Essays in Biochemistry (2018) EBC20180003
https://doi.org/10.1042/EBC20180003

PORTLAND PRESS

Review Article

Systems biology primer: the basic methods and approaches

Iman Tavassoly, Joseph Goldfarb and Ravi Iyengar

Department of Pharmacological Sciences and Systems Biology Center New York, Icahn School of Medicine at Mount Sinai, New York, NY 10029, U.S.A.

Correspondence: Ravi Iyengar (ravi.iyengar@mssm.edu)

Systems biology is an integrative discipline connecting the molecular components within a single biological scale and also among different scales (e.g. cells, tissues and organ systems) to physiological functions and organismal phenotypes through quantitative reasoning, computational models and high-throughput experimental technologies. Systems biology uses a wide range of quantitative experimental and computational methodologies to decode information flow from genes, proteins and other subcellular components of signaling, regulatory and functional pathways to control cell, tissue, organ and organismal level functions. The computational methods used in systems biology provide systems-level insights to understand interactions and dynamics at various scales, within cells, tissues, organs and organisms. In recent years, the systems biology framework has enabled research in quantitative and systems pharmacology and precision medicine for complex diseases. Here, we present a brief overview of current experimental and computational methods used in systems biology.

Introduction

In recent decades, our knowledge of the foundation of living organisms in terms of various components of cells, tissues and organ systems has been greatly expanded due to advances in technologies for high-throughput measurements such as genomics, transcriptomics, proteomics and metabolomics. In genetics and genomics, entire genomes of many organisms have been sequenced and the gene expression profiles comprehensively mapped. In biochemistry, mass spectrometry-based protein surveys have provided extensive lists of proteins and protein complexes, while molecular and cell biology have provided information on how proteins are organized to orchestrate the functions of subcellular systems such as cell organelles and cellular machinery components. Physiology has shed light on the complex functions of cells, tissues and organ systems. This enormous amount of information at different scales of organization can be used to obtain a new perspective that starts from genes and proteins, moves through subcellular interactions and pathways and ends in the physiology of cells, tissues and organ systems [1-4]. The availability of such multiscale information has catalyzed the formation of systems biology as a discipline in biomedical sciences. Systems biology is the study of molecular interactions at different levels, enabling the identification and description of the subcellular machinery that makes functional operational units in cells, tissues and organ systems resulting in physiological behaviors [5,6].

Historically, systems biology started by looking at cells, tissues and organ systems as complex biological systems [7]. The rapid development of genomics and sequencing technologies led to the uncovering of big datasets of basic components forming these complex systems [8,9]. Later, it was shown how interactions among molecular components of cells could give rise to functional behaviors that single components by themselves cannot [10-12]. One way to think of systems biology is that it provides a new and broader perspective of physiology. While physiology provides a description of functions in cells, tissues and organ systems using largely phenomenological approaches, systems biology integrates molecular biology and

Received: 19 July 2018
Revised: 22 August 2018
Accepted: 24 August 2018

Version of Record published:
04 October 2018

Essays in Biochemistry (2018) EBC20180003
https://doi.org/10.1042/EBC20180003

biochemistry of molecular components and their interactions and dynamics to understand how physiological functions arise and are controlled [1,13,14]. Systems biology integrates not only the molecular entities at a specific scale but also the connections among these molecular components at different scales. Integration of data is the core value in systems biology, in which the interactions of multiple components are treated as a single system. This integration can be applied at a single scale (e.g. the cellular level) to provide new systems-level insight, but also can be used to decode complex phenotypes at different scales. For example, systems biology is used to study the evolution of a cancer cell from a normal cell. This involves interactions among molecular components at the cell level. At the same time, systems biology can be used to integrate the interactions among cancer cells and the evolution of tumors. It is also capable of describing the interaction of different tissues such as blood vessels, tumors and the immune system to shed light on complex phenomena of cancer at the organ level [15-20].

Biological systems are multiscale, with multiple levels of organization and with multiple states at different times, and hence, systems-level analyses are particularly useful. Differences in scale of biological systems can be studied from molecular components to subcellular machinery (such as transcriptional and translational control machinery and cell motility machinery) and to cells, tissues, organ systems and whole organisms. In this systems-level view, as the organizational level of a system increases, it leads to new characteristics and capabilities [1,20]. Multiscale systems can be studied in two major ways: bottom-up and top-down. Both approaches have their advantages and disadvantages.

In a bottom-up approach, cellular and molecular components are studied as parts of a system that includes their interactions and dynamics leading to physiological functions. This approach has the ability to provide mechanistic insights into how different units work together to form a system. In this approach, however, as the system becomes bigger, the details may obscure the overall capabilities of the system [21]. In contrast, in the top-down approach, the system as a whole is studied, and the characteristics and potential capabilities of the system are discovered. This gives a big picture of the system, which can be comprehensive and integrative. In this approach, interactions among different units are often defined by correlation and the complexity of the biological systems often does not always allow one to make causal inferences [21]. The different experimental methods and computational approaches are summarized in Figure 1.

Genomic-wide analyses of single nucleotide polymorphisms, comprehensive transcriptomic profiling and deep proteomics that provide an extensive characterization of cellular proteins are all examples of top-down surveys that correlate molecular components with cellular, tissue or organismal level phenotypes. Although such relationships are often correlative, they can provide useful bookends for more mechanistic systems-level characterizations. In both bottom-up and top-down approaches, there are two main sets of tools: experimental tools and computational tools. Experimental studies in systems biology often start with omics, high-throughput technologies including genomics, transcriptomics, epigenomics, proteomics and metabolomics [2]. Such large datasets are analyzed by use of statistical models as well as graph theory-based models. In bottom-up approaches, low-throughput, but high fidelity experiments can provide a foundation for verification of predictions from computational models both qualitatively and quantitatively [22,23] (Figure 1).

One can also use a middle-out approach in systems biology, studying a higher level function by selecting only a limited number of lower level interactions deemed to be relevant to a specific phenotype. This approach considers modularity in systems biology and uses an approach like engineering methods that use only selected functionally vital components to build and understand a processing circuit or machine [24,25].

Systems level experimental analysis of cells

The systems-level analysis of cells requires information on all of the individual entities at different levels of cell function. Omics technologies provide such information and, in the process, yield vast amounts of data from genes, mRNAs, proteins and metabolites. These high-throughput methods measure many individual subcellular components that act as a system to control cell function. Genomics, which utilizes sequencing technologies and microarrays, can determine the sequence of genomes and characterize genomic determinants including single nucleotide polymorphisms (SNPs), indels and epigenetic regulatory sites (such as DNA methylation sites), affecting a specific phenotype or function in cells or organisms [26]. Transcriptomics measures the transcriptome of cells or tissues that consists of all RNA transcripts [27]. Epigenomics describes all epigenetic modifications such as DNA methylation and histone modifications in cells [28,29]. Proteomics, which often uses mass-spectrometry technologies, measures and catalogs proteins and post-translational modifications at a large scale [30]. Metabolomics is the large-scale study of metabolites in cells and tissues and uses liquid chromatography, mass-spectrometry and NMR technologies [31]. The information gained by such systems-wide surveys needs to be processed and organized to turn data into knowledge. The organizing and analyzing of large datasets are called Bioinformatics. Currently, there are many databases that store,

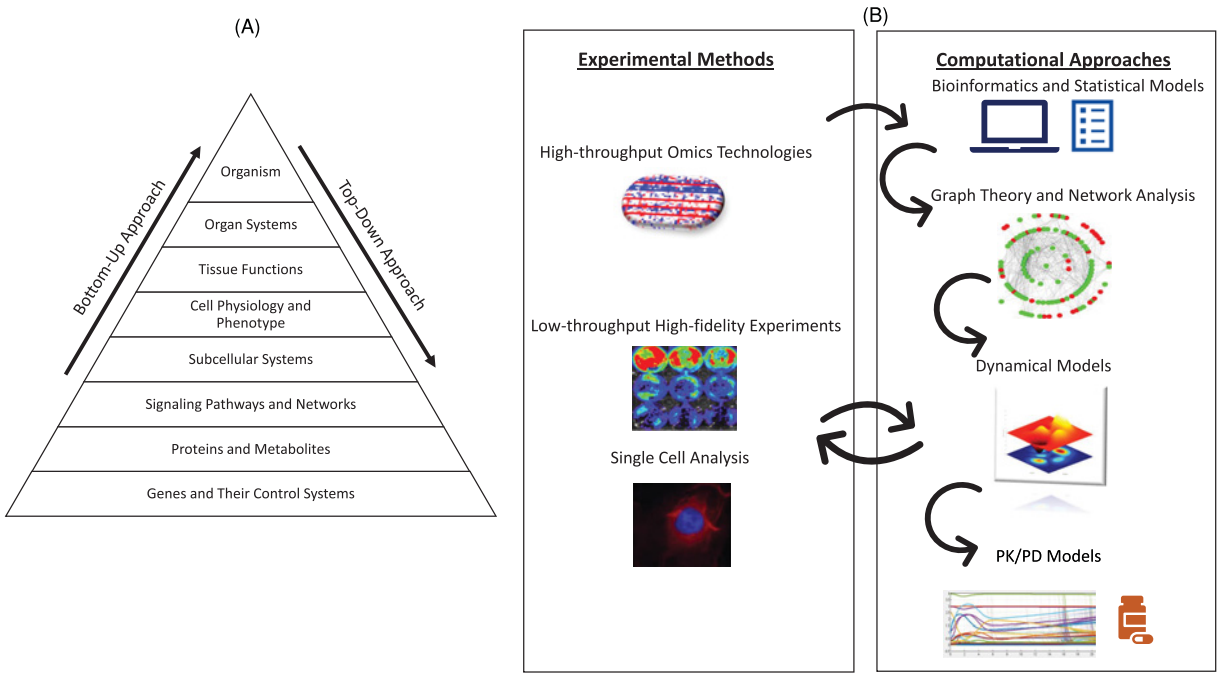

Figure 1. Systems biology approaches and methodologies

(**A**) Systems biology methodologies can be applied either in a bottom-up approach that puts small functional units together to make a system or in a top-down approach that starts from the global view of the system and then tries to study smaller subsystems. (**B**) Systems biology utilizes both experimental and computational frameworks to answer biological questions. Omics technology provides a platform to extract knowledge using bioinformatics, statistical methods and network analysis. The dynamical models of certain components in these networks must be verified by low-throughput, high-fidelity and single cell experiments that provide new strategies to improve and optimize the dynamical models. Dynamical models can be merged with PK/PD models to analyze therapeutic efficacies and design precision drug treatments.

and computational tools to analyze, these data, such as genomic characteristics including SNP profiles for diseases, mRNA profiles, protein networks etc. Systems biology integrates experiments and computational models to understand how systems function. Computation is a key feature that characterizes systems biology compared with classic biological disciplines such as biochemistry and cell biology. A good example of the use of large molecular datasets is Genome-wide Association Studies (GWAS), which is the process of finding variations in DNA sequence, usually SNPs, associated with increased risk of a specific disease or physiological state. GWAS is a useful map by which genomic data can be correlated with pathophysiological states. It can also contribute to understanding drug action and the discovery of new drug targets by evaluating genetic variations in response to drugs, and to progression of disease [32,33].

Qualitative methods include most of the omics technologies that produce large-scale, often comprehensive, lists of molecular components. Transcriptomics focuses on identifying all the mRNAs on a genome wide basis. As the cost of sequencing has come down dramatically in the past few years, transcriptomics measurements have moved from the use of microarray chips to sequencing methods [34]. Proteomics focuses on identifying proteins and their post-translational modifications using mass spectrometry [35]. Advances in computational identification of proteins from mass spectrometry data now allow for the identification of ~10,000 proteins per cell type [36]. Metabolomics uses mass spectrometry as well as NMR technologies to identify metabolites and track metabolic pathways [37]. Each of these omic technologies has advanced detailed experimental methods as well as specific informatics tools for transcriptomics [38], proteomics [39] and metabolomics [40]. The informatics tools are needed to analyze the large datasets to produce ranked lists of molecular entities that can be cast as pathways and networks to infer function.

From molecules to pathways and networks

Experimental omics studies produce large molecular datasets. Statistical methods are required to generate ranked lists of those molecular components (genes, mRNA, proteins etc.) involved in specific physiological or pathophysiological

Essays in Biochemistry (2018) EBC20180003
https://doi.org/10.1042/EBC20180003

states. Gene Set Enrichment Analysis (GSEA) is a statistical method to find potential molecular components responsible for phenotypes and functions based on those entities that are under- or over-represented in biological samples. The differentially expressed molecular entities (or, in general, differentially expressed biomarkers) are enriched using a specific ontology. An ontology is a set of structured terms with specific relationships that work like a classifier with hierarchical structure [41,42]. The ontology is a tool to find biological knowledge by association of data (genes or gene products) with biological processes, molecular functions and cellular components [41,42]. Several ontologies have been developed and used in systems biology including Gene Ontology (GO) and Molecular Biology of the Cell Ontology (MBCO). In addition, there are other bioinformatics tools such as the Kyoto Encyclopedia of Genes and Genomes (KEGG), Wikipathways, Reactome Pathway, Progeny Signatures and Broad Signatures to transform data into biological knowledge [42-48]. The results from GSEA yield knowledge about the pathways, including signaling pathways regulating the specific phenotype being studied.

Signaling pathways are the main systems that process information in cells. Signaling pathways receive signals from outside the cell and control cellular physiology in response to these signals. These pathways have many components each of which receives, transmits and transduces information to other components [13,14,23,49]. The flow of information, in the form of cellular signals, occurs in time and space and can be studied mathematically using dynamical systems theory and differential equations [14,16]. Receptors, which receive signals from outside the cell, and other intracellular signaling components, enable connectivity between signaling pathways within a network. The intracellular signaling components are information processing units, signal integrators and effectors that function as output devices that represent the cellular responses to extracellular signals [11,16].

In addition to linear pathways, GESA enables the construction and analysis of functional molecular networks. Networks are formed by interactions between molecular entities. These entities are called 'nodes' and the interactions between the entities are called 'edges'. Such interactions include direct binding leading to activation or inhibition of the downstream target and enzymatic activities [50-53].

Analysis of biological networks

A network is a set of nodes connected to each other via edges and mathematically defined as a graph. The structure and function of networks are studied by graph theory. Networks can be studied as computational units and systems, which provide insights into both their organization and functions [50,52,53]. In systems biology, the network nodes are cellular components and edges are reactions or interactions among these nodes. Viewing cell systems as networks is a helpful and practical way of understanding the functional organization of cells by analyzing network topology [10,18,50,52]. In cellular networks, there are cases when the relationships among nodes are conditional rather than fixed. Those networks where edges are defined in a probabilistic manner are called Bayesian networks [54]. Bayesian networks allow us to discover probabilistic relationships among molecular components and define the conditions that increase or decrease the probability of the relationships [55,56]. Networks can be represented as directed or undirected graphs. Undirected graphs represent the relationship among nodes without specifying hierarchy and are usually constructed from high-throughput large datasets Directed graphs represent not only the relationship among nodes but also the direction of signal propagation and hierarchy such as an upstream node regulating a downstream node. For example, in a directed graph of protein networks, inhibition or activation of a protein by another protein can be shown. There are many software packages and tools that enable the visualization of networks [57]. Visualization and analysis of cellular networks give a perspective on global organization of cell systems and help in identifying the key nodes in terms of connectivity. One of the properties of each network is the degree distribution, which is the probability distribution of all degrees of nodes within a network. The degree of a node is the number of edges via which it is connected to other nodes. A node with a degree much higher than average for the network is called a hub [58]. Hubs are not observed in random networks. Networks of real systems, such as cellular signaling systems, are organized differently from random networks. Real networks have a degree distribution that follows the power law and are called scale-free networks [59]. The robustness of a network and its sensitivity to perturbations are other properties of molecular networks that affect the functions of the system [50,60]. Perturbation, in terms of removing some of the nodes and measuring the resistance of the network to change, can be used to evaluate the robustness of a network. Scale-free networks are highly robust to random removal of nodes as there are few highly connected nodes [59]. These networks, however, are fragile to the specific removal of hubs [50,59,61].

Network dynamics

Decoding signal propagation and processing in molecular networks requires consideration of the temporal aspects of signal processing [13]. Network-based models have limited capabilities to capture temporal dynamics of the system,

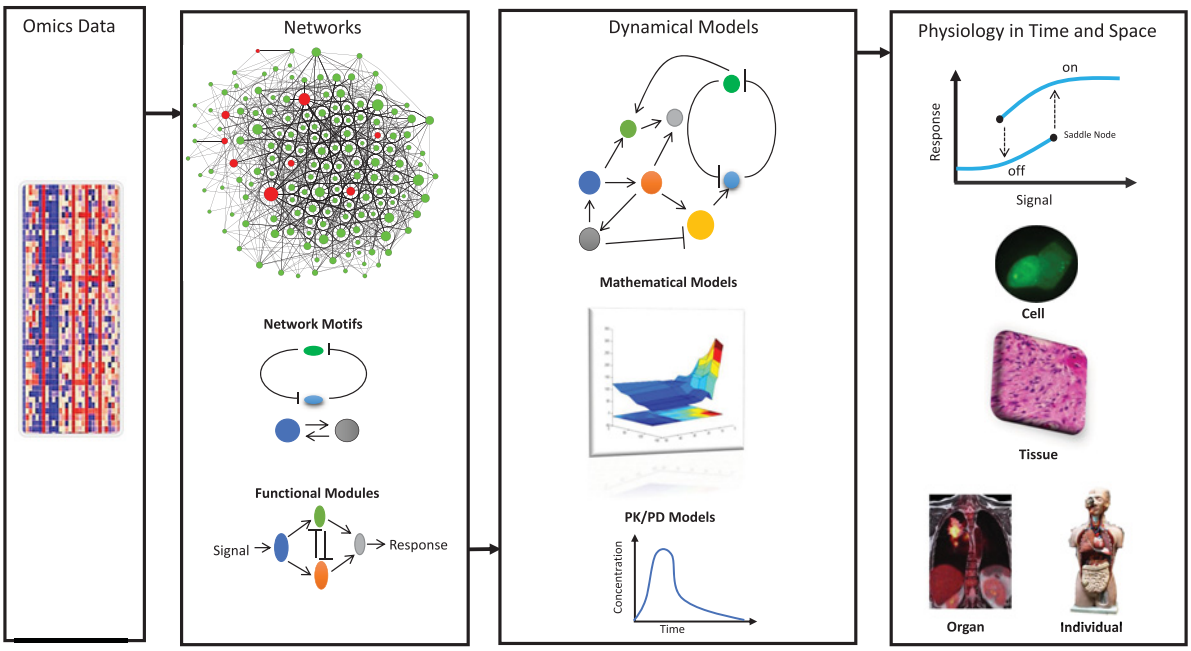

Figure 2. Computational methods in systems biology

Omics data are organized and analyzed using bioinformatics tools, and the resulting datasets are used to build networks. Within these networks of molecular interactions, the topological features of cellular wiring systems can be deduced. Network motifs and functional modules that are smaller sets of nodes and edges are commonly found in these networks and represent certain dynamical signal processing properties and carry out specific functional tasks. From these networks, sets of regulatory pathways (which include motifs, modules and feedback loops) are extracted to build dynamical models. These dynamical models are used for simulations to understand and predict the emergent behavior of the system in space and time. They also can be merged with PK/PD models to study drug action.

but temporal dynamics is essential for understanding systems behaviors at the cell and tissue level. Hence, network analysis needs to be combined with dynamic quantitative mathematical models. Dynamical models present a more accurate description of how a system progresses temporally and spatially. In fact, the networks within cells not only underlie structural and organizational aspects of cellular components but also can show emergent temporal properties defining cellular functions. Each network includes some functional modules that have a limited number of components that interact to receive a signal, to process it and then to transduce the signal to other modules. A network has a dynamic that can be studied by translating its components (nodes and edges) into a set of ordinary differential equations (ODEs) [29,30] (Figure 2). A simple way of translating a network to a mathematical formalism to study its dynamics is by using Boolean logic, which assigns a state of being 'on' (1) or 'off' (0) to each node. In addition to Boolean logic models and differential equation-based models, there are hybrid models, which use a combination of Boolean functions and differential equations, and fuzzy logic-based models that, in contrast with Boolean logic, represent nodal activity values between 0 and 1 [29,44-47].

Network motifs and functional modules

A motif is a set of a limited number of components that represent a certain dynamical behavior such as bistability or oscillations [49,62]. For instance, one of the motifs common in signaling pathways is mutual inhibition between two proteins. The structure of such a motif shows the emergent ability of this system in signal processing to make an on/off memory switch [10,62,63]. Network motifs, such as feedback and feedforward loops and bifan motifs, are recurrent and commonly found as subgraphs in biological networks [62,64,65]. A functional module consists of one or more such motifs with a particular function, such as signal integration or switch control in cells [49,62]. By extracting network motifs in molecular networks and surveying the dynamical behavior of functional modules, it is possible to decode temporal characteristics of signal transduction in cell systems. Methods from dynamical systems are used to

Essays in Biochemistry (2018) EBC20180003
https://doi.org/10.1042/EBC20180003

analyze ODE-based mathematical models of motifs, modules and networks [11]. The dynamical systems methods also provide a roadmap to design experiments for verifying the dynamical models, as the rate of molecular events in cells follow rules of dynamical systems [66] (Figure 2).

Dynamical models

Dynamical models are built by converting a network of interactions such as a gene regulatory network or a protein–protein interaction network to ODEs. Solving and analyzing these ODEs show the qualitative and time-course changes in the network as a dynamical system [67]. Often dynamical models do not have a unique solution, as defined by a single set of parameters. Such dynamical models are considered robust with more than one set of parameters and have a spectrum of parameter sensitivity [49,68]. Exploring these parameter spaces provides new information on the biological redundancies built into the system [49]. When the signal processing is done in different cellular compartments, compartmental dynamical models are built in which biochemical reactions within a compartment are represented as groups of ODEs [69]. These dynamical models can provide information regarding the state of the system. Changes between system states can provide knowledge about different types of activities a system is capable of. Such states can be at the cellular or tissue/organ or organismal levels. Bifurcation theory is a tool used to study states of dynamical systems that undergo qualitative or topological changes. For example, dynamical patterns such as bistability (switching between two stable states) and oscillations can be studied using saddle–node bifurcation or Hopf bifurcation [16,70]. Bifurcation analysis is a mathematically and computationally challenging task when the systems of ODEs become complex and is often used to study functional modules such as feedback loops [62].

When modeling cellular processes in time and space to understand the spatial organization of time-dependent cellular functions, partial differential equations (PDEs) are used [71,72]. PDEs can compute transitions in concentration and change in location of reactants and products. Solving PDEs is more challenging than ODEs because adding spatial parameters increases the complexity of the equations, and in PDEs one deals with multivariable functions in contrast with ODEs where the functions of a single variable are considered [73].

Solving ODEs in dynamical modeling can be done analytically or, more commonly, numerically. An analytical solution is expressed as a mathematical formalism that can readily be used to simulate time-courses of different components. Numerical solutions are based on obtaining numerical approximations for ODEs of the systems being studied. ODEs representing cellular and biological systems are usually very complex and cannot be solved analytically. They are most often solved numerically using different software packages and tools such as MATLAB, COPASI, Virtual Cell etc. [74-76]. PDE are typically solved numerically. Both MATLAB and Virtual Cell have PDE solvers.

If a system's temporal evolution is fully determined by specific initial conditions and reaction rates, then it can be modeled by a deterministic ODE or PDE model. However, many important cellular processes, such as gene expression, are stochastic, and modeling them requires stochastic modeling in contrast with deterministic ODE or PDE models. Heterogeneity is a main characteristic at all levels of biological systems. One way to include the heterogeneity of these systems in terms of probability distributions of intrinsic and extrinsic noise is stochastic modeling [77]. A stochastic dynamical model describes systems or functions in which the temporal evolution of the system is computed both by specific predictable reactions and some random variables and parameters. A common methodology for stochastic systems is the Gillespie algorithm [78]. In stochastic models, a master equation is implemented to control the evolution of the system such that a probabilistic function defines the next state of the system. The master equation basically defines the probabilistic distribution of all possible states that the system can have over time. The Gillespie algorithm makes it possible to simulate each bimolecular reaction while time or space intervals between reactions adhere to a probability distribution defined by the master equation [77-79].

Another aspect of dynamical models in systems biology is linking a dynamical model built for a single cell to the behavior of a population of cells, such as within a tumor [80]. In these cases, each cell can have a distinct parameter space with some parameters following probabilistic distributions. In such cases each cell may be simulated separately, and the behavior of the population computed from the average behavior of individual cells. [49,78,80,81]. Depending on the biological questions one wants to answer, the type of mathematical model chosen is deterministic or stochastic. The parameters in dynamical systems of cellular processes and signaling pathways need to be measured directly from experiments or estimated based on experiments. Although there are toy dynamical models that are built using arbitrary parameters, which are helpful to gain mechanistic insights into the system, the most common dynamical models in systems biology are plausible models in which parameters are measured or estimated by experiments. Identifiable dynamical models are made to explain the experimental data, and variables and parameters are specific to a certain system and fitted to experimental data from that system. These models are very common in quantitative systems pharmacology (QSP) and studies that involve drug actions [20,63,67,82]. In all dynamical models of cell

systems, thermodynamic constraints must be fulfilled [13,83]. These dynamical models allow one to study and predict physiological responses in space and time (Figure 2).

Pharmacokinetic/pharmacodynamic (PK/PD) models are commonly used in the study of drug action. Pharmacokinetic models are focused on drug disposition and availability whereas pharmacodynamics focuses on mechanisms of drug action. Combining PK/PD models with dynamical models of cellular regulatory systems can be used for predicting both therapeutic and adverse effects of drugs [20].

Strengths and limitations of different types of models

The different modeling approaches in systems biology have their own applications and limitations. They are chosen based on the system under study and the complexity of the problem being addressed, and the use of multiple models may be necessary to predict system behavior. When using high-throughput and quantitative experimental approaches, model types used include statistical models, networks and dynamical models. Statistical models, which are the first layer in top-down systems biology, deal with defining molecular datasets assigned to given phenotypes and functions. These models can deal with probabilistic relationships built upon correlations. This makes statistical models useful for clinical decision making because for most complex diseases, pathological phenotypes are associated with molecular markers like genes in a probabilistic manner. These models, however, do not enable the understanding of mechanisms underlying the development of phenotypes because they do not consider the nature and direction of interactions among components [84,85]. They cannot decode information flow from pathways or the dynamics of networks within the cells. Statistical models have a static view of biological functions, and systems evolving in time and space are not fully described. For example, statistical models are inadequate to describe the time-course of initiation of a disease phenotype or acquisition of treatment resistance. Mechanistic models are required to describe an integrative view of the pathological process [20,86]. Network-based models serve as representations of whole-cell interactions and their topologies. These topologies are a vital first step to understand the dynamics of cell systems in a flexible multiscale fashion. They represent all cellular components and their relationships as a global map for information transmission in cells, tissues and organs. Inside these networks, it is possible to search for functional modules by identifying hubs and network motifs. To truly understand computation within cells, we require both network models and dynamical models. However, lack of sufficient kinetic data often prevents us from building dynamical models at the level of large networks. We usually need to select the most important components, including functional modules and computational units, to make insightful and realistic dynamical models [16,51,85,87]. In addition, the assumptions and estimated parameters needed for the construction of dynamical models require that predictions from model simulations be experimentally verified.

Quantitative experimental methods for systems biology

Quantitative methods encompass a wide range of experiments that measure the quantity of cellular components such as protein concentrations and their temporal changes in different time scales. These include standard molecular biology and cell biology experiments as well as high-throughput experiments. These experiments can be based on a single cell or cell populations. Single-cell experiments are helpful to verify and explore parameter spaces of models designed at the cell level. Specifically, when a cell population is heterogeneous (for example cancer cells), each cell may have a different parameter space and responses to signals [22,23,81]. Over the past few years single-cell transcriptomics [88-90] has been developed to provide mRNA profiles in single cells. This approach has been very useful in mapping subtypes of classes of cells within tissues and organs. Conventional molecular biology experimental methods, such as Western blots for measuring protein concentrations, provide an average result from many cells in an often heterogeneous cell population [49]. Although both single cell and population experiments can be used, the ergodic nature of cellular events favors measuring single cell dynamics from a cell population [90].

Often it is not possible to measure the concentrations and kinetic parameters of all components of a system. Thus, some component parameters used in models are estimated based on data from other components. Finding kinetic parameters is often difficult due to limitations imposed by experimental design. Quantitative measurements, such as time-course experiments, involve many components with different kinetic parameters making it difficult to explicitly measure the kinetic parameters associated with individual molecular components. Quantitative characterization of molecular components, both with respect to kinetic parameters and concentrations within different cell types is an underdeveloped area of study.

One type of experiment helpful for building precise networks and models is using omics technologies at the single-cell level. Conventional omics methods provide a list of entities from a heterogeneous cell population. However, in single-cell transcriptomics, the mRNA concentrations of expressed genes are measured in each cell in a population

Essays in Biochemistry (2018) EBC20180003
https://doi.org/10.1042/EBC20180003

of cells. Although the number of genes identified by this method is ∼1000 per cell, the method of measurement, using 3′ unique molecular identifiers, counts each molecule of RNA and hence provides quantitative estimates of the different RNA species in each cell. These single-cell omics data are useful in describing the heterogeneity of cells in tissues and organs. Heterogeneity is an important consideration for building predictive models for complex tissues and diseases because the phenotypes are dependent on cells with different identities [91-93]. An example is the systems biology of cancer, in which both statistical and dynamical models are built to design therapeutic regimens for tumors and cell lines that contain many individual cells with heterogeneous expressions of genes and proteins [88]. Molecular information from single cells can be used to build models of cell populations by considering single cells with different identities as components, with each cell considered as a system of the biochemical and molecular network. Such an approach captures the diversity of cell subtypes in a tissue or organ system.

Artificial intelligence in systems biology

One of the main challenges in systems biology is to convert big data at different scales into actionable knowledge. This knowledge is vital to improve methodologies to study biological systems, to understand and diagnose diseases at various stages precisely and to design new therapeutic modalities focused on the individual. Mechanistic models, such as dynamical models that depend on the causality of relationships among components, can combine biological data from hypothesis-based experiments with mathematical modeling to produce predictive models. Often, such models also provide insight into mechanisms. The amount of information in biology and medicine is rapidly surpassing the current capability of building large-scale mechanistic models. An alternative way to generate predictive models from big data is through statistical models based on correlation. This process can benefit from artificial intelligence (AI) that uses statistical reasoning to detect unseen correlation, co-occurrence and dependencies in large-scale datasets [94].

In computer science, AI is a way of developing machine-based expert systems that can analyze data and predict new outcomes. Machine learning, deep learning and artificial neural networks are different approaches used in AI. Artificial neural networks were inspired by real brain neural networks and are capable of learning specific task-oriented classifications when trained by a training set [95]. Machine learning refers to a group of methods that analyze big datasets and, based on them, make predictions. Machine learning can be used in a supervised, unsupervised or semi-supervised manner. In supervised machine learning, training datasets in the form of labeled input/output relations are provided and a function is inferred that can be used to analyze new examples—to predict the output based on input data and classification. Unsupervised machine learning is when the data are not labeled, and the aim is to detect underlying patterns with no guide. Semi-supervised learning is a modality between supervised and unsupervised learning when there is limited labeled data [96-98]. Deep learning is a machine learning method that uses multilayer computational processing units for data representation and detection of intrinsic patterns in big data [99,100].

AI is a powerful tool for developing models and optimizing them. In AI, an algorithm and a dataset are used. The dataset usually has two properties, one is a large set of measurements (e.g. molecular signatures such as genes, proteins or metabolites) and the other is the resultant prediction (e.g. the resulting phenotype). The underlying algorithm, usually a statistical model, is trained using the dataset and then a test set is used to evaluate the predictions (Figure 3) [96,97,101]. This training and testing procedure can be optimized as a loop by using new data as training datasets. The algorithm adjusts itself to make better predictions, and thus the AI system learns as it is being used. Network building and analyses can benefit from AI methods to extract data from omics data in terms of finding interactions and relationships among molecular entities in given phenotypes, finding network motifs and functional modules, and decoding main pathways involving selected functions (Figure 3) [54,102]. Exploring the parameter spaces of dynamical models and sensitivity analysis also use machine learning approaches because they deal with big numerical datasets [103]. The predictive models arising from AI and machine learning are potentially powerful tools in precision medicine as they can extract genomic signatures related to drug treatment and therapeutic responses. In such cases, both large clinical and biomolecular datasets are used as training sets that are assigned to specific responses. The predictions of these models are tested and valid predictions are used as new data for making the training set bigger [101]. AI, for example, has been successfully used in predicting cancer outcomes based on molecular biomarkers and pathology [104]. The use of artificial intelligence in studies of basic biological systems, as well as in clinical data, are schematically shown in Figure 3.

Systems pharmacology and systems biomedicine

Systems-level insights serve as building blocks to advance medicine to a higher level of personalized and precision care. Currently, systems biology approaches are being implemented in both drug discovery and the practice of

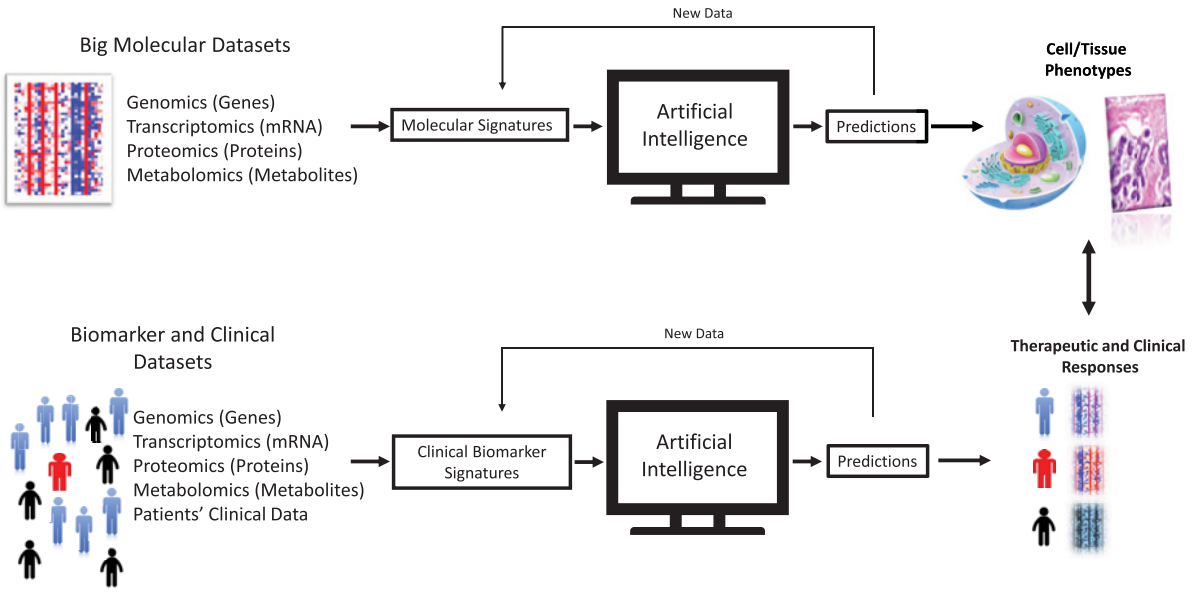

Figure 3. AI in systems biology and precision medicine

AI and machine learning approaches are helpful tools for finding molecular signatures from high-throughput measurement datasets and assigning them to higher level phenotypes and cellular functions. They also can be used to tailor treatment based on an individual patient's molecular markers for precision medicine. Clinical data and omics data from patients are used to extract clinical biomarkers for training sets and to build predictive models of the course of the disease and patient responses to treatment. Verified predictions are used as additional training sets to make the predictive models progressively more accurate.

medicine. Systems biology has enabled pharmacology to become a systems science. Systems pharmacology, of which pharmacogenomics is a part, has been shown to be useful for drug discovery and predicting therapeutic responses and drug adverse effects. While pharmacogenomics uses genomic data of drug metabolizing enzymes for prediction of drug responses and effects, QSP utilizes network and dynamical models integrated with pharmacodynamics and pharmacokinetics to find optimized therapeutics for specific patients with a given disease [20,85,105,106]. These advances enable adjustment of drug regimens and drug doses for individual patients based on their molecular markers [85,107].

With advances in biosensors that are able to collect time-course data from patients, liquid biopsies and biomarker discovery, the practice of precision medicine based on systems biology approaches seems feasible. These data collection tools provide the basic materials for predictive models using systems biology approaches. Biosensors can collect real-time data on the concentration of different components in the blood of a patient or record quantitative data on physiological signals such as heart rate and electrical activities of brain and heart [108,109]. Liquid biopsy collects cancer cells or other tissue components in fluids such as blood, saliva and urine that can be used for omics data and biomarker detection [110]. Advances in high-content image analysis that requires quantitative analysis of vast numbers of images such as pathology slides can, with the help of machine learning, provide an accurate diagnostic tool for predictive models of disease states and progression [111,112].

Genomic signatures in systems therapeutics

One of the recent advances in the field of precision medicine is the discovery of genomic signatures related to pathogenesis and therapeutic responses in different diseases [107,113-115]. Drug treatments change the gene expression profiles of cells, and measuring these changes before and after treatment *in vitro*, *in vivo* in animal models, and in patients, can bring new insights about genomic determinants of drug responses and drug adverse effects. Genomic signatures also can be used as prognostic markers for patients suffering from chronic diseases such as cancer and help in selecting individuals for specific treatment plans. For example, in the case of cancer immunotherapy, there has been

Essays in Biochemistry (2018) EBC20180003
https://doi.org/10.1042/EBC20180003

a major effort to detect the genomic determinates of responses to immunotherapy agents such as PD-L1 inhibitors. Currently, the main practice is based on the expression of PD-L1 protein in tumor tissue, but several genomic and clinical markers, both tumor genomic profiles and patients' immune system characteristics, have been found useful in guiding immunotherapy [116].

Perspective

By investigating qualitative as well as quantitative properties, both temporal and spatial, and emerging functions of molecular interactions in biological systems, we are able to understand many phenomena in cells, tissues/organs, and at the level of whole organisms. The transmission of information from genes to organismal behaviors, and complex phenotypes arising from molecular and cellular networks, can be explored using systems biology methodologies. Statistical, network and dynamical models are essential tools in systems biology leading to discoveries at various scales of biological organization. These discoveries are basic building blocks for future advances in medicine, leading to precision and individualization of treatment. Advances in computational and experimental methods, including faster and more accurate technologies, will enable systems biology to provide basic understanding of cells, tissues and organs, as well as future medical advances.

Summary

- Systems biology studies cells, tissues and organ systems as systems of interacting components.
- Omics technologies are the main sources of information on individual molecular entities in cells.
- Bioinformatics organizes the big data obtained from systems-wide surveys.
- Statistical methods enable analyses of big datasets based on high-throughput technologies that can then be used to decode pathways and molecular networks.
- Dynamical models of networks of cellular components help to explain the emergent properties of cell and tissue physiology in time and space.
- Systems biology methods can be used for drug discovery and development of systems pharmacology approaches.
- Artificial intelligence and machine learning approaches are used as tools in systems biology to link molecular datasets to phenotypes and physiological behaviors at the organismal level.
- Insights from systems biology studies are useful for the design of precision and individualized medicine protocols.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

Work leading to this review was funded by Systems Biology Center grant [GM-071558] from the National Institute of General Medical Sciences (NIGMS).

Abbreviations

AI, artificial intelligence; GO, Gene Ontology; GSEA, Gene Set Enrichment Analysis; GWAS, Genome-Wide Association Studies; KEGG, Kyoto Encyclopedia of Genes and Genomes; MBCO, Molecular Biology of the Cell Ontology; ODE, ordinary differential equation; PDE, partial differential equation; QSP, quantitative systems pharmacology; SNP, single nucleotide polymorphisms.

References

1 Kitano, H. (2002) Systems biology: a brief overview. *Science* **295**, 1662–1664, https://doi.org/10.1126/science.1069492
2 Hood, L., Heath, J.R., Phelps, M.E. and Lin, B. (2004) Systems biology and new technologies enable predictive and preventative medicine. *Science* **306**, 640–643, https://doi.org/10.1126/science.1104635
3 Wist, A.D., Berger, S.I. and Iyengar, R. (2009) Systems pharmacology and genome medicine: a future perspective. *Genome Med.* **1**, 11, https://doi.org/10.1186/gm11
4 Weng, G., Bhalla, U.S. and Iyengar, R. (1999) Complexity in biological signaling systems. *Science* **284**, 92–96, https://doi.org/10.1126/science.284.5411.92
5 Ideker, T., Galitski, T. and Hood, L. (2001) A new approach to decoding life: systems biology. *Annu. Rev. Genomics Hum. Genet.* **2**, 343–372, https://doi.org/10.1146/annurev.genom.2.1.343
6 Kirschner, M.W. (2005) The meaning of systems biology. *Cell* **121**, 503–504, https://doi.org/10.1016/j.cell.2005.05.005
7 Jensen, H.J. (1998) Self-Organized Criticality: Emergent Complex Behavior in Physical and Biological Systems. Cambridge University Press
8 Ideker, T., Thorsson, V., Ranish, J.A., Christmas, R., Buhler, J., Eng, J.K. et al. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929–934, https://doi.org/10.1126/science.292.5518.929
9 Consortium IHGS (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860, https://doi.org/10.1038/35057062
10 Bhalla, U.S. and Iyengar, R. (1999) Emergent properties of networks of biological signaling pathways. *Science* **283**, 381–387, https://doi.org/10.1126/science.283.5400.381
11 Tyson, J.J., Chen, K.C. and Novak, B. (2003) Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol.* **15**, 221–231, https://doi.org/10.1016/S0955-0674(03)00017-6
12 Alon, U., Surette, M.G., Barkai, N. and Leibler, S. (1999) Robustness in bacterial chemotaxis. *Nature* **397**, 168, https://doi.org/10.1038/16483
13 Tyson, J.J., Chen, K. and Novak, B. (2001) Network dynamics and cell physiology. *Nat. Rev. Mol. Cell Biol.* **2**, 908, https://doi.org/10.1038/35103078
14 Kholodenko, B.N. (2006) Cell-signalling dynamics in time and space. *Nat. Rev. Mol. Cell Biol.* **7**, 165, https://doi.org/10.1038/nrm1838
15 Clarke, R., Shajahan, A.N., Wang, Y., Tyson, J.J., Riggins, R.B., Weiner, L.M. et al. (2011) Endoplasmic reticulum stress, the unfolded protein response, and gene network modeling in antiestrogen resistant breast cancer. *Hormone Mol. Biol. Clin. Invest.* **5**, 35–44, https://doi.org/10.1515/HMBCI.2010.073
16 Tyson, J.J., Baumann, W.T., Chen, C., Verdugo, A., Tavassoly, I., Wang, Y. et al. (2011) Dynamic modelling of oestrogen signalling and cell fate in breast cancer cells. *Nat. Rev. Cancer* **11**, 523–532, Epub 2011/06/16. eng, https://doi.org/10.1038/nrc3081
17 Clarke, R., Cook, K.L., Hu, R., Facey, C.O., Tavassoly, I., Schwartz, J.L. et al. (2012) Endoplasmic reticulum stress, the unfolded protein response, autophagy, and the integrated regulation of breast cancer cell fate. *Cancer Res.* **72**, 1321–1331
18 Kreeger, P.K. and Lauffenburger, D.A. (2009) Cancer systems biology: a network modeling perspective. *Carcinogenesis* **31**, 2–8, https://doi.org/10.1093/carcin/bgp261
19 Anderson, A.R. and Chaplain, M. (1998) Continuous and discrete mathematical models of tumor-induced angiogenesis. *Bull. Math. Biol.* **60**, 857–899, https://doi.org/10.1006/bulm.1998.0042
20 Iyengar, R., Zhao, S., Chung, S.-W., Mager, D.E. and Gallo, J.M. (2012) Merging systems biology with pharmacodynamics. *Sci. Transl. Med.* **4**, P126ps7, https://doi.org/10.1126/scitranslmed.3003563
21 Bruggeman, F.J. and Westerhoff, H.V. (2007) The nature of systems biology. *Trends Microbiol.* **15**, 45–50, https://doi.org/10.1016/j.tim.2006.11.003
22 Batchelor, E., Loewer, A. and Lahav, G. (2009) The ups and downs of p53: understanding protein dynamics in single cells. *Nat. Rev. Cancer* **9**, 371, https://doi.org/10.1038/nrc2604
23 Purvis, J.E. and Lahav, G. (2013) Encoding and decoding cellular information through signaling dynamics. *Cell* **152**, 945–956, https://doi.org/10.1016/j.cell.2013.02.005
24 Noble, D. (2008) *The Music of Life: Biology Beyond Genes*, Oxford University Press
25 Walker, D.C. and Southgate, J. (2009) The virtual cell—a candidate co-ordinator for 'middle-out'modelling of biological systems. *Brief. Bioinform.* **10**, 450–461, https://doi.org/10.1093/bib/bbp010
26 Morozova, O. and Marra, M.A. (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics* **92**, 255–264, https://doi.org/10.1016/j.ygeno.2008.07.001
27 Wang, Z., Gerstein, M. and Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57, https://doi.org/10.1038/nrg2484
28 Epigenomics, E.V. (2006) Mapping the methylome. *Cell Cycle* **5**, 155–158, https://doi.org/10.4161/cc.5.2.2367
29 Jones, P.A. and Baylin, S.B. (2007) The epigenomics of cancer. *Cell* **128**, 683–692, https://doi.org/10.1016/j.cell.2007.01.029
30 Rual, J.-F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N. et al. (2005) Towards a proteome-scale map of the human protein–protein interaction network. *Nature* **437**, 1173, https://doi.org/10.1038/nature04209
31 German, J.B., Hammock, B.D. and Watkins, S.M. (2005) Metabolomics: building on a century of biochemistry to guide human health. *Metabolomics* **1**, 3–9, https://doi.org/10.1007/s11306-005-1102-8
32 Bush, W.S. and Moore, J.H. (2012) Genome-wide association studies. *PLoS Comput. Biol.* **8**, e1002822, https://doi.org/10.1371/journal.pcbi.1002822
33 Hirschhorn, J.N. and Daly, M.J. (2005) Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.* **6**, 95, https://doi.org/10.1038/nrg1521
34 Cloonan, N., Forrest, A.R., Kolle, G., Gardiner, B.B., Faulkner, G.J., Brown, M.K. et al. (2008) Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat. Methods* **5**, 613, https://doi.org/10.1038/nmeth.1223

35 Aebersold, R. and Mann, M. (2016) Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347, https://doi.org/10.1038/nature19949
36 Doll, S., Dreßen, M., Geyer, P.E., Itzhak, D.N., Braun, C., Doppler, S.A. et al. (2017) Region and cell-type resolved quantitative proteomic map of the human heart. *Nat. Commun.* **8**, 1469, https://doi.org/10.1038/s41467-017-01747-2
37 Higashi, R.M., Fan, T.W-.M., Lorkiewicz, P.K., Moseley, H.N.B and Lane, A.N. (2014) Stable isotope-labeled tracers for metabolic pathway elucidation. *Mass Spectrometry in Metabolomics. Methods in Molecular Biology (Methods and Protocols)*, vol. **1198**, Humana Press, New York, https://doi.org/10.1007/978-1-4939-1258-2˙11
38 Garber, M., Grabherr, M.G., Guttman, M. and Trapnell, C. (2011) Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat. Methods* **8**, 469, https://doi.org/10.1038/nmeth.1613
39 Tyanova, S., Temu, T. and Cox, J. (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301, https://doi.org/10.1038/nprot.2016.136
40 Kirpich, A.S., Ibarra, M., Moskalenko, O., Fear, J.M., Gerken, J., Mi, X. et al. (2018) SECIMTools: a suite of metabolomics data analysis tools. *BMC Bioinformatics* **19**, 151, https://doi.org/10.1186/s12859-018-2134-1
41 Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M. et al. (2000) Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25, https://doi.org/10.1038/75556
42 Hansen, J., Meretzky, D., Woldesenbet, S., Stolovitzky, G. and Iyengar, R. (2017) A flexible ontology for inference of emergent whole cell function from relationships between subcellular processes. *Sci. Rep.* **7**, 17689, https://doi.org/10.1038/s41598-017-16627-4
43 Kanehisa, M. and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30, https://doi.org/10.1093/nar/28.1.27
44 Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A. et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550, https://doi.org/10.1073/pnas.0506580102
45 Gundersen, G.W., Jones, M.R., Rouillard, A.D., Kou, Y., Monteiro, C.D., Feldmann, A.S. et al. (2015) GEO2Enrichr: browser extension and server app to extract gene sets from GEO and analyze them for biological functions. *Bioinformatics* **31**, 3060–3062, https://doi.org/10.1093/bioinformatics/btv297
46 Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z. et al. (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97, https://doi.org/10.1093/nar/gkw377
47 Schubert, M., Klinger, B., Klünemann, M., Sieber, A., Uhlitz, F., Sauer, S. et al. (2018) Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat. Commun.* **9**, 20, https://doi.org/10.1038/s41467-017-02391-6
48 Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P. and Mesirov, J.P. (2011) Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740, https://doi.org/10.1093/bioinformatics/btr260
49 Tavassoly, I., Parmar, J., Shajahan-Haq, A.N., Clarke, R., Baumann, W.T. and Tyson, J.J. (2015) Dynamic modeling of the interaction between autophagy and apoptosis in mammalian cells. *CPT Pharmacometrics Syst. Pharmacol.* **4**, 263–272, Epub 2015/04/17.eng, https://doi.org/10.1002/psp4.29
50 Barabasi, A.-L. and Oltvai, Z.N. (2004) Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101, https://doi.org/10.1038/nrg1272
51 Berger, S.I. and Iyengar, R. (2009) Network analyses in systems pharmacology. *Bioinformatics* **25**, 2466–2472, https://doi.org/10.1093/bioinformatics/btp465
52 Ideker, T. and Krogan, N.J. (2012) Differential network biology. *Mol. Syst. Biol.* **8**, 565, https://doi.org/10.1038/msb.2011.99
53 Zhao, S. and Iyengar, R. (2012) Systems pharmacology: network analysis to identify multiscale mechanisms of drug action. *Annu. Rev. Pharmacol. Toxicol.* **52**, 505–521, https://doi.org/10.1146/annurev-pharmtox-010611-134520
54 Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D.A. and Nolan, G.P. (2005) Causal protein-signaling networks derived from multiparameter single-cell data. *Science* **308**, 523–529, https://doi.org/10.1126/science.1105809
55 Needham, C.J., Bradford, J.R., Bulpitt, A.J. and Westhead, D.R. (2007) A primer on learning in Bayesian networks for computational biology. *PLoS Comput. Biol.* **3**, e129, https://doi.org/10.1371/journal.pcbi.0030129
56 Sachs, K., Gifford, D., Jaakkola, T., Sorger, P. and Lauffenburger, D.A. (2002) Bayesian network approach to cell signaling pathway modeling. *Sci STKE* **2002**, pe38–pe
57 Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D. et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504, https://doi.org/10.1101/gr.1239303
58 Han, J.-D.J., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., Zhang, L.V. et al. (2004) Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature* **430**, 88, https://doi.org/10.1038/nature02555
59 Albert, R. (2005) Scale-free networks in cell biology. *J. Cell Sci.* **118**, 4947–4957, https://doi.org/10.1242/jcs.02714
60 Hornung, G. and Barkai, N. (2008) Noise propagation and signaling sensitivity in biological networks: a role for positive feedback. *PLoS Comput. Biol.* **4**, e8, https://doi.org/10.1371/journal.pcbi.0040008
61 Barabási, A.-L. (2009) Scale-free networks: a decade and beyond. *Science* **325**, 412–413, https://doi.org/10.1126/science.1173299
62 Tyson, J.J. and Novák, B. (2010) Functional motifs in biochemical reaction networks. *Annu. Rev. Phys. Chem.* **61**, 219–240, https://doi.org/10.1146/annurev.physchem.012809.103457
63 Azeloglu, E.U. and Iyengar, R. (2015) Signaling networks: information flow, computation, and decision making. *Cold Spring Harbor Perspect. Biol.* **7**, a005934, https://doi.org/10.1101/cshperspect.a005934
64 Alon, U. (2007) Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* **8**, 450, https://doi.org/10.1038/nrg2102

65 Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. and Alon, U. (2002) Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827, https://doi.org/10.1126/science.298.5594.824
66 Kafri, R., Levy, J., Ginzberg, M.B., Oh, S., Lahav, G. and Kirschner, M.W. (2013) Dynamics extracted from fixed cells reveal feedback linking cell growth to cell cycle. *Nature* **494**, 480, https://doi.org/10.1038/nature11897
67 Neves, S.R. and Iyengar, R. (2002) Modeling of signaling networks. *Bioessays* **24**, 1110–1117, https://doi.org/10.1002/bies.1154
68 Gutenkunst, R.N., Waterfall, J.J., Casey, F.P., Brown, K.S., Myers, C.R. and Sethna, J.P. (2007) Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput. Biol.* **3**, e189, https://doi.org/10.1371/journal.pcbi.0030189
69 Yadaw, A.S., Siddiq, M.M., Rabinovich, V., Tolentino, R., Iyengar, R. and Hansen, J. (2018) Dynamic balance between vesicle transport and microtubule growth enables neurite growth. *bioRxiv* 153569, https://doi.org/10.1101/153569
70 Novák, B. and Tyson, J.J. (2008) Design principles of biochemical oscillators. *Nat. Rev. Mol. Cell Biol.* **9**, 981, https://doi.org/10.1038/nrm2530
71 Neves, S.R. and Iyengar, R. (2009) Models of spatially restricted biochemical reaction systems. *J. Biol. Chem.* **284**, 5445–5449, https://doi.org/10.1074/jbc.R800058200
72 Rangamani, P., Lipshtat, A., Azeloglu, E.U., Calizo, R.C., Hu, M., Ghassemi, S. et al. (2013) Decoding information in cell shape. *Cell* **154**, 1356–1369, https://doi.org/10.1016/j.cell.2013.08.026
73 Leung, A.W. (2013) *Systems of Nonlinear Partial Differential Equations: Applications to Biology and Engineering*, Springer Science & Business Media
74 Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N. et al. (2006) COPASI—a complex pathway simulator. *Bioinformatics* **22**, 3067–3074, https://doi.org/10.1093/bioinformatics/btl485
75 Schmidt, H. and Jirstrand, M. (2005) Systems Biology Toolbox for MATLAB: a computational platform for research in systems biology. *Bioinformatics* **22**, 514–515, https://doi.org/10.1093/bioinformatics/bti799
76 Loew, L.M. and Schaff, J.C. (2001) The Virtual Cell: a software environment for computational cell biology. *Trends Biotechnol.* **19**, 401–406, https://doi.org/10.1016/S0167-7799(01)01740-1
77 Szekely, T. and Burrage, K. (2014) Stochastic simulation in systems biology. *Comput. Struct. Biotechnol. J.* **12**, 14–25, https://doi.org/10.1016/j.csbj.2014.10.003
78 Gillespie, D.T. (2007) Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.* **58**, 35–55, https://doi.org/10.1146/annurev.physchem.58.032806.104637
79 Wilkinson, D.J. (2009) Stochastic modelling for quantitative description of heterogeneous biological systems. *Nat. Rev. Genet.* **10**, 122, https://doi.org/10.1038/nrg2509
80 Tavassoly, I. (2015) *Dynamics of Cell Fate Decision Mediated by the Interplay of Autophagy and Apoptosis in Cancer Cells: Mathematical Modeling and Experimental Observations*, Springer
81 Loewer, A. and Lahav, G. (2011) We are all individuals: causes and consequences of non-genetic heterogeneity in mammalian cells. *Current Opin. Genet. Dev.* **21**, 753–758, https://doi.org/10.1016/j.gde.2011.09.010
82 Azeloglu, E.U. and Iyengar, R. (2015) Good practices for building dynamical models in systems biology. *Sci. Signal* **8**, fs8-fs, https://doi.org/10.1126/scisignal.aab0880
83 Morohashi, M., Winn, A.E., Borisuk, M.T., Bolouri, H., Doyle, J. and Kitano, H. (2002) Robustness as a measure of plausibility in models of biochemical networks. *J. Theor. Biol.* **216**, 19–30, https://doi.org/10.1006/jtbi.2002.2537
84 Boran, A.D. and Iyengar, R. (2010) Systems approaches to polypharmacology and drug discovery. *Current Opin. Drug Discov. Dev.* **13**, 297
85 Hansen, J., Zhao, S. and Iyengar, R. (2011) Systems pharmacology of complex diseases. *Ann. N. Y. Acad. Sci.* 1245
86 Iyengar, R., Altman, R.B., Troyanskya, O. and FitzGerald, G.A. (2015) Personalization in practice. *Science* **350**, 282–283, https://doi.org/10.1126/science.aad5204
87 Eungdamrong, N.J. and Iyengar, R. (2004) Computational approaches for modeling regulatory cellular networks. *Trends Cell Biol.* **14**, 661–669, https://doi.org/10.1016/j.tcb.2004.10.007
88 Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H. et al. (2014) Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401
89 Saliba, A.-E., Westermann, A.J., Gorski, S.A. and Vogel, J. (2014) Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Res.* **42**, 8845–8860, https://doi.org/10.1093/nar/gku555
90 Shalek, A.K., Satija, R., Shuga, J., Trombetta, J.J., Gennert, D., Lu, D. et al. (2014) Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **510**, 363, https://doi.org/10.1038/nature13437
91 Sandberg, R. (2014) Entering the era of single-cell transcriptomics in biology and medicine. *Nat. Methods* **11**, 22, https://doi.org/10.1038/nmeth.2764
92 Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaublomme, J.T., Raychowdhury, R. et al. (2013) Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498**, 236, https://doi.org/10.1038/nature12172
93 Wang, D. and Bodovitz, S. (2010) Single cell analysis: the new frontier in 'omics'. *Trends Biotechnol.* **28**, 281–290, https://doi.org/10.1016/j.tibtech.2010.03.002
94 Baker, R.E., Peña, J.-M., Jayamohan, J. and Jérusalem, A. (2018) Mechanistic models versus machine learning, a fight worth fighting for the biological community? *Biol. Lett.* **14**, 20170660, https://doi.org/10.1098/rsbl.2017.0660
95 Baxt, W.G. (1995) Application of artificial neural networks to clinical medicine. *Lancet North Am. Ed.* **346**, 1135–1138, https://doi.org/10.1016/S0140-6736(95)91804-3
96 Deo, R.C. (2015) Machine learning in medicine. *Circulation* **132**, 1920–1930, https://doi.org/10.1161/CIRCULATIONAHA.115.001593
97 Leung, M.K., Delong, A., Alipanahi, B. and Frey, B.J. (2016) Machine learning in genomic medicine: a review of computational problems and data sets. *Proc. IEEE* **104**, 176–197, https://doi.org/10.1109/JPROC.2015.2494198

98 Sommer, C. and Gerlich, D.W. (2013) Machine learning in cell biology–teaching computers to recognize phenotypes. *J. Cell Sci.*, jcs. 123604, https://doi.org/10.1242/jcs.123604

99 Janowczyk, A. and Madabhushi, A. (2016) Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *J. Pathol. Informatics* **7**, https://doi.org/10.4103/2153-3539.186902

100 LeCun, Y., Bengio, Y. and Hinton, G. (2015) Deep learning. *Nature* **521**, 436, https://doi.org/10.1038/nature14539

101 Camacho, D.M., Collins, K.M., Powers, R.K., Costello, J.C. and Collins, J.J. (2018) Next-generation machine learning for biological networks. *Cell* **173**, 1581–1592, https://doi.org/10.1016/j.cell.2018.05.015

102 Kell, D.B. (2006) Metabolomics, modelling and machine learning in systems biology–towards an understanding of the languages of cells. *FEBS J.* **273**, 873–894, https://doi.org/10.1111/j.1742-4658.2006.05136.x

103 Moles, C.G., Mendes, P. and Banga, J.R. (2003) Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res.* **13**, 2467–2474, https://doi.org/10.1101/gr.1262503

104 Catto, J.W., Linkens, D.A., Abbod, M.F., Chen, M., Burton, J.L., Feeley, K.M. et al. (2003) Artificial intelligence in predicting bladder cancer outcome: a comparison of neuro-fuzzy modeling and artificial neural networks. *Clin. Cancer Res.* **9**, 4172–4177

105 Berger, S.I., Ma'ayan, A. and Iyengar, R. (2010) Systems pharmacology of arrhythmias. *Sci. Signal.* **3**, ra30, https://doi.org/10.1126/scisignal.2000723

106 van der Graaf, P.H. and Benson, N. (2011) Systems pharmacology: bridging systems biology and pharmacokinetics-pharmacodynamics (PKPD) in drug discovery and development. *Pharm. Res.* **28**, 1460–1464, https://doi.org/10.1007/s11095-011-0467-9

107 Zhao, S., Nishimura, T., Chen, Y., Azeloglu, E.U., Gottesman, O., Giannarelli, C. et al. (2013) Systems pharmacology of adverse event mitigation by drug combinations. *Sci. Transl. Med.* **5**, 206ra140–206ra140, https://doi.org/10.1126/scitranslmed.3006548

108 Asada, H.H., Shaltis, P., Reisner, A., Rhee, S. and Hutchinson, R.C. (2003) Mobile monitoring with wearable photoplethysmographic biosensors. *IEEE Eng. Med. Biol. Mag.* **22**, 28–40, https://doi.org/10.1109/MEMB.2003.1213624

109 Turner, A.P. (2000) Biosensor–sense and sensitivity. *Science* **290**, 1315–1317, https://doi.org/10.1126/science.290.5495.1315

110 Crowley, E., Di Nicolantonio, F., Loupakis, F. and Bardelli, A. (2013) Liquid biopsy: monitoring cancer-genetics in the blood. *Nat. Rev. Clin. Oncol.* **10**, 472, https://doi.org/10.1038/nrclinonc.2013.110

111 Pantanowitz, L. (2010) Digital images and the future of digital pathology. *J. Pathol. Informatics* **1**, https://doi.org/10.4103/2153-3539.68332

112 Jelinek, H.F., Rocha, A., Carvalho, T., Goldenstein, S. and Wainer, J. (2011) Machine learning and pattern classification in identification of indigenous retinal pathology. Engineering in Medicine and Biology Society, EMBC, 2011. *Annual Int. Conf. IEEE*

113 van Hassselt, J.C. and Iyengar, R. (2017) Systems pharmacology-based identification of pharmacogenomic determinants of adverse drug reactions using human iPSC-derived cell lines. *Curr. Opin. Syst Biol.* **4**, 9–15, https://doi.org/10.1016/j.coisb.2017.05.006

114 Tavassoly, I. and Iyengar, R. (2018) Analysis of sensitivity of genomic signatures of therapeutic responses of non-small cell lung cancer in patient-derived xenograft models. *AACR*, https://doi.org/10.1158/1538-7445

115 Tavassoly, I., Hu, Y., Zhao, S., Mariottini, C., Boran, A., Chen, Y. et al. (2018) Systems therapeutics analyses identify genomic signatures defining responsiveness to allopurinol and combination therapy for lung cancer. *bioRxiv.*, https://doi.org/10.1101/396697

116 Voong, K.R., Feliciano, J., Becker, D. and Levy, B. (2017) Beyond PD-L1 testing-emerging biomarkers for immunotherapy in non-small cell lung cancer. *Ann. Transl. Med.* **5**, https://doi.org/10.21037/atm.2017.06.48