ANNUAL REVIEWS

Annual Review of Biomedical Data Science

Network Analysis as a Grand Unifier in Biomedical Data Science

Patrick McGillivray,[1] Declan Clarke,[1] William Meyerson,[2] Jing Zhang,[1,2] Donghoon Lee,[2] Mengting Gu,[2,3] Sushant Kumar,[1] Holly Zhou,[1] and Mark Gerstein[1,2,3]

[1]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520, USA; email: mark@gersteinlab.org

[2]Program in Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut 06520, USA

[3]Department of Computer Science, Yale University, New Haven, Connecticut 06520, USA

Annu. Rev. Biomed. Data Sci. 2018. 1:153–80

First published as a Review in Advance on April 25, 2018

The *Annual Review of Biomedical Data Science* is online at biodatasci.annualreviews.org

https://doi.org/10.1146/annurev-biodatasci-080917-013444

Keywords

network analysis, molecular interaction, systems biology, cross-disciplinary research, network medicine

Abstract

Biomedical data scientists study many types of networks, ranging from those formed by neurons to those created by molecular interactions. People often criticize these networks as uninterpretable diagrams termed hairballs; however, here we show that molecular biological networks can be interpreted in several straightforward ways. First, we can break down a network into smaller components, focusing on individual pathways and modules. Second, we can compute global statistics describing the network as a whole. Third, we can compare networks. These comparisons can be within the same context (e.g., between two gene regulatory networks) or cross-disciplinary (e.g., between regulatory networks and governmental hierarchies). The latter comparisons can transfer a formalism, such as that for Markov chains, from one context to another or relate our intuitions in a familiar setting (e.g., social networks) to the relatively unfamiliar molecular context. Finally, key aspects of molecular networks are dynamics and evolution, i.e., how they evolve over time and how genetic variants affect them. By studying the relationships between variants in networks, we can begin to interpret many common diseases, such as cancer and heart disease.

1. INTRODUCTION

1.1. Networked Systems Are at the Core of Human Biology

A great diversity of networks are relevant to the field of biomedicine. Social networks model human interaction and may help explain pathways of disease transmission. Layers of neurons in the brain process sensory information, and the layered architecture of neuronal networks inspired the artificial neural networks used to identify patterns in data, including biomedical data sets (1). The circulatory system is a branching network of vessels that connects organs in the body. Vast networks of interacting molecules, in particular, are foundational to human health and disease, forming a functional base layer for several higher-order biological networks (**Figure 1*a***). Transfer of genetic information, cellular communication, and human metabolism are all mediated by complex pathways and networks of molecules.

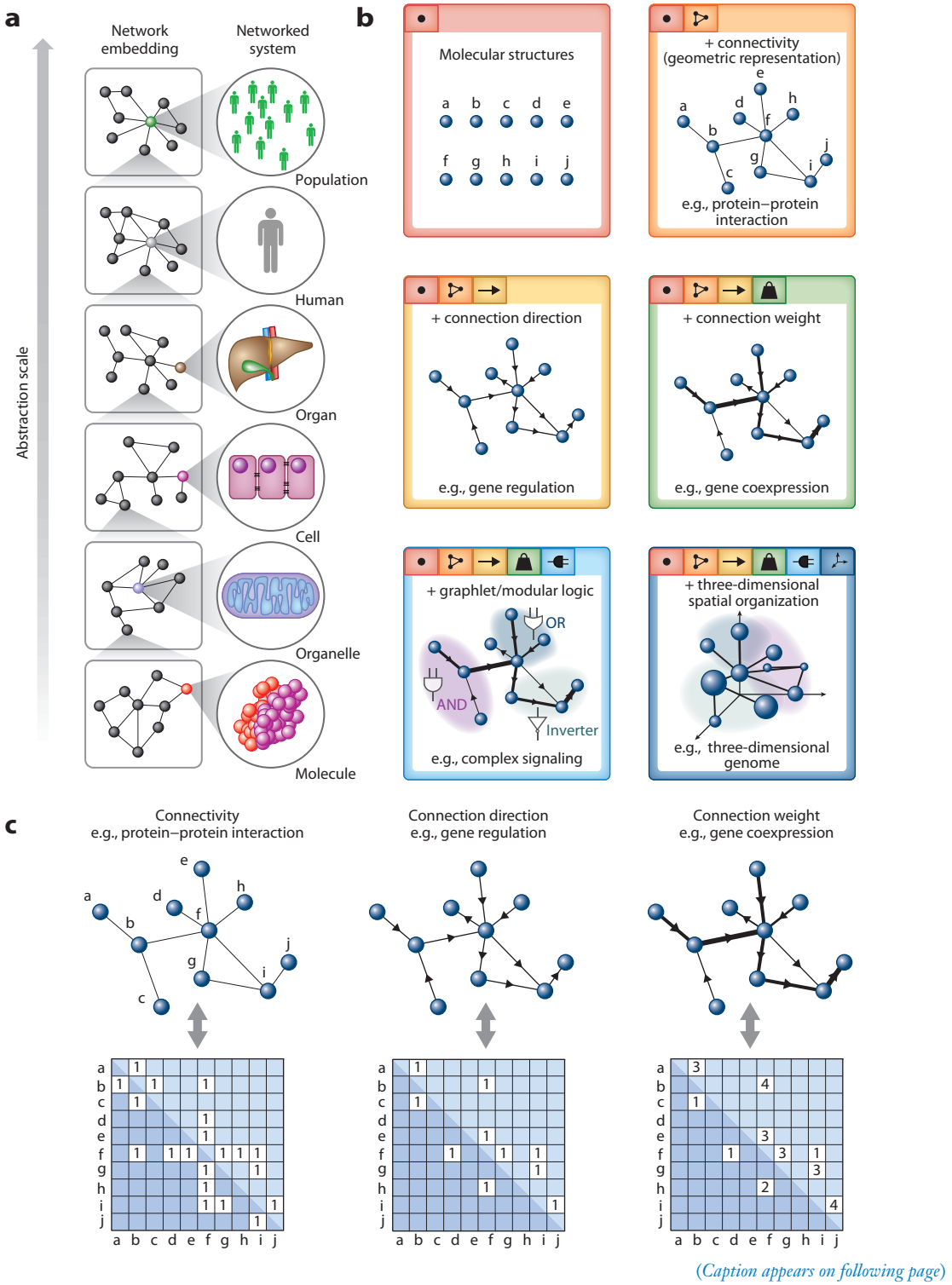
Networks are a powerful framework for understanding molecular interactions because of the breadth of network analysis techniques developed across diverse disciplines. Novel network analysis techniques like HotNet (2, 3) use algorithms similar to those first developed for studying belief propagation in social networks (4) to annotate function in molecular networks. Machine learning techniques like the deep neural network DeepBind (5) apply techniques refined for use in computer vision (6) to generate accurate network topology predictions from genomic sequences. Cross-disciplinary comparisons between networks have revealed that the gene regulatory network (GRN) of *Escherichia coli* is functionally robust compared to computer software networks that prioritize efficiency and reuse of basic functions (7). Like a social network, apparently distant immune cell types may be more closely connected through mutual acquaintances than they appear, and cross talk between immune cells may modulate the body's immune response (8).

Molecular networks can function in ways that are unfamiliar from a human perspective, and it can be challenging to develop intuitions about them. Because network analysis also applies to systems about which humans have well-developed intuitions, such as social networks and electrical wiring networks, by comparing molecular networks to familiar or more intuitive networks, we can gain knowledge and understanding about the molecular world.

Network analysis of large-scale molecular data has been used to identify critical pathways and proteins in GRNs (9), including molecular pathways affected by cancer (10). Off-target effects of prescription drugs have been predicted through a network model of metabolism (11). Insights into inflammatory diseases like asthma have been revealed by studying the structure and function of networks of inflammatory signaling molecules (12–14).

Molecular networks change and evolve over time with surprising dynamic complexity (15). Pro-inflammatory T cells of the immune system rewire their regulatory networks in autoimmune disease (16). The microbiome of the gut interacts with the human metabolome, and both change together in response to diabetes, pregnancy, or antibiotic treatment (17–19). Substantial changes in the epigenome are observed in human tissues according to cell type (20). Network rewiring may be both the cause and the consequence of changes to human health (21). Complete understanding of many molecular networks requires an understanding of these temporal features.

The temporal evolution of molecular networks allows them to perform logical operations and transmit complex signals (22). Exciting discoveries have been made related to the possibility of logic-based communication performed by networks. A Boolean model of GRN function has been used to successfully predict gene expression in embryonic development (23, 24). There is a possibility for future bioengineering of molecular interaction networks to perform complex logic and to intervene in disease processes (25, 26). A greater understanding of biological networks and their logical structures may eventually provide a platform for augmenting existing biological capabilities.



(Caption appears on following page)

Figure 1 (*Figure appears on preceding page*)

Network representations. (*a*) Molecular networks form a functional base layer for several higher-order biological networks, including networks of organelles (e.g., vesicular transport), cellular networks (e.g., neural), and population-scale networks (e.g., disease transmission). (*b*) Abstract network representations can be built through a progressive layering of information and logic, according to the network under study. For instance, the addition of directional information to a network may be particularly important when representing a gene regulatory network. (*c*) Matrices are useful for representing certain network variables, like the pattern of connections and connection weights.

Network analysis of biomedical data is not just a research technique but has also contributed to advances in understanding and practice in modern medicine. Many common diseases, including heart disease (27), schizophrenia (28, 29), diabetes (30), and cancer (31), are unlikely to be associated with a single molecular alteration but with multiple affected genes in critical molecular pathways. Gene expression panels used in clinical practice, like the 21-gene panel Oncotype Dx® that predicts breast cancer recurrence, identify molecular phenotypes as proxies for disease phenotypes (32). Disease transmission through social networks, as in the 2013 Ebola virus outbreak in West Africa (33) or the Zika virus spread in the Americas (34, 35), may be tracked through molecular signatures left by the virus as it spreads. These examples suggest the value of network analysis techniques to medicine.

1.2. Networks Leverage Abundant Biomedical Data

The Human Genome Project was an early big data and large-scale science project in biology (36). It was among the motivators for the development of the discipline of systems biology (37). When large-scale biology projects like the Human Genome Project produce a parts list of molecular structures and entities, systems biologists seek to understand how these parts are connected. Network theory became a foundational technique for making sense of these increasingly large data sets of connected biomolecules.

Molecular biology projects continue to expand in size and scope. Genome-scale network reconstructions of metabolic networks have been produced for hundreds of species and are constantly undergoing refinement (38, 39). The recently released BioPlex 2.0 is the largest protein–protein interaction network (PPI) ever built, with 56,000 listed interactions (40). Whole-genome sequencing projects like the 100,000 Genomes Project and the Genome Sequencing Program at the National Institutes of Health now seek to enroll hundreds of thousands of participants (41, 42). Researchers have presented visions for sequencing at even larger scales (43, 44), and the growth of big data in genomics may outpace big data growth in other data-intensive fields (45).

Networks produced from data of this scale have been likened to a hairball when visualized, suggesting their complexity (46). Identifying meaningful structure and function in these hairballs represents a challenge in the field of biology. The application and development of computational network approaches represents one of the most promising means of unraveling the complicated patterns of connection in these networks (47–49).

The importance of network techniques for analyzing large-scale molecular interaction data is further underlined by the need to integrate diverse sources of molecular data. The number of advanced functional molecular assays available to researchers continues to grow through projects like ENCODE (Encyclopedia of DNA Elements) (20), and new network-based approaches for integrating large-scale biological data are being developed (50). Integration of functional genomics data has been proposed as the clearest way forward to understanding the significance of human genetic variation (51, 52). Network approaches play a central role in the integration of these diverse sources of large-scale molecular interaction data.

1.3. Making Sense of Complexity in Biomolecular Networks

Complex biomolecular networks are incomprehensible in their raw, complete form. Finding meaning and understanding in a complex network requires focus, synthesis, and comparison. Most straightforwardly, networks become comprehensible by focusing on only some portion of the full network. A more scalable approach is to compute summary statistics about the network. Alternatively, networks can sometimes best be appreciated by comparison with other networks, including cross-disciplinary comparisons.

Networks are like maps in that both organize local information in a global context. This is analogous to a map of the world, where the architecture of cities cannot be appreciated at the scale of countries or continents. Large, complex biomolecular networks are best visualized with either reduced detail, restricted scale, or both, except when demonstrating the size of a data set. For example, although metabolism is an extremely complex process (**Figure 2**), glycolysis—the core subgraph of metabolism—is simple enough for a dedicated high school student to appreciate in an afternoon, while rich enough to convey principles of metabolism. In Section 2.4, we use logic gates as a case study to illustrate the interpretative utility of subnetworks. Premier online databases of biomolecular networks, such as the KEGG (Kyoto Encyclopedia of Genes and Genomes) Pathway Database (53), support interactive visualizations of networks.

A second way to understand a biomolecular network is through its summary network properties. Stanley Milgram famously discovered that between any two residents of the United States he studied, there are on average six degrees of separation (54). The short average path length of the American social network is an interesting property that helps us to appreciate how people are connected to each other and how ideas and infections can quickly spread. In the human PPI, one study found that the average path length is around 4.85 (55). This connectivity between proteins helps us appreciate why so many different proteins may be relevant to a given human trait or disease. Several summary measures that may be calculated for a network are provided in **Table 1**.

Some of the most interesting insights about biomolecular networks come from comparisons between them. We can fruitfully compare a biological network to a randomly generated network, a related biological network, or even a network from another discipline. Comparing a biomolecular network against randomized networks helps us appreciate which properties are fundamental to a network and which are merely expected by chance (see Section 3.3). Comparing a biomolecular network between healthy and diseased samples highlights changes that may be relevant to disease pathogenesis (see Section 2.3 for an application to cancer). These comparisons between healthy and diseased states can be made either at the level of individual edges that have been gained or lost or at the level of summary network properties, such as their overall connectedness or hierarchical properties. Comparing biological networks with man-made networks that have been designed for some function can inspire us to wonder, with due caution, whether the biological network has been evolutionarily designed to perform that function (see Section 4.3 for examples but Section 3.3 for challenges in making inferences about evolutionary forces in networks).

Biomolecular networks are so rich in information as to be unintelligible in raw form. Fundamentally, to understand something about a network, we need to process the information about biomolecular networks into human-sized chunks. These chunks can literally be subgraphs of a network, summary statistics about a network, or subgraphs or summary statistics that emerge as special when comparing two networks. Each of these three approaches for understanding networks represents a potential source for future progress in understanding networks. We can better visually navigate subgraphs of biological networks by borrowing techniques from interactive

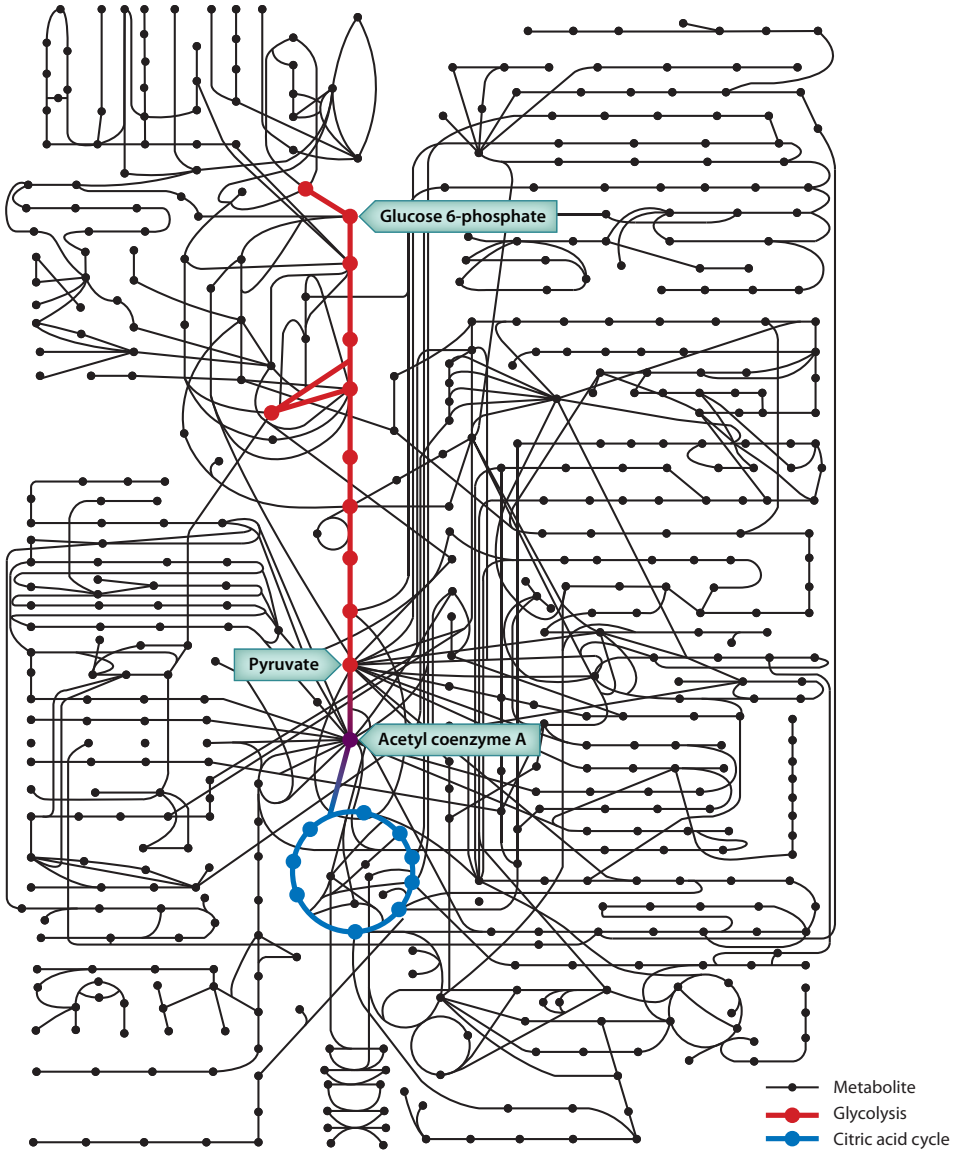


Figure 2

Glycolysis and the citric acid cycle. Despite the complexity of the complete human metabolic network, the core subgraphs of glycolysis and the citric acid cycle can be appreciated in their global context through selective focus. The network structure of glycolysis is linear, while that of the citric acid cycle is cyclical. The two subgraphs are deeply enmeshed within the other processes of metabolism. Adapted with permission from Reference 157.

cartography. We can enrich our repertoire of summary statistics in biological networks by reflecting on the kinds of patterns relevant to biology. Finally, we can devise more apt comparisons between biological networks by accumulating natural and interventional experiments and by employing state-of-the-art randomization techniques from network science.

Table 1 Commonly used network statistics and measures

Name	Basic description	Reference
Modularity	A measure of the strength of network partitioning. Apart from measuring degrees and paths, one can easily observe that social networks tend to have communities within them due to the relatively larger number of interactions between people in the same neighborhood, school, or workplace.	145
Betweenness	The number (or fraction) of shortest paths between a given node. High betweenness nodes are termed bottlenecks, and removal of these nodes could reduce the efficiency of communication between nodes.	146
Influence	A property of a node that measures its importance by taking into account the importance of its neighbors. The PageRank algorithm is a prominent example of this characteristic.	147
Missing links	Unobservable or missing connections. Link prediction makes use of known relationships or connections among nodes to identify missing links. High-throughput experiments can be noisy, and the resultant networks may contain spurious links; missing data are also very common. Methods for link prediction and denoising are therefore useful.	No primary reference available

2. MODELING A MOLECULAR INTERACTION NETWORK

2.1. Basic Features of an Abstract Molecular Interaction Network

Before discussing more advanced techniques for modeling and analyzing molecular interaction networks, we present a few widely used definitions and principles that serve as building blocks for more advanced methods.

In abstract form, networks consist of a set of nodes, with edges representing connections or relationships between them. In the context of molecular networks, the nodes of a network may represent a parts list of molecular entities, without labeled connections (**Figure 1*b***). If the pattern of connections (edges) between molecules is known, a network can be formed. Information and logic can be layered on such a basic network and can be tailored to the kind of network under study. For example, the direction of connections and the weight of connections may be important information for GRNs and gene coexpression networks, respectively.

Matrix representations of interaction network variables are also possible for some networks. Matrix representations of the connections, weight, and direction of connections in hypothetical interaction networks are shown in **Figure 1*c***.

These network variables (connections, direction, weight, time-dependent logic, and spatial geometry) are basic building blocks that network scientists use to describe molecular interaction networks. In addition to these basic building blocks, summarized in **Figure 1*b***, a pictorial glossary of network terminology is presented in **Figure 3**.

2.2. Incorporating Molecular Structure in a Network Model

Although there are advantages to abstract representations of molecular networks, there are also inherent limitations. For instance, protein–protein interactions are often represented as a PPI (**Figure 4*a*,*b***). Nodes in this network correspond to individual proteins and edges represent interactions between them. Such abstract representations are helpful for understanding the overall topological properties of the PPI. Furthermore, one can identify key proteins based on their connectivity in the network. However, such abstract representations do not provide any biophysical insight into interactions underlying protein–protein interactions.

To address this issue, various studies have integrated three-dimensional structural information data available for various biomolecules to produce structural interaction networks (SINs) (56–58) (**Figure 4*c***). Integration of structural information can help address key issues. For example, one

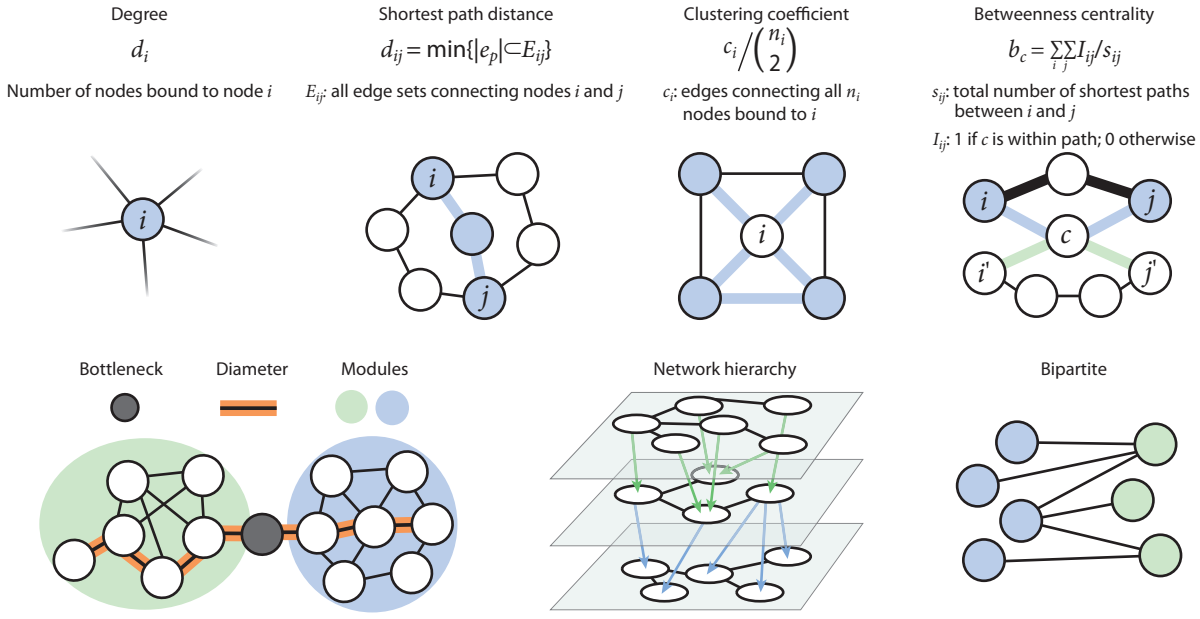


Figure 3

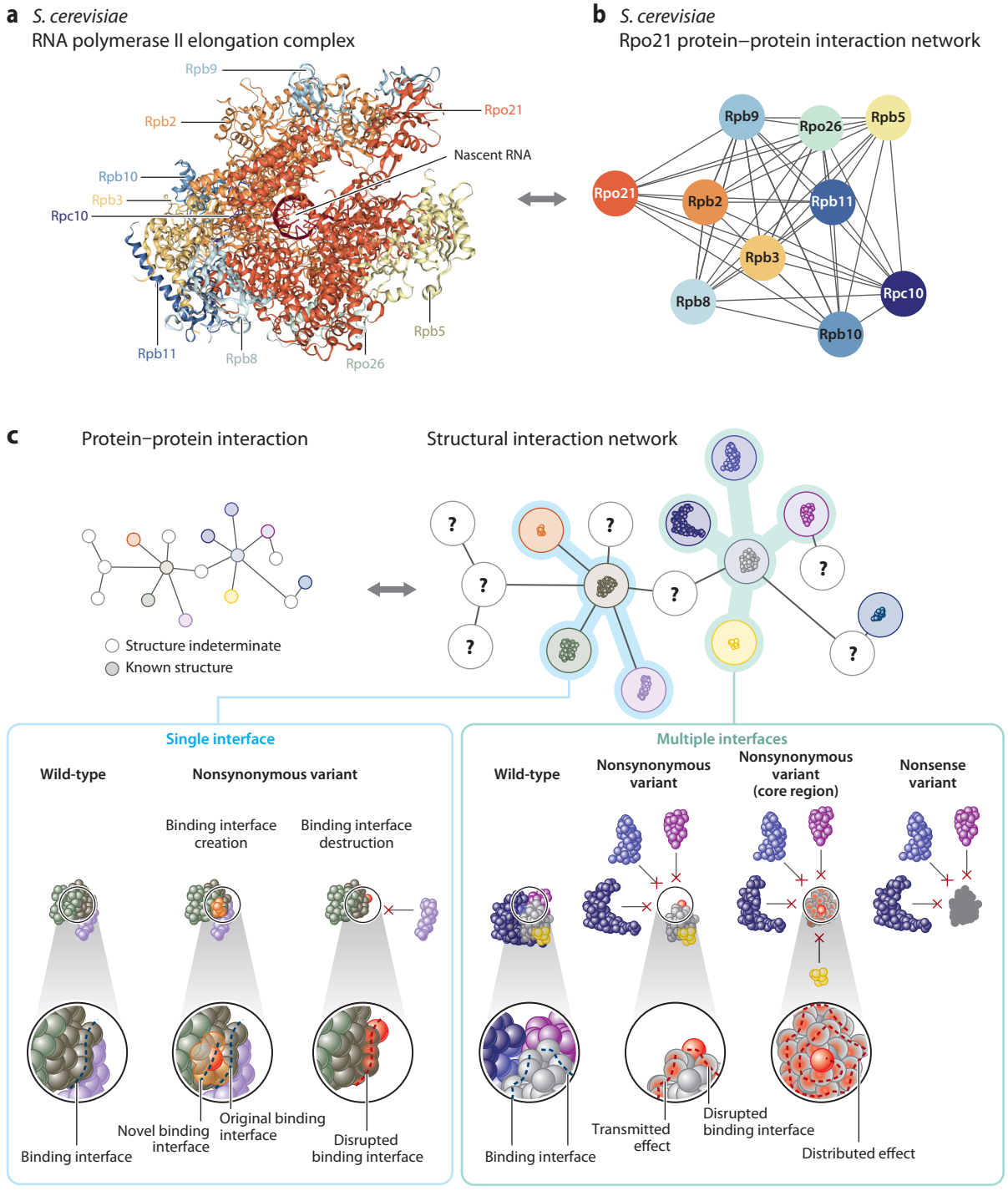
Pictorial glossary of common network concepts and measures. Many of these metrics (such as degree, clustering coefficient, and betweenness centrality) are used as measures of node importance or influence. Node and edge metrics may be used by algorithms to elucidate higher-order topological features of networks (such as modules and diameter). Hierarchical structures have been used to organize many types of systems, including regulatory networks.

can identify key residues or domains on the surface of proteins, which are involved in interactions. In addition, structural information is helpful for predicting binding affinities and kinetic constants of the underlying interactions. Furthermore, SINs are helpful for identifying obligate (permanent) or transient interactions in a network. Structural information can also help distinguish between simultaneous and exclusive interactions. These are key network properties, which cannot be addressed with a simple abstract representation of the network. Finally, integration of structural information can help in gaining a mechanistic understanding of the impact of rare or disease-associated mutations on protein–protein interactions (59). SINs can thus be used to prioritize variants in a disease cohort or rare deleterious variants in a population-level study.

2.3. Network Rewiring: The Time-Based Evolution of Molecular Networks

Biological networks are hardly static; they may evolve slowly over time or transform rapidly to adapt to an environmental change, either throughout development (60) or simply as a result of the accumulation of mutations. In the context of biological networks, rewiring refers to a complex reformation of interacting partners, such as genes, proteins, and other biologically relevant chemicals (**Figure 5*a***).

The central concepts of network rewiring are decades old. Prior efforts to understand network dynamics compared GRNs in varying conditions (15). However, the scope of these efforts was limited by data availability. The advent of large-scale genomic and proteomic surveys allowed for the creation of different types of biological networks, including PPIs and GRNs, in a variety of cellular contexts.



(Caption appears on following page)

Figure 4 (*Figure appears on preceding page*)

The molecular interaction network of the RNA polymerase II elongation complex in *Saccharomyces cerevisiae* can be represented structurally (*a*) or as an abstract molecular interaction network (*b*). The molecular structure information lost in an abstract network representation may be important for interpreting certain observed molecular network phenomena. Panel *a* adapted with permission from the Research Collaboratory for Structural Bioinformatics Protein Data Bank (identifier 1I6H) (158), visualized with NGL Viewer (159). Panel *b* adapted with permission from STRING v10 protein–protein interaction database, showing experimentally determined interactions (77). (*c*) Three-dimensional protein structure data can be mapped onto protein–protein interaction networks (PPIs) to construct structural interaction networks (SINs). SINs provide physical intuition and nuance for the interactions in a PPI. For instance, a SIN can help distinguish interactions involving single or multiple interfaces. This can be helpful for identifying permanent and transient interactions in the network. High-resolution definitions of various interactions are helpful when prioritizing disease-associated variants to gain mechanistic insights. For example, disease-associated nonsynonymous variants can either create or destroy a binding interface of an individual protein. This, in turn, will influence its interaction with other proteins in the network, which can drive disease progression. Furthermore, variants influencing the core and surface of proteins will affect interactions in different ways. For example, for a given protein, mutations on its surface will mostly affect interactions involving a particular interface, whereas those in the core may disrupt all interactions.

It remains difficult to measure the dynamic nature of biological networks. However, advanced biomolecular assays can provide clearer insight into how genes and proteins operate in a point-in-time snapshot (**Figure 5*b***). Researchers may then stitch these snapshots together to answer complex, time-dependent questions in systems biology (**Figure 5*c***).

For example, a survey of the regulatory dynamics of PPIs in both time and space allowed researchers to discover interesting global properties in the interactome network. In this study, they discovered two distinct types of hub proteins: party hubs, which interact with most of their partners simultaneously, and date hubs, which bind their different partners at different times or locations (61).

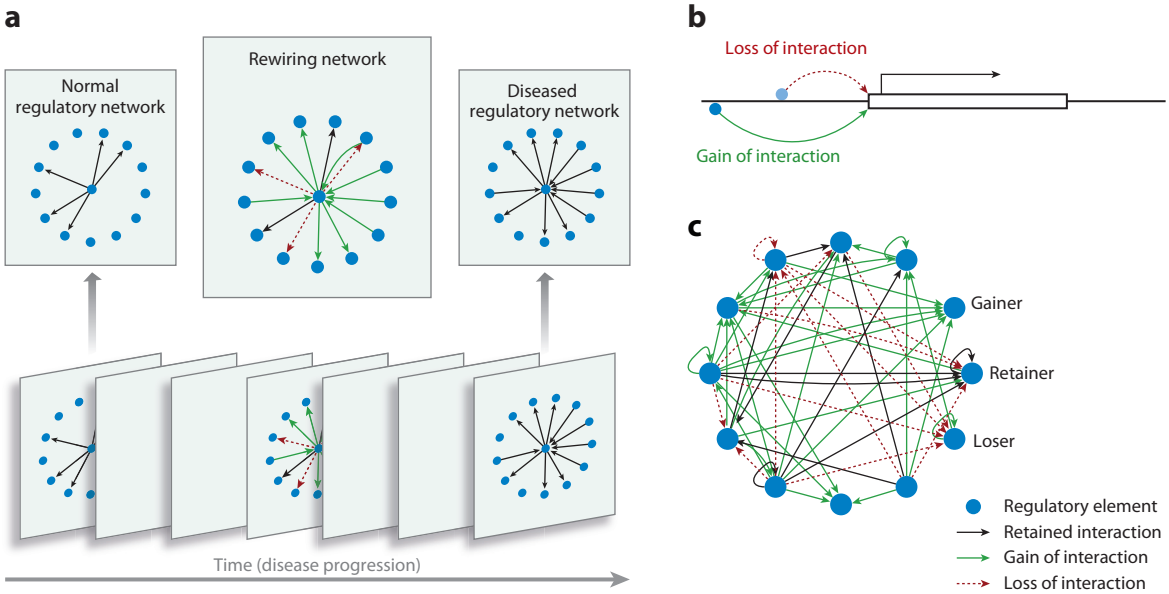


Figure 5

Network rewiring. (*a*) A schematic diagram illustrates the progression of a regulatory network from normal to a diseased state. The state of the regulatory network at a specific point in time is depicted as a snapshot. (*b*) Binding profiles of regulatory proteins can be used to infer both gain and loss of interaction in different cell states. (*c*) By reconstituting the time progression of the regulatory network, the resulting network rewiring can summarize the dynamic changes in regulatory elements.

Many studies have focused on the broadest timescale for network rewiring by linking the evolutionary changes of biological networks to diversity among species (62). In particular, it has been shown that regulatory changes in GRNs may account for species differentiation (63–66). However, researchers have also attempted to interpret network rewiring at much shorter timescales. It is possible to introduce an artificial perturbation into a network and examine the rewiring that results. One study of a bacterial GRN showed that a single perturbation can affect gene expression by four orders of magnitude greater than the scale of perturbation, altering up to approximately 70% of the transcriptome (67).

Rewiring is often the result of genetic mutation. A single mutation placed at a regulatory protein binding site can alter binding specificity, perturb its interacting neighbors, and consequently, have a detrimental downstream effect on the whole network. Naturally, many researchers have attempted to measure rewiring to infer the consequence to disease phenotype.

For example, cancer mutations can affect both downstream and upstream rewiring of the GRN, altering cell signaling and gene expression (68, 69). Measuring rewiring (i.e., target changing) of a GRN involves comparison of a network in two states: the reference (healthy) state and the evolved (diseased) state. Measuring the extent to which a gene is perturbed in a network has revealed tumor drivers and genes associated with patient prognosis (70). The regulatory interconnection between genes can be represented as the gain, loss, or retention of molecular interaction. As a result, network rewiring can change gene hierarchy, promoting or demoting the importance of a gene as regulator (71).

More recently, CRISPR genome-editing technology has been developed and widely applied in the field of genomics, allowing researchers to design more complex models to test the effects of cancer mutations. CRISPR could prove to be an excellent tool for both performing a high-throughput screening of network perturbation and experimentally validating the results of rewiring obtained via an integrative approach.

Rewiring may be viewed as an irreversible temporal evolution of a biomolecular network. However, when viewed at a much shorter timescale, biomolecular dynamics can be understood as concerted and responsive changes in a biomolecular network. Pairs of regulatory molecules can work collaboratively, competitively, or redundantly. More complex function—like the integration of a time-varying hormonal signal or a conditional cellular response to an environmental change—is enabled through the dynamic behavior of molecular networks. Molecular networks may even be compared to logic gates (72), with spatiotemporal information revealing their mode of operation.

2.4. Network Motifs, Network Logic, and Network Stability

At the evolutionary timescale, biological networks such as PPIs have evolved to maximize network efficiency, functionality, and stability. Network structure evolves alongside biological function and lays the foundation for complex network processes. Studies have shown that small, structurally stable network motifs are enriched in GRNs and perform various functions (73). Negative autoregulation motifs, for example, allow the use of strong promoters, which shorten the response time of stimuli-induced gene expression regulation. The autorepressive nature of these motifs allows cells to quickly attain stable protein product concentrations and reduce variation in protein levels among cells (74).

Another frequently observed motif in GRNs is the feedforward loop (**Figure 6*a*,*b***). Unlike direct stimuli that generate a rapid response, feedforward loops with AND gate logic require more persistent stimulation to activate both input components, thus filtering out brief spurious pulses of a signal. Combinations of network motifs enable more precise control of biological systems, including the temporal order of gene expression and circadian oscillations (75).

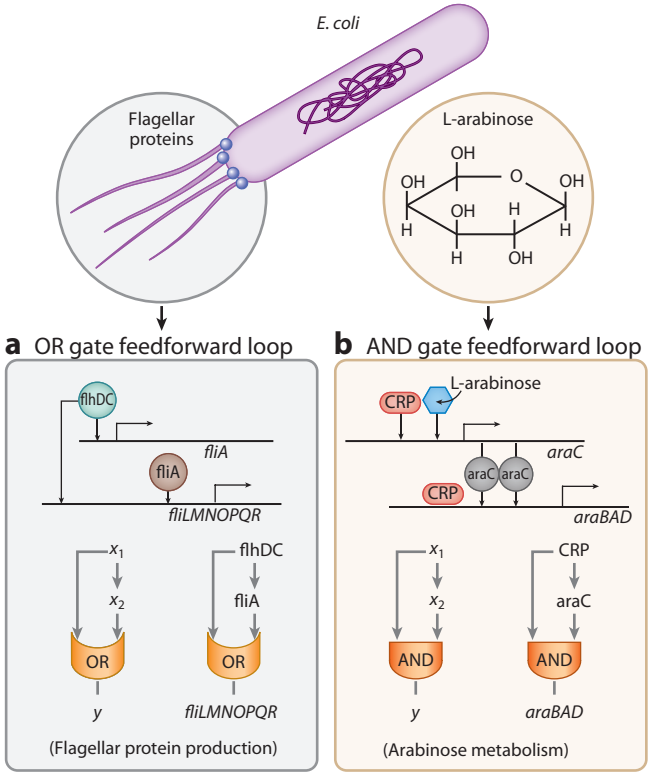


Figure 6

Feedforward loops (FFLs) are a frequently observed motif in molecular networks. (*a*) An example of a coherent FFL active in the regulation of flagellar protein production in *Escherichia coli*. The flhDC complex directs the production of fliA, which activates class 2 operon genes *fliLMNOPQR*. The flhDC complex also acts additively to activate *fliLMNOPQR*. (*b*) Also in *E. coli*, the presence of arabinose induces the formation of the araC-arabinose complex, which is essential to transcribe the *ara* operon. CRP (C-reactive protein) and cyclic AMP (adenosine monophosphate) are required in this process.

Biological networks have also developed structure to enhance stability. The molecular network, for example, is subjected to exogenous attacks or endogenous mutations that result in dysfunction. A cascading deleterious effect could propagate via links in the network. An observed feature of many molecular interaction networks is the duplication of extremely vital hubs. Multiple and repeated domains are enriched in hub proteins (76). While redundancy may lead to inefficiency, biological networks must balance between stability and energy loss.

3. TOOLS AND ALGORITHMS FOR NETWORK ANALYSIS

3.1. Network Prediction Using Machine Learning and Neural Networks

Network prediction methods have evolved in parallel with the evolution of large-scale biological experimentation. Experimental molecular interaction data contain both false positive and false negative interactions (77). Predictive algorithms attempt to identify these false positive and false negative cases and so address the limitations of experimental methods. For the well-studied case of PPI data, diverse predictions methods include predictions based on gene ordering and genetic sequences (78), network topology (79), Bayesian inference and machine learning methods (80),

measurements of structural similarity (81), and text mining (77). Network prediction methods can be combined to yield more accurate predictions, and a large body of literature is devoted to improving network predictions (82).

Machine learning methods, and neural networks in particular, have become popular methods for network prediction. Machine learning methods can predict relationships in networks without necessarily requiring strong assumptions about underlying interaction mechanisms (83). Dimensionality reductions make large genomic data sets more computationally tractable, and machine learning methods also allow diverse data types and a wide variety of molecular features to be integrated to form predictions (84). These attributes allow these methods to scale with increasing volumes of high-throughput molecular data and to accommodate new forms of data as they become available.

An example application for network prediction is the identification of DNA and RNA targets of regulatory proteins. An accurate understanding of GRNs is important for modeling networked biological processes and for determining the impact of genomic variants—particularly those variants in noncoding regions that do not directly affect protein structure. Predictive methods can integrate protein–DNA and protein–RNA interaction data from a variety of sources, including protein-binding microarray and chromatin immunoprecipitation (ChIP), while also tolerating bias and error latent in these data sources.

Conceptually straightforward methods for predicting the targets of DNA and RNA regulatory proteins count the frequency of sequence-based motifs identified in high-throughput experiments (85). It is also possible to compare candidate protein-binding sequences to those already categorized in databases or confidently identified in other species (85).

Recently developed neural network algorithms designed for predicting DNA–protein and RNA–protein interactions include DeepBind (5), DeepMotif (86), and TFImpute (87), a deep learning–based imputation method for transcription factor (TF) binding prediction. These convolutional neural networks aim to provide a better understanding of regulatory network structure and tools for researchers to prioritize mutations by their impact on protein binding sites (88).

DeepBind and DeepMotif take sequencing data from high-throughput experiments and perform a convolution of sequence-based protein-binding motifs to predict the sequence specificities of DNA-binding proteins and RNA-binding proteins (RBPs) (5, 86) (**Figure 7*a***). DeepBind improves upon prior motif-scanning algorithms by taking into account RBPs that recognize secondary or tertiary structural elements. It also recognizes higher-order structures that result from competitive or synergistic effects of protein binding (5).

To predict TF binding sites, TFImpute takes input data from combinations of cell lines and also considers low-affinity binding sites and repeat sequence symmetries (87). These features are designed to provide a more accurate model of TF–DNA binding specificity. Improvements to TFImpute over DeepBind and DeepMotif in TF binding site prediction were particularly notable in sequencing from cell types for which protein binding data through ChIP are not available. This suggests an application of predictive computational approaches to replace more expensive experiments that may have limited availability.

As experimental methods improve and evolve, computational biologists can expect to have greater quantities of high-quality data to work with. The predictive algorithms that will be most helpful in elucidating the complicated biological networks studied in systems biology will be those that can integrate diverse data sources while also scaling with increasing data set size.

3.2. Advances in Network Algorithms: Network Propagation Methods

In biology and other disciplines, networks have long been used to study complex associations within large data sets. In the context of biology, such data sets include physical interactions

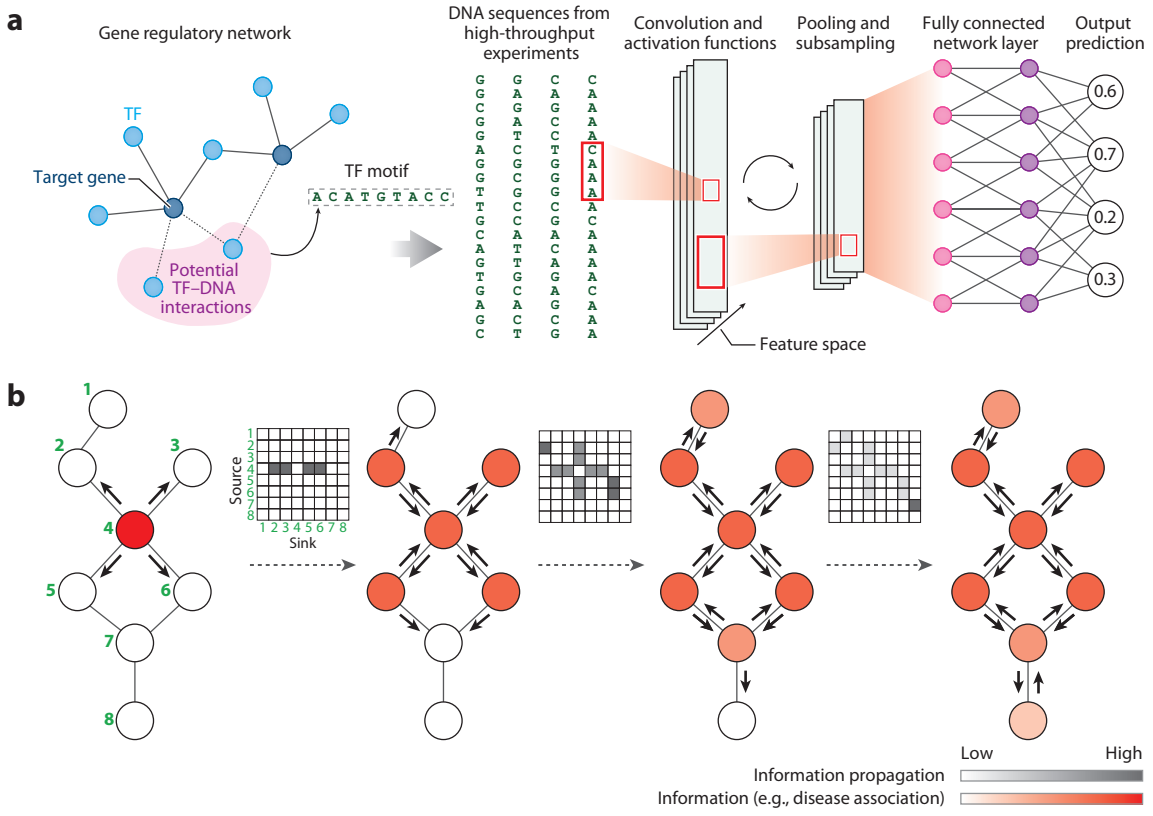


Figure 7

Network algorithms. (*a*) The general structure of a convolutional neural network with sample input and output (similar to DeepBind). Here we are trying to detect transcription factor (TF) binding sites. If we have high-throughput sequencing data containing sequences of potential TF binding sites, we can produce as output the probability that a particular sequence is a TF binding site. Training data consist of sequences with experimentally determined binding scores. The convolution layer performs feature extraction by convolving the input matrix with a convolution matrix called a kernel or feature detector. The resulting matrix is the feature map, which in this example would be sequence motifs. An activation function operation (e.g., rectified linear unit) introduces nonlinearity into the model. Pooling and subsampling reduce the dimensionality of the feature map; the depth of the feature map corresponds to the number of kernels used in the convolution step. The fully connected layer uses the feature maps to make predictions about the input. (*b*) A series of steps by which information (sometimes termed "heat" in networks literature) propagates through a network (*left to right*). This information originates in node 4 (often a gene believed to be disease-associated with high confidence) and subsequently flows to neighboring nodes 2, 3, 5, and 6. In the next step, this signal may partially flow back into node 4, as well as neighboring nodes 1 and 7, before eventually reaching node 8. Matrices represent the propagation of heat from source to sink nodes. When applied to large networks, the resultant distribution of heat throughout the network may enable one to assign well-defined modules.

between proteins (i.e., PPIs), regulatory relationships [e.g., associations between TFs and target genes or microRNAs (miRNAs) and their associated targets], or directed pathways of interacting cellular species. As these data sets grow, the associated networks used to describe them become more topologically complex. Positively identifying true signals in these networks can be difficult, given the noise and complexity that accompany them. Algorithmic frameworks have been recently developed to capture relationships between genes that are difficult to discern, as well as to identify subnetworks that may be dysregulated (**Table 2** offers a list of network partitioning and module detection techniques). Along these lines, algorithms based on network propagation have proven to be the most powerful (89) (**Figure 7*b***).

Table 2 Methods for network module identification and network partitioning

Name	Basic description	Reference
Girvan–Newman (GN)	Remove edges (in descending order of their betweenness) until the modularity of the current network partition is maximized (modularity evaluates a partition relative to that of a null model).	148
Gap statistic	Similar to GN, but with the intention of identifying K (i.e., the number of clusters) without a priori information about the ideal value of K.	149
Greedy optimization	Edges are successively introduced to nodes (to build the graph from scratch). The order in which edges are added is guided by the need to give the largest possible modularity jump at each stage.	150
Simulated annealing	Similar in nature to greedy modularity optimization, but with greater performance (although longer run times result from exhaustive searches).	140
GN with edge clustering coefficient	As a local measure, edge clustering coefficient is much faster than node-based GN.	151
Infomap	Searches for modular network communities by reducing module detection to an information compression problem.	152
Cfinder	Searches for communities that may overlap (i.e., share nodes). Such a case is common in social interaction networks.	153
Spectral clustering	Node eigenvectors (within a community) would need to have similar values if communities are well-defined with strong partitions.	154
Potts models	Minimizes a Hamiltonian function of a Potts-like spin model, wherein spin states designate community membership.	155
Fast modularity maximization	A coagulation-based method in which nodes may be appended to neighboring nodes (to build a conglomerated node), thereby forming a smaller and simpler network. This is iteratively performed until the modularity is optimized.	156

Generally speaking, the term "network propagation" refers to the analysis of networks by allowing some form of information to flow from one node to another via shared edges (90, 91). This information may traverse from node to node as a random walk, for instance. Edges may also be weighted (by the confidence of an interaction, for example) to influence the current of information traveling from one node to another.

Other approaches at inferring gene–gene associations include direct neighbors or shortest paths. Such methods may suffer from high rates of false positives or false negatives, whereas propagation-based methods may optimally capture known gene–gene associations. For instance, Ruffalo et al. (90) use propagation to positively identify cancer-associated genes using both somatic variant data and gene expression as the input to the original network. Such methods have also been used to identify cancer subtypes based on patient stratification (92) and in an array of other disease contexts (50, 93–95).

3.3. Causal Inference About Network Properties

Do the network properties of biological systems really matter for health and disease? We believe the answer is yes. For example, redundancy among paths between nodes within biological networks leads to robustness against genetic and environmental perturbations. Nonetheless, a more systematic approach to answering this question would involve an assessment of the evolutionary selection on network properties, which, despite significant progress, remains an unsolved problem and an area of active study in biological network science.

Conceptually, there are two steps to identify whether some biological network property of interest has been subject to evolutionary pressure. The first step is to show that the observed network property differs from neutral expectations. The second step is to show that the difference was a direct result of evolutionary optimization, rather than a side effect of the evolutionary optimization of some other property.

How does one show that an observed network property differs from expectations? To start, one identifies and models the mutational processes that generate network structure diversity. Then, one computes the network property of interest on hypothetically generated nulls. If the network property of interest falls at a very high or very low quantile among these null networks, then this is some evidence that that network property is not merely neutrally evolving. An approach like this identified that the exponential distribution of edges within PPIs is a simple consequence of known neutral patterns of gene duplication (and therefore lacks evidence for selection) (96). This approach is not tenable when we lack the detailed knowledge of mutational processes to accurately specify neutral models. An alternative approach derives neutral models by permuting elements of the observed network using general network techniques, but the biological relevance of such permutations is not obvious. A limitation of both approaches is that it may well be that the network property was evolutionarily optimized, with fitness costs for small departures in either direction from the optimized value, but if that value is an intermediate value, then it will not appear to be extreme compared to the neutral set.

It is even more challenging to show that some statistically significantly extreme network property is not simply the result of evolutionary selection on some other property. For example, the fact that the mammalian brain divides into two hemispheres is a foundational property of the brain that has a dramatic impact on network properties. If this inherent hemispheric structure in the brain is not considered, then many properties of human neural networks will incorrectly appear significantly different from null even if they merely represent random perturbations from this hemispheric structure (97). Furthermore, the fact that the brain divides into two hemispheres is only the most obvious global structural constraint on the brain; there are many layers of structure underneath this one in the brain and in biomolecular networks that constrain neutral variation more than our neutral models predict. This example illustrates the general principle that the fundamentality and causal impact of network properties are extremely difficult to infer and cannot be solved by any one network algorithm.

4. APPLICATIONS

4.1. Network Medicine: Clinical Application of Molecular Interaction Networks

Some diseases, like sickle cell anemia, are thought to be caused by single mutations or alterations of a single genetic locus (98). Complex diseases are conditions understood to have multiple determinants of severity, including genetic and environmental risk factors (99). This is similar to how complex traits like height are thought to arise from the interaction of multiple genetic loci (100). Complex diseases include prevalent conditions like heart disease (27), schizophrenia (28, 29), diabetes (30), and cancer (31). Single or multiple effectors in the same molecular pathway may cause a complex disease, or a disease may result from a more distributed network effect with multiple pathways involved (101).

Gene set enrichment analysis and other forms of pathway analysis directly address the possibility of pathway-driven diseases (102). Pathway analysis reveals that genetic variation in patients with autism affects many genes, but these genetic variants appear to organize into relatively few functional pathways (103, 104). In diabetes, many of the genes in the same pathway as the

transcriptional activator PGC-1α have independently been associated with diabetes (105). These results suggest that it may not be possible to fully understand such conditions except in the context of a network of interacting elements.

Even for so-called single-gene disorders—diseases that are understood to be caused by a single mutation of a single gene—the manifestations and severity of disease may depend on a network process. For example, cystic fibrosis is a congenital lung disease caused by a defect in the CFTR membrane protein channel, but the severity of the condition may depend on an associated miRNA regulatory network (106) and on the presence of disease-modifying genetic variants (107, 108). Disease-modifying variants and the influence of an individual's genetic background on disease expression are concepts from classical genetics that may be reframed in the context of network interactions between genes.

Network interactions between molecular contributors may also be measured as an epistatic effect, even when the involved pathways and interactions themselves are not known (109). Epistatic interaction is the contribution to a phenotype from interplay between multiple molecular partners (110). Epistatic effects are an important reason why molecular changes cannot always be studied in isolation from their network interactions: Interacting molecules may modulate the relative impact of their binding partners. Interactive epistatic effects on disease phenotype highlight a need for network analysis to understand disease pathogenesis—in cases where the source of these interactive effects between molecules is not known, subsequent identification through a systems-based analysis may be possible (111).

Network-based analyses have revealed shared molecular pathway alterations among diseases that were once thought distinct. Calcium channel pathway mutations are shared by five different psychiatric conditions: autism spectrum disorder, attention deficit-hyperactivity disorder, bipolar disorder, major depressive disorder, and schizophrenia (112). Cancers that are thought to be distinct based on organ system may share similar underlying gene and pathway alterations (31). This overlap of molecular phenotype among diseases that were once thought distinct may change how we think of disease and diagnosis. Rather than relying on established disease definitions, our understanding of disease may be shaped by a network definition of disease. Relationships between diseases may be better understood in the context of a global diseaseome (113).

Knowledge of molecular network architecture in health and disease may also lead to disease treatment. A network approach to drug discovery allows researchers to identify new target molecules through their network interactions and to minimize side effects by identifying the relationships among interacting molecules (114). The goal of multidrug therapy is to address the multiple networked molecular contributors to disease and has led to successful management of HIV, depression, and some forms of cancer (115–118). In addition to pharmacotherapy, bioengineering of interaction networks may be able to restore function to patients with certain diseases. For instance, an engineered gene network restored thyroid function in a mouse model of toxic diffuse goiter, commonly known as Graves disease (119).

4.2. Network Techniques in Cancer Genomics

Molecular networks have a particular relevance to cancer biology. Using a pathway- or network-based approach to analyzing mutational patterns, cancer types may be redefined or subcategorized. This approach, when performed as part of a broad molecular profiling strategy, has defined novel cancer subtypes for many cancers, including breast cancer (120), lung cancer (121), and kidney cancer (122). Significantly, the only route to diagnosis of metastatic cancer of unknown primary origin may be through analysis of the patterns of activity and cross talk defined by molecular profiling (123).

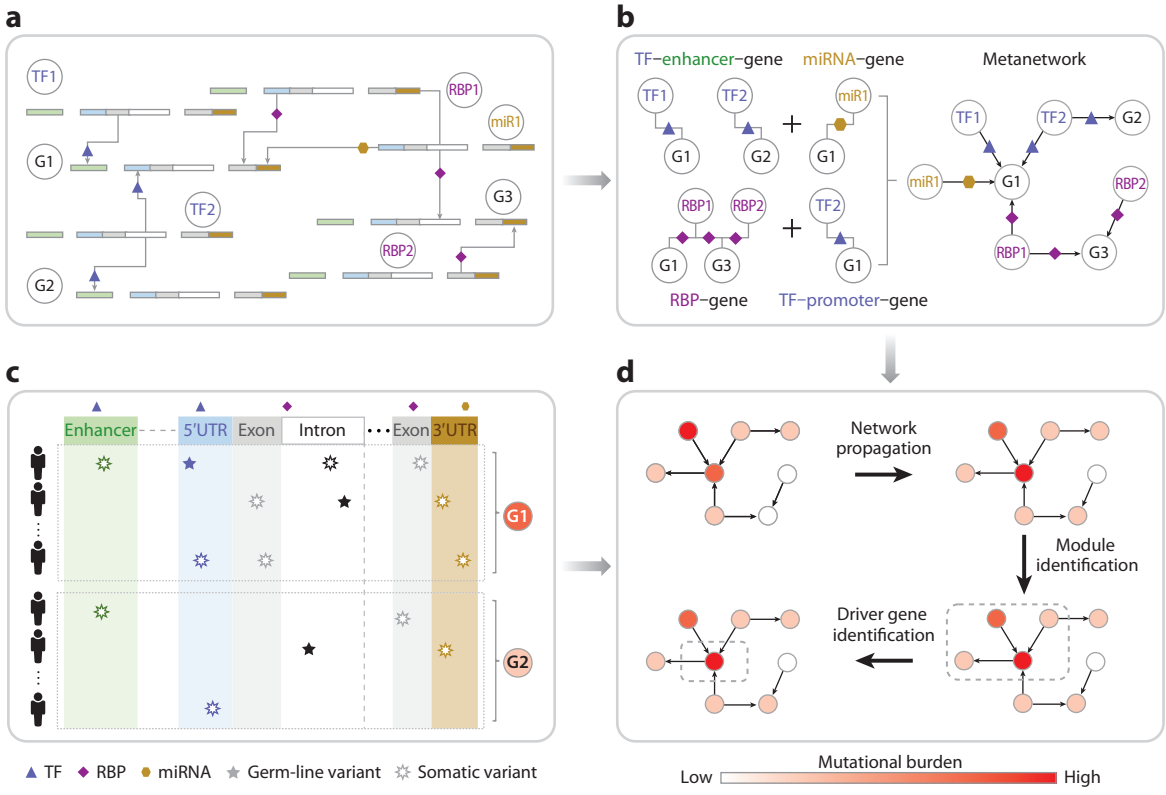


Figure 8

Cancer gene networks. (*a*,*b*) Gene interactions from multiple regulatory levels may be integrated together to form a metanetwork. (*c*) By pooling variants from multiple patients and mapping these mutations to extended gene regulatory regions, an aggregated mutational burden score can be defined. (*d*) Through techniques like network propagation, highly mutated subnetworks and key genes can be identified. Abbreviations: G, gene; miRNA, microRNA; RBP, RNA-binding protein; TF, transcription factor; UTR, untranslated region.

Regulatory networks may provide deep functional annotations to more accurately evaluate mutation impact and prioritize key mutations in cancer. For example, network centrality information has been used by researchers to pinpoint key cancer mutations (124, 125). TF and RBP networks may also provide insights to explain disease-specific expression patterns and help highlight key cancer regulators. For instance, by combining large-scale expression profiles from cancer patients with TF networks identified by ChIP-sequencing, it is possible to identify important TFs that drive tumor-to-normal differential expression (126, 127).

Integration of diverse sources of biological network data may be used to reveal novel cancer biology. Recent sequencing technologies have shown that key cancer-driving mutations are usually distributed across many regions of the genome (3). Integration of TF–gene, RBP–gene, miRNA–gene, and PPI data has been used obtain a systems-level view of cancer, highlighting key genes and mutations associated with tumorigenesis (**Figure 8*a*,*b***). In particular, by pooling data from multiple patients across extended gene regulatory regions, we can define a gene-level aggregated mutation effect (**Figure 8*c*,*d***). This reflects the overall mutation burden affecting each gene. Following this approach, many methods, including network propagation techniques, integrate mutational burden scores across multiple molecular networks and identify highly mutated pathways or subnetworks (2, 92, 128–130).

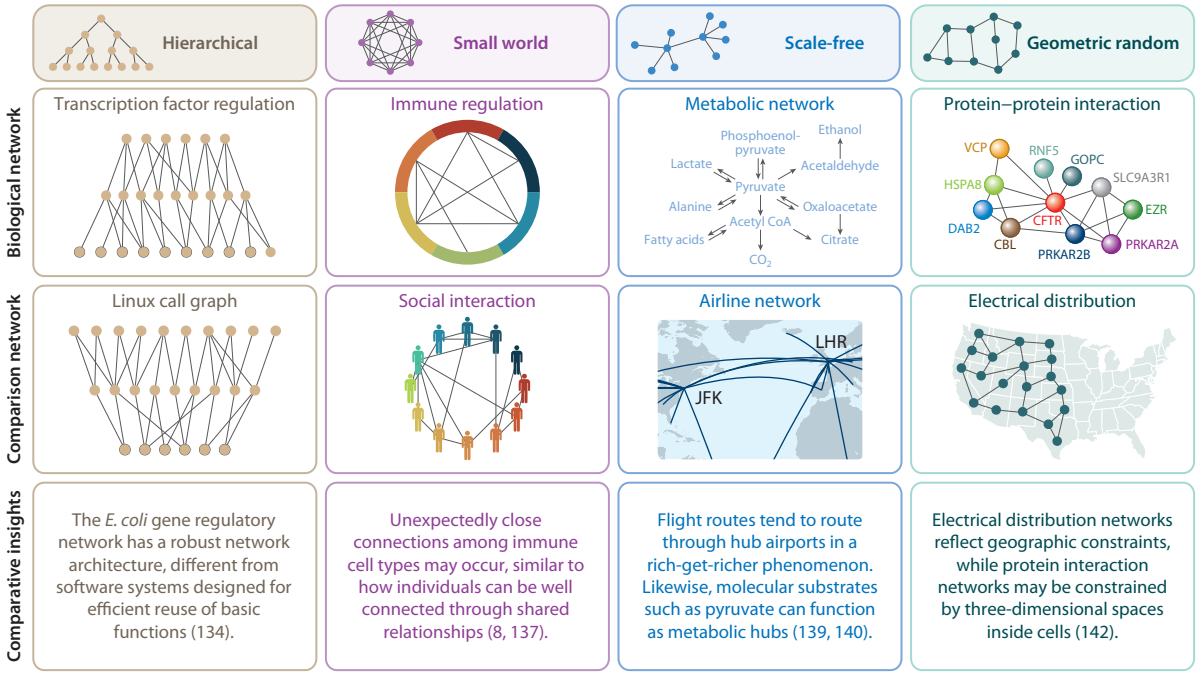


Figure 9

Cross-disciplinary network comparisons. By comparing networks across disciplines, we may learn more about the structure and function of both biological and human-made networks. For example, by comparing airline flight routes to the human metabolic network, we have learned that both follow a scale-free distribution (139, 140). A similar rich-get-richer evolutionary process may apply to both networks. Just as flight options are most easily expanded by connecting to an already well-connected airport, pyruvate and acetyl CoA (acetyl coenzyme A) may function as hub metabolites, facilitating molecular transitions between biochemical pathways.

Molecular network discovery may yield new cancer therapies. PD-L1 is a protein that helps regulate the body's immune response to cancer cell surface markers (131). It is the target for several cancer immunotherapies. There is interest in the protein CMTM6 because it has been shown to interact with PD-L1 and regulate its expression (132). Thus, perhaps CMTM6 will prove useful as a target for drug development. Knowledge of such pathways may result in the development of new cancer therapies and combination drug therapies that reduce the risk of developed resistance to cancer treatments (131, 133).

4.3. Cross-Disciplinary Comparisons Provide Insights into Molecular Interaction Networks

We may learn more about the mechanisms and function of molecular networks through cross-disciplinary comparison to networks found in other natural and human-made systems (**Figure 9**). The comparison of networks may reveal the evolutionary pressures that shape complex biology. Network attributes that vary among biological and nonbiological networks may highlight functional network architectures. Through such comparisons, advantages in biological network architecture may be identified and used to improve human-engineered systems through biomimicry.

A comparison of the transcriptional interaction network of the bacteria *E. coli* to the call graph of the Linux operating system demonstrated that the transcriptional network in *E. coli* has a robust architecture, with many network elements sharing overlapping functions (134). Conversely,

the Linux call graph is built on frequent reuse of many basic operating functions. An analysis of biological protein–DNA and protein–protein interactions in both *Saccharomyces cerevisiae* and *E. coli* to internet connectivity networks also favored the robustness of the biological networks (135).

Rieckmann et al. (8) recently conceptualized the human immune system as a social network. By mapping a social network architecture based on cytokine messages between cells, these researchers demonstrated unexpectedly close relationships between immune cell types. For example, neutrophils and naïve B cells were unexpectedly closely related, as were natural killer cells and memory T cells (136). It is intriguing to think that the discovered proximity of relationships in this small-world network may reflect how immune cells interact within the compartments of the human body (137).

Metabolic networks have been described as a type of scale-free network, meaning that the network is self-similar at each scale, with the degree of nodes following a power law. Air transport networks also have a network architecture that is classically described as scale free. Airports with many connecting flights are likely to gain additional flight routes due to the increase in travel options gained by connecting through a network hub (138, 139). This rich-get-richer process is thought to result in a scale-free network distribution. Metabolism appears organized around two central molecular hubs, pyruvate and acetyl-CoA (acetyl coenzyme A) (140). Just as flight options are most easily expanded by connecting to an already well-connected airport, pyruvate and acetyl-CoA may function as hub metabolites, facilitating molecular transitions between biochemical pathways.

Like metabolic networks, PPIs are also often thought of as scale-free networks, following this same rich-get-richer principle (141). However, researchers have also suggested that PPIs may be more similar to geometric networks based on their network topology (142). Examples of geometric networks include electrical grids connected based on the existing geographies of cities and wireless mesh networks connecting electronic devices based on spatial proximity. The observation that PPIs appear to have geometric network topology may be due to the spatial organization of molecules within the cell determining their interactions (142, 143). Geometric constraints within cells may also provide bio-inspired templates for efficient generation of geometric graphs. Such a possibility was demonstrated by comparing the growth of the single-celled organism *Physarum plasmodium* to the rail system in Tokyo (144).

5. DISCUSSION

We hope to have given the reader a sense of the strategic significance of network analysis techniques and interaction networks. We are convinced that because molecular interaction networks are the lowest common denominator in many higher-order biological systems, network analysis techniques will be a critical component of future advances in molecular biology and medicine. We further believe that there will be cross-disciplinary advantages to the investigation of molecular interaction networks, propelled by the need to adopt new network techniques to analyze large data sets and by the need to integrate diverse sources of biological data.

SUMMARY POINTS

1. Molecular interaction networks represent the base layer of function for many higher-order biological systems and have contributed to the development of biology, medicine, and data science (Sections 1.1 and 1.2).

2. Although large-scale molecular networks may at first appear uninterpretable, they can be understood in several straightforward ways. Complex networks can be understood by (*a*) focusing on some portion of the full network, (*b*) computing summary statistics about the network, or (*c*) comparing with other networks, including cross-disciplinary comparisons (Section 1.3).
3. Abstract network representations provide a useful platform for modeling network behavior (Section 2.1); however, not all interactions can be inferred without molecular structural information (Section 2.2).
4. The time dependency and computational capacity of interaction networks provide ways of maintaining homeostasis, and these same networks may also serve as the sensors and drivers of common diseases (Sections 2.3, 2.4, 4.1, and 4.2). In this review, we have given special emphasis to network applications in cancer genomics (Section 4.2).
5. New algorithms for understanding molecular interactions have revealed novel molecular relationships. Network prediction techniques, including deep learning models, may identify novel network structures through sophisticated pattern recognition performed on markers of molecular interaction (Section 3.1). Network propagation algorithms amplify important associations between molecules through a diffusion-like process (Section 3.2).
6. Related to our discussion of network algorithms, we observed that there is challenge in performing useful network comparisons and identifying causal network properties (Section 3.3). Networks can be compared to a null model of interaction (a random generative process) or to other biological or nonbiological networks.
7. Many disease processes arise through pathway or network phenomena and require an analysis of network properties to understand their pathology and identify treatment strategies (Sections 4.1 and 4.2).
8. The use of molecular interaction networks to make cross-disciplinary comparisons has led to greater understanding of networks in wide-ranging fields of study (Section 4.3).

FUTURE ISSUES

1. The identification of appropriate null comparisons for molecular interaction networks remains a challenge. Possible null comparisons include random network rewiring, random generative processes, and cross-disciplinary network analogies.
2. There is increasing opportunity to derive novel insight by incorporating three-dimensional structure and time dependency (e.g., network logic, network rewiring) into network models.
3. Recently popularized network algorithms that include machine learning techniques and network propagation methods will provide greater refinement to network predictions.
4. It will be important to design efficient, scalable algorithms for large search spaces that provide accurate approximations of actual network properties. Likewise, there is a need to define scalable approaches for integrating diverse molecular data sets, including functional genomics data.

5. We will see increasing use of network techniques in translational research and in application to clinical medicine. Network techniques will be used to analyze clinical data and identify correlations among clinical phenotypes. Redefinition of disease by molecular phenotype and molecular pathology will require substantial pathway and network analysis.
6. Experimentation with network engineering and network intervention in disease has the potential to yield new disease treatments.
7. Cross-disciplinary network science efforts will gain importance, such as molecular epidemiology (e.g., intersection of social networks, molecular networks, and epidemiology) and molecular phenotypic pathology (e.g., intersection of pathology and molecular networks).
8. The predictions of network analyses will require appropriate validations on a genomic scale.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This work was supported by funding from the National Human Genome Research Institute of the National Institutes of Health (grant number 5R01HG008126-02), by a National Institutes of Health Medical Scientist Training Program Training Grant (grant number T32GM007205), and by the AL Williams Professorship funds.

LITERATURE CITED

1. Hassabis D, Kumaran D, Summerfield C, Botvinick M. 2017. Neuroscience-inspired artificial intelligence. *Neuron* 95(2):245–58
2. Leiserson MDM, Vandin F, Wu H-T, Dobson JR, Eldridge JV, et al. 2015. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat. Genet.* 47(2):106–14
3. Vandin F, Upfal E, Raphael BJ. 2011. Algorithms for detecting significantly mutated pathways in cancer. *J. Comput. Biol.* 18(3):507–22
4. Pearl J. 1982. Reverend Bayes on inference engines: a distributed hierarchical approach. *Proc. AAAI Conf. Artif. Intell., 2nd, Pittsburgh, Pa., 18–20 Aug.*, pp. 133–36. Menlo Park, CA: AAAI
5. Alipanahi B, Delong A, Weirauch MT, Frey BJ. 2015. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat. Biotechnol.* 33(8):831–38
6. Krizhevsky A, Sutskever I, Hinton GE. 2012. ImageNet classification with deep convolutional neural networks. *Proc. Int. Conf. Neural Inf. Process. Syst., 25th, Lake Tahoe, Nev., 3–6 Dec.*, ed. F Pereira, CJC Burges, L Bottou, KQ Weinberger, pp. 1097–105. Red Hook, NY: Curran Assoc.
7. Yan K-K, Fang G, Bhardwaj N, Alexander RP, Gerstein M. 2010. Comparing genomes to computer operating systems in terms of the topology and evolution of their regulatory control networks. *PNAS* 107(20):9186–91
8. Rieckmann JC, Geiger R, Hornburg D, Wolf T, Kveler K, et al. 2017. Social network architecture of human immune cells unveiled by quantitative proteomics. *Nat. Immunol.* 18(5):583–93

9. Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan K-K, et al. 2012. Architecture of the human regulatory network derived from ENCODE data. *Nature* 489(7414):91–100
10. Khurana E, Fu Y, Colonna V, Mu XJ, Kang HM, et al. 2013. Integrative annotation of variants from 1092 humans: application to cancer genomics. *Science* 342(6154):1235587
11. Thiele I, Swainston N, Fleming RMT, Hoppe A, Sahoo S, et al. 2013. A community-driven global reconstruction of human metabolism. *Nat. Biotechnol.* 31(5):419–25
12. Venkatakrishnan AJ, Deupi X, Lebon G, Tate CG, Schertler GF, Babu MM. 2013. Molecular signatures of G-protein-coupled receptors. *Nature* 494(7436):185–94
13. Manglik A, Kim TH, Masureel M, Altenbach C, Yang Z, et al. 2015. Structural insights into the dynamic process of β_2-adrenergic receptor signaling. *Cell* 161(5):1101–11
14. Rosenbaum DM, Cherezov V, Hanson MA, Rasmussen SGF, Thian FS, et al. 2007. GPCR engineering yields high-resolution structural insights into β_2-adrenergic receptor function. *Science* 318(5854):1266–73
15. Luscombe NM, Babu MM, Yu H, Snyder M, Teichmann SA, Gerstein M. 2004. Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* 431(7006):308–12
16. Yosef N, Shalek AK, Gaublomme JT, Jin H, Lee Y, et al. 2013. Dynamic regulatory network controlling TH17 cell differentiation. *Nature* 496(7446):461–68
17. Koren O, Goodrich JK, Cullender TC, Spor A, Laitinen K, et al. 2012. Host remodeling of the gut microbiome and metabolic changes during pregnancy. *Cell* 150(3):470–80
18. Forslund K, Hildebrand F, Nielsen T, Falony G, Le Chatelier E, et al. 2015. Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature* 528(7581):262–66
19. Theriot CM, Koenigsknecht MJ, Carlson PE, Hatton GE, Nelson AM, et al. 2014. Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to *Clostridium difficile* infection. *Nat. Commun.* 5:3114
20. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* 518(7539):317–30
21. Schadt EE. 2009. Molecular networks as sensors and drivers of common human diseases. *Nature* 461(7261):218–23
22. Mangan S, Alon U. 2003. Structure and function of the feed-forward loop network motif. *PNAS* 100(21):11980–85
23. Tu S, Pederson T, Weng Z. 2013. Networking development by Boolean logic. *Nucleus* 4(2):89–91
24. Peter IS, Faure E, Davidson EH. 2012. Predictive computation of genomic logic processing functions in embryonic development. *PNAS* 109(41):16434–42
25. Moon TS, Lou C, Tamsir A, Stanton BC, Voigt CA. 2012. Genetic programs constructed from layered logic gates in single cells. *Nature* 491(7423):249–53
26. Fenno LE, Mattis J, Ramakrishnan C, Hyun M, Lee SY, et al. 2014. Targeting cells with single vectors using multiple-feature Boolean logic. *Nat. Methods* 11(7):763–72
27. Leeper NJ, Kullo IJ, Cooke JP. 2012. Genetics of peripheral artery disease. *Circulation* 125(25):3220–28
28. Sawa A, Snyder SH. 2002. Schizophrenia: diverse approaches to a complex disease. *Science* 296(5568):692–95
29. Ripke S, Neale BM, Corvin A, Walters JTR, Farh K-H, et al. 2014. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511(7510):421–27
30. Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, et al. 2016. The genetic architecture of type 2 diabetes. *Nature* 536(7614):41–47
31. Hoadley KA, Yau C, Wolf DM, Cherniack AD, Tamborero D, et al. 2014. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* 158(4):929–44
32. Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, et al. 2015. Prospective validation of a 21-gene expression assay in breast cancer. *N. Engl. J. Med.* 373(21):2005–14
33. Mate SE, Kugelman JR, Nyenswah TG, Ladner JT, Wiley MR, et al. 2015. Molecular evidence of sexual transmission of Ebola virus. *N. Engl. J. Med.* 373(25):2448–54
34. Grubaugh ND, Ladner JT, Kraemer MUG, Dudas G, Tan AL, et al. 2017. Genomic epidemiology reveals multiple introductions of Zika virus into the United States. *Nature* 546(7658):401–5

35. Faria NR, Azevedo RSS, Kraemer MUG, Souza R, Cunha MS, et al. 2016. Zika virus in the Americas: early epidemiological and genetic findings. *Science* 352(6283):345–49
36. Collins FS, Morgan M, Patrinos A. 2003. The Human Genome Project: lessons from large-scale biology. *Science* 300(5617):286–90
37. Chuang H-Y, Hofree M, Ideker T. 2010. A decade of systems biology. *Annu. Rev. Cell Dev. Biol.* 26:721–44
38. Monk J, Nogales J, Palsson BO. 2014. Optimizing genome-scale network reconstructions. *Nat. Biotechnol.* 32(5):447–52
39. Magnúsdóttir S, Heinken A, Kutt L, Ravcheev DA, Bauer E, et al. 2017. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat. Biotechnol.* 35(1):81–89
40. Huttlin EL, Bruckner RJ, Paulo JA, Cannon JR, Ting L, et al. 2017. Architecture of the human interactome defines protein communities and disease networks. *Nature* 545(7655):505–9
41. Marx V. 2015. The DNA of a nation. *Nature* 524(7566):503–5
42. Natl. Human Genome Res. Inst. 2016. *NIH genome sequencing program targets the genomic bases of common, rare disease*. News Release, Jan. 14, updated Sept. 3, Natl. Inst. Health, Washington, DC. **https://www.genome.gov/27563453/**
43. Erlich Y. 2015. A vision for ubiquitous sequencing. *Genome Res.* 25(10):1411–16
44. Shendure J, Aiden EL. 2012. The expanding scope of DNA sequencing. *Nat. Biotechnol.* 30(11):1084–94
45. Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, et al. 2015. Big Data: astronomical or genomical? *PLOS Biol.* 13(7):e1002195
46. Yan K-K, Wang D, Sethi A, Muir P, Kitchen R, et al. 2016. Cross-disciplinary network comparison: matchmaking between hairballs. *Cell Syst.* 2(3):147–57
47. Pržulj N, Malod-Dognin N. 2016. Network analytics in the age of big data. *Science* 353(6295):123–24
48. Benson AR, Gleich DF, Leskovec J. 2016. Higher-order organization of complex networks. *Science* 353(6295):163–66
49. Schatz MC. 2012. Computational thinking in the era of big data biology. *Genome Biol.* 13(11):177
50. Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, et al. 2014. Similarity network fusion for aggregating data types on a genomic scale. *Nat. Methods* 11(3):333–37
51. Drost M, Zonneveld JBM, van Dijk L, Morreau H, Tops CM, et al. 2010. A cell-free assay for the functional analysis of variants of the mismatch repair protein MLH1. *Hum. Mutat.* 31(3):247–53
52. Letai A. 2017. Functional precision cancer medicine—moving beyond pure genomics. *Nat. Med.* 23(9):1028–35
53. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45(D1):D353–61
54. Travers J, Milgram S. 1969. An experimental study of the small world problem. *Sociometry* 32(4):425–43
55. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, et al. 2005. A human protein-protein interaction network: a resource for annotating the proteome. *Cell* 122(6):957–68
56. Kim PM, Lu LJ, Xia Y, Gerstein MB. 2006. Relating three-dimensional structures to protein networks provides evolutionary insights. *Science* 314(5807):1938–41
57. Kim PM, Sboner A, Xia Y, Gerstein M. 2008. The role of disorder in interaction networks: a structural analysis. *Mol. Syst. Biol.* 4:179
58. Bhardwaj N, Abyzov A, Clarke D, Shou C, Gerstein MB. 2011. Integration of protein motions with molecular networks reveals different mechanisms for permanent and transient interactions. *Protein Sci.* 20(10):1745–54
59. Kumar S, Clarke D, Gerstein M. 2016. Localized structural frustration for evaluating the impact of sequence variants. *Nucleic Acids Res.* 44(21):10062–73
60. Kim M-S, Kim J-R, Cho K-H. 2010. Dynamic network rewiring determines temporal regulatory functions in *Drosophila melanogaster* development processes. *BioEssays* 32(6):505–13
61. Han J-DJ, Bertin N, Hao T, Goldberg DS, Berriz GF, et al. 2004. Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature* 430(6995):88–93
62. Nocedal I, Johnson AD. 2015. How transcription networks evolve and produce biological novelty. *Cold Spring Harb. Symp. Quant. Biol.* 80:265–74

63. Borneman AR, Gianoulis TA, Zhang ZD, Yu H, Rozowsky J, et al. 2007. Divergence of transcription factor binding sites across related yeast species. *Science* 317(5839):815–19
64. Schmidt D, Wilson MD, Ballester B, Schwalie PC, Brown GD, et al. 2010. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* 328(5981):1036–40
65. Shou C, Bhardwaj N, Lam HYK, Yan K-K, Kim PM, et al. 2011. Measuring the evolutionary rewiring of biological networks. *PLOS Comput. Biol.* 7(1):e1001050
66. Kim J, Kim I, Han SK, Bowie JU, Kim S. 2012. Network rewiring is an important mechanism of gene essentiality change. *Sci. Rep.* 2(1):900
67. Baumstark R, Hänzelmann S, Tsuru S, Schaerli Y, Francesconi M, et al. 2015. The propagation of perturbations in rewired bacterial gene networks. *Nat. Commun.* 6:10105
68. Creixell P, Schoof EM, Simpson CD, Longden J, Miller CJ, et al. 2015. Kinome-wide decoding of network-attacking mutations rewiring cancer signaling. *Cell* 163(1):202–17
69. Creixell P, Palmeri A, Miller CJ, Lou HJ, Santini CC, et al. 2015. Unmasking determinants of specificity in the human kinome. *Cell* 163(1):187–201
70. Grechkin M, Logsdon BA, Gentles AJ, Lee S-I. 2016. Identifying network perturbation in cancer. *PLOS Comput. Biol.* 12(5):e1004888
71. Bhardwaj N, Kim PM, Gerstein MB. 2010. Rewiring of transcriptional regulatory networks: Hierarchy, rather than connectivity, better reflects the importance of regulators. *Sci. Signal.* 3(146):ra79
72. Wang D, Yan K-K, Sisu C, Cheng C, Rozowsky J, et al. 2015. Loregic: a method to characterize the cooperative logic of regulatory factors. *PLOS Comput. Biol.* 11(4):e1004132
73. Prill RJ, Iglesias PA, Levchenko A. 2005. Dynamic properties of network motifs contribute to biological network organization. *PLOS Biol.* 3(11):e343
74. Rosenfeld N, Elowitz MB, Alon U. 2002. Negative autoregulation speeds the response times of transcription networks. *J. Mol. Biol.* 323(5):785–93
75. Alon U. 2007. Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* 8(6):450–61
76. Ekman D, Light S, Björklund AK, Elofsson A. 2006. What properties characterize the hub proteins of the protein-protein interaction network of *Saccharomyces cerevisiae*? *Genome Biol.* 7(6):R45
77. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, et al. 2015. STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43(D1):D447–52
78. Dandekar T, Snel B, Huynen M, Bork P. 1998. Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem. Sci.* 23(9):324–28
79. Clauset A, Moore C, Newman MEJ. 2008. Hierarchical structure and the prediction of missing links in networks. *Nature* 453(7191):98–101
80. Jansen R, Yu H, Greenbaum D, Kluger Y, Krogan NJ, et al. 2003. A Bayesian networks approach for predicting protein-protein interactions from genomic data. *Science* 302(5644):449–53
81. Zhang QC, Petrey D, Deng L, Qiang L, Shi Y, et al. 2012. Structure-based prediction of protein–protein interactions on a genome-wide scale. *Nature* 490(7421):556–60
82. Zahiri J, Bozorgmehr J, Masoudi-Nejad A. 2013. Computational prediction of protein–protein interaction networks: algorithms and resources. *Curr. Genom.* 14(6):397–414
83. Angermueller C, Pärnamaa T, Parts L, Stegle O. 2016. Deep learning for computational biology. *Mol. Syst. Biol.* 12(7):878
84. Park Y, Kellis M. 2015. Deep learning for regulatory genomics. *Nat. Biotechnol.* 33(8):825–26
85. Stergachis AB, Neph S, Sandstrom R, Haugen E, Reynolds AP, et al. 2014. Conservation of trans-acting circuitry during mammalian regulatory evolution. *Nature* 515(7527):365–70
86. Lanchantin J, Singh R, Lin Z, Qi Y. 2016. Deep Motif: visualizing genomic sequence classifications. arXiv:1605.01133 [cs.LG]
87. Qin Q, Feng J. 2017. Imputation for transcription factor binding predictions based on deep learning. *PLOS Comput. Biol.* 13(2):e1005403
88. Kelley DR, Snoek J, Rinn JL. 2016. Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks. *Genome Res.* 26(7):990–99
89. Cowen L, Ideker T, Raphael BJ, Sharan R. 2017. Network propagation: a universal amplifier of genetic associations. *Nat. Rev. Genet.* 18(9):551–62

90. Ruffalo M, Koyutürk M, Sharan R. 2015. Network-based integration of disparate omic data to identify "silent players" in cancer. *PLOS Comput. Biol.* 11(12):e1004595
91. Erten S, Bebek G, Koyutürk M. 2011. Vavien: an algorithm for prioritizing candidate disease genes based on topological similarity of proteins in interaction networks. *J. Comput. Biol.* 18(11):1561–74
92. Hofree M, Shen JP, Carter H, Gross A, Ideker T. 2013. Network-based stratification of tumor mutations. *Nat. Methods* 10(11):1108–15
93. Kim Y-A, Cho D-Y, Przytycka TM. 2016. Understanding genotype-phenotype effects in cancer via network approaches. *PLOS Comput. Biol.* 12(3):e1004747
94. Mazza A, Klockmeier K, Wanker E, Sharan R. 2016. An integer programming framework for inferring disease complexes from network data. *Bioinformatics* 32(12):i271–77
95. Nakka P, Raphael BJ, Ramachandran S. 2016. Gene and network analysis of common variants reveals novel associations in multiple complex diseases. *Genetics* 204(2):783–98
96. Pastor-Satorras R, Smith E, Solé RV. 2003. Evolving protein interaction networks through gene duplication. *J. Theor. Biol.* 222(2):199–210
97. Orsini C, Dankulov MM, Colomer-de-Simón P, Jamakovic A, Mahadevan P, et al. 2015. Quantifying randomness in real networks. *Nat. Commun.* 6:8627
98. Piel FB, Steinberg MH, Rees DC. 2017. Sickle cell disease. *N. Engl. J. Med.* 376(16):1561–73
99. Dempfle A, Scherag A, Hein R, Beckmann L, Chang-Claude J, Schäfer H. 2008. Gene-environment interactions for complex traits: definitions, methodological requirements and challenges. *Eur. J. Hum. Genet.* 16(10):1164–72
100. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, et al. 2014. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* 46(11):1173–86
101. Leiserson MD, Eldridge JV, Ramachandran S, Raphael BJ. 2013. Network analysis of GWAS data. *Curr. Opin. Genet. Dev.* 23(6):602–10
102. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, et al. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *PNAS* 102(43):15545–50
103. Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, et al. 2014. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am. J. Hum. Genet.* 94(5):677–94
104. Krishnan A, Zhang R, Yao V, Theesfeld CL, Wong AK, et al. 2016. Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. *Nat. Neurosci.* 19(11):1454–62
105. Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, et al. 2003. PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* 34(3):267–73
106. Ramachandran S, Karp PH, Jiang P, Ostedgaard LS, Walz AE, et al. 2012. A microRNA network regulates expression and biosynthesis of wild-type and ΔF508 mutant cystic fibrosis transmembrane conductance regulator. *PNAS* 109(33):13362–67
107. Guggino WB, Stanton BA. 2006. New insights into cystic fibrosis: molecular switches that regulate CFTR. *Nat. Rev. Mol. Cell Biol.* 7(6):426–36
108. Gu Y, Harley ITW, Henderson LB, Aronow BJ, Vietor I, et al. 2009. Identification of *IFRD1* as a modifier gene for cystic fibrosis lung disease. *Nature* 458(7241):1039–42
109. Hemani G, Shakhbazov K, Westra H-J, Esko T, Henders AK, et al. 2014. Detection and replication of epistasis influencing transcription in humans. *Nature* 508(7495):249–53
110. Carlborg Ö, Haley CS. 2004. Epistasis: too often neglected in complex trait studies? *Nat. Rev. Genet.* 5(8):618–25
111. Costanzo M, VanderSluis B, Koch EN, Baryshnikova A, Pons C, et al. 2016. A global genetic interaction network maps a wiring diagram of cellular function. *Science* 353(6306):aaf1420
112. Cross-Disord. Group Psychiatr. Genom. Consort. 2013. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* 381(9875):1371–79
113. Goh K-I, Cusick ME, Valle D, Childs B, Vidal M, Barabasi A-L. 2007. The human disease network. *PNAS* 104(21):8685–90

114. Csermely P, Korcsmáros T, Kiss HJM, London G, Nussinov R. 2013. Structure and dynamics of molecular networks: a novel paradigm of drug discovery. *Pharmacol. Ther.* 138(3):333–408
115. Mohamed S, Johnson GR, Chen P, Hicks PB, Davis LL, et al. 2017. Effect of antidepressant switching versus augmentation on remission among patients with major depressive disorder unresponsive to antidepressant treatment. *JAMA* 318(2):132–45
116. Lee MJ, Ye AS, Gardino AK, Heijink AM, Sorger PK, et al. 2012. Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. *Cell* 149(4):780–94
117. Keith CT, Borisy AA, Stockwell BR. 2005. Innovation: multicomponent therapeutics for networked systems. *Nat. Rev. Drug Discov.* 4(1):71–78
118. Günthard HF, Saag MS, Benson CA, del Rio C, Eron JJ, et al. 2016. Antiretroviral drugs for treatment and prevention of HIV infection in adults. *JAMA* 316(2):191–210
119. Saxena P, Charpin-El Hamri G, Folcher M, Zulewski H, Fussenegger M. 2016. Synthetic gene network restoring endogenous pituitary-thyroid feedback control in experimental Graves' disease. *PNAS* 113(5):1244–49
120. Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Kalicki-Veizer J, et al. 2012. Comprehensive molecular portraits of human breast tumours. *Nature* 490(7418):61–70
121. Collisson EA, Campbell JD, Brooks AN, Berger AH, Lee W, et al. 2014. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 511(7511):543–50
122. Cancer Genome Atlas Res. Netw., Linehan WM, Spellman PT, Ricketts CJ, Creighton CJ, et al. 2016. Comprehensive molecular characterization of papillary renal-cell carcinoma. *N. Engl. J. Med.* 374(2):135–45
123. Varadhachary GR, Raber MN. 2014. Cancer of unknown primary site. *N. Engl. J. Med.* 371(8):757–65
124. Khurana E, Fu Y, Chen J, Gerstein M. 2013. Interpretation of genomic variants using a unified biological network approach. *PLOS Comput. Biol.* 9(3):e1002886
125. Fu Y, Liu Z, Lou S, Bedford J, Mu XJ, et al. 2014. FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer. *Genome Biol.* 15(10):480
126. Jiang P, Freedman ML, Liu JS, Liu XS. 2015. Inference of transcriptional regulation in cancers. *PNAS* 112(25):7731–36
127. Falco MM, Bleda M, Carbonell-Caballero J, Dopazo J. 2016. The pan-cancer pathological regulatory landscape. *Sci. Rep.* 6(1):39709
128. Bashashati A, Haffari G, Ding J, Ha G, Lui K, et al. 2012. DriverNet: uncovering the impact of somatic driver mutations on transcriptional networks in cancer. *Genome Biol.* 13(12):R124
129. Jia P, Zhao Z. 2014. VarWalker: personalized mutation network analysis of putative cancer genes from next-generation sequencing data. *PLOS Comput. Biol.* 10(2):e1003460
130. Creighton CJ, Morgan M, Gunaratne PH, Wheeler DA, Gibbs RA, et al. 2013. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* 499(7456):43–49
131. Sharma P, Allison JP. 2015. The future of immune checkpoint therapy. *Science* 348(6230):56–61
132. Burr ML, Sparbier CE, Chan Y-C, Williamson JC, Woods K, et al. 2017. CMTM6 maintains the expression of PD-L1 and regulates anti-tumour immunity. *Nature* 549(7670):101–5
133. Zaretsky JM, Garcia-Diaz A, Shin DS, Escuin-Ordinas H, Hugo W, et al. 2016. Mutations associated with acquired resistance to PD-1 blockade in melanoma. *N. Engl. J. Med.* 375(9):819–29
134. Yan K-K, Fang G, Bhardwaj N, Alexander RP, Gerstein M. 2010. Comparing genomes to computer operating systems in terms of the topology and evolution of their regulatory control networks. *PNAS* 107(20):9186–91
135. Navlakha S, He X, Faloutsos C, Bar-Joseph Z. 2014. Topological properties of robust biological and computational networks. *J. R. Soc. Interface* 11(96):20140283
136. Bird L. 2017. Immune regulation: immune cell social networks. *Nat. Rev. Immunol.* 17(4):216
137. Bergthaler A, Menche J. 2017. The immune system as a social network. *Nat. Immunol.* 18(5):481–82
138. Guimerà R, Sales-Pardo M, Amaral LAN. 2007. Classes of complex networks defined by role-to-role connectivity profiles. *Nat. Phys.* 3(1):63–69
139. Guimerà R, Mossa S, Turtschi A, Amaral LAN. 2005. The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles. *PNAS* 102(22):7794–99

140. Guimerà R, Nunes Amaral LA. 2005. Functional cartography of complex metabolic networks. *Nature* 433(7028):895–900
141. Colizza V, Flammini A, Serrano MA, Vespignani A. 2006. Detecting rich-club ordering in complex networks. *Nat. Phys.* 2(2):110–15
142. Pržulj N, Corneil DG, Jurisica I. 2004. Modeling interactome: scale-free or geometric? *Bioinformatics* 20(18):3508–15
143. Wu Z, Menichetti G, Rahmede C, Bianconi G. 2015. Emergent complex network geometry. *Sci. Rep.* 5(1):10073
144. Tero A, Takagi S, Saigusa T, Ito K, Bebber DP, et al. 2010. Rules for biologically inspired adaptive network design. *Science* 327(5964):439–42
145. Newman MEJ. 2006. Modularity and community structure in networks. *PNAS* 103(23):8577–82
146. Freeman LC. 1977. A set of measures of centrality based on betweenness. *Sociometry* 40(1):35–41
147. Borgatti SP, Everett MG. 2006. A graph-theoretic perspective on centrality. *Soc. Netw.* 28(4):466–84
148. Newman MEJ, Girvan M. 2004. Finding and evaluating community structure in networks. *Phys. Rev. E* 69(2):26113
149. Tibshirani R, Walther G, Hastie T. 2001. Estimating the number of clusters in a data set via the gap statistic. *J. R. Stat. Soc. B* 63(2):411–23
150. Clauset A, Newman MEJ, Moore C. 2004. Finding community structure in very large networks. *Phys. Rev. E* 70(6):66111
151. Radicchi F, Castellano C, Cecconi F, Loreto V, Parisi D. 2004. Defining and identifying communities in networks. *PNAS* 101(9):2658–63
152. Rosvall M, Bergstrom CT. 2007. An information-theoretic framework for resolving community structure in complex networks. *PNAS* 104(18):7327–31
153. Palla G, Derényi I, Farkas I, Vicsek T. 2005. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435(7043):814–18
154. Donetti L, Muñoz MA. 2004. Detecting network communities: a new systematic and efficient algorithm. *J. Stat. Mech.* 2004(10):P10012
155. Ronhovde P, Nussinov Z. 2009. Multiresolution community detection for megascale networks by information-based replica correlations. *Phys. Rev. E* 80(1):16109
156. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. 2008. Fast unfolding of communities in large networks. *J. Stat. Mech.* 2008(10):P10008
157. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. 2008. *Molecular Biology of The Cell*. New York: Garland Sci. 5th ed.
158. Gnatt AL, Cramer P, Fu J, Bushnell DA, Kornberg RD. 2001. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* 292(5523):1876–82
159. Rose AS, Hildebrand PW. 2015. NGL Viewer: a web application for molecular visualization. *Nucleic Acids Res.* 43(W1):W576–79

Annual Review
of Biomedical Data
Science

Volume 1, 2018

Contents

Errata

An online log of corrections to *Annual Review of Biomedical Data Science* articles may be found at http://www.annualreviews.org/errata/biodatasci