9 The Network of Networks Involved in Human Disease

Celine Sin and Jörg Menche

9.1 Introduction

Human health and disease are influenced by an astounding complexity of intertwined processes that span many orders of magnitude in both space and time (Figure 9.1). At the molecular level, for example, the five primary nucleobases that form the basic units of the genetic code are relatively simple molecules consisting of only 12–16 atoms. Yet, lined up end to end, the human genome stretches over 3 meters in length, and it is these same nucleotides that encode all the processes of life from the simplest single-celled organisms up to the most complex animals. Some of the most fundamental chemical processes of life, such as the conversion of biochemical energy from nutrients into adenosine triphosphate (ATP), occur at the order of microseconds. On the other hand, decade-long processes such as malnutrition or environmental exposures are key to understanding some of the most common diseases, such as heart diseases or diabetes.

Over the last years, it has been increasingly recognized that the reductionist approach – investigating proteins or individual organs in isolation – is rarely sufficient for a complete understanding of pathobiological or potential therapeutic approaches [1]. Indeed, the many components of biological systems, from molecules to cells to complex organisms, interact in an intricate and tightly coordinated fashion. Disease states can be understood as perturbations out of balance from normal function, either from outside sources or abnormalities within the components or their interactions. While some diseases can be traced back to a single genetic defect, or invasion by specific microorganisms, such as viruses or bacteria, others develop from the combined effect of multiple factors. Contrary to the archetype of "one-gene, one-function, one-phenotype" proposed in the 1940s [2], it is increasingly evident that a particular gene abnormality not only impacts the activity of its particular gene product, but also trickles through the whole cascade of interactions between several sub- to supra-cellular processes [3]. In other words, a disease phenotype is rarely a consequence of an abnormality in a single gene, but reflective of a slew of perturbed pathobiological processes that interact in complex networks.

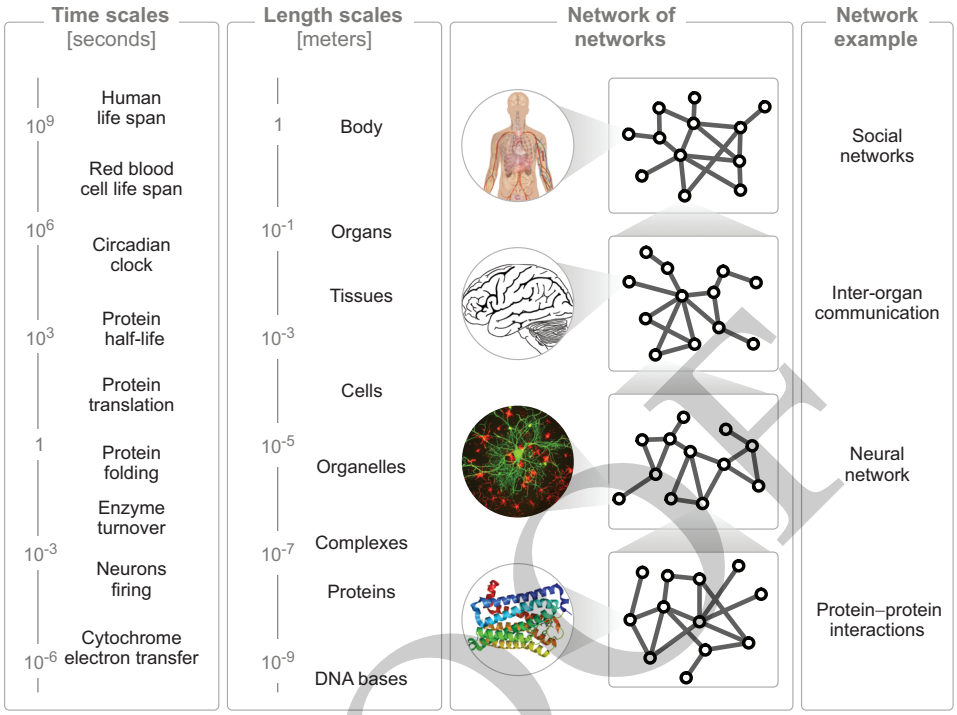

Figure 9.1 The network of networks in medicine. The processes involved in health and disease span a wide range of scales in both time and space. Different processes are relevant at different scales. These processes can be summarized by networks in which the actors are represented by nodes and their relationships as links. An entire network at one scale may be embedded as a single node into the next.

As the components relevant to health and disease span many orders of magnitudes, so do their interactions. We can conceptualize these multi-scale interactions as an embedding of networks into networks [4] (Figure 9.1). Proteins interact with each other in our cells, which in turn interact with other cells forming whole organs, and so on, all the way to social interactions among humans. Over the last two decades, these networks have been identified and mapped more and more accurately. Network science plays a crucial role in solving the next part of the puzzle, which is how to interpret and ultimately understand the functional, logical and dynamic aspects of how they contribute to diverse disease phenomena [5, 6]. This endeavor is inherently cross- and interdisciplinary, and network science can rely on a rich body of rigorous results from graph theory and statistical physics, as well as tools and concepts from computer science and sociology.

In the following, we will first give an overview of the most important biological network layers and their uses in biomedical research. We will then provide more details on how the emerging field of network medicine uses interactome networks as maps to study human disease. Finally, we will introduce a network-based methodology for elucidating the molecular mechanisms of specific diseases.

9.2 The Networks of Networks in Health and Disease

We can broadly classify the multitude of networks relevant in biomedical research into two main categories: (1) networks based on direct physical interactions and thereby representing the innate organization of many biological systems, such as the network of physical protein interactions or cell–cell contacts among neurons; (2) networks that serve as an analysis framework to investigate more abstract relationships, such as shared molecular mechanisms among diseases. With the increasing sophistication of -omic data collection, growing knowledge of biological processes, and access to population-wide health records, we are in an era of explosive data growth. In addition to being "big," these data are often noisy, and interpretation is often not straightforward. Networks provide a convenient and powerful toolbox to systematically analyze these complex systems in a comprehensive and holistic fashion.

9.2.1 Networks as Organizational Principle of Biological Systems

Following roughly the length scales of biological organization outlined in Figure 9.1 from smallest to largest, we start by discussing important molecular networks.

Metabolic Networks

Among the most comprehensive and best-studied physical interaction networks are metabolic networks. Metabolism is the collection of all the chemical processes in a cell, such as the conversion of one metabolite to the next. These processes provide both the energy and the building blocks that are essential for all life. Metabolic networks represent an integrated map of all these processes with metabolites as nodes (e.g. glucose, ATP, triglycerol), which are linked through reactions or enzymes converting one metabolite to the other. There are several curated databases of metabolic maps, such as the Human Recon 2.2 [7], containing 5324 metabolites, 7785 reactions and 1675 associated genes; the Edinburgh human metabolic network [8], with 3000 reactions and over 2000 associated genes; and the Kyoto Encyclopedia of Genes and Genomes (KEGG) [9], which also contains data on a wide range of other species.

Constructing context-dependent metabolic networks for different cell types or under different perturbations is particularly useful for the identification of essential pathways and prediction of cellular responses to different treatments. For example, a metabolic network of the human hepatocyte showed how the liver responds to the availability of different metabolites in order to maintain homeostasis in the blood [10].

Network analysis of metabolic networks from disease states is also informative to elucidate pathologies that result from deficiencies or excessive amplification of metabolic pathways. In [11], metabolic networks were supplemented with gene expression data to identify principle metabolic and regulatory nodes in type 2 diabetes mellitus. For different cancer cell lines, metabolic networks have been used to predict drug efficacy [12], and patient survivability [13].

Gene Regulatory Networks

Another essential dynamical process requiring tight regulation is gene expression. Compared to metabolism, our knowledge of gene regulatory processes is much

less complete. The most basic nodes in gene regulatory networks are transcription factors and their respective target DNA regulatory elements [14]. Network motifs can reveal important principles of regulatory mechanisms (e.g. autoregulation, feedforward loops) [15]. Commonly used databases containing experimentally verified genetic regulatory interactions include JASPAR [16] and TRANSFAC [17]. However, gene expression is not only regulated through transcription factors, but also through other interactions, such as between RNAs or RNA and DNA. For example, microRNA plays a considerable role in regulating mRNA concentrations by modulating mRNA stability and degradation [18]. Additionally, microRNA was shown to regulate pre-mRNA processing in the nucleus, assist in mRNA structure formation, and modulate mRNA–protein interactions [19]. Numerous microRNA are associated with the development of several diseases [20, 21]. Interactions between RNA and/or DNA species can be measured experimentally [22, 23] or predicted computationally [24, 25]. Databases containing computationally predicted interactions include TargetScan [25], PicTar [26], microRNA [27], miRWalk [28] and miRBase [29]; experimentally confirmed interactions are collected in TarBase [30] miRecords [31].

Protein–Protein Interaction Networks

The gene products that emerge from the regulatory process then engage in numerous physical interactions with each other to perform a multitude of molecular processes in the cell. Such protein–protein interactions are in general very specific, mediated through complementary "lock and key" interaction interfaces. As these interactions are integral to normal cell functioning, mutations affecting the interaction interfaces often have particularly strong effects.

Two main experimental techniques allow for systematic, large-scale mapping of protein–protein interactions: binding affinity purifications coupled to mass spectrometry [32, 33] and yeast two-hybrid (Y2H) assays [6, 34]. In the first method, cell lines are engineered to produce a "bait" protein with a tag that can be captured by beads. These proteins (and their interaction partners) are then captured and identified through mass spectrometry. Since often a whole complex is collected, translation of the data into direct pairwise interactions is often difficult. Yet, the interactions revealed by this method are specific to a particular biologically relevant condition. On the other hand, Y2H assays map out precise binary protein interactions. However, not all interactions found are biologically relevant; for example, two respective interacting proteins might never be expressed at the same time in the same cell [35]. Additionally, there are many small-scale studies using a myriad of methods, such as co-immunoprecipitation, X-ray crystallography or nuclear magnetic resonance. Computational predictions of protein interactions use features of amino acid sequences [36–39], gene fusion [40], or phylogenetic trees [41]. Each of these sources of protein interactions have strengths and limitations in terms of comprehensiveness, noise and biases [42], such as biases in the selection of protein pairs [43] or experimental biases, for example toward highly expressed genes [34].

There are several online depositories of protein–protein interaction data. EMBL-EBI maintains the manually curated "IntAct" molecular interaction database containing 57 857 proteins and 275 145 interactions from over 5000 publications [44, 45].

"Reactome" is another manually curated database containing also interactions between protein and nucleic acids, small molecules and macromolecules [46, 47]. The Human Integrated Protein Protein Interaction rEference (HIPPIE) collects data from several primary databases and offers a confidentiality score for each reported interaction [48]. The Search Tool for Recurring Instances of Neighboring Genes (STRING) database contains even more interactions by including predictions, for example based on co-expression or automated text mining [49].

Protein interaction networks have found numerous applications, ranging from elucidating basic principles of cellular organization [50] to the prediction of disease genes [51–53] or of the therapeutic effect of drugs [54] (see Sections 9.3 and 9.4).

Neuronal Networks

The networks considered so far occur, for the most part, within individual cells. An important example of inter-cellular networks in which different cells communicate with each other are networks of neurons. Neurons exchange signals either chemically, through the release of neurotransmitters at synapses, or electrically via the flow of ions through gap junctions. The first, and so far only, completely mapped network of all neural connections of an organism was published as early as 1986 for the roundworm *C. elegans* and contains around 9000 chemical and around 800 electrical connections between around 300 neurons [55]. Partial neural networks are available, for example, for mice [56, 57], and there are also first, very coarse-grained data sets available for humans [58]. The ultimate goal of mapping the complete human "connectome" of all our brain cells will likely remain out of reach for many years to come [59].

Network analysis of these connectomes has contributed to our understanding of basic principles of neural function [60] and cognition [61], as well as psychopathologies [62].

The Immune System

Similar to the nervous system that interfaces to many other body parts, but perhaps even more diverse in terms of participating organs, cell types and molecules, is the immune system. To meet the constant challenges of internal and external threats, ranging from tumor cells to bacterial infections, our immune system orchestrates a multitude of cells with often highly specialized functions. The nodes in the "social network" of the immune system therefore represent cells, links represent communication through signaling molecules, such as cell-surface receptors or secreted molecules [63]. This network can be mapped using mass spectrometry by measuring the proteomes of immune cell populations and comparing the levels of intracellular and secreted proteins between different stimuli. A recent study identified over 180 000 high-confidence interactions between 460 receptors and 300 ligands in such a fashion [64]. A network analysis revealed several principles of intra-cellular communication in the immune system. For example, lineages that are developmentally less related to each other tend to have a higher number of interactions and different immune cells exhibit pronounced differences in their communication patterns after being activated. Immune networks constructed from large-scale text mining have also been used to predict cytokine disease associations [65].

Population Networks

While the "social network" served merely as a metaphor in the case of the immune system, the actual relationships between humans are also important for a number of diseases, most notably for the spread of contagious diseases, such as viral or bacterial infections. The first mathematical models of diseases spreading among individuals of a population were formulated in 1760 by the Swiss mathematician and physicist Daniel Bernoulli [66]. As maps of the networks on which diseases propagate become available, in particular global transportation maps and networks of social interactions, these models are becoming increasingly accurate [67]. Network-based epidemiological models can help us understand global propagation patterns observed in recent pandemic outbreaks, identify the source of an outbreak, predict future highly affected areas or design effective immunization or prevention strategies [68, 69].

Interestingly, the spread of diseases through social contacts is not only limited to diseases that are transmitted through viruses or bacteria. It has been shown that also obesity [70], the tendency to start to smoke [71] or general happiness in life [72] may propagate along social connections between people.

9.2.2 Networks as Data Analysis Tools

All networks reviewed above build on direct, often physical relationships between entities ranging from molecules to people. However, networks can also be used to characterize more abstract, less direct relationships. In the following we introduce important examples of such networks that represent non-physical, yet still biologically highly relevant relationships.

Co-expression Networks

As mentioned above, our knowledge of gene regulatory networks remains scarce as they are highly context-dependent, involve a large number of diverse molecules and are therefore difficult to assess experimentally in a comprehensive fashion. A much more easily accessible quantity that may serve as a proxy to study gene regulatory programs is co-expression. Two genes are co-expressed if their respective expression levels correlate strongly under different experimental conditions, such as over time or under different stimuli. Genome-wide gene expression can be assessed using RNAseq technology, enabling the construction of large-scale co-expression networks [73, 74]. A comprehensive database of expression data across many tissues, cell types and conditions is collected and curated by the GTEx consortium [75]. In contrast to gene regulatory networks, co-expression networks do not imply a causal relationship between genes. Yet, co-expression networks can still be used to identify groups of genes that are more broadly functionally related – for example controlled by the same transcriptional regulatory program, or members of the same pathway or protein complex [76]. Analyses of these networks have identified commonly affected pathways in autism spectrum disorder [77], Alzheimer's disease [78] and inflammatory bowel disease [79]. Candidate biomarkers for myocardial infarction [80] and several cancers [81, 82] have also been identified using co-expression networks.

Genetic Interaction Networks

Another important indirect relationship between genes is given by genetic interactions. Most generally, these interactions describe the phenomena of observing an unexpected phenotype upon simultaneous mutations in two genes. More specifically, two genes are said to have a negative genetic interaction if mutations in the genes individually are not lethal, but become lethal when simultaneously mutated. Conversely, genes are said to have a positive genetic interaction if a mutation in one gene "rescues" a lethal mutation in another [83].

Genetic interactions can be evaluated by creating gene-deletion mutants for the genes of interest. Large-scale screens have been performed in yeast [83–85]. The most comprehensive screen in human haploid cells identified approximately 2000 essential genes, revealed genes regulating the secretory pathway and generated new insights into Golgi apparatus homeostasis [86]. There are also several more specialized screens in human cells, focusing on tumor suppressor genes [87, 88] and cancer drug targets [89], for example.

In addition to studying the large-scale functional organization of genes, genetic interactions also hold great promise for concrete therapeutic applications. For example, a positive genetic interaction with the BLM helicase complex was recently shown to rescue the Fanconi anemia (FA) phenotype caused by a loss of function mutation in the FA gene that leads to defective DNA damage repair [90]. Furthermore, genetic interaction networks also hold special promise for studying complex diseases, such as cancer, that result from a number of genetic mutations (and environmental factors, potentially) that impact several subcellular systems. In [88], the genetic interaction network was used to identify potential chemotherapeutic drug targets.

Co-perturbation Networks

The concept of genetic interactions can be generalized from gene inactivation to arbitrary perturbations. Co-perturbation networks thus encapsulate the information from perturbation biology screens, with nodes representing genes and edges again representing significant correlations among the response of the system toward perturbations in the two respective genes. Examples of such perturbations that are assessed in high-throughput measurements range from RNAi [91] or CRISPR [92] to drug treatment [93–95]. Commonly used readouts of the cellular response include gene expression through RNAseq technology [96] and high-resolution fluorescent microscopy [97]. Co-perturbation networks have been used, for example, to predict drug targets [96], elucidate molecular mechanisms of drugs [98, 99] or infer pathway activity from gene expression [100–102].

Disease Networks

Diseases, while having diverse causes, development and manifestations, often share a number of similar characteristics. These relationships among diseases may occur at several scales and can be systematically investigated using disease networks: on the molecular level (e.g. sharing common genetic origin), on the phenotypic level (e.g. sharing common clinical signs and symptoms) and on the population level (e.g. having frequent co-occurrence in patients). The first comprehensive map of the human "diseaseome" was presented in [103], linking 1377 diseases based on their shared genetic

associations as reported in the OMIM database [104]. The resultant network clearly showed that very few diseases could be regarded as isolated entities and directly attributed to a single distinct origin. Instead, the majority of diseases fall into highly connected clusters of disease groups, with overlapping molecular roots. The network properties within and around disease clusters are also predictive of disease characteristics: diseases occupying a more central position in the disease network tend to be more prevalent and have higher mortality rates [105].

Genetic overlap among diseases extends to the physical interactions among the respective gene products, to the resultant gene expression profiles, to the cells, to the organ systems, and eventually to the organism. Thus, it is a logical progression to expect similar findings at the phenotypic scale. In [106], a disease network based on the similarity of clinical symptoms was built using the annotated Medical Subject Headings (MeSH) metadata [107]. Indeed, two diseases sharing similar symptoms tended to also share similar defects on protein interactions, if not the genetic associations directly. The study further revealed that the degree of localization of the associated genes on the underlying protein interaction network is indicative of the diversity of clinical manifestations. That is, diseases that are more localized on the protein interaction network tend to have narrower clinical presentation. Comparison of protein interaction networks and disease networks from different classes of disease (e.g. complex diseases, Mendelian diseases, cancer) revealed interesting differences between diseases with difference inheritance modes [108–110].

Disease relationships at the population level can be evaluated through co-morbidity networks. Co-morbidity describes the tendency of certain diseases to co-occur in the same patient. Such data can be mined from patient records – in [111], a disease network built from 30 million patient records showed a relationship between disease progression patterns to topological properties of the respective disease. Highly central, highly connected diseases are associated with a higher mortality rate, and patients tend to be affected by peripheral diseases before developing more central ones. More recently, similar differences in disease progression patterns related to age and sex have been characterized [112]. Co-morbidity networks have been used to explore possibilities for drug repurposing [113], evaluate potential drug side-effects [114], identify biomarkers [115], and disentangle genetic and environmental factors of diseases [116].

9.3 Molecular Networks as Maps

Despite the large diversity of networks introduced in the previous section, they exhibit certain universal features and patterns (Figure 9.2). (1) The position of an individual node within the network is often related to its importance within the represented biological system. (2) The local connectivity among a group of nodes can be associated with shared (patho-)biological functions. (3) The network distance between groups of nodes often indicates their degree of relatedness. Taken together, these features form the basis of viewing molecular networks as maps. In this section, we review this important metaphor in more detail using the interactome as an example – that is, the integrated network of molecular interactions in the cell. The tools and concepts that we will introduce can be readily applied to other networks as well.

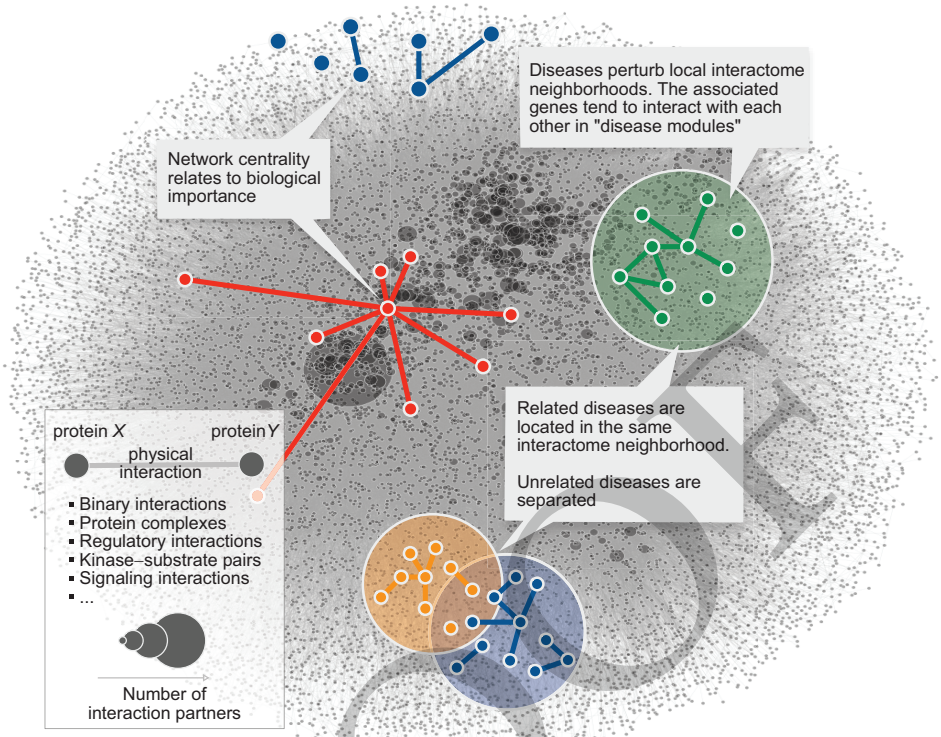

Figure 9.2 The interactome as a map. The interactome represents all biologically relevant molecular interactions in the cell. The protein–protein interaction network shown here contains 13 460 proteins connected by 141 296 interactions [117]. The annotations on top illustrate the basic findings that inspired the analogy between networks and maps.

9.3.1 Basic Interactome Properties

As introduced above, the various molecular interactomes can be broadly categorized into direct physical and more indirect functional relationships. Focusing on the physical interactions, in particular on protein–protein interactions, we can identify three main sources: (1) interactions curated from the scientific literature, mainly derived from specific, small-scale experiments; (2) interactions from systematic, proteome-scale mapping efforts; and (3) interactions from computational predictions. Figure 9.2 gives a visual impression of a manually curated interactome network focusing only on physical interactions with direct experimental evidence [117]. It contains 13 460 proteins connected by 141 296 interactions between them. On average, each protein has approximately 21 interaction partners, but in this network, as well as in many complex networks, the number of interaction partners per node ("degree," k) vary widely: while the majority of nodes have only a few neighbors (more than 2000 proteins have only one interaction partner), there are also a handful of nodes with hundreds of connections, such as *GRB2* (degree $k = 872$), *YWHAZ* ($k = 502$) and *TP53* ($k = 450$), the so-called "hubs." The heterogeneous distribution of degrees,

and in particular the existence of hubs, have a profound effect on many network properties. Hubs connect many distinct parts of the network, shortening the distance between nodes, also known as the "small world effect" [118]. In some cases of scale-free networks whose degree distribution approximates $P(k)\ k^{-\gamma}$, one can even observe "ultra small world effects" [119]. In the interactome, it takes an average of fewer than four hops ($\langle d \rangle = 3.6$) to move from any protein to any other protein. Networks that have such high degree of connectedness tend to be very resilient to random failure – that is, the structure of the network is preserved despite random removal of nodes or edges [120–123]. While these networks are robust against random failure, they are also particularly vulnerable to targeted attacks of the hubs [124]. In the interactome, the removal of the ∼30% of the most highly connected nodes is sufficient to completely destroy the network, leaving only disconnected fragments.

9.3.2 Node Localization in the Interactome

Given the vulnerability of such networks to attacks targeted toward highly connected nodes, we expect that proteins that serve as hubs in a biological network also have higher biological importance. Indeed, as first shown in yeast (*Saccharomyces cerevisiae*) [50] and later confirmed in human cell lines [125], the protein products of essential genes – the genes that are crucial for survival – tend to be hubs located toward the center of the interactome. Conversely, less essential genes tend to have fewer interactions and are situated more in the periphery of the interactome. These findings were later extended and refined for disease-associated genes, revealing specific topological properties that differ between classes of diseases (e.g. complex diseases, Mendelian diseases, cancer) and inheritance modes (autosomal dominant or recessive). Cancer-driver genes, for example, are often highly central, while recessive disease genes tend to be located toward the periphery [110].

9.3.3 Neighborhoods in the Interactome

Beyond measures of centrality and connectedness of individual nodes, many important structural connection patterns between a group of nodes have been identified. For example, "network modules," that is nodes that are densely connected among themselves but only sparsely connected to the rest of the network, often perform a certain function together [126–128]. Similarly, shared pathway membership, co-localization in the cell and co-expression [33, 34] have been found to be associated with specific interactome neighborhoods. In addition to functional relationships, network modules have also been identified with disease-related processes, showing that genes implicated in the same disease tend to be more connected to each other than expected by chance [129]. A systematic study on ∼300 complex diseases revealed that currently available interactome networks offer sufficient coverage to identify these disease neighborhoods, thereby confirming a fundamental hypothesis of interactome-based approaches to human disease [117].

There are, however, subtle differences between the connectivity patterns of functionally related proteins and proteins implicated in the same disease. Genes that jointly perform a biological task are often much more densely connected than genes

associated with the same disease [53]. An interpretation of this empirical finding is that *dys*function is typically distributed among several, only loosely connected, functional modules on the interactome. This has important implications for the design of network-based algorithms that aim to identify genes with a certain function or dysfunction. While functional associations may be identified using so-called "community detection" algorithms that target dense node groups [130], the identification of disease-associated genes requires different strategies, as reviewed in Section 9.4.

The relationships among nodes within a certain network neighborhood can be generalized to relationships between neighborhoods. A study of more than 44 000 disease pairs identified a network measure for the interactome-based distance of two disease modules, allowing systematic distinguishing of separated or overlapping disease pairs [117]. While overlapping disease modules correspond to diseases with significant molecular similarities, as well as related symptoms and elevated co-morbidity, diseases whose modules are separated lack any detectable molecular or clinical relationships. These findings were later extended, showing that the network distance of the targets of a particular drug to a disease module is predictive for drug efficacy [54] and can be used for identifying drug repurposing candidates [131, 132].

9.4 Disease Module Analysis

The local aggregation of both physiological and pathobiological processes within interactome networks represents a fundamental biological organization principle that forms the basis for many important applications, ranging from the prediction of protein function to disease gene identification and drug target prioritization. In this section, we will briefly review the process of disease module analysis. A disease module is loosely defined as the comprehensive set of cellular components and their interactions that are associated with a particular disease. Most commonly, a disease module is identified with a connected subgraph within the interactome [3] (Figure 9.3). For most common diseases, such as cardiovascular diseases, cancer or diabetes mellitus, there are hundreds of genes known to be involved. Yet, despite these impressive advances, in particular fueled by sequencing technology, we are still far from a complete understanding of their molecular determinants. For example, more than 2000 genes are estimated to be involved in intellectual disabilities, yet our current knowledge includes only around 800 genes [133]. The basic idea of disease module analysis is to use the connectivity patterns observed among known disease-associated genes to systematically scan their respective interactome neighborhood for genes with a yet unknown important role for the disease. This principle has been applied successfully to a broad range of diseases, from rare Mendelian disorders [134], to cancer [135] and other complex disorders, like metabolic [136], inflammatory [137] or developmental diseases [138].

9.4.1 Seed Cluster Construction

The starting point of the disease module analysis process is the construction of a suitable interactome network (see above for resources) and the curation of known disease-associated genes ("seed genes") for the particular disease of interest. There are

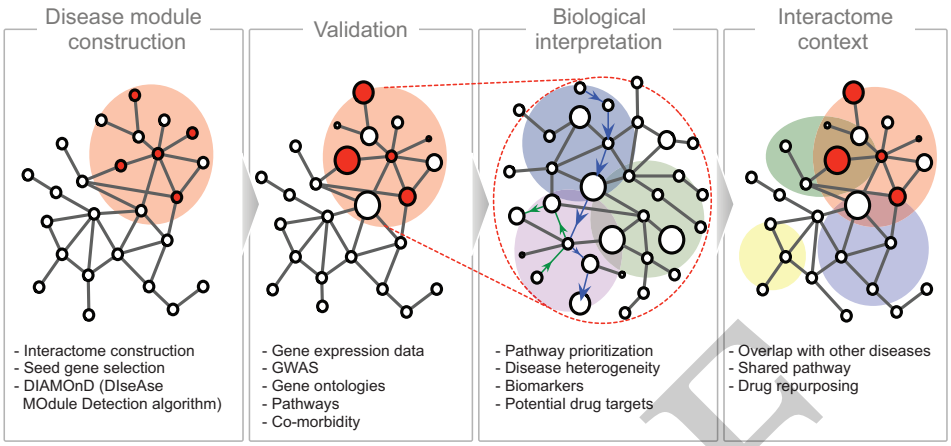

Figure 9.3 Disease module analysis. Overview of the different steps involved in constructing and analyzing the interactome module of a particular disease. The disease module represents all cellular components and their interactions that are responsible for a certain disease. Springer Nature, 2010.

several comprehensive ressources of known disease genes, including OMIM [104], the GWAS catalogue [139] and DisGeNET [140]. The reported disease associations cover a wide spectrum from rare variants with a known and experimentally validated functional mechanism to GWAS variants of rather small effect size and unknown mechanism. Other associations may not be of genetic origin at all, such as differential gene expression or associations inferred solely from text mining. Given this broad variety in possible interaction data and gene–disease associations, a certain trade-off between using only highest-confidence data and achieving the highest possible coverage is unavoidable. We recommend to experiment with different solutions, ideally guided by a domain expert of the specific disease of interest. While there is no simple recipe to tackle this challenging problem of setting up the initial interactome and seed genes, the network perspective can provide some guidance: if the seed cluster is not significantly localized on the network, it is unlikely that a network-based expansion algorithm will be able to identify the relevant neighborhood around this cluster. The connectivity of the initial seed cluster, for example in terms of the statistical significance of the largest connected component size, may therefore provide a rough indication as to whether a particular combination of interactome and seed gene data meets the minimal criteria for a meaningful disease module analysis [141].

9.4.2 Network-Based Disease Gene Prioritization

There are numerous algorithms that scan the network neighborhood around a given set of seed genes. They can be broadly classified into three major categories: (1) connectivity based methods, (2) path-based methods and (3) diffusion-based methods.

Connectivity-Based Methods
These methods build directly on the observation that disease genes tend to interact with each other. Early approaches considered all direct neighbors of the seed cluster as potential candidate genes [142]. More recent refinements of this idea take the degree heterogeneity of the interactome into account [143] or include more advanced relationships among groups of nodes, such as graphlets [144] or connectivity significance [53, 141].

Path-Based Methods
The genes in the neighborhood of the seed cluster can also be ranked according to their network distance to the respective seed genes. There are several possibilities for how to quantify the network distance between sets of nodes, such as using different weighted averages as implemented in [145]. Another option is to search for a set of candidate nodes that collectively minimize the path lengths within the seed cluster, such as using minimum spanning tree (or "Steiner tree") approaches [146–148]. These approaches extend the set of seed genes using a minimal number of additional nodes and edges required for connecting all seed genes in a single connected component.

Diffusion-Based Methods
The closeness of candidate genes to the seed cluster can also be assessed using dynamic approaches, such as diffusion processes [134, 135, 149–152]. A widely used choice is the random walk with restart (RWR) [153]. Starting from the seed genes, a random walker wanders along randomly chosen links of the network until returning to a randomly chosen seed gene (with restart probability r per time step) and starting the process all over. After sufficient iterations of this process, the frequencies with which the individual nodes in the network are visited will converge to a stationary value. This value can then be used to rank the nodes from most related (highest frequency) to least related (lowest frequency) to the seed cluster. The restarting probability r can be used to tune the influence of the seed genes on the diffusive process, from free diffusion (no influence of the seed genes, $r = 0$) to no diffusion at all (walker never leaves the seed genes, $r = 1$).

9.4.3 Validation of the Disease Module

Depending on the chosen prioritization algorithm, the outcome is most commonly given by a ranked list of genes. The next step in the disease module analysis process is to assess the relevance of the identified genes and choose a cutoff, that is, the number of additional candidate genes that will be integrated into the final module. We recommend a complementary strategy that uses cross-validation methods and enrichment with independent biological data.

Cross-Validation of Prediction Performance
Similar to other class prediction tasks, we can use k-fold cross-validation to evaluate the performance of the chosen algorithm. First, the set of original seed genes is randomly divided into k groups. Next, we remove one of the k subgroups from the

seed genes and repeat the prioritization procedure using only the remaining $k - 1$ groups as a modified seed gene pool. The performance of the algorithm can then be evaluated from its ability to retrieve the genes from the original seed gene set that was left out. In contrast to many other classification tasks, disease gene prioritization lacks clear true negatives, that is, genes known not to be involved in the disease. Standard performance measures, such as receiver operating characteristic curves, therefore need to be interpreted with some caution. Some approximations have been proposed for true negative genes, such as genes that are essential or unlikely to be involved in a particular disease due to their expression pattern.

Enrichment with Independent Biological Data

A more biologically motivated approach for estimating the relevance of the identified candidate genes is to use independent biological data and test from enrichment. Examples of such complementary data include (1) genes corresponding to GWAS loci of relevant studies, (2) genes found to be differentially expressed in a respective case/control study, (3) genes involved in biological processes or pathways that are known to be relevant to the diseases (expert curated), (4) genes involved in the same pathways or processes as the seed genes (unbiased enrichment) or (5) genes implicated in diseases that show high co-morbidity with the disease under investigation. In [141], for example, the authors used a sliding window approach to assess how the enrichment decreases along the list of ranked candidate genes. This also allows both for evaluating how relevant the candidate genes are, as well as choosing a cutoff rank, beyond which the candidate genes show no strong biological signal.

9.4.4 Interpreting the Disease Module

In the last, and arguably most important, step of the disease module analysis we turn to the biological interpretation of the identified genes and their interactions. We can distinguish two main perspectives, one that focuses on the biological mechanisms contained within the module, and one evaluating the interactome context of the entire module.

Elucidating the Molecular Mechanisms of the Disease

Utilizing the diverse biological data collected at the validation step, we can now extract a more detailed picture of the biological processes that are contained within the disease module. Following a strategy proposed in [141], we first combine the various layers of evidence per gene into a single score. For example, a gene that has been identified in a GWAS study, that is linked to differential expression and that is also known to be involved in a highly co-morbid disease may receive a higher score than a gene with fewer lines of evidence supporting its involvement in the disease. This can be achieved by first ranking all genes separately for each data set and then combining the scores, using for example the so-called Borda count [154] or other methods. Next, we can use the integrated score for prioritizing pathways or biological processes contained in the module, for example through the average score of the contained genes. Alternatively, one can use standard tools for gene set enrichment analyses in [155, 156], or more advanced network-based methods such as in [157, 158].

Context within the Interactome

As discussed above, the network-based relationships between modules may offer biological insights that can also be explored as part of a disease module analysis. Overlaps with other disease modules may reveal mechanisms of frequent co-morbidities, or indicate potential drug repurposing options. A network-based analysis may further identify potential submodules, for example for stratifying patients into subgroups for more personalized treatment.

9.5 Discussion and Outlook

In light of the complex networks of networks that are involved in human disease, this chapter could only offer a first glimpse into the emerging field of network medicine. Important research questions remain for each network introduced above. For most systems, no complete mapping exists yet, on either the node level or the edge level. Parallel to efforts to obtain more and more complete maps, increasingly sophisticated network analysis methods are being developed, hopefully allowing us to gain deeper insights into the complex relationships across the different scales that govern health and disease. Indeed, a major challenge that we are only beginning to address is to go beyond individual networks and include the intricate relationships between them. First steps in this direction consider multi-layer networks, in which different layers represent different relationships among entities or multiplex networks, in which different kinds of interactions are introduced between different types of nodes [159–161]. Such approaches have already been used to study the spread of epidemics by representing different modalities of contact as different layers [162–164]. The different layers may, for example, represent online social networks, public transit use or flight paths between cities, each characterized by different spreading rates and mechanisms. These promising first results indicate the potential that such approaches may hold also for integrating molecular and cellular networks.

References

[1] Greene JA, Loscalzo J. Putting the patient back together-social medicine, network medicine, and the limits of reductionism. *New Eng J Med*. 2017;377(25):2493.

[2] Beadle GW, Tatum EL. Genetic control of biochemical reactions in *Neurospora*. *PNAS*. 1941;27(11):499–506.

[3] Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet*. 2011;12(1):56–68.

[4] McGillivray P, Clarke D, Meyerson W, et al. Network analysis as a grand unifier in biomedical data science. *Ann Rev Biomed Data Sci*. 2018;1:153–180.

[5] Lazebnik Y. Can a biologist fix a radio? Or, what I learned while studying apoptosis. *Cancer Cell*. 2002;2(3):179–182.

[6] Vidal M, Cusick ME, Barabási A-L. Interactome networks and human disease. *Cell*. 2011;144(6):986–998.

[7] Swainston N, Smallbone K, Hefzi H, et al. Recon 2.2: from reconstruction to model of human metabolism. *Metabolomics*. 2016;12:109.

[8] Ma H, Sorokin A, Mazein A, et al. The Edinburgh human metabolic network reconstruction and its functional analysis. *Mol Syst Biol*. 2007;3:135.

[9] Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*. 2017;45(D1):D353–D361.

[10] Gille C, Bölling C, Hoppe A, et al. HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology. *Mol Syst Biol*. 2010;6:411.

[11] Zelezniak A, Pers TH, Soares S, Patti ME, Patil KR. Metabolic network topology reveals transcriptional regulatory signatures of type 2 diabetes. *PLoS Comput Biol*. 2010;6(4):e1000729.

[12] Folger O, Jerby L, Frezza C, et al. Predicting selective drug targets in cancer through metabolic networks. *Mol Syst Biol*. 2011;7: 501.

[13] Breitkreutz D, Hlatky L, Rietman E, Tuszynski JA. Molecular signaling network complexity is correlated with cancer patient survivability. *PNAS*. 2012;109(23):9209–9212.

[14] Hecker M, Lambeck S, Toepfer S, et al. Gene regulatory network inference: data integration in dynamic models – a review. *Biosystems*. 2009;96(1): 86–103.

[15] Lee TI, Rinaldi NJ, Robert F, et al. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science*. 2002;298(5594):799–804.

[16] Mathelier A, Fornes O, Arenillas DJ, et al. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res*. 2016;44(D1):D110–D115.

[17] Wingender E, Dietze P, Karas H, Knüppel R. TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Res*. 1996;24(1): 238–241.

[18] Huntzinger E, Izaurralde E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat Rev Genet*. 2011;12(2):99–110.

[19] Filipowicz W, Bhattacharyya SN, Sonenberg N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet*. 2008;9(2):102–114.

[20] Ardekani AM, Naeini MM. The role of microRNAs in human diseases. *Avicenna J Med Biotechnol*. 2010;2(4):161–179.

[21] Lu M, Zhang Q, Deng M, et al. An analysis of human microRNA and disease associations. *PLoS One*. 2008;3(10):e3420.

[22] Heyn J, Hinske LC, Ledderose C, Limbeck E, Kreth S. Experimental miRNA target validation. *Methods Mol Biol*. 2013;936:83–90.

[23] Kuhn DE, Martin MM, Feldman DS, et al. Experimental validation of miRNA targets. *Methods*. 2008;44(1):47–54.

[24] Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R. Fast and effective prediction of microRNA/target duplexes. *RNA*. 2004;10(10):1507–1517.

[25] Agarwal V, Bell GW, Nam J-W, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *Elife*. 2015;4.

[26] Krek A, Grün D, Poy MN, et al. Combinatorial microRNA target predictions. *Nat Genet*. 2005;37(5):495–500.

[27] Betel D, Wilson M, Gabow A, Marks DS, Sander C. The microRNA.org resource: targets and expression. *Nucleic Acids Res*. 2008;36(Database issue):D149–D153.

[28] Dweep H, Sticht C, Pandey P, Gretz N. miRWalk-database: prediction of possible miRNA binding sites by "walking" the genes of three genomes. *J Biomed Inform*. 2011;44(5):839–847.

[29] Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. *Nucleic Acids Res*. 2008;36(Database issue):D154–D158.

[30] Sethupathy P, Corda B, Hatzigeorgiou AG. TarBase: a comprehensive database of experimentally supported animal microRNA targets. *RNA*. 2006;12(2): 192–197.

[31] Xiao F, Zuo Z, Cai G, et al. mirecords: an integrated resource for microRNA–target interactions. *Nucleic Acids Res*. 2009;37(suppl 1):D105–D110.

[32] Huttlin EL, Ting L, Bruckner RJ, et al. The BioPlex network: a systematic exploration of the human interactome. *Cell*. 2015;162(2):425–440.

[33] Huttlin EL, Bruckner RJ, Paulo JA, et al. Architecture of the human interactome defines protein communities and disease networks. *Nature*. 2017;545(7655): 505–509.

[34] Rolland T, Tasan M, Charloteaux B, et al. A proteome-scale map of the human interactome network. *Cell*. 2014;159(5):1212–1226.

[35] De Las Rivas J, Fontanillo C. Protein–protein interactions essentials: key concepts to building and analyzing interactome networks. *PLoS Comput Biol*. 2010;6 (6):e1000807.

[36] Ofran Y, Rost B. Predicted protein–protein interaction sites from local sequence information. *FEBS Lett*. 2003;544(1–3):236–239.

[37] Gallet X, Charloteaux B, Thomas A, Brasseur A. A fast method to predict protein interaction sites from sequences. *J Mol Biol*. 2000;302:917–926.

[38] Yan C, Dobbs D, Honavar V. A two-stage classifier for identification of protein–protein interface residues. *Bioinformatics*. 2004;20(suppl 1):i371–i378.

[39] Deng M, Mehta S, Sun F, Chen T. Inferring domain–domain interactions from protein–protein interactions. *Genome Res*. 2002;12(10):1540–1548.

[40] Marcotte CJV, Marcotte EM. Predicting functional linkages from gene fusions with confidence. *Appl Bioinformatics*. 2002;1(2):93–100.

[41] Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *PNAS*. 1999;96(8):4285–4288.

[42] Hakes L, Pinney JW, Robertson DL, Lovell SC. Protein–protein interaction networks and biology – what's the connection? *Nat Biotechnol*. 2008;26(1): 69–72.

[43] Gillis J, Ballouz S, Pavlidis P. Bias tradeoffs in the creation and analysis of protein–protein interaction networks. *J Proteomics*. 2014;100:44–54.

[44] Kerrien S, Aranda B, Breuza L, et al. The IntAct molecular interaction database in 2012. *Nucleic Acids Res*. 2012;40(Database issue):D841–D846.

[45] Hermjakob H, Montecchi-Palazzi L, Lewington C, et al. IntAct: an open source molecular interaction database. *Nucleic Acids Res*. 2004;32(Database issue):D452–D455.

[46] Croft D, Mundo AF, Haw R, et al. The reactome pathway knowledgebase. *Nucleic Acids Res*. 2014;42(Database issue):D472–D477.

[47] Fabregat A, Jupe S, Matthews L, et al. The reactome pathway knowledgebase. *Nucleic Acids Res*. 2018;46(D1):D649–D655.

[48] Alanis-Lobato G, Andrade-Navarro MA, Schaefer MH. HIPPIE v2.0: enhancing meaningfulness and reliability of protein–protein interaction networks. *Nucleic Acids Res*. 2017;45(D1):D408–D414.

[49] Szklarczyk D, Morris JH, Cook H, et al. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res*. 2017;45(D1):D362–D368.

[50] Jeong H, Mason SP, Barabási A-L, Oltvai ZN. Lethality and centrality in protein networks. *Nature*. 2001;411(6833):41–42.

[51] Lim J, Hao T, Shaw C, et al. A protein–protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. *Cell*. 2006;125(4):801–814.

[52] Vanunu O, Magger O, Ruppin E, Shlomi T, Sharan R. Associating genes and protein complexes with disease via network propagation. *PLoS Comput Biol*. 2010;6(1):e1000641.

[53] Ghiassian SD, Menche J, Barabási A-L. A DIseAse MOdule detection (DIAMOnD) algorithm derived from a systematic analysis of connectivity patterns of disease proteins in the human interactome. *PLoS Comput Biol*. 2015;11(4):e1004120.

[54] Guney E, Menche J, Vidal M, Barábasi A-L. Network-based in silico drug efficacy screening. *Nat Commun*. 2016;7:10331.

[55] White JG, Southgate E, Thomson JN, Brenner S. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philos Trans R Soc Lond B Biol Sci*. 1986;314(1165):1–340.

[56] Briggman KL, Helmstaedter M, Denk W. Wiring specificity in the direction-selectivity circuit of the retina. *Nature*. 2011;471(7337):183.

[57] Bock DD, Lee W-CA, AM Kerlin, et al. Network anatomy and in vivo physiology of visual cortical neurons. *Nature*. 2011;471(7337):177.

[58] Glasser MF, Coalson TS, Robinson EC, et al. A multi-modal parcellation of human cerebral cortex. *Nature*. 2016;536(7615):171–178.

[59] Sporns O. The human connectome: origins and challenges. *Neuroimage*. 2013;80:53–61.

[60] Yan G, Vértes PE, Towlson EK, et al. Network control principles predict neuron function in the *Caenorhabditis elegans* connectome. *Nature*. 2017;550(7677):519.

[61] Seidlitz J, Váša F, Shinn M, et al. Morphometric similarity networks detect microscale cortical organization and predict inter-individual cognitive variation. *Neuron*. 2018;97(1):231–247.

[62] Xia CH, Ma Z, Ciric R, et al. Linked dimensions of psychopathology and connectivity in functional brain networks. *Nat Commun*. 2018;9(1):3003.

[63] Bergthaler A, Menche J. The immune system as a social network. *Nat Immunol*. 2017;18(5):481.

[64] Rieckmann JC, Geiger R, Hornburg D, et al. Social network architecture of human immune cells unveiled by quantitative proteomics. *Nat Immunol*. 2017;18(5):583.

[65] Kveler K, Starosvetsky E, Ziv-Kenet A, et al. Immune-centric network of cytokines and cells in disease context identified by computational mining of PubMed. *Nature Biotechnol*. 2018;36(7):651–659.

[66] Bernoulli D. Essai dune nouvelle analyse de la mortalité causée par la petite vérole et des avantages de linoculation pour la prévenir. *Histoire de lAcad Roy Sci(Paris) avec Mém des Math et Phys and Mém*. 1760;1:1–45.

[67] Pastor-Satorras R, Castellano C, Van P Mieghem, Vespignani A. Epidemic processes in complex networks. *Rev Mod Phys*. 2015;87(3):925–979.

[68] Longini IM, Jr, Nizam A, Xu S, et al. Containing pandemic influenza at the source. *Science*. 2005;309(5737):1083–1087.

[69] Granell C, Gómez S, Arenas A. Dynamical interplay between awareness and epidemic spreading in multiplex networks. *Phys Rev Lett*. 2013;111(12):128701.

[70] Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. *N Engl J Med*. 2007;357(4):370–379.

[71] Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. *N Engl J Med*. 2008;358(21):2249–2258.

[72] Fowler JH, Christakis NA. Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study. *BMJ*. 2008;337:a2338.

[73] Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*. 2005;4:Article 17.

[74] De Smet R, Marchal K. Advantages and limitations of current network inference methods. *Nat Rev Microbiol*. 2010;8(10):717–729.

[75] GTEx Consortium. Human genomics: the Genotype-Tissue expression (GTEx) pilot analysis – multitissue gene regulation in humans. *Science*. 2015;348(6235):648–660.

[76] Weirauch MT. Gene coexpression networks for the analysis of DNA microarray data. In *Applied Statistics for Network Biology*. Wein: Wiley-VCH Verlag, 2011, pp. 215–250.

[77] Parikshak NN, Swarup V, Belgard TG, et al. Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature*. 2016; 540:423.

[78] Zhang B, Gaiteri C, Bodea L-G, et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell*. 2013;153(3):707–720.

[79] Peters LA, Perrigoue J, Mortha A, et al. A functional genomics predictive network model identifies regulators of inflammatory bowel disease. *Nat Genet*. 2017;49:1437.

[80] Zhang S, Liu W, Liu X, Qi J, Deng C. Biomarkers identification for acute myocardial infarction detection via weighted gene co-expression network analysis. *Medicine*. 2017;96(47):e8375.

[81] Zhang J, Xiang Y, Ding L, et al. Using gene co-expression network analysis to predict biomarkers for chronic lymphocytic leukemia. *BMC Bioinformatics*. 2010;11(suppl. 9): S5.

[82] Wang L-X, Li Y, Chen G-Z. Network-based co-expression analysis for exploring the potential diagnostic biomarkers of metastatic melanoma. *PLoS One*. 2018;13(1):e0190447.

[83] Costanzo M, VanderSluis B, Koch EN, et al. A global genetic interaction network maps a wiring diagram of cellular function. *Science*. 2016;353(6306).

[84] Szappanos B, Kovács K, Szamecz B, et al. An integrated approach to characterize genetic interaction networks in yeast metabolism. *Nat Genet*. 2011;43 (7):656–662.

[85] Tong AHY, Lesage G, Bader GD, et al. Global mapping of the yeast genetic interaction network. *Science*. 2004;303(5659):808–813.

[86] Blomen VA, Májek P, Jae LT, et al. Gene essentiality and synthetic lethality in haploid human cells. *Science*. 2015;350 (6264):1092–1096.

[87] Wang T, Yu H, Hughes NW, et al. Gene essentiality profiling reveals gene networks and synthetic lethal interactions with oncogenic ras. *Cell*. 2017;168(5):890–903.e15.

[88] Srivas R, Shen JP, Yang CC, et al. A network of conserved synthetic lethal interactions for exploration of precision cancer therapy. *Mol Cell*. 2016;63(3):514–525.

[89] Han K, Jeng EE, Hess GT, et al. Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. *Nat Biotechnol*. 2017;35(5):463.

[90] Moder M, Velimezi G, Owusu M, et al. Parallel genome-wide screens identify synthetic viable interactions between the BLM helicase complex and Fanconi anemia. *Nat Commun*. 2017;8(1):1238.

[91] Kim D, Rossi J. RNAi mechanisms and applications. *Biotechniques*. 2008;44(5): 613–616.

[92] Doench JG, Fusi N, Sullender M, et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol*. 2016;34(2):184–191.

[93] Kubicek S, Gilbert JC, Fomina-adlin DY, et al. Chromatin-targeting small molecules cause class-specific transcriptional changes in pancreatic endocrine cells. *PNAS*. 2012;109(14):5364–5369.

[94] Bansal M, Yang J, Karan C, et al. A community computational challenge to predict the activity of pairs of compounds. *Nat Biotechnol*. 2014;32(12):1213–1222.

[95] Markt P, Dürnberger G, Colinge J, Kubicek S. CLOUD: CeMM library of unique drugs. *J Cheminform*. 2012;4(Suppl 1):P23.

[96] Isik Z, Baldow C, Cannistraci CV, Schroeder M. Drug target prioritization by perturbed gene expression and network information. *Sci Rep*. 2015;5:17417.

[97] Bray M-A, Singh S, Han H, et al. Cell painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat Protocol*. 2016;11(9):1757.

[98] Zhang F, Gao B, Xu L, et al. Allele-specific behavior of molecular networks: understanding small-molecule drug response in yeast. *PLoS One*. 2013;8(1):e53581.

[99] Noh H, Shoemaker JE, Gunawan R. Network perturbation analysis of gene transcriptional profiles reveals protein targets and mechanism of action of drugs and influenza A viral infection. *Nucleic Acids Res*. 2018;46(6):e34.

[100] Schubert M, Klinger B, Münemann Kl, et al. Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat Commun*. 2018;9(1):20.

[101] Dorel M, Klinger B, Sieber A, et al. Modelling signalling networks from perturbation data. *bioRxiv*, 2018.

[102] Molinelli EJ, Korkut A, Wang W, et al. Perturbation biology: inferring signaling networks in cellular systems. *PLoS Comput Biol*. 2013;9(12):e1003290.

[103] Goh K-I, Cusick ME, Valle D, et al. The human disease network. *PNAS*. 2007;104(21):8685–8690.

[104] Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res*. 2015;43(Database issue):D789–D798.

[105] Lee D-S, Park J, Kay K, et al. The implications of human metabolic network topology for disease comorbidity. *PNAS*. 2008;105(29):9880–9885.

[106] Zhou X, Menche J, Barabási A-L, Sharma A. Human symptoms–disease network. *Nat Commun*. 2014;5.

[107] NIH. Medical subject headings. www.nlm.nih.gov/mesh.

[108] Barrenas F, Chavali S, Holme P, Mobini R, Benson M. Network properties of complex human disease genes identified through genome-wide association studies. *PLoS One*. 2009;4(11):e8090.

[109] Zhang M, Zhu C, Jacomy A, Lu LJ, Jegga AG. The orphan disease networks. *Am J Hum Genet*. 2011;88(6):755–766.

[110] Piñero J, Berenstein A, Gonzalez-Perez A, Chernomoretz A, Furlong LI. Uncovering disease mechanisms through network biology in the era of next generation sequencing. *Sci Rep*. 2016;6:24570.

[111] Hidalgo CA, Blumm N, Barabási A-L, Christakis NA. A dynamic network approach for the study of human phenotypes. *PLoS Comput Biol*. 2009;5(4):e1000353.

[112] Chmiel A, Klimek P, Thurner S. Spreading of diseases through comorbidity networks across life and gender. *New J Phys*. 2014;16(11):115013.

[113] Hu JX, Thomas CE, Brunak S. Network biology concepts in complex disease comorbidities. *Nat Rev Genet*. 2016;8.

[114] Duran-Frigola M, Rossell D, Aloy P. A chemo-centric view of human health and disease. *Nat Commun*. 2014;5:5676.

[115] Gomez-Cabrero D, Menche J, Vargas C, et al. From comorbidities of chronic obstructive pulmonary disease to identification of shared molecular mechanisms by data integration. *BMC Bioinformat*. 2016;17(Suppl. 15):441.

[116] Klimek P, Aichberger S, Thurner S. Disentangling genetic and environmental risk factors for individual diseases from multiplex comorbidity networks. *Sci Rep*. 2016;6:39658.

[117] Menche J, Sharma A, Kitsak M, et al. Uncovering disease–disease relationships through the incomplete interactome. *Science*. 2015;347(6224):1257601.

[118] Watts DJ, Strogatz SH. Collective dynamics of "small-world" networks. *Nature*. 1998;393(6684):440–442.

[119] Cohen R, Havlin S. Scale-free networks are ultrasmall. *Phys Rev Lett*. 2003;90(5):058701.

[120] Callaway DS, Newman ME, Strogatz SH, Watts DJ. Network robustness and fragility: percolation on random graphs. *Phys Rev Lett*. 2000;85(25): 5468.

[121] Newman ME, Strogatz SH, Watts DJ. Random graphs with arbitrary degree distributions and their applications. *Phys Rev E*. 2001;64(2):026118.

[122] Cohen R, Erez K, Ben-Avraham D, Havlin S. Resilience of the internet to random breakdowns. *Phys Rev Lett*. 2000;85(21):4626.

[123] Dorogovtsev SN, Mendes JF. *Evolution of Networks: From Biological Nets to the Internet and WWW* . Oxford: Oxford University Press, 2003.

[124] Albert R, Jeong H, Barabási A-L. Error and attack tolerance of complex networks. *Nature*. 2000;406(6794):378–382.

[125] Blomen VA, Májek P, Jae LT, et al. Gene essentiality and synthetic lethality in haploid human cells. *Science*. 2015;350(6264):1092–1096.

[126] Spirin V, Mirny LA. Protein complexes and functional modules in molecular networks. *PNAS*. 2003;100(21):12123–12128.

[127] Hartwell LH, Hopfield JJ, Leibler S, Murray AW. From molecular to modular cell biology. *Nature*. 1999;402(6761 Suppl.):C47–C52.

[128] Barabási A-L, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet*. 2004;5(2):101–113.

[129] Feldman I, Rzhetsky A, Vitkup D. Network properties of genes harboring inherited disease mutations. *PNAS*. 2008;105(11):4323–4328.

[130] Fortunato S. Community detection in graphs. *Phys Rep*. 2010;486(3–5):75–174.

[131] Langhauser F, Casas AI, Guney E, et al. A diseasome cluster-based drug repurposing of soluble guanylate cyclase activators from smooth muscle relaxation to direct neuroprotection. *NPJ Syst Biol Appl*. 2018;4(1):8.

[132] Cheng F, Desai RJ, Handy DE, et al. Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nat Commun*. 2018;9(1):2691.

[133] Vissers LELM, Gilissen C, Veltman JA. Genetic studies in intellectual disability and related disorders. *Nat Rev Genet*. 2016;17(1):9–18.

[134] Smedley D, Köhler S, Czeschik JC, et al. Walking the interactome for candidate prioritization in exome sequencing studies of Mendelian diseases. *Bioinformatics*. 2014;30(22):3215–3222.

[135] Leiserson MDM, Vandin F, Wu H-T, et al. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat Genet*. 2015;47(2):106–114.

[136] Chen Y, Zhu J, Lum PY, et al. Variations in DNA elucidate molecular networks that cause disease. *Nature*. 2008;452(7186):429–435.

[137] Peters LA, Perrigoue J, Mortha A, et al. A functional genomics predictive network model identifies regulators of inflammatory bowel disease. *Nat Genet*. 2017;49(10):1437–1449.

[138] Krishnan A, Zhang R, Yao V, et al. Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. *Nat Neurosci*. 2016;19(11):1454–1462.

[139] MacArthur J, Bowler E, Cerezo M, et al. The new NHGRI-EBI catalog of published genome-wide association studies (GWAS catalog). *Nucleic Acids Res*. 2016;45(D1):D896–D901.

[140] Piñero J, Bravo A, Queralt-Rosinach N, et al. Disgenet: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res*. 2016;gkw943.

[141] Sharma A, Menche J, Huang CC, et al. A disease module in the interactome explains disease heterogeneity, drug response and captures novel pathways and genes in asthma. *Hum Mol Genet*. 2015;24(11):3005–3020.

[142] Oti M, Snel B, Huynen MA, Brunner HG. Predicting disease genes using protein–protein interactions. *J Med Genet*. 2006;43(8):691–698.

[143] Erten S, Bebek G, Ewing RM, Koyut Mürk, et al. DADA: degree-aware algorithms for network-based disease gene prioritization. *BioData Min*. 2011;4(1).

[144] Wang X-D, Huang J-L, Yang L, et al. Identification of human disease genes from interactome network using graphlet interaction. *PLoS One*. 2014;9(1):e86142.

[145] Guney E, Oliva B. Exploiting protein–protein interaction networks for genome-wide disease–gene prioritization. *PLoS One*. 2012;7(9):e43557.

[146] Bailly-Bechet M, Borgs C, Braunstein A, et al. Finding undetected protein associations in cell signaling by belief propagation. *PNAS*. 2011;108(2):882–887.

[147] Tuncbag N, McCallum S, Huang S-SC, Fraenkel E. SteinerNet: a web server for integrating "omic" data to discover hidden components of response pathways. *Nucleic Acids Res*. 2012;40(Web Server issue):W505–W509.

[148] Tuncbag N, Gosline SJC, Kedaigle A, et al. Network-based interpretation of diverse high-throughput datasets through the omics integrator software package. *PLoS Comput Biol*. 2016;12(4):e1004879.

[149] Krauthammer M, Kaufmann CA, Gilliam TC, Rzhetsky A. Molecular triangulation: bridging linkage and molecular-network information for identifying candidate genes in Alzheimer's disease. *PNAS*. 2004;101(42):15148–15153.

[150] Vanunu O, Magger O, Ruppin E, Shlomi T, Sharan R. Associating genes and protein complexes with disease via network propagation. *PLoS Comput Biol*. 2010;6(1):e1000641.

[151] Vandin F, Upfal E, Raphael BJ. Algorithms for detecting significantly mutated pathways in cancer. *J Comput Biol*. 2011;18(3):507–522.

[152] Cowen L, Ideker T, Raphael BJ, Sharan R. Network propagation: a universal amplifier of genetic associations. *Nat Rev Genet*. 2017;18:551–562.

[153] Köhler S, Bauer S, Horn D, Robinson PN. Walking the interactome for prioritization of candidate disease genes. *Am J Hum Genet*. 2008;82(4):949–958.

[154] Van Erp M, Schomaker L. Variants of the Borda count method for combining ranked classifier hypotheses. In *Seventh International Workshop on Frontiers in Handwriting Recognition*, 2000.

[155] Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *PNAS*. 2005;102(43):15545–15550.

[156] Kuleshov MV, Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016;44(W1):W90–W97.

[157] Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One*. 2010;5(11):e13984.

[158] Glaab E, Baudot A, Krasnogor N, Schneider R, Valencia A. Enrichnet: network-based gene set enrichment analysis. *Bioinformatics*. 2012;28(18):i451–i457.

[159] Cozzo E, Baños RA, Meloni S, Moreno Y. Contact-based social contagion in multiplex networks. *Phys Rev E*. 2013;88(5):050801.

[160] Boccaletti S, Bianconi G, Criado R, et al. The structure and dynamics of multilayer networks. *Phys Rep*. 2014;544(1):1–122.

[161] Kivelä M, Arenas A, Barthelemy M, et al. Multilayer networks. *J Complex Netw*. 2014;2(3):203–271.

[162] Min Y, Hu J, Wang W, et al. Diversity of multilayer networks and its impact on collaborating epidemics. *Phys Rev E*. 2014;90(6):062803.

[163] Chen X, Wang R, Tang M, et al. Suppressing epidemic spreading in multiplex networks with social-support. *New J Phys*. 2018;20(1):013007.

[164] Guo Q, Jiang X, Lei Y, et al. Two-stage effects of awareness cascade on epidemic spreading in multiplex networks. *Phys Rev E*. 2015;91(1):012822.