Chapter 3

Interactome Networks

Anne-Ruxandra Carvunis[1,2], Frederick P. Roth[1,3], Michael A. Calderwood[1,2], Michael E. Cusick[1,2], Giulio Superti-Furga[4] and Marc Vidal[1,2]

[1]*Center for Cancer Systems Biology (CCSB) and Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA,* [2]*Department of Genetics, Harvard Medical School, Boston, MA 02115, USA,* [3]*Donnelly Centre for Cellular & Biomolecular Research, University of Toronto, Toronto, Ontario M5S-3E1, Canada, & Samuel Lunenfeld Research Institute, Mt. Sinai Hospital, Toronto, Ontario M5G-1X5, Canada,* [4]*Research Center for Molecular Medicine of the Austrian Academy of Sciences, 1090 Vienna, Austria*

Chapter Outline

INTRODUCTION

Life Requires Systems

What is Life? The answer to the question posed by Schrödinger in a short but incisive book published in 1944 remains elusive more than seven decades later. Perhaps a less ambitious, but more pragmatic question could be: what does Life require? Biologists agree on at least four fundamental requirements, among which three are palpable and easily demonstrable, and a fourth is more intangible (Figure 3.1). First, Life requires *chemistry*. Biomolecules, including metabolites, proteins and nucleic acids, mediate the most elementary functions of biology. Life also requires *genes* to encode and 'reproduce' biomolecules. For most organisms, *cells* provide the fundamental medium in which biological processes take place. The fourth requirement is *evolution* by natural selection. Classically, these 'four great ideas of biology' [1] have constituted the main intellectual framework around which biologists formulate biological questions, design experiments, interpret data, train younger generations of scientists and attempt to design new therapeutic strategies.

The next question then should be: even if we fully understood each of these four basic requirements of biology, would we be anywhere near a complete understanding of how Life works? Would we be able to fully explain genotype–phenotype relationships? Would we be able to fully predict biological behaviors? How close would we be to curing or alleviating suffering from human diseases? It is becoming clear that even if we knew everything there is to know about the four currently accepted requirements of biology, the answer to 'What Life is' would remain elusive.

The main reason is that biomolecules do not function in isolation, nor do cells, organs or organisms, or even ecosystems and sociological groups. Rather, biological entities are involved in intricate and dynamic interactions, thereby forming 'complex systems'. In the last decade, novel biological questions and answers have surfaced, or resurfaced, pointing to *systems* as a fifth fundamental

Handbook of Systems Biology Concepts and Insights. http://dx.doi.org/10.1016/B978-0-12-385944-0.00003-4

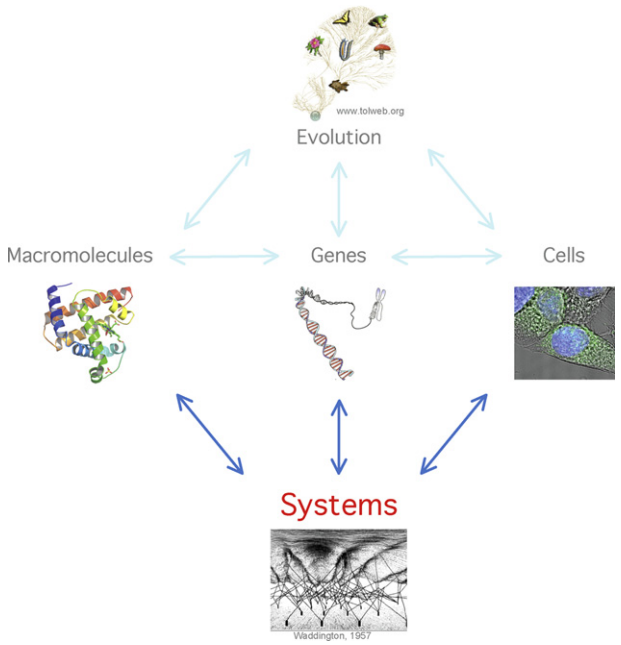

FIGURE 3.1 **Systems as a fifth requirement for Life.**

requirement for Life [2]. Although conceptual, systems may turn out to be as crucial to biology as chemistry, genes, cells or evolution (Figure 3.1).

Cells as Interactome Networks

The study of biological systems, or 'systems biology', originated more than half a century ago, when a few pioneers initially formulated a theoretical framework according to which multiscale dynamic complex systems formed by interacting biomolecules could underlie cellular behavior. To explain cellular differentiation, Delbrück hypothesized the existence of positive feedback circuits required for 'bistability', a model in which systems would remain stably activated after having been turned on, and conversely, remain steadily inactive once turned off [3]. Empirical evidence for feedback regulation in biology first emerged in the 1950s. The Umbarger and Pardee groups uncovered enzymatic feedback inhibition [4,5], and Novick and Weiner described the positive feedback circuit regulating the *lac* operon [6]. Monod and Jacob subsequently proposed how negative feedback circuits could account for homeostasis and other oscillatory phenomena observed in many biological processes [7]. These teleonomic arguments were later formalized by René Thomas and others in terms of requirements for cellular and whole organism differentiation based on positive and negative feedback circuits of regulation, using Boolean modeling as powerful simplifications of cellular systems [8] (see Chapter 10).

Equally enlightening theoretical systems properties were imagined beyond small-scale regulatory mechanisms composed of just a few molecules. Waddington introduced the metaphor of 'epigenetic landscape', whereby cells respond to genetic, developmental and environmental cues by following paths across a landscape containing peaks and valleys dictated by interacting genes and gene products [9]. This powerful idea, together with theoretical models of 'randomly constructed genetic nets' by Kauffman [10], suggested that a cellular system could be described in terms of 'states' resulting from particular combinations of genes, gene products, or metabolites, all considered either active or inactive at any given time. Complex wiring diagrams of functional and logical interconnectivity between biomolecules and genes acting upon each other could be imagined to depict how systems 'travel' from state to state over time throughout a 'state space' determined by intricate, sophisticated combinations of genotype, systems properties and environmental conditions. These concepts, elaborated at a time when the molecular components of biology were poorly described, remained largely ignored by molecular biologists until recently (see Chapter 15).

Over the past two decades, scientific knowledge of the biomolecular components of biology has dramatically increased. In particular, sequencing and bioinformatics have allowed prediction of coding and non-coding gene products at genome scale. Transcriptome sequencing approaches have revealed the existence of transcripts that had escaped prediction and which often remain of unknown function (see Chapter 2). Additionally, the list of known molecular components of cellular systems, including nucleic acids, gene products and metabolites, is lengthening and becoming increasingly detailed. With these advances came a humbling realization, best summed up as 'too much data, too few drugs' (see Chapter 8). It has become clearer than ever that knowing everything there is to know about each biomolecule in the cell is not sufficient to predict how the cell will react as a whole to particular external or internal perturbations.

Functional interactions, perhaps more so than individual components, mediate the fundamental requirements of the cell. Consequently, one needs to consider biological phenomena as the product of ensembles of interacting components with emergent properties that go beyond those of their individual components considered in isolation. One needs to step back and measure, model, and eventually perturb nearly all functional interactions between cellular components to fully understand how cellular systems work. In analogy to the word genome, the union of all interactions between all cellular components is termed the 'interactome'. Our working hypothesis is that interactomes exhibit local and global properties that relate to biology in general, and to genotype–phenotype relationships in particular.

Interactome Networks and Genotype–Phenotype Relationships

Since drafts of a composite reference human genome sequence were released 10 years ago, powerful technological developments, such as next-generation sequencing, have started a true revolution in genomics [11–14]. With time, most human genotypic variations will be described, together with large numbers of phenotypic associations. Unfortunately, such knowledge cannot translate directly into new mechanistic understanding or therapeutic strategies, in part because the 'one-gene/one-enzyme/one-function' concept developed by Beadle and Tatum oversimplifies genotype–phenotype relationships [15]. In fact, Beadle and Tatum themselves state so in the introduction of their groundbreaking paper that initiated reductionism in molecular biology: 'An organism consists essentially of an integrated system of chemical reactions controlled in some manner by genes. Since the components of such systems are likely to be interrelated in complex ways, it would appear that there must exist orders of directedness of gene control ranging from simple one-to-one relations to relations of great complexity.'

So-called complex traits provide the most compelling evidence of complexity between genotypes and phenotypes in human disease. Genome-wide association studies have revealed many more contributing loci than originally anticipated, with some loci contributing as little as a few percent to the heritability of the phenotype(s) of interest. Simple Mendelian traits are not immune to complex genotype–phenotype relationships either. Incomplete penetrance, variable expressivity, differences in age of onset, and modifier mutations are more frequent than generally appreciated. These discrepancies in the one-gene/one-function model appear across all phyla. In worms, for example, where self-fertilization is possible and growth conditions are easily controllable, it is not uncommon to observe that a significant proportion of mutant animals exhibit a near wild-type phenotype [16,17] (see Chapter 19).

Genome-wide functional genomic and proteomic experiments also point to greater complexity than anticipated, leading one to ask fundamental questions about gene function and evolution. How could a linear view of gene function account for the seemingly small proportions of essential genes [18]? How do genes become essential during evolution in the first place? How to explain that genes that are genetically required for particular biologic processes are not necessarily transcriptionally regulated during those same processes [19]? How to account for increasing reports of 'protein moonlighting', where specific gene products appear to be necessary for multiple biological processes [20–22]? We argue that viewing and modeling cells as interactome networks will help unravel the complexity of genotype–phenotype relationships, including susceptibility to human disease (Figure 3.2).

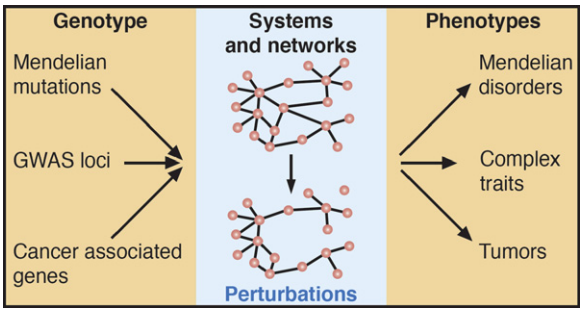

FIGURE 3.2 Interactome networks and genotype–phenotype relationships.

Mapping and Modeling Interactome Networks

In the path towards deciphering the mechanisms underlying genotype–phenotype relationships, it helps to abstractly simplify the complexity of interactomes by modeling. Interactomes can be effectively modeled as network representations of biological relationships among biomolecules. This abstraction converts a complex web of biological relationships into a graph, allowing the application of intuition and mathematical concepts of graph theory.

The power of graph theory in revealing emergent properties of complex systems is exemplified by a social science experiment of the 1960s. Stanley Milgram attempted to follow the path of letters sent through the mail in order to measure the average number of 'degrees of separation' between people, thereby providing a network representation of human society. He famously found that humans are connected on average at a distance of six degrees from each other, and our vision of the human population on earth immediately became that of a 'small-world' [23]. From politics to social media to modern journalism, Milgram's discovery still resonates today, and probably will for years to come.

In a network representation of interactomes, nodes represent biomolecules and an edge between two nodes indicates a biological relationship between the corresponding biomolecules. In the cell, multiple types of biomolecule, e.g., genes, proteins, RNAs, regulatory elements, or small-molecule metabolites, can be connected by multiple types of physical or functional biological relationship. These can be combined into 'multicolor' interactome network representations, where node colors represent biomolecule types and edge colors represent biological relationship types, refining complex biological processes [24]. Binding of transcription factors to DNA regulatory elements (physical relationship), regulation of

target genes by transcription factors or micro-RNAs (direct functional relationship), and similarity of expression profiles of genes across multiple conditions (profile-based functional relationship) form interrelated interactome networks. Study of these interactome networks, individually or together, is contributing key insights into the cellular control of gene expression (see Chapters 2 and 4).

Material flow, such as in metabolic reactions, and information flow, such as in transduction pathways, can be represented mathematically by edge direction (see Chapters 4, 5 and 11), while edge thickness, or weight, can symbolize the relative strength of biological relationships. Additionally, interactome network models can incorporate logic (see Chapters 10 and 11) or dynamics (see Chapters 12, 13, 14 and 16). Eventually, the aim is to understand how different interactomes are integrated together to form the cellular systems that we believe underlie genotype–phenotype relationships.

For this aim to be reached, complete single-color interactome network maps first need to be assembled. Physical interactions between proteins, or protein interactions, constitute the fundamental backbone of the cell and are instrumental for most biological processes, including signaling, differentiation and cell fate determination. This chapter describes the mapping and modeling of protein–protein interactome networks, where edges connect pairs of proteins that physically associate with one another directly or indirectly.

TOWARDS A REFERENCE PROTEIN–PROTEIN INTERACTOME MAP

Most individual proteins execute their biological functions by interacting with one or several other proteins. Protein interactions can form large protein complexes such as the proteasome, in which ~50 protein subunits act together to degrade other proteins and play a key role in cell protein homeostasis. The existence of such molecular machines, performing functions that no single protein can assume, demonstrates that protein–protein interactomes exhibit emergent properties beyond the sum of all individual protein interactions.

There is no such thing as a typical protein interaction. Protein interactions occur *in vivo* with a wide range of dissociation constants and dynamics. Proteins associated by strong and permanent interactions with low dissociation constants tend to form protein complexes. Protein interactions may also be simultaneously strong and transient, when controlled by the expression level of either interacting partner using 'just-in-time' assembly, by a change in subcellular location of one protein or the other, or by conformation changes induced by post-translational modification. For example, GTP hydrolysis controls the interaction between the α and β subunits of G-proteins, which in turn rapidly switches entire signaling pathways on or off [25]. Many protein interactions are weak and transient with high dissociation constants, as are associations between membrane receptors and extracellular matrix proteins that assist cellular motility [25].

Forthcoming models of protein–protein interactomes will undoubtedly involve sophisticated network representations integrating weighted and dynamic protein interactions [26]. For protein interactions such as those involved in signaling, interactome network models can also incorporate edge direction. Many protein interactions, such as subunit–subunit interactions within protein complexes, are best described with undirected edges [27–29]. It is not possible yet to assemble a proteome-scale interactome network model that integrates strength, dynamics and direction of edges because available technology is only beginning to allow experimental measurements of such interaction properties. Even a catalog of all possible protein interactions has not yet been compiled for any single species. Today's challenge lies in obtaining nearly complete but static, undirected and unweighted reference protein–protein interactome maps.

Strategies for Large-Scale Protein–Protein Interactome Mapping

Three fundamentally different but complementary strategies have been deployed towards this goal: i) curation of protein interaction data already available in the scientific literature [30]; ii) computational predictions of protein interactions based on available orthogonal information, such as sequence similarity or the co-presence of genes in sequenced genomes [31]; and iii) systematic, unbiased high-throughput experimental mapping strategies applied at the scale of whole genomes or proteomes [32].

Literature-curated interactome maps present the advantage of using already available, experimentally derived information, but are limited by the inherently variable quality of the curation process [33–35]. A randomly chosen set of literature-curated protein interactions supported by a single publication was shown to be approximately three times less reproducible than a reference set of manually curated protein interactions supported by multiple publications [36]. Another caveat of literature-curated interactome maps is that they mostly derive from hypothesis-driven research, which often focuses on a few proteins deemed to be scientifically interesting [37]. Some 'star proteins', such as the cancer-associated product of the *TP53* gene [38], are interrogated for protein interactions much more often than other proteins, resulting in an artificial increase of their apparent connectivity relative to

other proteins. For these reasons, literature-curated maps cannot be viewed as representative samples of the underlying interactome, and inferring systems-level properties from literature-curated protein–protein interactome maps can be misleading [39]. Nevertheless, literature-curated protein–protein interactome maps are instrumental in deriving hypotheses about focused biological mechanisms.

Computational predictions have the advantage of being applicable at genome or proteome scale for only a moderate cost. We discuss the numerous computational strategies that have been designed to predict protein interactions in the section of this chapter entitled 'Drawing inferences from interactome networks'. In brief, computational predictions apply 'rules' learned from current knowledge to infer new protein interactions. Albeit potent, this approach is also intrinsically limiting since the rules governing biological systems in general, and protein interactions in particular, remain largely undiscovered. Therefore, predicted protein–protein interactome maps, like literature-curated interactome maps, should be handled with caution when modeling biological systems.

High-throughput experimental interactome mapping approaches attempt to describe unbiased, systematic and well-controlled biophysical interactions. Two complementary approaches are currently in widespread use for high-throughput experimental interactome mapping (Figure 3.3): i) testing all combinations of protein pairs encoded by a given genome to find all binary protein interactions that can take place among them and uncover the 'binary interactome'; and ii) interrogating *in vivo* protein complexes in one or several cell line(s) or tissue(s) to expose the 'co-complex interactome'. Binary interactome maps contain mostly direct physical interactions, an unknown proportion of which may never take place *in vivo* despite being biophysically true. On the other hand, co-complex interactome datasets are composed of many indirect, and some direct, associations that mostly do take place *in vivo*.

Large-Scale Binary Interactome Mapping

The technologies that enabled large-scale binary interactome mapping were first developed in the 1990s by a few groups [40–45]. The following years saw significant progress towards assembling binary protein– protein interactome maps for model organisms such as the yeast *Saccharomyces cerevisiae* [39,46–49], the worm *Caenorhabditis elegans* [50–53], the fly *Drosophila melanogaster* [54], and most recently the plant *Arabidopsis thaliana* [55–57]. Similar efforts have been deployed to map the human binary interactome [36,58–62].

Large-scale binary interactome mapping is amenable to only a few existing experimental assays [47,48,63] and is carried out primarily by ever- improving variants on the yeast two-hybrid (Y2H) system [64,65]. The Y2H system is based on the reconstitution of a yeast transcription factor through the expression of two hybrid proteins, one fusing the DNA-binding (DB) domain to a protein X (DB-X) and the other fusing the activation domain (AD) to a protein Y (AD-Y) [65]. In the last 20 years the technique has been streamlined to increase throughput and quality controlled to avoid foreseeable artifacts [32,43,45,53,66–69]. Today, Y2H can interrogate hundreds of millions of protein pairs for binary interactions, in a manner that is both highly efficient and highly reliable.

The contemporary Y2H-based high-throughput binary interactome mapping pipeline consists of two essential stages: primary screening and secondary verification [64,70,71]. Large collections of cloned genes are transferred into DB-X and AD-Y expression vectors, then efficiently screened using either a pooling or a pairwise strategy [49,51,64,72–74]. All protein pairs identified in

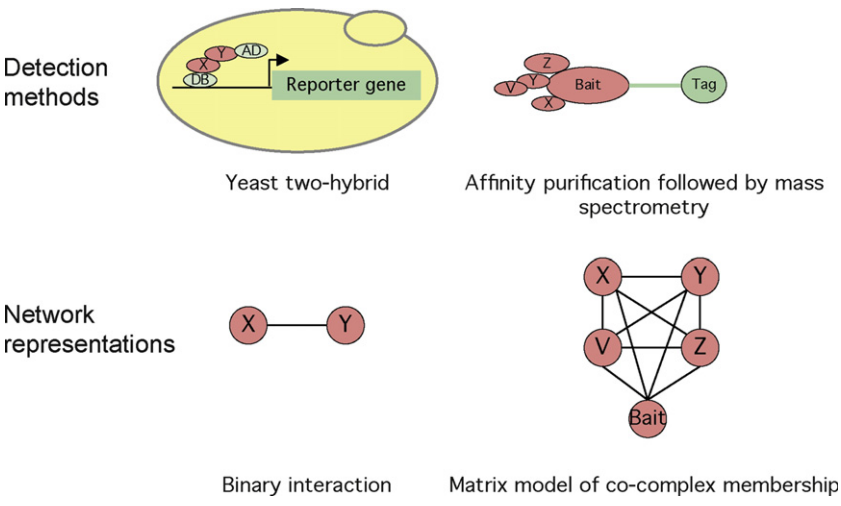

FIGURE 3.3 **Binary and co-complex protein-interaction detection methods and network representations.**

the primary screen are re-tested independently by Y2H using fresh yeast cells (verification). In our most recent Y2H-based binary interactome mapping efforts, DB-X/AD-Y pairs that score positive in at least three out of four replicates during the verification step were considered high-quality verified Y2H interactions [55,57]. Further sequencing can be undertaken to confirm the identity of the reported protein pairs [55]. Ultimately, this pipeline can systematically generate large numbers of highly reproducible binary protein interactions.

Reproducibility, however, does not necessarily guarantee high dataset quality. Technical artifacts could result in protein pairs appearing reproducibly positive despite actually being false positives. To obtain a quantitative estimate of the quality of a protein−protein interactome dataset, an integrated empirical framework for quality control of interactome mapping was proposed in 2009 [36]. At the heart of this endeavor was the recognition that high-throughput interactome mapping needed to be rigorously calibrated, like any well-controlled reliable small-scale experiment. Using this framework, protein−protein interactome maps generated with the mapping pipeline described here were shown to have high precision (80−100% of reported protein pairs are true positives) but low sensitivity (~10−15% of all interrogated true positive interactions are captured in the experiment) [36,39,52,55]. Because of this low sensitivity, binary interactome maps generated so far represent small fractions of the underlying true interactomes. This explains why only a marginal number of protein interactions are found in multiple binary interactome maps assembled independently for the same organism [36,39,55].

How far along is the journey towards a complete binary protein−protein interactome reference map? The answer requires an estimation of the size of such a reference map for any given species. Many statistical methods have been designed to this end, often based on dataset overlap and hypergeometric distributions [36,75−82]. Mapping of the binary interactome of the model organism *S. cerevisiae* is estimated to be the closest to completion, with ~6−30% coverage already obtained (~2900 binary protein interactions of demonstrated high technical quality detected, out of an estimated total of 10 000−45 000, assuming one splicing isoform per gene) [39,46,49]. However, most of the task remains to be accomplished.

How can this daunting challenge be overcome? Inspiration is drawn from the history of genome sequencing, which underwent a disruptive shift in the late 1990s. Like sequencing at that time, Y2H-based mapping is currently seeing more efficient automation, stricter quality control and innovative technology development which together are increasing productivity while reducing cost [36,62,83]. It is unlikely that Y2H-based mapping alone will be sufficient to complete a comprehensive reference binary protein−protein interactome map. No single interaction assay may ever be capable of doing so [84,85]. Individual protein interaction assays seem optimized for the detection of a certain subtype of binary interactions, although the biochemical and structural biases of these assays remain poorly understood. Intuitively, interactions involving membrane proteins would be expected to perform better in the cell membrane environment of the split-ubiquitin system than in the nuclear environment of the Y2H system [65,47]. It will therefore be necessary to join forces and use multiple assays to fully map the reference binary interactome of any organism. We are confident that this challenge can be overcome within the next decade.

Large-Scale Co-Complex Interactome Mapping

To fully map the reference interactome, it is operationally helpful to go beyond binary protein interactions and identify protein complexes within cells [86]. Protein complexes typically contain five to six different proteins, within a wide range from two to hundreds in a variety of stoichiometries [87]. The concentration and binding affinity of the protein subunits determine complex assembly according to the law of mass action [88]. Two proteins in isolation may have only weak or no propensity to form a binary interaction. Owing to cooperative and allosteric effects, a third protein may have a high affinity to both simultaneously, so that the resulting protein complex is considerably more stable than the sum of its component affinities [89]. Hence, even if a reference binary protein−protein interactome had been fully mapped, co-complex interactome network maps would still provide novel protein interactions and bring a fundamentally different view of interactome organization. The characterization of entire protein complexes, as they assemble in cells, is a necessary route to gather information on gene function and biological systems [90−94].

The most common methodologies currently used for the mapping of co-complex interactomes rely on protein complex purification followed by identification of constituent proteins by mass spectrometry. These experiments necessitate a trade-off between throughput, reproducibility and physiological setting. Cellular proteins range in their abundance up to seven orders of magnitude in humans [95] and five in yeast [96]. Protein complexes therefore need to be purified from the soup of cellular extracts without losing too many components, while at the same time avoiding those proteins that are extremely abundant and co-purify artificially [97−100]. A fraction of protein complexes may consist of dedicated elements, but most complexes also include abundant proteins that participate in several other complexes, such as chaperone proteins. Purification

strategies tailored for individual complexes typically make use of high-affinity antibodies directed against a specific complex member, or use other affinity matrices such as DNA, RNA, metabolites, or drugs, inspired by the specific biochemical properties of the complex. These approaches have the important advantage of targeting endogenous, natural forms of the complexes, but they are not readily amenable to proteome scale.

Proteome-scale co-complex interactome mapping employs a variety of strategies that attach epitope tags to bait proteins. These include DNA engineering as well as post-translational protein engineering [101–106]. The purified protein complexes are then systematically treated with proteases to release peptides, which then are fractionated by one- or two-dimensional liquid chromatography. Amino acid sequences are then imputed based on the mass and charge of the resulting peptides using mass spectrometry readouts. If applied rigorously and with attention to statistical significance, proteins containing these peptide sequences can be derived with a low false discovery rate. Proteins may also go undetected for a number of reasons, leading to false negatives in co-complex interactome maps. Quantitative estimation of dataset quality using a framework analogous to the one implemented for binary interactome maps [36] has now been implemented once for a *D. melanogaster* co-complex interactome map, which demonstrated high sensitivity [107].

How should we interpret the long lists of proteins that typically are the readouts of the mass spectrometry analyses of co-purified proteins? All successful bioinformatics approaches to assign co-complex memberships to co-purified proteins rely on network analyses, considering each protein as a node and each co-purification relationship as an edge. Algorithms can isolate subnetworks that are highly interconnected or completely interconnected, and then compute affiliation to one subnetwork compared to overall frequency to determine the most likely co-complex associations [91,108–110]. High-quality datasets are obtained from multiple redundant purifications over a single search space, which may encompass whole genomes, as was done for *S. cerevisiae, Escherichia coli, Mycoplasma pneumonia,* and *D. melanogaster* [87,91,104,107,111,112], or selected pathways and subnetworks as was done for human [113–118]. The more redundant the dataset, the more reliable complex prediction will be, leading to ever finer granularity of the resolution of the map.

When two proteins belong to the same protein complex they may not necessarily be in direct physical contact (Figure 3.3). Hence, edges in a network representation of co-complex interactomes have a very different meaning than edges in a network representation of binary interactomes. Most literature-curation databases are struggling to design the appropriate infrastructure that will allow users to distinguish intuitively between these two types of edge [119]. Curation, storage, representation and analysis of co-complex interactome data are key challenges in computational systems biology.

Regardless, the emerging picture of mass spectrometry-derived co-complex interactome maps is that of a modular organization. Protein complexes appear to assemble from a limited number of core modules, with small subcomplexes as well as individual proteins binding to the core modules and to each other. Modular organization creates the possibility of achieving functional diversity through combinatorial effects while maintaining highly interdependent central parts of molecular machines invariable [120–123].

DRAWING INFERENCES FROM INTERACTOME NETWORKS

Combining analyses of network topology with exogenous data integration can help make sense of complex systems. This ability is well illustrated in a famous study of high-school friendships across the USA [124]. In the topological structure of networks where nodes are students and edges are friendships, communities of tightly linked high-school friends emerge (see Chapter 9). When nodes in these topological communities are colored with ethnicity information, the extent of ethnic segregation in each high school is revealed (Figure 3.4). Similarly, binary and co-complex protein–protein interactome network maps provide 'scaffold' information about cellular systems. When interactome maps are analyzed in terms of topology and integrated with orthogonal functional information, the resulting knowledge allows investigators to imagine novel hypotheses and answer basic questions of biology (Figure 3.4) [125,126].

Refining and Extending Interactome Network Models

A major aim of the analysis of interactome network maps is to obtain better representations of the underlying interactome itself, since available maps are imperfect and incomplete. Given topological and exogenous biological data, which proteins are most likely to interact with a given protein of interest for which few or no protein interactions have been described? Which binary interactions and co-complex associations are the most reliable, and therefore worthy of mechanistic follow-up?

When there are multiple sources of experimental evidence supporting a particular protein interaction, the evidence can be combined to generate a confidence score for this interaction. This integration can be restricted to

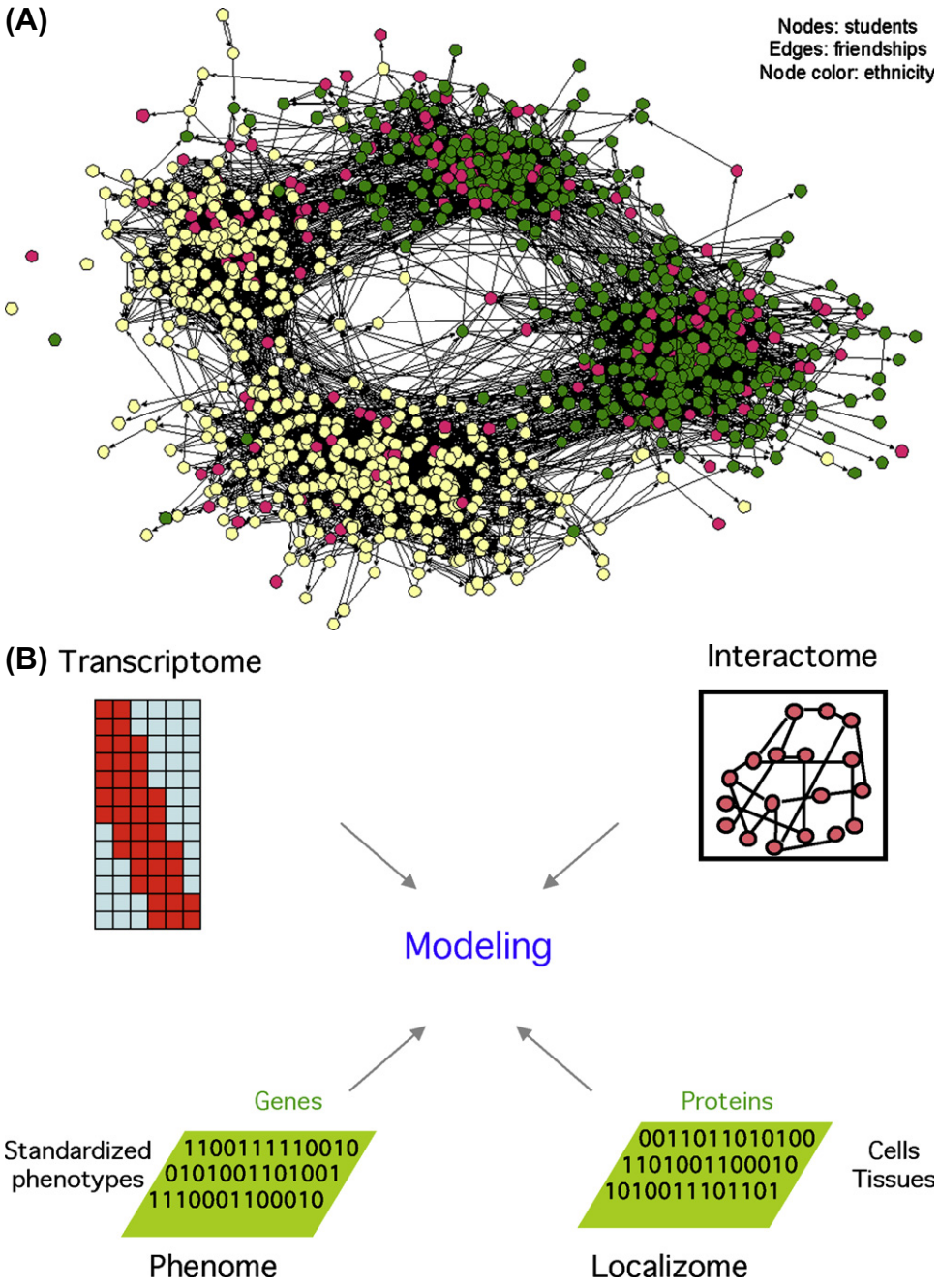

FIGURE 3.4 **Drawing inferences from complex networks through modeling.**

experimental evidence that directly relates to the protein interaction [84], or can include a broader range of biological relationships. Methods are emerging to score the likelihood that a given protein pair will physically interact, given all the other known single-color relationships about that protein pair. Prediction strategies that model single-color relationships as being independent of one another have proved worthwhile [54,127,128]. More sophisticated approaches that explicitly model the interdependency between multiple interaction types further improve the quality of interaction predictions [128–133].

With such confidence scores, interactome networks can be represented as probabilistic networks in which each edge is assigned a weight representing the posterior probability that the edge is real. Sampling many 'deterministic' subgraphs from the probabilistic network measures the fraction of sampled networks containing a path between

a known protein complex and a candidate protein member of that complex, thus predicting new members of partially defined protein complexes [133,134].

The topology of interactome networks can be exploited to predict novel protein interactions. Network motifs are patterns of interconnection involving more than two nodes [135,136]. Triangles and larger densely connected subgraphs are more frequent in most protein−protein interactome networks than would be expected by chance, in turn rendering candidate protein interactions that complete many new triangles in the network more likely to be true interactions [137,138]. Accordingly, identification of densely connected subgraphs in an interactome network can help identify protein complexes [108,139−142]. Other methods exploiting network topology parameters have proved useful in predicting physical interaction networks of other organisms [54].

A promising strategy for interaction prediction is to produce multicolor network motifs derived by integrating protein−protein interactome networks with genetic interaction networks or phenotypic profile similarity networks (see Chapters 5 and 6). Protein interactions tend to connect genes with related phenotypes, as was first discerned in small-scale studies of protein interactions, and later demonstrated for large-scale interactome maps and systematic phenotyping data [50,143−147]. It follows that genetic interaction profiles can be used to predict protein complexes [18,143,146,148−150]. On the basis of the strength of the physical and genetic interactions of a particular protein pair it is possible to assess the likelihood of that protein pair operating either within a protein complex or connecting two functionally related protein complexes [151]. The predictive power of these integrative approaches lies in the systems organization through which interactomes underlie genotype−phenotype relationships.

Predicting Gene Functions, Phenotypes and Disease Associations

Early in the implementation of Y2H, Raf kinase was imputed as an oncogene based on its specific interaction with H-Ras [152]. This type of reasoning is behind the principles of ‘guilt-by-association’ and ‘guilt-by-profiling’, whereby a functional annotation can be transferred from one gene/protein to another ‘across’ biological relationships, or ‘across’ profiles of biological relationships [153−157]. Function prediction methods based on these principles either make assumptions about the independence of evidence types, or model the interdependencies between edge types [158]. Protein A, of unknown function, can be said with some likelihood to be involved in the same biological process as protein B, of known function, if A and B belong to the same protein complex, or if their profile of membership in modules identified from the interactome network are similar [141]. Protein interactions, mostly transferred by orthology from human, allow the most precise predictions of gene function in mouse genes among several data types [159].

Topological modules in interactome networks most likely correspond to specific biological processes or functions [160]. It follows that identifying modules containing genes/proteins of both known and unknown function can help assign function to uncharacterized genes/proteins. Combining topological modules with other genomic or proteomic data in multicolor networks brings about more biologically coherent units [128,161−165].

That protein interactions mediate protein functions, and that protein interactions tend to connect genes/proteins with related phenotypes just as they tend to connect genes/proteins with related functions, suggests that protein interactions can be used to predict new disease genes. Mutations causative for ataxia, a neurodegenerative disorder, affect proteins that share interacting partners. A subset of these shared partners have been found to be associated with neurodegeneration in animal models [59,166]. The ability of interactome maps to highlight new candidate disease genes and disease modifier genes had been anticipated in the early large-scale binary interactome maps [60,61]. With large-scale interactome maps available for human, various computational efforts systematically prioritize potential human disease genes based on the patterns of protein interactions [93,167−171]. Current efforts to predict human gene function and disease phenotype are now striving to combine several orthogonal large-scale genomic and proteomic data types [172,173].

Progress in this area will depend increasingly on efforts to establish benchmark data to allow rigorous comparisons among the evolving methods. Benchmarking has happened preliminarily for gene function prediction [159]. For prediction of phenotype or disease most current methods rely upon a handful of known ‘training examples’ − small sets of genes known to be associated with the phenotype or disease. This strategy has its worth, but eventually methods that can predict disease genes in the absence of known examples will also be needed. Genome-wide association (GWA) studies provide an emerging example where predictions can be attempted without training examples. Although GWA studies serve to identify a genomic locus associated with a disease, they often cannot pinpoint which single gene of several or many resident within the locus is the actual disease gene. Where multiple GWA loci are linked to a single disorder or trait, the subset of genes that exhibit between-locus protein interactions, or other types of biological relationships, may be the most likely to be the causal disease genes [168,169,174].

It may be possible to predict phenotypes and prioritize disease genes based on local network topology alone. One

particularly informative and well-studied network topology parameter is node connectivity, or 'degree'. Protein–protein interactome networks follow a power-law distribution, where most nodes have a low connectivity and a few nodes, the network 'hubs', have a high connectivity (see Chapter 9). This 'scale-free' degree distribution is also observed in many other networks, the internet being a particularly noteworthy example, and has important consequences for network robustness. The overall structure of a scale-free network is hardly affected by the removal of random nodes, but is highly sensitive to removal of hubs [175]. This scale-free behavior has clear applications in the design of electrical power grids, but what does it mean for protein–protein interactome networks?

About a decade ago, it was reported that hubs in the protein–protein interactome of *S. cerevisiae* are preferentially essential, meaning that knockouts of the corresponding genes tend to be unviable [176]. This initial observation was biased by the inclusion of literature-curated interactome maps, where 'star' proteins that have an artificially high connectivity also tend to be essential [39,176]. There is instead unbiased evidence of a correlation between degree and essentiality in co-complex interactome networks, but not in binary interactome networks [39,177]. Deeper examination clearly demonstrated a correlation between connectivity in an interactome network map and functional pleiotropy in *S. cerevisiae* [39]. Thus, the connectivity of a protein does relay information about the phenotype of its corresponding gene. Together with other topological information, network connectivity may one day be used to predict new disease genes, as suggested by the observation that proteins associated with cancer are preferentially hubs in the human interactome [178].

Assigning Functions to Individual Interactions, Protein Complexes and Network Motifs

Functional genomics experiments and function prediction algorithms are typically designed to uncover the biological roles of genes and gene products. We argue that if interactome networks underlie genotype–phenotype relationships, then edges (protein interactions) should be associated with functions and phenotypes just as nodes (proteins) are.

Fanconi anemia (FA) is a rare chromosome instability disorder associated with congenital defects and susceptibility to cancer. Of the 13 genes genetically associated with FA, seven encode members of a core FA protein complex [179,180]. This example and others show high interconnectivity between proteins associated to a particular disease, which suggests that the disease phenotype may result from genetically induced perturbation of protein interactions [59,166,178,181–183]. Similarly, when a single gene is linked to multiple disorders, it often seems to be because distinct mutations of this gene affect specific individual interactions with different partners [94,184]. It follows that looking for 'disease interactions' rather than for 'disease genes' should assist the delineation of disease mechanisms and aid efforts to rationally interrupt disease progression.

From an interactome network perspective the effects of genetic variations are traditionally modeled as complete loss of gene products ('node removal'). While such interpretations are generally suitable for nonsense or frameshift mutations occurring early in the protein, large insertions or deletions, or complete gene knockouts, the node removal model may not readily apply to truncations that preserve specific autonomous protein domains, or to single amino acid substitutions. Such genetic variations could instead lead to perturbations of specific interactions ('edge removal'), or 'edgetic perturbations' [184,185].

The systematic isolation of genetic variants associated with edgetic perturbations, or edgetic alleles, and their characterization *in vivo*, represent a promising strategy for investigating the function of specific interactions, particularly with regard to human disease [185]. Two complementary strategies, 'forward and reverse edgetics', reminiscent of the time-tested dichotomy of forward and reverse genetics [186], allow systematic investigation of the phenotypic outcomes of perturbations of specific binary protein interactions [187]. Taking a set of mutations in a gene associated with particular phenotypes, such as disease-associated mutations, the forward edgetics approach uses Y2H to determine the interaction defects of proteins where the mutations have been introduced by site-directed mutagenesis [184]. Reciprocally, reverse edgetics starts from a set of interactions for a protein of interest, and aims to systematically isolate alleles encoding proteins with desired specific interaction defects by reverse Y2H selections [44,68,188]. The edgetic alleles that are thus selected can be reintroduced *in vivo* into a model organism to investigate the phenotypic consequences of specifically altering the corresponding molecular interaction(s) [185].

Besides individual edges, higher-level topological structures such as network motifs can also be associated to specific biological functions [135,136]. Different types of networks exhibit distinctive profiles of the relative abundance of network motifs, so network motif profiles can be used to characterize and compare networks. Neuron networks or regulatory networks are enriched in feedforward loops, whereas food webs are enriched in bi-parallel motifs [135,136,189]. These distinctive enrichments suggest that interactions that are part of

feedforward loops are likely to be involved in information processing, while interactions that are part of bi-parallel motifs are likely to be involved in energy transfer. We anticipate that analytical tools utilizing network motifs, particularly multicolor network motifs, will help in deciphering the function of many interactome network edges and local structures.

Multicolor triangle motifs containing two nodes linked by protein interactions with both of these nodes connected to a third node by a genetic interaction appear enriched in the *S. cerevisiae* interactome [24]. Consequently, if A, B and C are three genes such that the translation products of A and B physically interact, and the A and C genes are linked by a genetic interaction, then a genetic interaction between B and C can be predicted [131]. This particular motif suggests a 'compensatory complex' theme wherein two proteins/complexes/processes function in parallel. An excellent example of compensatory complexes is the pairing of endoplasmic reticulum (ER) protein-translocation sub-complex [190] and the Gim complex [191], with each complex connected to the other by multiple genetic interactions [24,150]. The Gim complex facilitates the folding of actin and tubulin components of the cytoskeleton. The genetic interactions between the Gim complex and ER protein translocation suggests that defects in moving proteins into the ER are ameliorated by a fully functioning cytoskeleton, whereas the trafficking of protein via lipid vesicles requires the cytoskeleton to act as a 'molecular train track'.

TOWARDS DYNAMIC INTERACTOMES

Towards Cell-Type and Condition-Specific Interactomes

The cell interior is a constantly changing environment. Biomolecules and cellular processes respond dynamically to intra- and extracellular cues. Available protein−protein interactome maps are, regrettably, mostly static, representing the union of protein interactions that may occur in all locations, times and environments. Analysis of protein−protein interactome networks will continue to contribute profound insights to systems biology only by reaching the temporal and spatial resolution necessary to dynamically model coordination of biological processes across the cell and the organism.

For a protein complex to be active at the right time and place in the cell, and at a controlled concentration, the cell has to undertake a large number of parallel and successive decisions. For each complex subunit, the cognate gene needs to be transcribed (chromatin opening, initiation and elongation of transcription …) and the mRNA processed, exported, and translated. Complex subunits often also need to be post-translationally modified, controlled in quality and quantity by chaperones, and actively targeted to the required site of action. There, protein complex assembly can require a specific order of addition to reach stability. Because each step leading to a protein complex is potentially subject to regulation, co-complex interactome networks are dense with accumulated information on the cell dynamics.

Interrogating the dynamics of complex assembly at proteome scale is not yet feasible experimentally. It is, however, becoming possible to compare proteome expression across cell types, thanks to technological innovations developed throughout the last decade, such as stable isotope labeling by amino acids in cell culture [192] (see Chapter 1). We can now interrogate interactome networks for node dynamics (at least partially), but not yet for edge dynamics. The first dynamic measures of protein complex membership successfully followed a single protein, GRB2, as it dynamically associated with multiple complexes [193]. Pending increases in throughput and further advances, the modeling of co-complex interactome network dynamics will need to rely on computational analyses. Empirical measurements of binary interaction dynamics are also lacking. The LUMIER technology has paved the way [194], but most binary interactome maps remain static, and attempts at dynamical modeling also rely on computational analyses.

Computational integration of interactome maps and expression profiles can identify biological conditions whereby two proteins that can interact, according to an interactome map, are also co-present, according to their expression profiles. This additional knowledge allows the inference of spatiotemporal 'interaction territories' marking where or when the interaction can take place, e.g., during cell cycle or organism development [52,122]. To what extent is the expression of interacting proteins transcriptionally coordinated in cellular systems? Physically interacting proteins are more likely to exhibit similar expression patterns than would be expected by chance [39,129,195]. Most interacting proteins are not co-expressed, however, and some pairs are even anti-correlated in expression. Interactome dynamics therefore appear to be under tight transcriptional control, with most protein interactions being transient.

Transient protein interactions involved in signaling and intercomplex connections are enriched in binary interactome maps compared to co-complex interactome maps [39]. Members of a given protein complex can be co-regulated by a common transcription factor, when a transcription factor is connected by transcription regulatory edges directed towards several interacting and co-expressed proteins, forming 'regulonic complexes' [24,196,197]. In response to extracellular perturbations, protein complexes generally remain stable, but the functional connections between these complexes are

substantially reorganized, as reflected by genetic interaction changes [151,198,199].

By overlaying transcriptome patterns with a binary interaction network, Han and colleagues discovered two types of highly connected hub protein. On the one hand, 'party' hubs are strongly co-expressed with most of their interacting partners; on the other hand, 'date' hubs connect to different partners at different times or contexts [200]. These interactome dynamics differences are reflected in the structures of hub proteins from the two groups. Relative to date hubs, where different interacting partners may utilize the same surface of interaction at different times, party hubs tend to contain less disordered regions and to display more interaction interfaces at their surface, as would be expected for proteins with many simultaneous interacting partners [201–203]. The observation that date hubs are more strongly associated with breast cancer phenotypes was used to develop a co-expression signature that strongly differentiated breast cancer patients on the basis of disease outcome [204].

Despite these initial successes, interactome dynamics modeling will need to move beyond computational analyses. Cell-type-specific transcriptome data may provide intuitive approximation of protein expression levels, but such estimations are bound to be imprecise. Detection of a transcript does not necessarily imply that the corresponding protein is present and stable, and the absence of a transcript does not necessarily imply absence of the corresponding protein, as proteins can remain stable long past transcript degradation and can transit from cell to cell. Relative protein concentrations must also be considered when modeling interactome dynamics. Protein concentrations influence the affinity of proteins for one another due to mass action, and the effect of cell crowding on protein–protein interactomes remains unexplored. The intracellular environment also affects protein interactions. Proteins can be restricted to particular organelles bound by membranes, as are the nuclei or mitochondria, or localized in less sharply delimited regions such as nucleoli. Cellular localization data are available at genomic scale for several organisms [205,206], but information about the dynamic movements of proteins across cellular compartments is lacking.

These caveats limit the scope of computational approaches in modeling interactome dynamics. Experimentally measuring protein–protein interactomes at high resolution both in space, across subcellular locations and across cell and tissue types, and in time, for example through the course of development, may still appear a distant goal, but this goal deserves to be actively pursued. Conversely, evidence that the expression of interacting proteins is tightly regulated shows that co-expression should not be used as a benchmark for protein interaction reliability.

Evolutionary Dynamics of Protein–Protein Interactome Networks

A central hypothesis of systems biology is that genotype–phenotype relationships are mediated through physical and functional interactions between genes and gene products that form intricate molecular networks within cells. Genotype–phenotype relationships are also governed by natural selection. Hence, understanding the principles driving the evolution of molecular networks would contribute to a deeper understanding of Life. Is there a 'core interactome' shared by every form of Life? Do the constraints of natural selection enforce constraints on interactome network structure? Do interactomes grow over evolutionary time? Does interactome complexity scale with organism complexity? Are interactomes more stable or more variable than genomes? Given these fundamental questions [207], it is not surprising that the evolutionary dynamics of protein–protein interactome networks have been a focus of investigation ever since the first large-scale protein–protein interactome maps appeared.

If protein–protein interactomes were evolutionarily stable systems, interactions between orthologous protein pairs from distinct species should be largely conserved. However, the observed fraction of interactions corresponding to such 'interolog' pairs is consistently low across several species [53,208–211]. The incompleteness of available interactome maps, and/or the difficulties of orthology mapping, may explain these apparently low proportions of conserved interactions. Still, as even these low proportions would not be expected at random, they are consistent with natural selection acting on the conservation of at least a subset of interactions throughout evolutionary time. A complementary interpretation would be that protein–protein interactomes are evolutionarily dynamic systems, constantly changing under the action of natural selection.

Cross-species comparisons indicate that $\sim 10^{-5}$ interactions are lost or gained per protein pair per million years, leaving aside the interactome remodeling that necessarily follows gene death and gene birth events [25,212,213]. This corresponds to approximately 10^3 interaction changes per 10^6 years in the evolution of the human lineage. Different types of protein interaction are rewired at different rates. Transient interactions appear more evolutionarily volatile than the more lasting interactions forming protein complexes [213–215], and protein–peptide interactions appear to change more rapidly than interactions between long proteins [212]. Evolutionary variation is observed even for protein complexes participating in the cell cycle. These complexes are globally conserved across several yeast species, but differ in their regulatory subunit composition and timing of assembly [216]. Incidentally, this dynamic rewiring of interactome networks during evolution is bound to limit the reliability of predictions of

protein interactions based on sequence similarity measures across species [217].

Single amino acid changes can result in edgetic perturbations of protein–protein interactome networks [184]. Fixation of such sequence changes under selective constraints is expected to shape protein interaction interfaces. In agreement, the sequences of hub proteins appear under tighter constraints than the sequences of non-hub proteins, with intra-module hubs significantly more constrained than inter-module hubs [202,218,219]. The yeast protein Pbs2, which endogenously interacts specifically with a single yeast SH3 domain, is able to promiscuously interact with many non-yeast SH3 domains [220]. At an equivalent level of sequence similarity, protein interactions are more conserved within species, when considering paralogous protein pairs originating from duplication events, than across species when considering orthologous protein pairs [217]. It seems that tinkering with interaction interfaces and specificity causes protein interactions to co-evolve dynamically within biological systems.

Immediately following gene duplication events, paralogous proteins are expected to have identical protein sequences and to share all of their interactors. Empirical observations have revealed that the fraction of interactors shared by paralogous proteins decreases over evolutionary time, likely reflecting the well-described functional divergence of retained paralogous proteins [221]. Such evolutionary dynamics may explain the origin of the scale-free degree distribution that protein–protein interactome networks invariably follow, via an evolutionary version of the 'rich-get-richer' principle [222,223] (see Chapter 9). These evolutionary dynamics may also lead to an elevated clustering in protein–protein interactome networks if self-interactions are taken into account, as their duplication enables the formation of novel complexes of paralogous proteins [224,225] (see Chapter 9). The proteasome complex likely evolved from an ancestral homodimeric interaction through multiple successive duplication events [25,226].

Attempts to estimate the interactome rewiring rate following duplication events have yielded conflicting results [212,213,221–223,227–230]. These contradictions may be reconciled by a model according to which rewiring does not occur at a constant rate, but rather in a rapid-then-slow fashion [55]. Similar rapid-then-slow dynamics characterize protein sequence divergence following duplication events, likely reflecting relaxed-then-tight selective constraints on the function of the duplicated proteins. Signatures of neo-functionalization, sub-functionalization and asymmetric edge-specific divergence have been observed in protein–protein interactome networks [219,231–233]. Edgetic rewiring of protein– protein interactome networks following duplication events thus appears associated with the Darwinian selection of the functions of the corresponding proteins.

How to reconcile the dynamic rewiring of protein–protein interactome networks with the existence of universal processes found within all forms of Life? Beyond the union of individual interactions, interactome networks exhibit higher-level organizational properties, such as signaling pathways, or other types of functional module. Several pathways and modules appear evolutionarily conserved, as measured by orthology-based network alignment algorithms [234,235]. Similarly, topology-based network alignment algorithms have revealed considerable similarities in the local wiring of cellular networks across evolutionarily distant organisms [236]. The global network topology of binary interactomes of organisms as diverse as humans, plants, worms and yeasts appear qualitatively similar, characterized by a scale-free distribution of degrees and small-world structures [55] (see Chapter 9). Likewise, the estimated ratio of interacting pairs among all possible protein pairs in these organisms, with genomes encoding anywhere from 6000 to 30 000 proteins, appears surprisingly stable, with 5–10 interactions per 10 000 protein pairs [55]. It is possible that these high-level systems properties are ultimately the object of evolutionary conservation and so unify all forms of Life.

In summary, natural selection seems to shape the dynamic evolution of protein–protein interactome networks. Regulatory interactome networks seem to evolve faster than protein–protein interactome networks [213,228]. More refined models of the evolution of biological systems, including population size effects and the concept of 'genotype networks', are being investigated [237] (see chapter by A. Wagner). Life could be perceived as a system containing genotypes and phenotypes, with genotypes shaping phenotypes through the prism of interactomes, and phenotypes shaping genotypes through the feedback of evolution by natural selection.

CONCLUDING REMARKS

It is becoming increasingly clear that protein–protein interactome mapping and modeling will be key to understanding cellular systems and genotype–phenotype relationships. In this chapter we have described the state-of-the-art experimental and computational strategies currently used to detect and predict binary and co-complex protein interactions at proteome scale, and outlined the major achievements of the field so far. We covered the new concepts that will need to emerge, and the new technologies that will need to be developed, so that complete reference protein–protein interactome maps can materialize for several organisms in the near future. With such maps in hand, the principles governing interactome dynamics will be deciphered and causal paths between genotype and phenotype will be drawn.

An additional layer of complexity lies in the number of protein isoforms resulting from alternative splicing of transcripts for each individual gene. Protein−protein interactome maps available so far have mostly disregarded isoforms, opting for a gene-centered approach for simplicity and because differentiating between protein isoforms is technically challenging. Isoforms of the same protein may exhibit distinct combinations of protein interaction interfaces, leading to distinct local interactome networks. It will therefore be crucial to differentiate isoforms in future interactome mapping and modeling. This will undoubtedly shed light on protein interactome dynamics, as isoform expression is expected to be highly regulated across different cell types and conditions.

Mechanistic understanding of biological systems will also require quantitative estimation of interaction strength. To this end, systematic measures of the affinity of proteins for each other, in binary as well as in higher-order interactions, would generate a tremendous impetus to mathematically model biological processes. It will also be essential to achieve the systematic integration of three-dimensional structural data, whether derived experimentally or by computational modeling [238,239]. Eventually, three-dimensional mapping of the sequence variations found in populations and their association with traits may allow the almost seamless reconstruction of genotype−phenotype relationships through edgetic modeling of protein−protein interactomes.

ACKNOWLEDGEMENTS

We thank past and current members of the Vidal Lab and the Center for Cancer Systems Biology (CCSB) for their help and constructive discussions over the course of developing our binary interaction mapping strategies, framework, and protocols. We thank Dr David E. Hill for insightful editing of this book chapter and Dr Robin Lee for providing an image of cells. This work was supported by National Human Genome Research Institute grants R01-HG001715 awarded to M.V. D.E.H, and F.P.R, P50-HG004233 awarded to M.V., R01-HG006061 to M.V., D.E.H. and M.E.C., RC4-HG006066 to M.V and D.E.H.; National Heart, Lung and Blood Institute grant U01-HL098166 (M.V. subcontract); National Institute of Environmental Health Sciences R01-ES015728 to M.V.; National Cancer Institute grant R33-CA132073 to M.V.; National Science Foundation PGRP grant DBI-0703905 to M.V. and D.E.H.; and by Institute Sponsored Research funds from the Dana-Farber Cancer Institute Strategic Initiative awarded to M.V. and CCSB. M.V. is a 'Chercheur Qualifié Honoraire' from the Fonds de la Recherche Scientifique (FRS-FNRS, French Community of Belgium).

REFERENCES

[1] Nurse P. The great ideas of biology. Clin Med 2003;3(6):560−8.
[2] Vidal M. A unifying view of 21st century systems biology. FEBS Lett 2009;583(24):3891−4.
[3] Delbrück M. Unités biologiques douées de continuité génétique. Paris: Editions du Centre National de la Recherche Scientifique; 1949. 33−35.
[4] Pardee AB, Yates RA. Control of pyrimidine biosynthesis in *Escherichia coli* by a feed-back mechanism. J Biol Chem 1956;221(2):757−70.
[5] Umbarger HE. Evidence for a negative-feedback mechanism in the biosynthesis of isoleucine. Science 1956;123(3202):848.
[6] Novick A, Weiner M. Enzyme induction as an all-or-none phenomenon. Proc Natl Acad Sci U S A 1957;43(7):553−66.
[7] Monod J, Jacob F. Teleonomic mechanisms in cellular metabolism, growth, and differentiation. Cold Spring Harb Symp Quant Biol 1961;26:389−401.
[8] Thomas R. Boolean formalization of genetic control circuits. J Theor Biol 1973;42(3):563−85.
[9] Waddington CH. The Strategy of the Genes. London: George Allen & Unwin; 1957.
[10] Kauffman SA. Metabolic stability and epigenesis in randomly constructed genetic nets. J Theor Biol 1969;22(3):437−67.
[11] Lander ES, Linton LM, et al. Initial sequencing and analysis of the human genome. Nature 2001;409(6822):860−921.
[12] Pettersson E, Lundeberg J, et al. Generations of sequencing technologies. Genomics 2009;93(2):105−11.
[13] Schuster SC. Next-generation sequencing transforms today's biology. Nat Methods 2008;5(1):16−8.
[14] Venter JC, Adams MD, et al. The sequence of the human genome. Science 2001;291(5507):1304−51.
[15] Beadle GW, Tatum EL. 'Genetic control of biochemical reactions in *Neurospora*. Proc Natl Acad Sci U S A 1941;27(11):499−506.
[16] Burga A, Casanueva MO, et al. Predicting mutation outcome from early stochastic variation in genetic interaction partners. Nature 2011;480(7376):250−3.
[17] Horvitz HR, Sulston JE. Isolation and genetic characterization of cell-lineage mutants of the nematode *Caenorhabditis elegans*. Genetics 1980;96(2):435−54.
[18] Giaever G, Chu AM, et al. Functional profiling of the *Saccharomyces cerevisiae* genome. Nature 2002;418(6896):387−91.
[19] Begley TJ, Rosenbach AS, et al. Damage recovery pathways in *Saccharomyces cerevisiae* revealed by genomic phenotyping and interactome mapping. Mol Cancer Res 2002;1(2):103−12.
[20] Cusick ME, Klitgord N, et al. Interactome: gateway into systems biology. Hum Mol Genet 2005;14(Spec No. 2):R171−81.
[21] Huberts DH, van der Klei IJ. Moonlighting proteins: an intriguing mode of multitasking. Biochim Biophys Acta 2010;1803(4):520−5.
[22] Jeffery CJ. Moonlighting proteins. Trends Biochem. Sci 1999;24(1):8−11.
[23] Milgram S. The small world problem. Psychol Today 1967;1(1):60−7.
[24] Zhang LV, King OD, et al. Motifs, themes and thematic maps of an integrated *Saccharomyces cerevisiae* interaction network. J Biol 2005;4(2):6.
[25] Levy ED, Pereira-Leal JB. Evolution and dynamics of protein interactions and networks. Curr Opin Struct Biol 2008;18(3):349−57.

[26] Kaushansky A, Allen JE, et al. Quantifying protein−protein interactions in high throughput using protein domain microarrays. Nat Protoc 2010;5(4):773−90.
[27] Linding R, Jensen LJ, et al. Systematic discovery of *in vivo* phosphorylation networks. Cell 2007;129(7):1415−26.
[28] Ma'ayan A, Jenkins SL, et al. Formation of regulatory patterns during signal propagation in a mammalian cellular network. Science 2005;309(5737):1078−83.
[29] Ptacek J, Devgan G, et al. Global analysis of protein phosphorylation in yeast. Nature 2005;438(7068):679−84.
[30] Roberts PM. Mining literature for systems biology. Brief Bioinform 2006;7(4):399−406.
[31] Marcotte E, Date S. Exploiting big biology: integrating large-scale biological data for function inference. Brief Bioinform 2001;2(4):363−74.
[32] Walhout AJ, Vidal M. Protein interaction maps for model organisms. Nat Rev Mol Cell Biol 2001;2(1):55−62.
[33] Cusick ME, Yu H, et al. Literature-curated protein interaction datasets. Nat Methods 2009;6(1):39−46.
[34] Turinsky AL, Razick S, et al. Literature curation of protein interactions: measuring agreement across major public databases. Database 2010 (Oxford) 2010: baq026.
[35] Turinsky AL, Razick S, et al. Interaction databases on the same page. Nat Biotechnol 2011;29(5):391−3.
[36] Venkatesan K, Rual JF, et al. An empirical framework for binary interactome mapping. Nat Methods 2009;6(1):83−90.
[37] Edwards AM, Isserlin R, et al. Too many roads not taken. Nature 2011;470(7333):163−5.
[38] Vousden KH, Lane DP. p53 in health and disease. Nat Rev Mol Cell Biol 2007;8(4):275−83.
[39] Yu H, Braun P, et al. High-quality binary protein interaction map of the yeast interactome network. Science 2008;322(5898): 104−10.
[40] Bartel PL, Roecklein JA, et al. A protein linkage map of *Escherichia coli* bacteriophage T7. Nat Genet 1996;12(1):72−7.
[41] Finley Jr RL, Brent R. Interaction mating reveals binary and ternary connections between *Drosophila* cell cycle regulators. Proc Natl Acad Sci U S A 1994;91(26):12980−4.
[42] Fromont-Racine M, Rain JC, et al. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. Nat Genet 1997;16(3):277−82.
[43] Vidal M. The reverse two-hybrid system. In: Fields S, Bartel P, editors. The yeast two-hybrid system. New York, NY: Oxford University Press; 1997. p. 109−47.
[44] Vidal M, Brachmann RK, et al. Reverse two-hybrid and one-hybrid systems to detect dissociation of protein−protein and DNA−protein interactions. Proc Natl Acad Sci U S A 1996;93(19):10315−20.
[45] Walhout AJ, Vidal M. A genetic strategy to eliminate self-activator baits prior to high-throughput yeast two-hybrid screens. Genome Res 1999;9(11):1128−34.
[46] Ito T, Chiba T, et al. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci U S A 2001;98(8):4569−74.
[47] Miller JP, Lo RS, et al. Large-scale identification of yeast integral membrane protein interactions. Proc Natl Acad Sci U S A 2005; 102(34):12123−8.
[48] Tarassov K, Messier V, et al. An *in vivo* map of the yeast protein interactome. Science 2008;320(5882):1465−70.
[49] Uetz P, Giot L, et al. A comprehensive analysis of protein−protein interactions in *Saccharomyces cerevisiae*. Nature 2000;403(6770): 623−7.
[50] Li S, Armstrong CM, et al. A map of the interactome network of the metazoan *C. elegans*. Science 2004;303(5657):540−3.
[51] Reboul J, Vaglio P, et al. *C. elegans* ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression. Nat Genet 2003;34(1):35−41.
[52] Simonis N, Rual JF, et al. Empirically controlled mapping of the *Caenorhabditis elegans* protein−protein interactome network. Nat Methods 2009;6(1):47−54.
[53] Walhout AJ, Sordella R, et al. Protein interaction mapping in *C. elegans* using proteins involved in vulval development. Science 2000;287(5450):116−22.
[54] Giot L, Bader JS, et al. A protein interaction map of *Drosophila melanogaster*. Science 2003;302(5651):1727−36.
[55] Arabidopsis interactome mapping consortium. Evidence for network evolution in an *Arabidopsis* interactome map. Science 2011;333(6042):601−7.
[56] Lalonde S, Sero A, et al. A membrane protein/signaling protein interaction network for *Arabidopsis* version AMPv2. Front Physiol 2010;1:24.
[57] Mukhtar MS, Carvunis AR, et al. Independently evolved virulence effectors converge onto hubs in a plant immune system network. Science 2011;333(6042):596−601.
[58] Colland F, Jacq X, et al. Functional proteomics mapping of a human signaling pathway. Genome Res 2004;14(7):1324−32.
[59] Lim J, Hao T, et al. A protein−protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. Cell 2006;125(4):801−14.
[60] Rual JF, Venkatesan K, et al. Towards a proteome-scale map of the human protein−protein interaction network. Nature 2005; 437(7062):1173−8.
[61] Stelzl U, Worm U, et al. A human protein−protein interaction network: a resource for annotating the proteome. Cell 2005; 122(6):957−68.
[62] Yu H, Tardivo L, et al. Next-generation sequencing to generate interactome datasets. Nat Methods 2011;8(6):478−80.
[63] Zhu H, Bilgin M, et al. Global analysis of protein activities using proteome chips. Science 2001;293(5537):2101−5.
[64] Dreze M, Monachello D, et al. High-quality binary interactome mapping. Methods Enzymol 2010;470:281−315.
[65] Fields S, Song O. A novel genetic system to detect protein-protein interactions. Nature 1989;340(6230):245−6.
[66] Durfee T, Becherer K, et al. The retinoblastoma protein associates with the protein phosphatase type 1 catalytic subunit. Genes Dev 1993;7(4):555−69.
[67] Gyuris J, Golemis E, et al. Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2. Cell 1993;75(4):791−803.
[68] Vidal M, Braun P, et al. Genetic characterization of a mammalian protein−protein interaction domain by using a yeast reverse two-hybrid system. Proc Natl Acad Sci U S A 1996;93(19):10321−6.
[69] Walhout AJ, Vidal M. High-throughput yeast two-hybrid assays for large-scale protein interaction mapping. Methods 2001;24(3): 297−306.
[70] Armstrong CM, Li S, Vidal M. Modular scale yeast two-hybrid screening. In: Celis JE, editor. Cell biology: A laboratory handbook, 3rd ed. San Diego, CA: Academic Press; 2005. p. IV. 39. 1−9.

[71] Walhout AJ, Boulton SJ, et al. Yeast two-hybrid systems and protein interaction mapping projects for yeast and worm. Yeast 2000;17(2):88−94.
[72] Grove CA, De Masi F, et al. A multiparameter network reveals extensive divergence between *C. elegans* bHLH transcription factors. Cell 2009;138(2):314−27.
[73] Lamesch P, Li N, et al. hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes. Genomics 2007;89(3):307−15.
[74] Rual JF, Hirozane-Kishikawa T, et al. Human ORFeome version 1.1: a platform for reverse proteomics. Genome Res 2004; 14(10B):2128−35.
[75] D'Haeseleer P, Church GM. Estimating and improving protein interaction error rates. Proc IEEE Comput Syst Bioinform Conf 2004:216−23.
[76] Grigoriev A. On the number of protein−protein interactions in the yeast proteome. Nucleic Acids Res 2003;31(14):4157−61.
[77] Hart GT, Ramani AK, et al. How complete are current yeast and human protein-interaction networks? Genome Biol 2006; 7(11):120.
[78] Huang H, Jedynak BM, et al. Where have all the interactions gone? Estimating the coverage of two-hybrid protein interaction maps. PLoS Comput Biol 2007;3(11):e214.
[79] Reguly T, Breitkreutz A, et al. Comprehensive curation and analysis of global interaction networks in *Saccharomyces cerevisiae*. J Biol 2006;5(4):11.
[80] Sambourg L, Thierry-Mieg N. New insights into protein−protein interaction data lead to increased estimates of the *S. cerevisiae* interactome size. BMC Bioinformatics 2010;11:605.
[81] Sprinzak E, Sattath S, et al. How reliable are experimental protein−protein interaction data? J Mol Biol 2003;327(5):919−23.
[82] Stumpf MP, Thorne T, et al. Estimating the size of the human interactome. Proc Natl Acad Sci U S A 2008;105(19):6959−64.
[83] Worseck JM, Grossmann A, et al. A stringent yeast two-hybrid matrix screening approach for protein−protein interaction discovery. Methods Mol Biol 2012;812:63−87.
[84] Braun P, Tasan M, et al. An experimentally derived confidence score for binary protein−protein interactions. Nat Methods 2009; 6(1):91−7.
[85] Chen YC, Rajagopala SV, et al. Exhaustive benchmarking of the yeast two-hybrid system. Nat Methods 2010;7(9):667−8.
[86] Alberts B. The cell as a collection of protein machines: preparing the next generation of molecular biologists. Cell 1998;92(3): 291−4.
[87] Gavin AC, Bosche M, et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 2002;415(6868):141−7.
[88] Kuriyan J, Eisenberg D. The origin of protein interactions and allostery in colocalization. Nature 2007;450(7172):983−90.
[89] Williamson JR. Cooperativity in macromolecular assembly. Nat Chem Biol 2008;4(8):458−65.
[90] Fraser HB, Plotkin JB. Using protein complexes to predict phenotypic effects of gene mutation. Genome Biol 2007;8(11): R252.
[91] Gavin AC, Aloy P, et al. Proteome survey reveals modularity of the yeast cell machinery. Nature 2006;440(7084):631−6.
[92] Ideker T, Sharan R. Protein networks in disease. Genome Res 2008;18(4):644−52.
[93] Lage K, Karlberg EO, et al. A human phenome-interactome network of protein complexes implicated in genetic disorders. Nat Biotechnol 2007;25(3):309−16.
[94] Wang PI, Marcotte EM. It's the machine that matters: predicting gene function and phenotype from protein networks. J Proteomics 2010;73(11):2277−89.
[95] Beck M, Schmidt A, et al. The quantitative proteome of a human cell line. Mol Syst Biol 2011;7:549.
[96] Ghaemmaghami S, Huh WK, et al. Global analysis of protein expression in yeast. Nature 2003;425(6959):737−41.
[97] Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature 2003;422(6928):198−207.
[98] Gingras AC, Gstaiger M, et al. Analysis of protein complexes using mass spectrometry. Nat Rev Mol Cell Biol 2007;8(8): 645−54.
[99] Kocher T, Superti-Furga G. Mass spectrometry-based functional proteomics: from molecular machines to protein networks. Nat Methods 2007;4(10):807−15.
[100] Puig O, Caspary F, et al. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. Methods 2001;24(3):218−29.
[101] Burckstummer T, Bennett KL, et al. An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells. Nat Methods 2006;3(12):1013−9.
[102] de Boer E, Rodriguez P, et al. Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. Proc Natl Acad Sci U S A 2003; 100(13):7480−5.
[103] Glatter T, Wepf A, et al. An integrated workflow for charting the human interaction proteome: insights into the PP2A system. Mol Syst Biol 2009;5:237.
[104] Ho Y, Gruhler A, et al. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. Nature 2002;415(6868):180−3.
[105] Poser I, Sarov M, et al. BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. Nat Methods 2008;5(5):409−15.
[106] Rigaut G, Shevchenko A, et al. A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotechnol 1999;17(10):1030−2.
[107] Guruharsha KG, Rual JF, et al. A protein complex network of *Drosophila melanogaster*. Cell 2011;147(3):690−703.
[108] Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 2003;4:2.
[109] Choi H, Larsen B, et al. SAINT: probabilistic scoring of affinity purification-mass spectrometry data. Nat Methods 2011;8(1): 70−3.
[110] Hart GT, Lee I, et al. A high-accuracy consensus map of yeast protein complexes reveals modular nature of gene essentiality. BMC Bioinformatics 2007;8:236.
[111] Hu P, Janga SC, et al. Global functional atlas of *Escherichia coli* encompassing previously uncharacterized proteins. PLoS Biol 2009;7(4):e1000096.
[112] Kuhner S, van Noort V, et al. Proteome organization in a genome-reduced bacterium. Science 2009;326(5957):1235−40.
[113] Behrends C, Sowa ME, et al. Network organization of the human autophagy system. Nature 2010;466(7302):68−76.

[114] Bouwmeester T, Bauch A, et al. A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway. Nat Cell Biol 2004;6(2):97–105.
[115] Brehme M, Hantschel O, et al. Charting the molecular network of the drug target Bcr-Abl. Proc Natl Acad Sci U S A 2009;106(18): 7414–9.
[116] Ewing RM, Chu P, et al. Large-scale mapping of human protein–protein interactions by mass spectrometry. Mol Syst Biol 2007;3:89.
[117] Sowa ME, Bennett EJ, et al. Defining the human deubiquitinating enzyme interaction landscape. Cell 2009;138(2): 389–403.
[118] Vermeulen M, Eberl HC, et al. Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell 2010;142(6):967–80.
[119] Orchard S, Kerrien S, et al. Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. Nat Methods 2012;9(4):345–50.
[120] Aloy P, Bottcher B, et al. 'Structure-based assembly of protein complexes in yeast. Science 2004;303(5666):2026–9.
[121] Bork P, Jensen LJ, et al. Protein interaction networks from yeast to human. Curr Opin Struct Biol 2004;14(3):292–9.
[122] de Lichtenberg U, Jensen LJ, et al. Dynamic complex formation during the yeast cell cycle. Science 2005;307(5710): 724–7.
[123] Schadt EE, Sachs A, et al. Embracing complexity, inching closer to reality. Sci STKE 2005;295:e40.
[124] Moody J. Race, school integration, and friendship segregation in America. Am J Sociol 2001;107(3):679–716.
[125] Vidal M. A biological atlas of functional maps. Cell 2001;104(3): 333–9.
[126] Vidal M. Interactome modeling. FEBS Lett 2005;579(8):1834–8.
[127] Bader JS, Chaudhuri A, et al. Gaining confidence in high-throughput protein interaction networks. Nat Biotechnol 2004;22(1):78–85.
[128] Jansen R, Yu H, et al. A Bayesian networks approach for predicting protein–protein interactions from genomic data. Science 2003;302(5644):449–53.
[129] Jansen R, Greenbaum D, et al. Relating whole-genome expression data with protein–protein interactions. Genome Res 2002;12(1): 37–46.
[130] Qi Y, Klein-Seetharaman J, et al. Random forest similarity for protein–protein interaction prediction from multiple sources. Pac Symp Biocomput 2005:531–42.
[131] Wong SL, Zhang LV, et al. Combining biological networks to predict genetic interactions. Proc Natl Acad Sci U S A 2004; 101(44):15682–7.
[132] Yan H, Venkatesan K, et al. A genome-wide gene function prediction resource for *Drosophila melanogaster*. PLoS ONE 2010;5(8):e12139.
[133] Zhang LV, Wong SL, et al. Predicting co-complexed protein pairs using genomic and proteomic data integration. BMC Bioinformatics 2004;5:38.
[134] Asthana S, King OD, et al. Predicting protein complex membership using probabilistic network reliability. Genome Res 2004; 14(6):1170–5.
[135] Milo R, Shen-Orr S, et al. Network motifs: simple building blocks of complex networks. Science 2002;298(5594):824–7.
[136] Shen-Orr SS, Milo R, et al. Network motifs in the transcriptional regulation network of *Escherichia coli*. Nat Genet 2002;31(1): 64–8.
[137] Albert I, Albert R. Conserved network motifs allow protein–protein interaction prediction. Bioinformatics 2004;20(18): 3346–52.
[138] Goldberg DS, Roth FP. Assessing experimentally derived interactions in a small world. Proc Natl Acad Sci U S A 2003;100(8): 4372–6.
[139] King OD. Comment on 'Subgraphs in random networks'. Phys Rev E Stat Nonlin Soft Matter Phys 2004;70(5 Pt 2):058101.
[140] Li W, Liu Y, et al. Dynamical systems for discovering protein complexes and functional modules from biological networks. IEEE/ACM Trans Comput Biol Bioinform 2007;4(2):233–50.
[141] Rives AW, Galitski T. Modular organization of cellular networks. Proc Natl Acad Sci U S A 2003;100(3):1128–33.
[142] Spirin V, Mirny LA. Protein complexes and functional modules in molecular networks. Proc Natl Acad Sci U S A 2003;100(21): 12123–8.
[143] Boulton SJ, Gartner A, et al. Combined functional genomic maps of the *C. elegans* DNA damage response. Science 2002;295(5552): 127–31.
[144] Dezso Z, Oltvai ZN, et al. Bioinformatics analysis of experimentally determined protein complexes in the yeast *Saccharomyces cerevisiae*. Genome Res 2003;13(11):2450–4.
[145] Ge H, Walhout AJ, et al. Integrating 'omic' information: a bridge between genomics and systems biology. Trends Genet 2003; 19(10):551–60.
[146] Gunsalus KC, Ge H, et al. Predictive models of molecular machines involved in *Caenorhabditis elegans* early embryogenesis. Nature 2005;436(7052):861–5.
[147] Walhout AJ, Reboul J, et al. Integrating interactome, phenome, and transcriptome mapping data for the *C. elegans* germline. Curr Biol 2002;12(22):1952–8.
[148] Costanzo M, Baryshnikova A, et al. The genetic landscape of a cell. Science 2010;327(5964):425–31.
[149] Piano F, Schetter AJ, et al. Gene clustering based on RNAi phenotypes of ovary-enriched genes in *C. elegans*. Curr Biol 2002;12(22):1959–64.
[150] Tong AH, Lesage G, et al. Global mapping of the yeast genetic interaction network. Science 2004;303(5659):808–13.
[151] Bandyopadhyay S, Kelley R, et al. 'Functional maps of protein complexes from quantitative genetic interaction data. PLoS Comput Biol 2008;4(4):e1000065.
[152] Vojtek AB, Hollenberg SM, et al. Mammalian Ras interacts directly with the serine/threonine kinase Raf. Cell 1993;74(1):205–14.
[153] Karaoz U, Murali TM, et al. Whole-genome annotation by using evidence integration in functional-linkage networks. Proc Natl Acad Sci U S A 2004;101(9):2888–93.
[154] Lee I, Date SV, et al. A probabilistic functional network of yeast genes. Science 2004;306(5701):1555–8.
[155] Letovsky S, Kasif S. Predicting protein function from protein/protein interaction data: a probabilistic approach. Bioinformatics 2003;19(Suppl. 1):i197–204.
[156] Oliver S. Guilt-by-association goes global. Nature 2000; 403(6770):601–3.
[157] Schwikowski B, Uetz P, et al. A network of protein–protein interactions in yeast. Nat Biotechnol 2000;18(12):1257–61.

[158] Tian W, Zhang LV, et al. Combining guilt-by-association and guilt-by-profiling to predict *Saccharomyces cerevisiae* gene function. Genome Biol 2008;9(Suppl. 1):S7.
[159] Pena-Castillo L, Tasan M, et al. A critical assessment of *Mus musculus* gene function prediction using integrated genomic evidence. Genome Biol 2008;9(Suppl. 1):S2.
[160] Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet 2004;5(2):101−13.
[161] Bader GD, Hogue CW. Analyzing yeast protein−protein interaction data obtained from different sources. Nat Biotechnol 2002;20(10):991−7.
[162] Bar-Joseph Z, Gerber GK, et al. Computational discovery of gene modules and regulatory networks. Nat Biotechnol 2003; 21(11):1337−42.
[163] Ihmels J, Friedlander G, et al. Revealing modular organization in the yeast transcriptional network. Nat Genet 2002;31(4): 370−7.
[164] Stuart JM, Segal E, et al. A gene-coexpression network for global discovery of conserved genetic modules. Science 2003; 302(5643):249−55.
[165] Tornow S, Mewes HW. Functional modules by relating protein interaction networks and gene expression. Nucleic Acids Res 2003;31(21):6283−9.
[166] Lim J, Crespo-Barreto J, et al. Opposing effects of polyglutamine expansion on native protein complexes contribute to SCA1. Nature 2008;452(7188):713−8.
[167] Aerts S, Lambrechts D, et al. Gene prioritization through genomic data fusion. Nat Biotechnol 2006;24(5):537−44.
[168] Franke L, van Bakel H, et al. Reconstruction of a functional human gene network, with an application for prioritizing positional candidate genes. Am J Hum Genet 2006;78(6): 1011−25.
[169] George RA, Liu JY, et al. Analysis of protein sequence and interaction data for candidate disease gene prediction. Nucleic Acids Res 2006;34(19):e130.
[170] Oti M, Snel B, et al. Predicting disease genes using protein−protein interactions. J Med Genet 2006;43(8):691−8.
[171] Tiffin N, Adie E, et al. Computational disease gene identification: a concert of methods prioritizes type 2 diabetes and obesity candidate genes. Nucleic Acids Res 2006;34(10): 3067−81.
[172] Lee I, Blom UM, et al. Prioritizing candidate disease genes by network-based boosting of genome-wide association data. Genome Res 2011;21(7):1109−21.
[173] Tasan M, Drabkin HJ, et al. A resource of quantitative functional annotation for *Homo sapiens* genes. G3 (Bethesda) 2012;2(2): 223−33.
[174] Raychaudhuri S, Plenge RM, et al. Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. PLoS Genet 2009;5(6):e1000534.
[175] Albert R, Jeong H, et al. Error and attack tolerance of complex networks. Nature 2000;406(6794):378−82.
[176] Jeong H, Mason SP, et al. Lethality and centrality in protein networks. Nature 2001;411(6833):41−2.
[177] Zotenko E, Mestre J, et al. Why do hubs in the yeast protein interaction network tend to be essential: reexamining the connection between the network topology and essentiality. PLoS Comput Biol 2008;4(8):e1000140.
[178] Goh KI, Cusick ME, et al. The human disease network. Proc Natl Acad Sci U S A 2007;104(21):8685−90.
[179] D'Andrea AD. Susceptibility pathways in Fanconi's anemia and breast cancer. N Engl J Med 2010;362(20):1909−19.
[180] Wang W. Emergence of a DNA-damage response network consisting of Fanconi anaemia and BRCA proteins. Nat Rev Genet 2007;8(10):735−48.
[181] Camargo LM, Collura V, et al. Disrupted in Schizophrenia 1 interactome: evidence for the close connectivity of risk genes and a potential synaptic basis for schizophrenia. Mol Psychiatry 2007;12(1):74−86.
[182] Kaltenbach LS, Romero E, et al. Huntingtin interacting proteins are genetic modifiers of neurodegeneration. PLoS Genet 2007;3(5):e82.
[183] Sakai Y, Shaw CA, et al. Protein interactome reveals converging molecular pathways among autism disorders. Sci Transl Med 2011;3(86). 86ra49.
[184] Zhong Q, Simonis N, et al. Edgetic perturbation models of human inherited disorders. Mol Syst Biol 2009;5:321.
[185] Dreze M, Charloteaux B, et al. 'Edgetic' perturbation of a *C.elegans* BCL2 ortholog. Nat Methods 2009;6(11):843−9.
[186] Griffiths A, Miller J, et al. An Introduction to Genetic Analysis. New York, W. H. Freeman; 2000.
[187] Charloteaux B, Zhong Q, et al. Protein-protein interactions and networks: forward and reverse edgetics. Methods Mol Biol 2011;759:197−213.
[188] Endoh H, Walhout AJ, et al. A green fluorescent protein-based reverse two-hybrid system: application to the characterization of large numbers of potential protein-protein interactions. Methods Enzymol 2000;328:74−88.
[189] Milo R, Itzkovitz S, et al. Superfamilies of evolved and designed networks. Science 2004;303(5663):1538−42.
[190] Hanein D, Matlack KE, et al. Oligomeric rings of the Sec61p complex induced by ligands required for protein translocation. Cell 1996;87(4):721−32.
[191] Geissler S, Siegers K, et al. A novel protein complex promoting formation of functional alpha- and gamma-tubulin. EMBO J 1998;17(4):952−66.
[192] Choudhary C, Mann M. Decoding signalling networks by mass spectrometry-based proteomics. Nat Rev Mol Cell Biol 2010; 11(6):427−39.
[193] Bisson N, James DA, et al. Selected reaction monitoring mass spectrometry reveals the dynamics of signaling through the GRB2 adaptor. Nat Biotechnol 2011;29(7):653−8.
[194] Barrios-Rodiles M, Brown KR, et al. High-throughput mapping of a dynamic signaling network in mammalian cells. Science 2005;307(5715):1621−5.
[195] Ge H, Liu Z, et al. Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. Nat Genet 2001;29(4):482−6.
[196] Simonis N, Gonze D, et al. Modularity of the transcriptional response of protein complexes in yeast. J Mol Biol 2006;363(2): 589−610.
[197] Simonis N, van Helden J, et al. Transcriptional regulation of protein complexes in yeast. Genome Biol 2004;5(5):R33.
[198] Hannum G, Srivas R, et al. Genome-wide association data reveal a global map of genetic interactions among protein complexes. PLoS Genet 2009;5(12):e1000782.

[199] Ideker T, Krogan NJ. Differential network biology. Mol Syst Biol 2012;8:565.
[200] Han JD, Bertin N, et al. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature 2004;430(6995):88–93.
[201] Ekman D, Light S, et al. What properties characterize the hub proteins of the protein–protein interaction network of *Saccharomyces cerevisiae*? Genome Biol 2006;7(6):R45.
[202] Kim PM, Sboner A, et al. The role of disorder in interaction networks: a structural analysis. Mol Syst Biol 2008;4:179.
[203] Kim SH, Yi SV. Correlated asymmetry of sequence and functional divergence between duplicate proteins of *Saccharomyces cerevisiae*. Mol Biol Evol 2006;23(5):1068–75.
[204] Taylor IW, Linding R, et al. Dynamic modularity in protein interaction networks predicts breast cancer outcome. Nat Biotechnol 2009;27(2):199–204.
[205] Huh WK, Falvo JV, et al. Global analysis of protein localization in budding yeast. Nature 2003;425(6959):686–91.
[206] Kumar A, Agarwal S, et al. Subcellular localization of the yeast proteome. Genes Dev 2002;16(6):707–19.
[207] Kiemer L, Cesareni G. Comparative interactomics: comparing apples and pears? Trends Biotechnol 2007;25(10):448–54.
[208] Cesareni G, Ceol A, et al. Comparative interactomics. FEBS Lett 2005;579(8):1828–33.
[209] Gandhi TK, Zhong J, et al. Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. Nat Genet 2006;38(3):285–93.
[210] Matthews LR, Vaglio P, et al. Identification of potential interaction networks using sequence-based searches for conserved protein–protein interactions or 'interologs'. Genome Res 2001; 11(12):2120–6.
[211] Suthram S, Sittler T, et al. The *Plasmodium* protein network diverges from those of other eukaryotes. Nature 2005;438(7064): 108–12.
[212] Beltrao P, Serrano L. Specificity and evolvability in eukaryotic protein interaction networks. PLoS Comput Biol 2007;3(2):e25.
[213] Shou C, Bhardwaj N, et al. Measuring the evolutionary rewiring of biological networks. PLoS Comput Biol 2011;7(1):e1001050.
[214] Roguev A, Bandyopadhyay S, et al. Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. Science 2008;322(5900):405–10.
[215] Teichmann SA. The constraints protein–protein interactions place on sequence divergence. J Mol Biol 2002;324(3):399–407.
[216] Jensen LJ, Jensen TS, et al. Co-evolution of transcriptional and post-translational cell-cycle regulation. Nature 2006;443(7111): 594–7.
[217] Mika S, Rost B. Protein-protein interactions more conserved within species than across species. PLoS Comput Biol 2006;2(7):e79.
[218] Fraser HB. Modularity and evolutionary constraint on proteins. Nat Genet 2005;37(4):351–2.
[219] Kim PM, Lu LJ, et al. Relating three-dimensional structures to protein networks provides evolutionary insights. Science 2006; 314(5807):1938–41.
[220] Zarrinpar A, Park SH, et al. Optimization of specificity in a cellular protein interaction network by negative selection. Nature 2003;426(6967):676–80.
[221] Wagner A. The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. Mol Biol Evol 2001;18(7):1283–92.
[222] Pastor-Satorras R, Smith E, et al. Evolving protein interaction networks through gene duplication. J Theor Biol 2003;222(2): 199–210.
[223] Vazquez A. Growing network with local rules: preferential attachment, clustering hierarchy, and degree correlations. Phys Rev E Stat Nonlin Soft Matter Phys 2003;67(5 Pt 2):056104.
[224] Musso G, Zhang Z, et al. Retention of protein complex membership by ancient duplicated gene products in budding yeast. Trends Genet 2007;23(6):266–9.
[225] Pereira-Leal JB, Levy ED, et al. Evolution of protein complexes by duplication of homomeric interactions. Genome Biol 2007; 8(4):R51.
[226] Bochtler M, Ditzel L, et al. The proteasome. Annu Rev Biophys Biomol Struct 1999;28:295–317.
[227] Ispolatov I, Krapivsky PL, et al. Duplication-divergence model of protein interaction network. Phys Rev E Stat Nonlin Soft Matter Phys 2005;71(6 Pt 1):061911.
[228] Maslov S, Sneppen K, et al. Upstream plasticity and downstream robustness in evolution of molecular networks. BMC Evol Biol 2004;4:9.
[229] Presser A, Elowitz MB, et al. The evolutionary dynamics of the *Saccharomyces cerevisiae* protein interaction network after duplication. Proc Natl Acad Sci U S A 2008;105(3):950–4.
[230] Wagner A. How the global structure of protein interaction networks evolves. Proc Biol Sci 2003;270(1514):457–66.
[231] Conant GC, Wagner A. Asymmetric sequence divergence of duplicate genes. Genome Res 2003;13(9):2052–8.
[232] He X, Zhang J. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. Genetics 2005;169(2):1157–64.
[233] Wagner A. Asymmetric functional divergence of duplicate genes in yeast. Mol Biol Evol 2002;19(10):1760–8.
[234] Sharan R, Ideker T. Modeling cellular machinery through biological network comparison. Nat Biotechnol 2006;24(4):427–33.
[235] Wuchty S, Oltvai ZN, et al. Evolutionary conservation of motif constituents in the yeast protein interaction network. Nat Genet 2003;35(2):176–9.
[236] Kuchaiev O, Milenkovic T, et al. Topological network alignment uncovers biological function and phylogeny. J R Soc Interface 2010;7(50):1341–54.
[237] Fernandez A, Lynch M. Non-adaptive origins of interactome complexity. Nature 2011;474(7352):502–5.
[238] O'Donoghue SI, Goodsell DS, et al. Visualization of macromolecular structures. Nat Methods 2010;7(Suppl. 3):S42–55.
[239] Stein A, Mosca R, et al. Three-dimensional modeling of protein interactions and complexes is going 'omics. Curr Opin Struct Biol 2011;21(2):200–8.