectively they need to be informed by appropriate high quality data. As we have seen in the discussion of the case studies, compromises had to be made due to either the lack of data, their poor quality, or both. This applies in particular to detailed dynamic models like the one in [29]. In order to calibrate such models accurately, detailed time course observations with coherent measurements at different scales are needed, but are currently not available in most or all cases. It can be expected that with technological progress data collection will become cheaper, more abundant, and of higher quality. New imaging techniques at the molecular and tissue level will allow a representation of events at a much higher scale of resolution, which can inform increasingly detailed and complex models.

Increasing computing power will enable us to simulate and analyze complex models in a short enough time frame to be useful as a clinical tool.

In [24] the authors lay out a vision for the future: “Two decades from now, having fully charted the wiring diagrams of every cellular signaling pathway, it will be possible to lay out the complete “integrated circuit of the cell.” We will then be able to apply the tools of mathematical modeling to explain how specific genetic lesions serve to reprogram this integrated circuit in each of the constituent cell types so as to manifest cancer.” We have attempted in this review to show that in the field of cancer, the rapid progress of systems biology and its embrace of diverse mathematical and statistical modeling methods has brought this vision within grasp. Even more, quantitative tools are helping to translate insights about parts of this integrated circuit to clinical practice in the form of new diagnostic tools, modified or new treatments, and potential new drug targets for cancer treatment. The role of systems biology in translational cancer medicine is bound to increase, driven by integrated teams of mathematical modelers, cancer biologists, and clinicians. By working together in a transdisciplinary fashion, the needs of the clinic can directly inspire needed advances in fundamental understanding, enabled by quantitative tools.

Acknowledgments

This work was supported in part by grants T32 CA079448 (V.H.), R37 DK042412 (F.M.T.), R01CA120170 (V.S), and R01 GM080219 (P. M.). P. Mendes also acknowledges financial support from the U.K. BBSRC and EPSRC. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health. Thanks are due to M. Hatakeyama, the lead author of [19], for providing Mathematica code to generate Figure 1C.

VIII. REFERENCES

1. Wood LD, et al. The Genomic Landscapes of Human Breast and Colorectal Cancers. *Science* 2007;318(5853):1108–1113. [PubMed: 17932254]
2. Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. *Nat Med* 2004;10(8):789–799. [PubMed: 15286780]
3. Hornberg JJ, et al. Cancer: A systems biology disease. *Biosystems* 2006;83(23):81–90. [PubMed: 16426740]
4. Ahn AC, et al. The Limits of Reductionism in Medicine: Could Systems Biology Offer an Alternative? *PLoS Medicine* 2006;3(6):e208. [PubMed: 16681415]
5. Ahn AC, et al. The Clinical Applications of a Systems Approach. *PLoS Medicine* 2006;3(7):e209. [PubMed: 16683861]
6. Stiles B, et al. Essential Role of AKT-1/Protein Kinase B{alpha} in PTEN-Controlled Tumorigenesis. *Mol. Cell. Biol* 2002;22(11):3842–3851. [PubMed: 11997518]
7. Rodriguez-Viciana P, et al. Activation of phosphoinositide 3-kinase by interaction with Ras and by point mutation. *EMBO J* 1996;15(10):2442–2451. [PubMed: 8665852]
8. Tokunaga E, et al. Akt is frequently activated in HER2/neu-positive breast cancers and associated with poor prognosis among hormone-treated patients. *International Journal of Cancer* 2006;118(2):284–289.
9. Datta SR, et al. Akt Phosphorylation of BAD Couples Survival Signals to the Cell-Intrinsic Death Machinery. *Cell* 1997;91(2):231–241. [PubMed: 9346240]
10. Cardone MH, et al. Regulation of Cell Death Protease Caspase-9 by Phosphorylation. *Science* 1998;282(5392):1318–1321. [PubMed: 9812896]
11. Brunet A, et al. Akt Promotes Cell Survival by Phosphorylating and Inhibiting a Forkhead Transcription Factor. *Cell* 1999;96(6):857–868. [PubMed: 10102273]
12. Mayo LD, Donner DB. A phosphatidylinositol 3-kinase/Akt pathway promotes translocation of Mdm2 from the cytoplasm to the nucleus. *Proceedings of the National Academy of Sciences* 2001;98(20):11598–11603.

13. Inoki K, et al. TSC2 is phosphorylated and inhibited by Akt and suppresses mTOR signalling. *Nat Cell Biol* 2002;4(9):648–657. [PubMed: 12172553]
14. Elstrom RL, et al. Akt Stimulates Aerobic Glycolysis in Cancer Cells. *Cancer Res* 2004;64(11):3892–3899. [PubMed: 15172999]
15. Kane LP, et al. Induction of NF- κ B by the Akt/PKB kinase. *Current Biology* 1999;9(11):601–605. [PubMed: 10359702]
16. Lerebours F, et al. NF-kappa B genes have a major role in inflammatory breast cancer. *BMC Cancer* 2008;8
17. Futagami S, et al. COX-2 and CCR2 induced by CD40 ligand and MCP-1 are linked to VEGF production in endothelial cells. *Prostaglandins, Leukotrienes and Essential Fatty Acids* 2008;78(2):137–146.
18. Li Z, et al. A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells. *PNAS* 2003;100:8164–8169. [PubMed: 12808131]
19. Hatakeyama M, et al. A computational model on the modulation of mitogen-activated protein kinase (MAPK) and Akt pathways in heregulin-induced ErbB signalling. *Biochem. J* 2003;373(2):451–463. [PubMed: 12691603]
20. Stolovitzky G, Monroe D, Califano A. Dialogue on reverse-engineering assessment and methods: the DREAM of high-throughput pathway inference. *Ann N Y Acad Sci* 2007;1115:1–22. [PubMed: 17925349]
21. Hartemink AJ. Reverse engineering gene regulatory networks. *Nat Biotechnol* 2005;23(5):554–5. [PubMed: 15877071]
22. Laubenbacher R, Stigler B. A computational algebra approach to the reverse engineering of gene regulatory networks. *J Theor Biol.* 2004
23. Gardner TS, et al. Inferring Genetic Networks and Identifying Compound Mode of Action via Expression Profiling. *Science* 2003;301(5629):102–105. [PubMed: 12843395]
24. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell* 2000;100:57–70. [PubMed: 10647931]
25. de Pillis LG, Gu W, Radunskey AE. Mixed immunotherapy and chemotherapy of tumors: modeling, applications and biological interpretations. *J Theor Biol* 2006;238(4):841–62. [PubMed: 16153659]
26. Pujana MA, et al. Network modeling links breast cancer susceptibility and centrosome dysfunction. *Nat Genet* 2007;39(11):1338–1349. [PubMed: 17922014]
27. Ergun A, et al. A network biology approach to prostate cancer. *Mol Syst Biol* 2007;3
28. Abdollahi A, et al. Transcriptional network governing the angiogenic switch in human pancreatic cancer. *Proc. Nat. Acad. Sci* 2007;104(31):12890–12895. [PubMed: 17652168]
29. Ribba B, Colin T, Schnell S. A multiscale mathematical model of cancer, and its use in analyzing irradiation therapies. *Theor. Biol. Med. Modelling* 2006;3:7.
30. Haeno H, Iwasa Y, Michor F. The Evolution of Two Mutations During Clonal Expansion. *Genetics* 2007;177(4):2209–2221. [PubMed: 18073428]
31. Chuang H-Y, et al. Network-based classification of breast cancer metastasis. *Mol. Sys. Biol* 2007;3
32. Christopher R, et al. Data-Driven Computer Simulation of Human Cancer Cell. *Annals of the New York Academy of Sciences* 2004;1020(1):132–153. [PubMed: 15208190]
33. Mani KM, et al. A systems biology approach to prediction of oncogenes and molecular perturbation targets in B-cell lymphomas. *Mol Syst Biol* 2008;4
34. Shachar R, et al. A systems-level approach to mapping the telomere length maintenance gene circuitry. *Mol Syst Biol* 2008;4:172. [PubMed: 18319724]
35. Qi JP, et al. A mathematical model of P53 gene regulatory networks under radiotherapy. *Biosystems* 2007;90(3):698–706. [PubMed: 17512110]
36. Birtwistle MR., et al. *Mol Syst Biol.* EMBO and Nature Publishing Group; 2007. Ligand-dependent responses of the ErbB signaling network: experimental and modeling analyses.
37. Athale CA, Deisboeck TS. The effects of EGF-receptor density on multiscale tumor growth patterns. *Journal of Theoretical Biology* 2006;238(4):771–779. [PubMed: 16126230]
38. Huang PH, et al. Quantitative analysis of EGFRvIII cellular signaling networks reveals a combinatorial therapeutic strategy for glioblastoma. *Proc. Nat. Acad. Sci* 2007;104(31):12867–12872. [PubMed: 17646646]

39. Reyman S, Borlak J. Transcription profiling of lung adenocarcinomas of c-myc-transgenic mice: identification of the c-myc regulatory gene network. *BMC Sys Biol* 2008;2:46.
40. Kolch W. Ras/Raf signalling and emerging pharmacotherapeutic targets. *Exp Opin Pharmacother* 2002;3:709–718.
41. Kholodenko, BN., et al. Quantification of Short Term Signaling by the Epidermal Growth Factor Receptor. 1999. p. 30169-30181.
42. Hendriks BS, Wiley HS, Lauffenburger D. HER2-Mediated Effects on EGFR Endosomal Sorting: Analysis of Biophysical Mechanisms. *Biophys. J* 2003;85(4):2732–2745. [PubMed: 14507736]
43. Kumar N, et al. Modeling HER2 Effects on Cell Behavior from Mass Spectrometry Phosphotyrosine Data. *PLoS Computational Biology* 2007;3(1):e4. [PubMed: 17206861]
44. Castoria G, et al. Integrating signals between cAMP and MAPK pathways in breast cancer. *Frontiers in Bioscience* 2008;13:1318–1327. [PubMed: 17981632]
45. Kiyatkin A, et al. Scaffolding Protein Grb2-associated Binder 1 Sustains Epidermal Growth Factor-induced Mitogenic and Survival Signaling by Multiple Positive Feedback Loops. *J. Biol. Chem* 2006;281(29):19925–19938. [PubMed: 16687399]
46. Stites EC, et al. Network Analysis of Oncogenic Ras Activation in Cancer. *Science* 2007;318(5849):463–467. [PubMed: 17947584]
47. Sveiczzer A, Tyson J, Novak B. Modelling the fission yeast cell cycle. *Brief Funct Genomic Proteomic* 2004;2(4):298–307. [PubMed: 15163365]
48. Haberichter T, et al. A systems biology dynamical model of mammalian G1 cell cycle progression. *Mol Syst Biol* 2007;3:84. [PubMed: 17299420]
49. Watt F, Frye M, Benitah S. MYC in mammalian epidermis: how can an oncogene stimulate differentiation? *Nature Reviews Cancer* 2008;8(3):234–242.
50. Bottani S, Grammaticos B. Analysis of a minimal model for p53 oscillations. *J Theor Biol* 2007;249(2):235–245. [PubMed: 17850824]
51. Geva-Zatorsky N, et al. Oscillations and variability in the p53 systems. *Mol Syst Biol* 2006;22006:0033
52. Anderson AR, et al. Tumor morphology and phenotypic evolution driven by selective pressure from the microenvironment. *Cell* 2006;127:905–915. [PubMed: 17129778]
53. Jiang Y, et al. A Multiscale Model for Avascular Tumor Growth. *Biophys. J* 2005;89(6):3884–3894. [PubMed: 16199495]
54. Zhang L, Athale CA, Deisboeck TS. Development of a three-dimensional agent-based tumor model: simulating gene-protein interaction profiles, cell phenotypes, and multicellular patterns in brain cancer. *J Theor Biol* 2007;244:96–107. [PubMed: 16949103]
55. Anderson AR, Quaranta V. Integrative mathematical oncology. *Nature Reviews Cancer* 2008;8:227–234.
56. Mueller MM, Fusenig NE. Friends or foes-bipolar effects of the tumor stroma in cancer. *Nature Reviews Cancer* 2004;4:839–849.
57. Weigelt B, Peterse JL, van't Veer LJ. Breast cancer metastasis: markers and models. *Nat Rev Cancer* 2005;5(8):591–602. [PubMed: 16056258]
58. van de Vijver MJ, et al. A Gene-Expression Signature as a Predictor of Survival in Breast Cancer. *N Engl J Med* 2002;347(25):1999–2009. [PubMed: 12490681]
59. Wang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 2005;365(9460):671–9. [PubMed: 15721472]
60. Ramani AK, et al. Consolidating the set of known human protein-protein interactions in preparation for large-scale mapping of the human interactome. *Genome Biol* 2005;6(5):R40. [PubMed: 15892868]
61. Rual JF, et al. Towards a proteome-scale map of the human protein-protein interaction network. *Nature* 2005;437(7062):1173–8. [PubMed: 16189514]
62. van't Veer LJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415(6871):530–536. [PubMed: 11823860]
63. Lefebvre C, et al. A context-specific network of protein-DNA and protein-protein interactions reveals new regulatory motifs in human B-cells. *Lecture Notes in Bioinf* 2007;4532:42–56.

64. O'Brian K, Remm M, Sonnhammer E. Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic Acids Res* 2005;33(Database issue):D476–D480. [PubMed: 15608241]
65. Margolin A, et al. ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 2006;7(Suppl 1):S1–7.
66. Djavani M, et al. Early blood profiles of virus infection in a monkey model for Lassa Fever. *J Virology* 2007;81(15):7960–73. [PubMed: 17522210]
67. Djavani M, et al. Gene expression in primate liver during viral hemorrhagic fever. *J Virology*. 2008under review

Figure 1.

Different views of the Akt pathway/network. Figure 1A. represents a schematic view of Akt and the other types of molecules it interacts with, based on information available in the literature. Figure 1B. was constructed using Ingenuity Pathway and data from [66,67]. Figure 1C. was obtained using an implementation of the differential equations model in [19]. The simulated concentration curves of Akt (...), ERK (—), and MEK (--) for the first 30 minutes after the stimulation by a high dose of heregulin (HRG).

Figure 2.

The systems biology strategy. The goal of a systems biology approach is to construct predictive mathematical models of biological systems. This figure depicts the general strategy to be followed. The starting point is existing biological knowledge about the system in question. It is used to construct a general mathematical framework for a model. The model might contain several unknown biological parameters, such as kinetic rate constants or diffusion rates. The model is refined through known parameter values and experimental data, to which the model can be fitted through simulation. At this stage, the model is used to generate biological hypotheses that can be tested experimentally. Through an iterative process of hypothesis and experiment, the model is refined until it accurately captures the key biological features of the system to be modeled.

Table 1

Systems biology studies related to cancer. This Table highlights several different systems biology approaches that have been used in the recent literature, and provides examples of the kinds of information that can be derived in such studies

Reference	Input	Method	Kinetic?	Output
1	Data from human/mouse studies	ODE	yes	Response to chemo/immunotherapy
2	Gene expression/healthy human	Statistics	no	links HMMR expression and breast cancer
3	Microarray data metastatic and non-metastatic prostate tumors	Mode-of-action network identification	no	Genes active in the advancement of prostate cancer
4	Gene expression from both healthy and diseased pancreas tissue	Network inference using significant genes/interactions	no	Collection of key players in angiogenesis including PPAR δ
5	Kinetics and interactions obtained from the literature	A multiscale model, involving different dynamic models	Yes	Tumor characteristics are important in predicting the efficacy of radiation therapy
6	Kinetics and interactions obtained from the literature	Probabilistic model for genetic mutations	yes	Probability for obtaining 2 mutations in a tumor cell varies greatly across tumors
7	Microarray data from breast cancer patients	Network motif selection using statistical analysis	no	Subnetworks that discriminate between metastatic and non-metastatic breast cancer
8	Kinetics and interactions obtained from the literature	ODE-based model of reactions in a cancer cell	yes	Tool in drug discovery and development
9	Microarray data from multiple phenotypes of lymphoma	Network inference using statistics/information theory	no	Phenotype-specific collections of dysregulated genes and interactions
10	Telomere maintenance genes	Statistics/graph theory	no	Network involved in telomere maintenance
11	modeling/kinetics/ biology of p53 gene regulatory networks	ODE model of p53 gene regulatory network	yes	Simulation of DSB generation/repair, ATM activation, and p53-Mdm2 feedback
12	Experimental data from MCF7 breast cancer cell line	ODE-based model of ErbB signaling	yes	Kinetic relationship of EGF and HRG act. of ErbB receptors to ERK and Akt signaling
13	Kinetics and interactions	ODE/agent-based multiscale model	yes	EGFR affects brain tumor dynamics
14	MS-based phosphoproteomics of glioblastoma cell lines	Network inference using clustering	no	EGFRvIII activated pathways and c-Met receptor as a target in glioblastoma.
15	Microarray data from lung tissue of c-Myc transgenic mice	Statistical network inference using in positional weight matrices	no	Transcriptional regulation occurring in c-Myc overexpression