Interactome Networks and Human Disease

Marc Vidal[1,2,*], **Michael E. Cusick**[1,2], and **Albert-László Barabási**[1,3,4,*]

[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] Center for Complex Network Research (CCNR) and Departments of Physics, Biology and Computer Science, Northeastern University, Boston, MA 02115, USA

[4] Department of Medicine, Brigham and Women s Hospital, Harvard Medical School, Boston, MA 02115, USA

Abstract

Complex biological systems and cellular networks may underlie most genotype to phenotype relationships. Here we review basic concepts in network biology, discussing different types of interactome networks and the insights that can come from analyzing them. We elaborate on why interactome networks are important to consider in biology, how they can be mapped and integrated with each other, what global properties are starting to emerge from interactome network models, and how these properties may relate to human disease.

Introduction

Since the advent of molecular biology, considerable progress has been made in the quest to understand the mechanisms that underlie human disease, particularly for genetically inherited disorders. Genotype-phenotype relationships, as summarized in the Online Mendelian Inheritance in Man (OMIM) database (Amberger et al., 2009), include mutations in more than 3,000 human genes known to be associated with one or more of over 2,000 human disorders. This is a truly astounding number of genotype-phenotype relationships considering that a mere three decades have passed since the initial description of Restriction Fragment Length Polymorphisms (RFLPs) as molecular markers to map genetic loci of interest (Botstein et al., 1980), only two decades since the announcement of the first positional cloning experiments of disease-associated genes using RFLPs (Amberger et al., 2009), and just one decade since the release of the first reference sequences of the human genome (Lander et al., 2001; Venter et al., 2001). For complex traits, the information gathered by recent Genome Wide Association studies suggests high-confidence genotype-phenotype associations between close to 1,000 genomic loci and one or more of over one hundred diseases, including diabetes, obesity, Crohn s disease and hypertension (Altshuler et al., 2008). The discovery of genomic variations involved in cancer, inherited in the germline or acquired somatically, is equally striking, with hundreds of human genes found linked to cancer (Stratton et al., 2009). In light of new powerful technological developments such as

*Correspondence: marc_vidal@dfci.harvard.edu (M.V.), alb@neu.edu (A.L.B.).

next-generation sequencing, it is easily imaginable that a catalog of nearly all human genomic variations, whether deleterious, advantageous, or neutral, will be available within our lifetime.

Despite the natural excitement emerging from such a huge body of information, daunting challenges remain. Practically, the genomic revolution has thus far seldom translated directly into the development of new therapeutic strategies, and the mechanisms underlying genotype-phenotype relationships remain only partially explained. Assuming that with time most human genotypic variations will be described together with phenotypic associations, there would still be major problems to fully understand and model human genetic variations and their impact on diseases.

To understand why, consider the "one-gene/one-enzyme/one-function" concept originally framed by Beadle and Tatum (Beadle and Tatum, 1941), which holds that simple, linear connections are expected between the genotype of an organism and its phenotype. But the reality is that most genotype-phenotype relationships arise from a much higher underlying complexity. Combinations of identical genotypes and nearly identical environments do not always give rise to identical phenotypes. The very coining of the words "genotype" and "phenotype" by Johannsen more than a century ago derived from observations that inbred isogenic lines of bean plants grown in well-controlled environments give rise to pods of different size (Johannsen, 1909). Identical twins, although strikingly similar, nevertheless often exhibit many differences (Raser and O'Shea, 2005). Likewise, genotypically indistinguishable bacterial or yeast cells grown side-by-side can express different subsets of transcripts and gene products at any given moment (Elowitz et al., 2002; Blake et al., 2003; Taniguchi et al., 2010). Even straightforward Mendelian traits are not immune to complex genotype-phenotype relationships. Incomplete penetrance, variable expressivity, differences in age of onset, and modifier mutations are more frequent than generally appreciated (Perlis et al., 2010).

We argue, along with others, that the way beyond these challenges is to decipher the properties of biological systems, and in particular, those of molecular networks taking place within cells. As is becoming increasingly clear, biological systems and cellular networks are governed by specific laws and principles, the understanding of which will be essential for a deeper comprehension of biology (Nurse, 2003; Vidal, 2009).

Accordingly, our goal is to review key aspects of how complex systems operate inside cells. Particularly, we will review how by interacting with each other, genes and their products form complex networks within cells. Empirically determining and modeling cellular networks for a few model organisms and for human has provided a necessary scaffold towards understanding the functional, logical and dynamical aspects of cellular systems. Importantly, we will discuss the possibility that phenotypes result from perturbations of the properties of cellular systems and networks. The link between network properties and phenotypes, including susceptibility to human disease, appears to be at least as important as that between genotypes and phenotypes (Figure 1).

Cells as Interactome Networks

Systems biology can be said to have originated more than half a century ago, when a few pioneers initially formulated a theoretical framework according to which multi-scale dynamic complex systems formed by interacting macromolecules could underlie cellular behavior (Vidal, 2009). These theoretical systems biology ideas were elaborated upon at a time when there was little knowledge of the exact nature of the molecular components of biology, let alone any detailed information on functional and biophysical interactions between them. While greatly inspirational to a few specialists, systems concepts remained

largely ignored by most molecular biologists, at least until empirical observations could be gathered to validate them. Meanwhile, theoretical representations of cellular organization evolved steadily, closely following the development of ever improving molecular technologies. The organizational view of the cell changed from being merely a "bag of enzymes" to a web of highly inter-related and interconnected organelles (Robinson et al., 2007). Cells can accordingly be envisioned as complex webs of macromolecular interactions, the full complement of which constitutes the "interactome" network. At the dawn of the 21st century, with most components of cellular networks having been identified, the basic ideas of systems and *network biology* are ready to be experimentally tested and applied to relevant biological problems.

Mapping Interactome Networks

Network science deals with complexity by "simplifying" complex systems, summarizing them merely as components (nodes) and interactions (edges) between them. In this simplified approach, the functional richness of each node is lost. Despite or even perhaps because of such simplifications, useful discoveries can be made. As regards cellular systems, the nodes are metabolites and macromolecules such as proteins, RNA molecules and gene sequences, while the edges are physical, biochemical and functional interactions that can be identified with a plethora of technologies. One challenge of network biology is to provide maps of such interactions using systematic and standardized approaches and assays that are as unbiased as possible. The resulting "interactome" networks, the networks of interactions between cellular components, can serve as scaffold information to extract global or local graph theory properties. Once shown to be statistically different from randomized networks, such properties can then be related back to a better understanding of biological processes. Potentially powerful details of each interaction in the network are left aside, including functional, dynamic and logical features, as well as biochemical and structural aspects such as protein post-translational modifications or allosteric changes. The power of the approach resides precisely in such simplification of molecular detail, which allows modeling at the scale of whole cells.

Early attempts at experimental proteome-scale interactome network mapping in the mid-1990s (Finley and Brent, 1994; Bartel et al., 1996; Fromont-Racine et al., 1997; Vidal, 1997) were inspired by several conceptual advances in biology. The biochemistry of metabolic pathways had already given rise to cellular scale representations of metabolic networks. The discovery of signaling pathways and cross-talk between them, as well as large molecular complexes such as RNA polymerases, all involving innumerable physical protein-protein interactions, suggested the existence of highly connected webs of interactions. Finally, the rapidly growing identification of many individual interactions between transcription factors and specific DNA regulatory sequences involved in the regulation of gene expression raised the question of how transcriptional regulation is globally organized within cells.

Three distinct approaches have been used since to capture interactome networks: i) compilation or curation of already existing data available in the literature, usually obtained from one or just a few types of physical or biochemical interactions (Roberts, 2006); ii) computational predictions based on available "orthogonal" information apart from physical or biochemical interactions, such as sequence similarities, gene-order conservation, co-presence and co-absence of genes in completely sequenced genomes and protein structural information (Marcotte and Date, 2001); and iii) systematic, unbiased high-throughput experimental mapping strategies applied at the scale of whole genomes or proteomes (Walhout and Vidal, 2001). These approaches, though complementary, differ greatly in the possible interpretations of the resulting maps. Literature-curated maps present the advantage

of using already available information, but are limited by the inherently variable quality of the published data, the lack of systematization, and the absence of reporting of negative data (Cusick et al., 2009; Turinsky et al., 2010). Computational prediction maps are fast and efficient to implement, and usually include satisfyingly large numbers of nodes and edges, but are necessarily imperfect because they use indirect information (Plewczynski and Ginalski, 2009). While high-throughput maps attempt to describe unbiased, systematic and well-controlled data, they were initially more difficult to establish, although recent technological advances suggest that near completion can be reached within a few years for highly reliable, comprehensive protein-protein interaction and gene regulatory network maps for human (Venkatesan et al., 2009).

The mapping and analysis of interactome networks for model organisms was instrumental in getting to this point. Such efforts provided, and will continue to provide, both necessary pioneering technologies and crucial conceptual insights. As with other aspects of biology, advancements in mapping of interactome networks would have been minimal without a focus on model organisms (Davis, 2004). The field of interactome mapping has been helped by developments in several model organisms, primarily the yeast *Saccharomyces cerevisiae*, the fly *Drosophila melanogaster*, and the worm *Caenorhabditis elegans* (Figure 2). For instance, genome-wide resources such as collections of all, or nearly all, open reading frames (ORFeomes) were first generated for these model organisms, both because their genomes are the best annotated and because there are fewer complications, such as the high number of splice variants in human and other mammals. ORFeome resources allow efficient transfer of large numbers of ORFs into vectors suitable for diverse interactome mapping technologies (Hartley et al., 2000; Walhout et al., 2000b). Moreover, gene ablation technologies, knockouts (for yeast) and knockdowns by RNAi (for worms and flies) and transposon insertions (for plants), were discovered in and are being applied genome-wide for these model organisms (Mohr et al., 2010).

Metabolic Networks

Metabolic network maps attempt to comprehensively describe all possible biochemical reactions for a particular cell or organism (Schuster et al., 2000; Edwards et al., 2001). In many representations of metabolic networks, nodes are biochemical metabolites and edges are either the reactions that convert one metabolite into another or the enzymes that catalyze these reactions (Jeong et al., 2000; Schuster et al., 2000) (Figure 2). Edges can be directed or undirected, depending on whether a given reaction is reversible or not. In specific cases of metabolic network modeling, the converse situation can be used, with nodes representing enzymes and edges pointing to "adjacent" pairs of enzymes for which the product of one is the substrate of the other (Lee et al., 2008).

Although large metabolic pathway charts have existed for decades (Kanehisa et al., 2008), nearly complete metabolic network maps required the completion of full genome sequencing together with accurate gene annotation tools (Oberhardt et al., 2009). Network construction is manual with computational assistance, involving: i) the meticulous curation of large numbers of publications, each describing experimental results regarding one or several metabolic reactions characterized from purified or reconstituted enzymes, and ii) when necessary, the compilation of predicted reactions from studies of orthologous enzymes experimentally characterized in other species. Assembly of the union of all experimentally demonstrated and predicted reactions gives rise to proteome-scale network maps (Mo and Palsson, 2009). Such maps have been compiled for numerous species, predominantly prokaryotes and unicellular eukaryotes (Oberhardt et al., 2009), and full-scale metabolic reconstructions are now underway for human as well (Ma et al., 2007). Metabolic network

maps are likely the most comprehensive of all biological networks, although considerable gaps will remain to be filled in by direct experimental investigations.

Protein-Protein Interaction Networks

In protein-protein interaction network maps, nodes represent proteins and edges represent a physical interaction between two proteins. The edges are non-directed, as it cannot be said which protein binds the other, that is, which partner functionally influences the other (Figure 2). Of the many methodologies that can map protein-protein interactions, two are currently in wide use for large-scale mapping. Mapping of *binary* interactions is primarily carried out by ever improving variations of the yeast two-hybrid (Y2H) system (Fields and Song, 1989; Dreze et al., 2010). Mapping of membership in protein complexes, providing *indirect associations* between proteins, is carried out by affinity- or immuno- purification to isolate protein complexes, followed by some form of mass spectrometry (AP/MS) to identify protein constituents of these complexes (Rigaut et al., 1999; Charbonnier et al., 2008). While Y2H datasets contain mostly *direct* binary interactions, AP/MS co-complex data sets are composed of direct interactions mixed with a preponderance of *indirect* associations. Accordingly, the graphs generated by these two approaches exhibit different global properties (Seebacher and Gavin, 2011), such as the relationships between gene essentiality and the number of interacting proteins (Yu et al., 2008).

In the past decade significant steps have been taken towards the generation of comprehensive protein-protein interaction network maps. Comprehensive efforts using Y2H technologies to generate interactome maps began with the model organisms *S. cerevisiae*, *C. elegans* and *D. melanogaster* (Ito et al., 2000; Uetz et al., 2000; Walhout et al., 2000a; Ito et al., 2001; Giot et al., 2003; Reboul et al., 2003; Li et al., 2004), and eventually included human (Colland et al., 2004; Rual et al., 2005; Stelzl et al., 2005; Venkatesan et al., 2009). Comprehensive mapping of co-complex membership by high-throughput AP/MS was initially undertaken in yeast (Gavin et al., 2002; Ho et al., 2002), rapidly progressing to ever improving completeness and quality thereafter (Gavin et al., 2006; Krogan et al., 2006). For technical reasons future comprehensive AP/MS efforts will stay focused on unicellular organisms such as yeast (Collins et al., 2007) and mycoplasma (Kuhner et al., 2009), whereas Y2H efforts are more readily implemented for complex multicellular organisms (Seebacher and Gavin, 2011).

In their early implementations, systematic and comprehensive interaction network mapping efforts met with skepticism regarding their accuracy (von Mering et al., 2002), analogous to the original concerns over whether automated high-throughput genome sequencing efforts might have considerably lower accuracy than dedicated efforts carried out cumulatively in many laboratories. Only after the emergence of rigorous statistical tests to estimate sequencing accuracy could high-throughput sequencing efforts reach their full potential (Ewing et al., 1998). Analogously, an empirical framework recently propagated for protein interaction mapping (Venkatesan et al., 2009) now allows the estimation of overall accuracy and sensitivity for maps obtained using high-throughput mapping approaches.

Four critical parameters need to be estimated: completeness (the number of physical protein pairs actually tested in a given search space); assay sensitivity (which interactions can and cannot be detected by a particular assay); sampling sensitivity (the fraction of all detectable interactions found by a single implementation of any interaction assay); and precision (the proportion of true biophysical interactors). Careful consideration of these parameters offers a quantitative idea of the completeness and accuracy of a particular high-throughput interaction map (Yu et al., 2008; Simonis et al., 2009; Venkatesan et al., 2009), and allows comparison of multiple maps as long as standardized framework parameters are used. In

contrast, comparing the results of small-scale experiments available in literature curated databases is not possible, as there is simply no way to control for accuracy, reproducibility, and sensitivity. The binary interactome empirical framework offers a way to estimate the size of interactome networks, which in turn is essential to define a roadmap to reach completion for the interactome mapping efforts of any species of interest. While originally established for protein-protein interaction mapping, similar empirical frameworks can be applied more broadly to mapping of other types of interactome networks (Costanzo et al., 2010).

Gene Regulatory Networks

In most gene regulatory network maps, nodes are either a transcription factor or a putative DNA regulatory element, and directed edges represent the physical binding of transcription factors to such regulatory elements. Edges can be said to be incoming (transcription factor binds a regulatory DNA element) or outgoing (regulatory DNA element bound by a transcription factor) (Figure 2). Currently, two general approaches are amenable to large-scale mapping of gene regulatory networks. In yeast one-hybrid (Y1H) approaches, a putative *cis*-regulatory DNA sequence, commonly a suspected promoter region, is used as bait to capture transcription factors that bind to that sequence (Deplancke et al., 2004). In chromatin immunoprecipitation (ChIP) approaches, antibodies raised against transcription factors of interest, or against a peptide tag used in fusion with potential transcription factors, are used to immunoprecipitate potentially interacting cross-linked DNA fragments (Lee et al., 2002). As Y1H proceeds from genes and captures associated proteins it is said to be "gene-centric", whereas ChIP strategies are "protein-centric" in that they proceed from transcription factors and attempt to capture associated gene regions (Walhout, 2006). The two approaches are complementary. The Y1H system can discover novel transcription factors but relies on having known, or at least suspected, regulatory regions; ChIP methods can discover novel regulatory motifs but rely on the availability of reagents specific to transcription factors of interest, which themselves depend on accurate predictions of transcription factors (Reece-Hoyes et al., 2005; Vaquerizas et al., 2009)

Large-scale Y1H networks have been produced for *C. elegans* (Vermeirssen et al., 2007; Grove et al., 2009). Large-scale ChIP-based networks have been produced for yeast (Lee et al., 2002) and have been carried out for mammalian tissue culture cells as well (Cawley et al., 2004).

In addition to transcription factor activities, overall gene transcript levels are also regulated post-transcriptionally by micro RNAs (miRNAs), short non-coding RNAs that bind to complementary cis-regulatory RNA sequences usually located in 3 untranslated regions (UTRs) of target mRNAs (Lee et al., 2004; Ruvkun et al., 2004). miRNAs are not expected to act as master regulators, but rather act post-transcriptionally to fine-tune gene expression by modulating the levels of target mRNAs. Complex networks are formed by miRNAs interacting with their targets. In such networks, nodes are either a miRNA or a target 3 UTR, and edges represent the complementary annealing of the miRNA to the target RNA. Edges can be said to be incoming (miRNA binds a 3 UTR element) or outgoing (3 UTR element bound by a miRNA) (Martinez et al., 2008). The targets of miRNAs are generally computationally predicted, as experimental methodologies to map miRNA/3 UTR interactions at high-throughput are just coming online (Karginov et al., 2007; Guo et al., 2010; Hafner et al., 2010). Since transcription factors regulate the expression of miRNAs, it is however possible to combine Y1H methods with computationally predicted miRNA/3 UTR interactions, a strategy which was used to derive a large-scale miRNA network in *C. elegans* (Martinez et al., 2008) and which could be extended to other genomes.

Integrating Interactome Networks with other Cellular Networks

The three interactome network types considered so far, metabolic, protein-protein interaction and gene regulatory networks, are composed of physical or biochemical interactions between macromolecules. The corresponding network maps provide crucial "scaffold" information about cellular systems, on top of which additional layers of functional links can be added to fine-tune the representation of biological reality (Figure 3) (Vidal, 2001). Networks composed of functional links, although strikingly different in terms of what the edges represent, can nevertheless complement what can be learned from interactome network maps in powerful ways, and *vice versa*. Networks of functional links represent a category of cellular networks that can be derived from indirect, or "conceptual", interactions where links between genes and gene products are reported based upon functional relationships or similarities, independently of physical macromolecular interactions. We consider three types of functional networks that have been mapped thus far at the scale of whole genomes and used together with physical interactome networks to interrogate the complexities of genotype-to-phenotype relationships.

Transcriptional Profiling Networks

Gene products that function together in common signaling cascades or protein complexes are expected to show greater similarities in their expression patterns than random sets of gene products. But how does this expectation translate at the level of whole proteomes and transcriptomes? How do transcriptome states correlate globally with interactome networks? Since the original description of microarray and DNA chip techniques and more recently *de novo* RNA sequencing using next-generation sequencing technologies, vast compendiums of gene expression datasets have been generated for many different species across a multitude of diverse genetic and environmental conditions. This type of information can be thought of as matrices comprising all genes of an organism against all conditions that this organism has been exposed to within a given expression compendium (Vidal, 2001). In the resulting co-expression networks, nodes represent genes, and edges link pairs of genes that show correlated co-expression above a set threshold (Kim et al., 2001; Stuart et al., 2003). For any correlation measurement, Pearson Correlation Coefficients (PCCs) being commonly used, the threshold between what is considered co-expressed and not co-expressed needs to be set using appropriate titration procedures (Stuart et al., 2003; Gunsalus et al., 2005).

Integration attempts in yeast, combining physical protein-protein interaction maps with co-expression profiles, revealed that interacting proteins are more likely to be encoded by genes with similar expression profiles than non-interacting proteins (Ge et al., 2001; Grigoriev, 2001; Jansen et al., 2002; Kemmeren et al., 2002). These observations were subsequently confirmed in many other organisms (Ge et al., 2003). Beyond the fundamental aspect of finding significant overlaps between interaction edges in interactome networks and co-expression edges in transcription profiling networks, these observations have been used to estimate the overall biological significance of interactome datasets. While correlations can be statistically significant over huge datasets, still many valid biologically relevant protein-protein interactions correspond to pairs of genes whose expression is uncorrelated or even anti-correlated. Co-expression similarity links need not be perfectly overlapping with physical interactions of the corresponding gene products and *vice versa*.

In another example of what co-expression networks can be used for, preliminary steps have been taken to delineate gene regulatory networks from co-expression profiles (Segal et al., 2003; Amit et al., 2009). Such network constructions provide verifiable hypotheses about how regulatory pathways operate.

Phenotypic Profiling Networks

Perturbations of genes that encode functionally related products often confer similar phenotypes. Systematic use of gene knock-out strategies developed for yeast (Giaever et al., 2002) and knock-down approaches using RNA interference (RNAi) for *C. elegans, Drosophila* and recently human (Mohr et al., 2010), are amenable to the perturbation of (nearly) all genes and subsequent testing of a wide variety of standardized phenotypes. As with transcriptional profiling networks, this type of information can be thought of as matrices comprising all genes of an organism and all phenotypes tested within a given phenotypic profiling compendium. In the resulting phenotypic similarity or "phenome" network, nodes represent genes, and edges link pairs of genes that show correlated phenotypic profiles above a set threshold. Here again, titration is needed to decide on the threshold between what is considered phenotypically similar and what is not (Gunsalus et al., 2005).

The earliest evidence that phenotypic profiling or "phenome" networks might help in interpreting protein-protein interactome networks was obtained in studies of the *C. elegans* DNA damage response and hermaphrodite germline (Boulton et al., 2002; Piano et al., 2002; Walhout et al., 2002). The physical basis of phenome networks is not yet completely defined, though there are strong overlaps between correlated phenotypic profiles and physical protein-protein interactions (Walhout et al., 2002; Gunsalus et al., 2005). Overlapping three network types, binary interactions, co-expression, and phenotype profiling, produce integrated networks with high predictive power, as demonstrated for *C. elegans* early embryogenesis (Walhout et al., 2002; Gunsalus et al., 2005). Integration of transcriptional regulatory networks with these other network types has also been undertaken in worm (Grove et al., 2009).

Comprehensive genome-wide phenome networks are now a reality for the yeast *S. cerevisiae* (Giaever et al., 2002)**,** and are expected to be further developed for *C. elegans* (Sönnichsen et al., 2005) and *Drosophila* (Mohr et al., 2010). Now that RNAi reagents are available for nearly all genes of mouse and human (Root et al., 2006), phenome maps for cell lines of these organisms should soon follow.

Genetic Interaction Networks

Pairs of functionally related genes tend to exhibit genetic interactions, defined by comparing the phenotype conferred by mutations in pairs of genes (double mutants) to the phenotype conferred by either one of these mutations alone (single mutants). Genetic interactions are classified as negative, *i.e.* aggravating, synthetic sick or lethal, when the phenotype of double mutants is significantly worse than expected from that of single mutants, and positive, *i.e.* alleviating or suppressive, when the phenotype of double mutants is significantly better than that expected from the single mutants (Mani et al., 2008). Though finding genetic interactions has been crucial to geneticists for decades (Sturtevant, 1956; Novick et al., 1989), only in the last ten years has functional genomics advanced sufficiently to allow systematic high-throughput mapping of genetic interactions to give rise to large-scale networks (Boone et al., 2007).

Two general strategies have been followed for the systematic mapping of genetic interactions in yeast. Synthetic genetic arrays (SGA) and derivative methodologies use high-density arrays of double mutants by mating pairs from an available set of single mutants (Tong et al., 2001; Boone et al., 2007). Alternative strategies take advantage of sequence barcodes embedded in a set of yeast deletion mutants (Giaever et al., 2002; Beltrao et al., 2010) to measure the relative growth rate in a population of double mutants by hybridization to anti-barcode microarrays (Pan et al., 2004; Boone et al., 2007). These two approaches

seem to capture similar aspects of genetic interactions, as the overlap between the two types of datasets is significant (Costanzo et al., 2010).

Patterns of genetic interactions can be used to define a kind of network that is similar to phenotypic profiling or phenome networks. As with transcriptional and phenotypic profiling networks, this type of information can be thought of as matrices comprising all genes of an organism and the genes with which they exhibit a genetic interaction. In such "genetic interaction profiling" networks, edges functionally link two genes based on high similarities of genetic interaction profiles. Here again, predictive models of biological processes can be obtained when such networks are combined with other types of interactome networks.

Integration of genetic interaction networks with other types of interactome network maps provides potentially powerful models. While genetic interactions do not necessarily correspond to physical interactions between the corresponding gene products (Mani et al., 2008), interesting patterns emerge between the different datasets. Because they tend to reveal pairs of genes involved in parallel pathways or in different molecular machines, negative genetic interactions tend *not* to correlate with either protein associations in protein complexes or with binary protein-protein interactions (Beltrao et al., 2010; Costanzo et al., 2010). In contrast, positive genetic interactions tend to point to pairs of gene products physically associated with each other. This trend is usually explained by loss of either one or two gene products working together in a molecular complex resulting in similar effects (Beltrao et al., 2010).

Graph Properties of Networks

A critical realization over the past decade is that the structure and evolution of networks appearing in natural, technological, and social systems over time follows a series of basic and reproducible organizing principles. Theoretical advances in network science (Albert and Barabasi, 2002), paralleling advances in high-throughput efforts to map biological networks, have provided a conceptual framework with which to interpret large interactome network maps. Full understanding of the internal organization of a cell requires awareness of the constraints and laws that biological networks follow. We summarize several principles of network theory that have immediate applications to biology.

Degree Distribution and Hubs

Any empirical investigation starts with the same question: could the investigated phenomena have emerged by chance, or could random effects account for them? The earliest network models assumed that complex networks are wired randomly, such that any two nodes are connected by a link with the same probability *p*. This Erdos-Renyi model generates a network with a Poisson degree distribution, which implies that most nodes have approximately the same degree, that is, the same number of links, while nodes that have significantly more or fewer links than any average node are exceedingly rare or altogether absent. In contrast, many real networks, from the World Wide Web to social networks, are scale-free (Barabási and Albert, 1999), which means that their degree distribution follows a power law rather than the expected Poisson distribution. In a scale-free network most nodes have only a few interactions, and these coexist with a few highly connected nodes, the hubs, that hold the whole network together. This scale-free property has been found in all organisms for which protein-protein interaction and metabolic network maps exist, from yeast to humans (Barabási and Oltvai, 2004; Seebacher and Gavin, 2011). Regulatory networks, however, show a mixed behavior. The outgoing degree distribution, corresponding to how many different genes a transcription factor can regulate, is scale-free, meaning that some master regulators can regulate hundreds of genes. In contrast, the incoming degree distribution, corresponding to how many transcription factors regulate a

specific gene, best fits an exponential model (Deplancke et al., 2006), indicating that genes that are simultaneously regulated by large numbers of transcription factors are exponentially rare.

Gene Duplication as the Origin of the Scale-Free Property

The scale-free topology of biological networks likely originates from gene duplication. While the principle applies from metabolic to regulatory networks, it is best illustrated in protein-protein interaction networks, where it was first proposed (Pastor-Satorras et al., 2003; Vázquez et al., 2003). When cells divide and the genome replicates, occasionally an extra copy of one or several genes or chromosomes gets produced. Immediately following a duplication event, both the original protein and the new extra copy have the same structure, so both interact with the same set of partners. Consequently, each of the protein partners that interacted with the ancestor gains a new interaction. This process results in a rich-get-richer phenomenon (Barabási and Albert, 1999), where proteins with a large number of interactions tend to gain links more often, as it is more likely that they interact with a duplicated protein. This mechanism has been shown to generate hubs (Pastor-Satorras et al., 2003; Vázquez et al., 2003), and so could be responsible for the scale-free property of protein-protein interaction networks.

The Role of Hubs

Network biology attempts to identify global properties in interactome network graphs, and subsequently relate such properties to biological reality by integrating various functional datasets. One of the best examples where this approach was successful is in defining the role of hubs. In the model organisms *S. cerevisiae* and *C. elegans,* hub proteins were found to: i) correspond to essential genes (Jeong et al., 2001), ii) be older and have evolved more slowly (Fraser et al., 2002), iii) have a tendency to be more abundant (Ivanic et al., 2009), and iv) have a larger diversity of phenotypic outcomes resulting from their deletion compared to the deletion of less connected proteins (Yu et al., 2008). While the evidence attributed to some of these findings has been debated (Jordan et al., 2003; Yu et al., 2008; Ivanic et al., 2009), the special role of hub proteins in model organisms led to the expectation that, in humans, hubs should preferentially encode disease-related genes. Indeed, up-regulated genes in lung squamous cell carcinoma tissues tend to have a high degree in protein-protein interaction networks (Wachi et al., 2005), and cancer-related proteins have, on average, twice as many interaction partners as non-cancer proteins in protein-protein interaction networks (Jonsson and Bates, 2006). A cautionary note is necessary: since disease-related proteins tend to be more avidly studied their higher connectivity may be partly rooted in investigative biases. Therefore this type of finding needs to be appropriately controlled using systematic proteome-wide interactome network maps.

Understanding the role of hubs in human disease requires distinguishing between essential genes and disease-related genes (Goh et al., 2007). Some human genes are essential for early development, such that mutations in them often lead to spontaneous abortions. The protein products of mouse *in utero* essential genes show a strong tendency to be associated with hubs and to be expressed in multiple tissues (Goh et al., 2007). Non-essential disease genes tend not to be encoded by hubs and tend to be tissue specific. These differences can be best appreciated from an evolutionary perspective. Mutations that disrupt hubs may have difficulty propagating in the population, as the host may not survive long enough to have offspring. Only mutations that impair functionally and topologically peripheral genes can persist, becoming responsible for heritable diseases, particularly those that manifest in adulthood.

Date and Party Hubs

Another success in uncovering the functional consequences of the topology of interactome networks was provided by the discovery of date and party hubs (Han et al., 2004). Upon integrating protein-protein interaction network data with transcriptional profiling networks for yeast, at least two classes of hubs can be discriminated. Party hubs are highly co-expressed with their interacting partners while date hubs appear to be more dynamically regulated relative to their partners (Han et al., 2004). In other words, date hubs interact with their partners at different times and/or different conditions, whereas party hubs seem to interact with their partners at all times or conditions tested (Seebacher and Gavin, 2011). Despite the preponderance of evidence in its favor, the date and party hubs concept remains a subject of debate, (Agarwal et al., 2010), attributable to the necessity to appropriately calibrate co-expression and protein-protein interaction hub thresholds when analyzing new transcriptome and interactome datasets (Bertin et al., 2007).

Fundamentally, date hubs preferentially connect functional modules to each other, whereas party hubs preferentially act inside functional modules, hence they are occasionally called inter-module and intra-module hubs, respectively (Han et al., 2004; Taylor et al., 2009). Date hubs are less evolutionarily constrained than party hubs (Fraser, 2005; Ekman et al., 2006; Bertin et al., 2007). Party hubs contain fewer and shorter regions of intrinsic disorder than do date hubs (Ekman et al., 2006; Singh et al., 2006; Kahali et al., 2009) and contain fewer linear motifs (short binding motifs and post-translational modification sites) than do date hubs (Taylor et al., 2009). Initially explored in a yeast interactome (Han et al., 2004; Ekman et al., 2006), the distinction between date and party hubs can be recapitulated in human interactomes as well (Taylor et al., 2009).

Motifs

There has been considerable attention paid in recent years to network motifs, which are characteristic network patterns, or subgraphs, in biological networks that appear more frequently than expected given the degree distribution of the network (Milo et al., 2002). Such subgraphs have been found to be associated with desirable (or undesirable) biological function (or dysfunction). Hence identification and classification of motifs can offer information about the various network subgraphs needed for biological processes. It is now commonly understood that motifs constitute the basic building blocks of cellular networks (Milo et al., 2002; Yeger-Lotem et al., 2004).

Originally identified in transcriptional regulatory networks of several model organisms (Milo et al., 2002; Shen-Orr et al., 2002), motifs have been subsequently identified in interactome networks and in integrated composite networks (Yeger-Lotem et al., 2004; Zhang et al., 2005). Different types of networks exhibit different motif profiles, suggesting a means for network classification (Milo et al., 2004; Zhang et al., 2005). The high degree of evolutionary conservation of motif constituents within interaction networks (Wuchty et al., 2003), combined with the convergent evolution that is seen in the transcription regulatory networks of diverse species towards the same motif types (Barabási and Oltvai, 2004), makes a strong argument that motifs are of direct biological relevance. Classification of several highly significant motifs of two, three, and four nodes, with descriptors like coherent feed forward loop or single-input module, has shown that specific types of motifs carry out specific dynamic functions within cells (Alon, 2007; Shoval and Alon, 2010).

Topological, functional and disease modules

Most biological networks have a rather uneven organization. Many nodes are part of locally dense neighborhoods, or *topological modules*, where nodes have a higher tendency to link to nodes within the same local neighborhood than to nodes outside of it (Ravasz et al., 2002).

A region of the global network diagram that corresponds to a potential topological module can be identified by network clustering algorithms which are blind to the function of individual nodes. These topological modules are often believed to carry specific cellular functions, hence leading to the concept of a *functional module*, an aggregation of nodes of similar or related function in the same network neighborhood. Interest is increasing in *disease modules,* which represent groups of network components whose disruption results in a particular disease phenotype in humans (Barabási et al., 2010).

There is a tacit assumption, based on evidence in the biological literature**,** that cellular components forming topological modules have closely related functions, thus corresponding to functional modules. New potentially powerful methods to identify topological and functional clusterings continue to be described (Ahn et al., 2010). Such modules can serve as hypothesis building tools to identify "regions" of the interactome likely involved in particular cellular functions or disease (Barabási et al., 2010).

Networks and Human Diseases

Having reviewed why biological networks are important to consider, how they can be mapped and integrated with each other, and what global properties are starting to emerge from such models, we next return to our original question: to what extent do biological systems and cellular networks underlie genotype-phenotype relationships in human disease? We attempt to provide answers by covering four recent advancements in network biology: i) studies of global relationships between human disorders, associated genes and interactome networks, ii) predictions of new human disease-associated genes using interactome models, iii) analyses of network perturbations by pathogens, and iv) emergence of node removal versus edge-specific or "edgetic" models to explain genotype-phenotype relationships.

Global Disease Networks

One of the main predictions derived from the hypothesis that human disorders should be viewed as perturbations of highly interlinked cellular networks is that diseases should not be independent from each other, but should instead be themselves highly interconnected. Such potential cellular network-based dependencies between human diseases has led to the generation of various global disease network maps, which link disease phenotypes together if some molecular or phenotypic relationships exist between them. Such a map was built using known gene-disease associations collected in the OMIM database (Goh et al., 2007), where nodes are diseases and two diseases are linked by an edge if they share at least one common gene, mutations in which are associated with these diseases. In the obtained disease network more than 500 human genetic disorders belong to a single interconnected main giant component, consistent with the idea that human diseases are much more connected to each other than anticipated. The flip side of this representation of connectivity is a network of disease-associated genes linked together if mutations in these genes are known to be responsible for at least one common disorder. Providing support for our general hypothesis that perturbations in cellular networks underlie genotype-phenotype relationships, such disease associated gene networks overlap significantly with human protein-protein interactome network maps (Goh et al., 2007).

Additional types of connectivity between large numbers of human diseases can be found in "co-morbidity" networks where diseases are linked to each other when individuals who were diagnosed for one particular disease are more likely to have also been diagnosed for the other (Rzhetsky et al., 2007; Hidalgo et al., 2009). Diabetes and obesity represent probably the best known disease pair with such significant co-morbidity. While co-morbidity can have multiple origins, ranging from environmental factors to treatment side effects, its potential molecular origin has attracted considerable attention. A network biology

interpretation would suggest that the molecular defects responsible for one of a pair of diseases can "spread along" the edges in cellular networks, affecting the activity of related gene products and causing or affecting the outcome of the other disease (Park et al., 2009).

Predicting Disease Related genes using Interactome Networks

If cellular networks underlie genotype-phenotype relationships, then network properties should be predictive of novel, yet to be identified human disease-associated genes. In an early example, it was shown that the products of a few dozen ataxia-associated genes occupy distinct "locations" in the human interactome network, in that the number of edges separating them is on average much lower than for random sets of gene products (Lim et al., 2006). Physical protein-protein interactome network maps can indeed generate lists of genes potentially enriched for new candidate disease genes or modifier genes of known disease genes (Lim et al., 2006; Oti et al., 2006; Fraser and Plotkin, 2007).

Integration of various interactome and functional relationship networks have also been applied to reveal genes potentially involved in cancer (Pujana et al., 2007). Integrating a co-expression network, seeded with four well-known breast cancer associated genes, together with genetic and physical interactions yielded a breast cancer network model out of which candidate cancer susceptibility and modifier genes could be predicted (Pujana et al., 2007). Integrative network modeling strategies are applicable to other types of cancer and other types of disease (Ergun et al., 2007; Wu et al., 2008; Lee et al., 2010).

Network Perturbations by Pathogens

Pathogens, particularly viruses, have evolved sophisticated mechanisms to perturb the intracellular networks of their hosts to their advantage. As obligate intracellular pathogens, viruses must intimately rewire cellular pathways to their own ends to maintain infectivity. Since many virus-host interactions happen at the level of physical protein-protein interactions, systematic maps capturing viral-host physical protein-protein interactions, or "virhostome" maps, have been obtained using Y2H for Epstein-Barr virus (Calderwood et al., 2007), hepatitis C virus (de Chassey et al., 2008), several herpesviruses (Uetz et al., 2006), influenza virus (Shapira et al., 2009) and others (Mendez-Rios and Uetz, 2010), and by co-AP/MS methodologies for HIV (Jäger et al., 2010). An eminent goal is to find perturbations in network properties of the host network, properties that would not be made evident by small-scale investigations focused on one or a handful of viral proteins. For instance, it has been found several times now that viral proteins preferentially target hubs in host interactome networks (Calderwood et al., 2007; Shapira et al., 2009). The many host targets identified in virhostome screens are now getting biologically validated by RNAi knock-down and transcriptional profiling, leading to detailed maps of the interactions underlying viral-host relationships (Shapira et al., 2009).

Another impetus for mapping virhostome networks is that virus protein interactions can act as surrogates for human genetic variations, inducing disease states by influencing local and global properties of cellular networks. The inspiration for this concept emerged from classical observations such as the binding of Adenovirus E1A, HPV E7, and SV40 Large T antigen to the human retinoblastoma protein, which is the product of a gene in which mutations lead to a predisposition to retinoblastoma and other types of cancers (DeCaprio, 2009). This hypothesis will soon be tested globally by systematic investigations of how host networks, including physical interaction, gene regulatory and genetic interaction networks, are perturbed upon viral infection. Pathogen-host interaction mapping projects are also in their first iterations, with similar goals of identifying emergent global properties and disease surrogates. As microbial pathogens can have thousands of gene products relative to much

smaller numbers for most viruses, such projects will require considerably more effort and time.

Edgetics

Our underlying premise throughout has been that phenotypic variations of an organism, particularly those that result in human disease, arise from perturbations of cellular interactome networks. These alterations range from the complete loss of a gene product, through the loss of some but not all interactions, to the specific perturbation of a single molecular interaction while retaining all others. In interactome networks these alterations range from node removal at one end and edge-specific or "edgetic" perturbations at the other (Zhong et al., 2009). The consequences on network structure and function are expected to be radically dissimilar for node removal versus edgetic perturbation. Node removal not only disables the function of a node but also disables *all* the interactions of that node with other nodes, disrupting in some way the function of *all* of the neighboring nodes. An edgetic disruption, removing one or a few interactions but leaving the rest intact and functioning, has subtler effects on the network, though not necessarily on the resulting phenotype (Madhani et al., 1997). The distinction between node removal and edgetic perturbation models can provide new clues on mechanisms underlying human disease, such as the different classes of mutations that lead to dominant versus recessive modes of inheritance (Zhong et al., 2009).

The idea that the disruption of specific protein interactions can lead to human disease (Schuster-Bockler and Bateman, 2008) complements canonical gene loss/perturbation models (Botstein and Risch, 2003), and is poised to explain confounding genetic phenomena such as genetic heterogeneity.

Matching the edgetic hypothesis to inherited human diseases, approximately half of ~50,000 Mendelian alleles available in the Human Gene Mutation Database can be modeled as potentially edgetic if one considers deletions and truncating mutations as node removal, and in-frame point mutations leading to single amino-acid changes and small insertions and deletions as edgetic perturbations (Zhong et al., 2009). This number is probably a good approximation, since thus far disease-associated genes predicted to bear edgetic alleles using this model have been experimentally confirmed (Zhong et al., 2009). Consistent with the edgetic hypothesis, for genes associated with multiple disorders and for which predicted protein interaction domains are available, it was shown that putative edgetic alleles responsible for different disorders tend to be located in different interaction domains, consistent with different edgetic perturbations conferring strikingly different phenotypes.

Conclusion

A comprehensive catalog of sequence variations among the ~7 billion human genomes present on earth might soon become available. This information is and will continue to revolutionize biology in general and medicine in particular for many decades and perhaps centuries to come. The prospects of predictive and personalized medicine are enormous. However, it should be kept in mind that genome variations merely constitute variations in the "parts list", and often fail to provide a description of the mechanistic consequences on cellular functions.

Here we have summarized why considering perturbations of biological networks within cells is crucial to help interpret how genome variations relate to phenotypic differences. Given their high levels of complexity, it is no surprise that interactome networks have not yet been mapped completely. The data and models accumulated in the last decade point to clear directions for the next decade. We envision that with more interactome datasets of

increasingly high quality, the trends reviewed here will be fine-tuned. The global properties observed so far and those yet to be uncovered will "make sense" of the enormous body of information encompassed in the human genome.

Acknowledgments

References

Agarwal S, Deane CM, Porter MA, Jones NS. Revisiting date and party hubs: novel approaches to role assignment in protein interaction networks. PLoS Comput Biol. 2010; 6:e1000817. [PubMed: 20585543]

Ahn YY, Bagrow JP, Lehmann S. Link communities reveal multiscale complexity in networks. Nature. 2010; 466:761–764. [PubMed: 20562860]

Albert R, Barabasi AL. Statistical mechanics of complex networks. Rev Mod Physics. 2002; 74:47–97.

Alon U. Network motifs: theory and experimental approaches. Nat Rev Genet. 2007; 8:450–461. [PubMed: 17510665]

Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. Science. 2008; 322:881–888. [PubMed: 18988837]

Amberger J, Bocchini CA, Scott AF, Hamosh A. McKusick's Online Mendelian Inheritance in Man (OMIM). Nucleic Acids Res. 2009; 37:D793–796. [PubMed: 18842627]

Amit I, Garber M, Chevrier N, Leite AP, Donner Y, Eisenhaure T, Guttman M, Grenier JK, Li W, Zuk O, et al. Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. Science. 2009; 326:257–263. [PubMed: 19729616]

Barabási AL, Albert R. Emergence of scaling in random networks. Science. 1999; 286:509–512. [PubMed: 10521342]

Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet. 2010; 12:56–68.

Barabási AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet. 2004; 5:101–113. [PubMed: 14735121]

Bartel PL, Roecklein JA, SenGupta D, Fields S. A protein linkage map of *Escherichia coli* bacteriophage T7. Nat Genet. 1996; 12:72–77. [PubMed: 8528255]

Beadle GW, Tatum EL. Genetic control of biochemical reactions in *Neurospora*. Proc Natl Acad Sci USA. 1941; 27:499–506. [PubMed: 16588492]

Beltrao P, Cagney G, Krogan NJ. Quantitative genetic interactions reveal biological modularity. Cell. 2010; 141:739–745. [PubMed: 20510918]

Bertin N, Simonis N, Dupuy D, Cusick ME, Han JD, Fraser HB, Roth FP, Vidal M. Confirmation of organized modularity in the yeast interactome. PLoS Biol. 2007; 5:e153. [PubMed: 17564493]

Blake WJ, MKA, Cantor CR, Collins JJ. Noise in eukaryotic gene expression. Nature. 2003; 422:633–637. [PubMed: 12687005]

Boone C, Bussey H, Andrews BJ. Exploring genetic interactions and networks with yeast. Nat Rev Genet. 2007; 8:437–449. [PubMed: 17510664]

Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat Genet. 2003; 33(Suppl):228–237. [PubMed: 12610532]

Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet. 1980; 32:314–331. [PubMed: 6247908]

Boulton SJ, Gartner A, Reboul J, Vaglio P, Dyson N, Hill DE, Vidal M. Combined functional genomic maps of the *C. elegans* DNA damage response. Science. 2002; 295:127–131. [PubMed: 11778048]

Calderwood MA, Venkatesan K, Xing L, Chase MR, Vazquez A, Holthaus AM, Ewence AE, Li N, Hirozane-Kishikawa T, Hill DE, et al. Epstein-Barr virus and virus human protein interaction maps. Proc Natl Acad Sci USA. 2007; 104:7606–7611. [PubMed: 17446270]

Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, Kampa D, Piccolboni A, Sementchenko V, Cheng J, Williams AJ, et al. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell. 2004; 116:499–509. [PubMed: 14980218]

Charbonnier S, Gallego O, Gavin AC. The social network of a cell: recent advances in interactome mapping. Biotechnol Annu Rev. 2008; 14:1–28. [PubMed: 18606358]

Colland F, Jacq X, Trouplin V, Mougin C, Groizeleau C, Hamburger A, Meil A, Wojcik J, Legrain P, Gauthier JM. Functional proteomics mapping of a human signaling pathway. Genome Res. 2004; 14:1324–1332. [PubMed: 15231748]

Collins SR, Kemmeren P, Zhao XC, Greenblatt JF, Spencer F, Holstege FC, Weissman JS, Krogan NJ. Towards a comprehensive atlas of the physical interactome of *Saccharomyces cerevisiae*. Mol Cell Proteomics. 2007; 6:439–450. [PubMed: 17200106]

Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JL, Toufighi K, Mostafavi S, et al. The genetic landscape of a cell. Science. 2010; 327:425–431. [PubMed: 20093466]

Cusick ME, Yu H, Smolyar A, Venkatesan K, Carvunis AR, Simonis N, Rual JF, Borick H, Braun P, Dreze M, et al. Literature-curated protein interaction datasets. Nat Methods. 2009; 6:39–46. [PubMed: 19116613]

Davis RH. The age of model organisms. Nat Rev Genet. 2004; 5:69–76. [PubMed: 14708017]

de Chassey B, Navratil V, Tafforeau L, Hiet MS, Aublin-Gex A, Agaugue S, Meiffren G, Pradezynski F, Faria BF, Chantier T, et al. Hepatitis C virus infection protein network. Mol Syst Biol. 2008; 4:230. [PubMed: 18985028]

DeCaprio JA. How the Rb tumor suppressor structure and function was revealed by the study of Adenovirus and SV40. Virology. 2009; 384:274–284. [PubMed: 19150725]

Deplancke B, Dupuy D, Vidal M, Walhout AJ. A Gateway-compatible yeast one-hybrid system. Genome Res. 2004; 14:2093–2101. [PubMed: 15489331]

Deplancke B, Mukhopadhyay A, Ao W, Elewa AM, Grove CA, Martinez NJ, Sequerra R, Doucette-Stamm L, Reece-Hoyes JS, Hope IA, et al. A gene-centered *C. elegans* protein-DNA interaction network. Cell. 2006; 125:1193–1205. [PubMed: 16777607]

Dreze M, Monachello D, Lurin C, Cusick ME, Hill DE, Vidal M, Braun P. High-quality binary interactome mapping. Methods Enzymol. 2010; 470:281–315. [PubMed: 20946815]

Edwards JS, Ibarra RU, Palsson BO. In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. Nat Biotechnol. 2001; 19:125–130. [PubMed: 11175725]

Ekman D, Light S, Bjorklund AK, Elofsson A. What properties characterize the hub proteins of the protein-protein interaction network of *Saccharomyces cerevisiae?* Genome Biol. 2006; 7:R45. [PubMed: 16780599]

Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. Science. 2002; 297:1183–1186. [PubMed: 12183631]

Ergun A, Lawrence CA, Kohanski MA, Brennan TA, Collins JJ. A network biology approach to prostate cancer. Mol Syst Biol. 2007; 3:82. [PubMed: 17299418]

Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 1998; 8:175–185. [PubMed: 9521921]

Fields S, Song O. A novel genetic system to detect protein-protein interactions. Nature. 1989; 340:245–246. [PubMed: 2547163]

Finley RL Jr, Brent R. Interaction mating reveals binary and ternary connections between *Drosophila* cell cycle regulators. Proc Natl Acad Sci USA. 1994; 91:12980–12984. [PubMed: 7809159]

Fraser HB. Modularity and evolutionary constraint on proteins. Nat Genet. 2005; 37:351–352. [PubMed: 15750592]

Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. Evolutionary rate in the protein interaction network. Science. 2002; 296:750–752. [PubMed: 11976460]

Fraser HB, Plotkin JB. Using protein complexes to predict phenotypic effects of gene mutation. Genome Biol. 2007; 8:R252. [PubMed: 18042286]

Fromont-Racine M, Rain JC, Legrain P. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. Nat Genet. 1997; 16:277–282. [PubMed: 9207794]

Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dumpelfeld B, et al. Proteome survey reveals modularity of the yeast cell machinery. Nature. 2006; 440:631–636. [PubMed: 16429126]

Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature. 2002; 415:141–147. [PubMed: 11805826]

Ge H, Liu Z, Church GM, Vidal M. Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. Nat Genet. 2001; 29:482–486. [PubMed: 11694880]

Ge H, Walhout AJ, Vidal M. Integrating 'omic' information: a bridge between genomics and systems biology. Trends Genet. 2003; 19:551–560. [PubMed: 14550629]

Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, Dow S, Lucau-Danila A, Anderson K, Andre B, et al. Functional profiling of the *Saccharomyces cerevisiae* genome. Nature. 2002; 418:387–391. [PubMed: 12140549]

Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao YL, Ooi CE, Godwin B, Vitols E, et al. A protein interaction map of *Drosophila melanogaster*. Science. 2003; 302:1727–1736. [PubMed: 14605208]

Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL. The human disease network. Proc Natl Acad Sci USA. 2007; 104:8685–8690. [PubMed: 17502601]

Grigoriev A. A relationship between gene expression and protein interactions on the proteome scale: analysis of the bacteriophage T7 and the yeast *Saccharomyces cerevisiae*. Nucleic Acids Res. 2001; 29:3513–3519. [PubMed: 11522820]

Grove CA, De Masi F, Barrasa MI, Newburger DE, Alkema MJ, Bulyk ML, Walhout AJ. A multiparameter network reveals extensive divergence between *C. elegans* bHLH transcription factors. Cell. 2009; 138:314–327. [PubMed: 19632181]

Gunsalus KC, Ge H, Schetter AJ, Goldberg DS, Han JD, Hao T, Berriz GF, Bertin N, Huang J, Chuang LS, et al. Predictive models of molecular machines involved in *Caenorhabditis elegans* early embryogenesis. Nature. 2005; 436:861–865. [PubMed: 16094371]

Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature. 2010; 466:835–840. [PubMed: 20703300]

Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr, Jungkamp AC, Munschauer M, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell. 2010; 141:129–141. [PubMed: 20371350]

Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJ, Cusick ME, Roth FP, et al. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature. 2004; 430:88–93. [PubMed: 15190252]

Hartley JL, Temple GF, Brasch MA. DNA cloning using in vitro site-specific recombination. Genome Res. 2000; 10:1788–1795. [PubMed: 11076863]

Hidalgo CA, Blumm N, Barabasi AL, Christakis NA. A dynamic network approach for the study of human phenotypes. PLoS Comput Biol. 2009; 5:e1000353. [PubMed: 19360091]

Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P, Bennett K, Boutilier K, et al. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. Nature. 2002; 415:180–183. [PubMed: 11805837]

Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci USA. 2001; 98:4569–4574. [PubMed: 11283351]

Ito T, Tashiro K, Muta S, Ozawa R, Chiba T, Nishizawa M, Yamamoto K, Kuhara S, Sakaki Y. Toward a protein-protein interaction map of the budding yeast: A comprehensive system to

examine two-hybrid interactions in all possible combinations between the yeast proteins. Proc Natl Acad Sci USA. 2000; 97:1143–1147. [PubMed: 10655498]

Ivanic J, Yu X, Wallqvist A, Reifman J. Influence of protein abundance on high-throughput protein-protein interaction detection. PLoS ONE. 2009; 4:e5815. [PubMed: 19503833]

Jäger S, Gulbahce N, Cimermancic P, Kane J, He N, Chou S, D'Orso I, Fernandes J, Jang G, Frankel AD, et al. Purification and characterization of HIV-human protein complexes. Methods. 2010; 53:13–19. [PubMed: 20708689]

Jansen R, Greenbaum D, Gerstein M. Relating whole-genome expression data with protein-protein interactions. Genome Res. 2002; 12:37–46. [PubMed: 11779829]

Jeong H, Mason SP, Barabasi AL, Oltvai ZN. Lethality and centrality in protein networks. Nature. 2001; 411:41–42. [PubMed: 11333967]

Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL. The large-scale organization of metabolic networks. Nature. 2000; 407:651–654. [PubMed: 11034217]

Johannsen, W. Elemente der exakten Erblichkeitslehre (Jena, Gustav Fischer). 1909.

Jonsson PF, Bates PA. Global topological features of cancer proteins in the human interactome. Bioinformatics. 2006; 22:2291–2297. [PubMed: 16844706]

Jordan IK, Wolf YI, Koonin EV. No simple dependence between protein evolution rate and the number of protein-protein interactions: only the most prolific interactors tend to evolve slowly. BMC Evol Biol. 2003; 3:1. [PubMed: 12515583]

Kahali B, Ahmad S, Ghosh TC. Exploring the evolutionary rate differences of party hub and date hub proteins in *Saccharomyces cerevisiae* protein-protein interaction network. Gene. 2009; 429:18–22. [PubMed: 18973798]

Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, et al. KEGG for linking genomes to life and the environment. Nucleic Acids Res. 2008; 36:D480–484. [PubMed: 18077471]

Karginov FV, Conaco C, Xuan Z, Schmidt BH, Parker JS, Mandel G, Hannon GJ. A biochemical approach to identifying microRNA targets. Proc Natl Acad Sci USA. 2007; 104:19291–19296. [PubMed: 18042700]

Kemmeren P, van Berkum NL, Vilo J, Bijma T, Donders R, Brazma A, Holstege FC. Protein interaction verification and functional annotation by integrated analysis of genome-scale data. Mol Cell. 2002; 9:1133–1143. [PubMed: 12049748]

Kim SK, Lund J, Kiraly M, Duke K, Jiang M, Stuart JM, Eizinger A, Wylie BN, Davidson GS. A gene expression map for *Caenorhabditis elegans*. Science. 2001; 293:2087–2092. [PubMed: 11557892]

Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis AP, et al. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. Nature. 2006; 440:637–643. [PubMed: 16554755]

Kuhner S, van Noort V, Betts MJ, Leo-Macias A, Batisse C, Rode M, Yamada T, Maier T, Bader S, Beltran-Alvarez P, et al. Proteome organization in a genome-reduced bacterium. Science. 2009; 326:1235–1240. [PubMed: 19965468]

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. Initial sequencing and analysis of the human genome. Nature. 2001; 409:860–921. [PubMed: 11237011]

Lee DS, Park J, Kay KA, Christakis NA, Oltvai ZN, Barabasi AL. The implications of human metabolic network topology for disease comorbidity. Proc Natl Acad Sci USA. 2008; 105:9880–9885. [PubMed: 18599447]

Lee I, Lehner B, Vavouri T, Shin J, Fraser AG, Marcotte EM. Predicting genetic modifier loci using functional gene networks. Genome Res. 2010; 20:1143–1153. [PubMed: 20538624]

Lee R, Feinbaum R, Ambros V. A short history of a short RNA. Cell. 2004; 116:S89–92. [PubMed: 15055592]

Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, et al. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. Science. 2002; 298:799–804. [PubMed: 12399584]

Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, Vidalain PO, Han JD, Chesneau A, Hao T, et al. A map of the interactome network of the metazoan *C. elegans*. Science. 2004; 303:540–543. [PubMed: 14704431]

Lim J, Hao T, Shaw C, Patel AJ, Szabo G, Rual JF, Fisk CJ, Li N, Smolyar A, Hill DE, et al. A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. Cell. 2006; 125:801–814. [PubMed: 16713569]

Ma H, Sorokin A, Mazein A, Selkov A, Selkov E, Demin O, Goryanin I. The Edinburgh human metabolic network reconstruction and its functional analysis. Mol Syst Biol. 2007; 3:135. [PubMed: 17882155]

Madhani HD, Styles CA, Fink GR. MAP kinases with distinct inhibitory functions impart signaling specificity during yeast differentiation. Cell. 1997; 91:673–684. [PubMed: 9393860]

Mani R, St Onge R, Hartman Jt, Giaever G, Roth F. Defining genetic interaction. Proc Natl Acad Sci USA. 2008; 105:3461–3466. [PubMed: 18305163]

Marcotte E, Date S. Exploiting big biology: integrating large-scale biological data for function inference. Brief Bioinform. 2001; 2:363–374. [PubMed: 11808748]

Martinez NJ, Ow MC, Barrasa MI, Hammell M, Sequerra R, Doucette-Stamm L, Roth FP, Ambros VR, Walhout AJ. A *C. elegans* genome-scale microRNA network contains composite feedback motifs with high flux capacity. Genes Dev. 2008; 22:2535–2549. [PubMed: 18794350]

Mendez-Rios J, Uetz P. Global approaches to study protein-protein interactions among viruses and hosts. Future Microbiol. 2010; 5:289–301. [PubMed: 20143950]

Milo R, Itzkovitz S, Kashtan N, Levitt R, Shen-Orr S, Ayzenshtat I, Sheffer M, Alon U. Superfamilies of evolved and designed networks. Science. 2004; 303:1538–1542. [PubMed: 15001784]

Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. Network motifs: simple building blocks of complex networks. Science. 2002; 298:824–827. [PubMed: 12399590]

Mo ML, Palsson BO. Understanding human metabolic physiology: a genome-to-systems approach. Trends Biotechnol. 2009; 27:37–44. [PubMed: 19010556]

Mohr S, Bakal C, Perrimon N. Genomic screening with RNAi: results and challenges. Annu Rev Biochem. 2010; 79:37–64. [PubMed: 20367032]

Novick P, Osmond BC, Botstein D. Suppressors of yeast actin mutations. Genetics. 1989; 121:659–674. [PubMed: 2656401]

Nurse P. The great ideas of biology. Clin Med. 2003; 3:560–568. [PubMed: 14994716]

Oberhardt MA, Palsson BO, Papin JA. Applications of genome-scale metabolic reconstructions. Mol Syst Biol. 2009; 5:320. [PubMed: 19888215]

Oti M, Snel B, Huynen MA, Brunner HG. Predicting disease genes using protein-protein interactions. J Med Genet. 2006; 43:691–698. [PubMed: 16611749]

Pan X, Yuan DS, Xiang D, Wang X, Sookhai-Mahadeo S, Bader JS, Hieter P, Spencer F, Boeke JD. A robust toolkit for functional profiling of the yeast genome. Mol Cell. 2004; 16:487–496. [PubMed: 15525520]

Park J, Lee DS, Christakis NA, Barabasi AL. The impact of cellular networks on disease comorbidity. Mol Syst Biol. 2009; 5:262. [PubMed: 19357641]

Pastor-Satorras R, Smith E, Sole RV. Evolving protein interaction networks through gene duplication. J Theor Biol. 2003; 222:199–210. [PubMed: 12727455]

Perlis RH, Smoller JW, Mysore J, Sun M, Gillis T, Purcell S, Rietschel M, Nothen MM, Witt S, Maier W, et al. Prevalence of incompletely penetrant Huntington's disease alleles among individuals with major depressive disorder. Am J Psychiatry. 2010; 167:574–579. [PubMed: 20360314]

Piano F, Schetter AJ, Morton DG, Gunsalus KC, Reinke V, Kim SK, Kemphues KJ. Gene clustering based on RNAi phenotypes of ovary-enriched genes in *C. elegans*. Curr Biol. 2002; 12:1959–1964. [PubMed: 12445391]

Plewczynski D, Ginalski K. The interactome: predicting the protein-protein interactions in cells. Cell Mol Biol Lett. 2009; 14:1–22. [PubMed: 18839074]

Pujana MA, Han JDJ, Starita LM, Stevens KN, Tewari M, Ahn JS, Rennert G, Moreno V, Kirchhoff T, Gold B, et al. Network modeling links breast cancer susceptibility and centrosome dysfunction. Nat Genet. 2007; 39:1338–1349. [PubMed: 17922014]

Raser JM, O'Shea EK. Noise in gene expression: origins, consequences, and control. Science. 2005; 309:2010–2013. [PubMed: 16179466]

Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL. Hierarchical organization of modularity in metabolic networks. Science. 2002; 297:1551–1555. [PubMed: 12202830]

Reboul J, Vaglio P, Rual JF, Lamesch P, Martinez M, Armstrong CM, Li S, Jacotot L, Bertin N, Janky R, et al. *C. elegans* ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression. Nat Genet. 2003; 34:35–41. [PubMed: 12679813]

Reece-Hoyes JS, Deplancke B, Shingles J, Grove CA, Hope IA, Walhout AJ. A compendium of *Caenorhabditis elegans* regulatory transcription factors: a resource for mapping transcription regulatory networks. Genome Biol. 2005; 6:R110. [PubMed: 16420670]

Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B. A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotechnol. 1999; 17:1030–1032. [PubMed: 10504710]

Roberts PM. Mining literature for systems biology. Brief Bioinform. 2006; 7:399–406. [PubMed: 17032698]

Robinson CV, Sali A, Baumeister W. The molecular sociology of the cell. Nature. 2007; 450:973–982. [PubMed: 18075576]

Root DE, Hacohen N, Hahn WC, Lander ES, Sabatini DM. Genome-scale loss-of-function screening with a lentiviral RNAi library. Nat Methods. 2006; 3:715–719. [PubMed: 16929317]

Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, et al. Towards a proteome-scale map of the human protein-protein interaction network. Nature. 2005; 437:1173–1178. [PubMed: 16189514]

Ruvkun G, Wightman B, Ha I. The 20 years it took to recognize the importance of tiny RNAs. Cell. 2004; 116:S93–96. [PubMed: 15055593]

Rzhetsky A, Wajngurt D, Park N, Zheng T. Probing genetic overlap among complex human phenotypes. Proc Natl Acad Sci USA. 2007; 104:11694–11699. [PubMed: 17609372]

Schuster-Bockler B, Bateman A. Protein interactions in human genetic diseases. Genome Biol. 2008; 9:R9. [PubMed: 18199329]

Schuster S, Fell DA, Dandekar T. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nat Biotechnol. 2000; 18:326–332. [PubMed: 10700151]

Seebacher J, Gavin A-C. SnapShot: Protein-protein interaction networks. Cell. 2011 this issue.

Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D, Friedman N. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. Nat Genet. 2003; 34:166–176. [PubMed: 12740579]

Shapira SD, Gat-Viks I, Shum BOV, Dricot A, de Grace MM, Wu L, Gupta PB, Hao T, Silver SJ, Root DE, et al. A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection. Cell. 2009; 139:1255–1267. [PubMed: 20064372]

Shen-Orr SS, Milo R, Mangan S, Alon U. Network motifs in the transcriptional regulation network of *Escherichia coli*. Nat Genet. 2002; 31:64–68. [PubMed: 11967538]

Shoval O, Alon U. SnapShot: network motifs. Cell. 2010; 143:326–e321. [PubMed: 20946989]

Simonis N, Rual JF, Carvunis AR, Tasan M, Lemmons I, Hirozane-Kishikawa T, Hao T, Sahalie JM, Venkatesan K, Gebreab F, et al. Empirically controlled mapping of the *Caenorhabditis elegans* protein-protein interactome network. Nat Methods. 2009; 6:47–54. [PubMed: 19123269]

Singh GP, Ganapathi M, Dash D. Role of intrinsic disorder in transient interactions of hub proteins. Proteins. 2006; 66:761–765. [PubMed: 17154416]

Sönnichsen B, Koski LB, Walsh A, Marschall P, Neumann B, Brehm M, Alleaume AM, Artelt J, Bettencourt P, Cassin E, et al. Full-genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans*. Nature. 2005; 434:462–469. [PubMed: 15791247]

Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S, et al. A human protein-protein interaction network: a resource for annotating the proteome. Cell. 2005; 122:957–968. [PubMed: 16169070]

Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature. 2009; 458:719–724. [PubMed: 19360079]

Stuart JM, Segal E, Koller D, Kim SK. A gene-coexpression network for global discovery of conserved genetic modules. Science. 2003; 302:249–255. [PubMed: 12934013]

Sturtevant AH. A highly specific complementary lethal system in *Drosophila melanogaster*. Genetics. 1956; 41:118–123. [PubMed: 17247604]

Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, Emili A, Xie XS. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. Science. 2010; 329:533–538. [PubMed: 20671182]

Taylor IW, Linding R, Warde-Farley D, Liu Y, Pesquita C, Faria D, Bull S, Pawson T, Morris Q, Wrana JL. Dynamic modularity in protein interaction networks predicts breast cancer outcome. Nat Biotechnol. 2009; 27:199–204. [PubMed: 19182785]

Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, Page N, Robinson M, Raghibizadeh S, Hogue CW, Bussey H, et al. Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science. 2001; 294:2364–2368. [PubMed: 11743205]

Turinsky AL, Razick S, Turner B, Donaldson IM, Wodak SJ. Literature curation of protein interactions: measuring agreement across major public databases. Database (Oxford). 2010:baq026. [PubMed: 21183497]

Uetz P, Dong YA, Zeretzke C, Atzler C, Baiker A, Berger B, Rajagopala SV, Roupelieva M, Rose D, Fossum E, et al. Herpesviral protein networks and their interaction with the human proteome. Science. 2006; 311:239–242. [PubMed: 16339411]

Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, et al. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. Nature. 2000; 403:623–627. [PubMed: 10688190]

Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. A census of human transcription factors: function, expression and evolution. Nat Rev Genet. 2009; 10:252–263. [PubMed: 19274049]

Vázquez A, Flammini A, Maritan A, Vespignani A. Modeling of protein interaction networks. ComPlexUs. 2003; 1:38–44.

Venkatesan K, Rual JF, Vazquez A, Stelzl U, Lemmens I, Hirozane-Kishikawa T, Hao T, Zenkner M, Xin X, Goh KI, et al. An empirical framework for binary interactome mapping. Nat Methods. 2009; 6:83–90. [PubMed: 19060904]

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, et al. The sequence of the human genome. Science. 2001; 291:1304–1351. [PubMed: 11181995]

Vermeirssen V, Barrasa MI, Hidalgo CA, Babon JA, Sequerra R, Doucette-Stamm L, Barabasi AL, Walhout AJ. Transcription factor modularity in a gene-centered *C. elegans* core neuronal protein-DNA interaction network. Genome Res. 2007; 17:1061–1071. [PubMed: 17513831]

Vidal, M. The reverse two-hybrid system. In: Bartels, P.; Fields, S., editors. The Yeast Two-Hybrid System. New York: Oxford University Press; 1997. p. 109-147.

Vidal M. A biological atlas of functional maps. Cell. 2001; 104:333–339. [PubMed: 11239391]

Vidal M. A unifying view of 21st century systems biology. FEBS Lett. 2009; 583:3891–3894. [PubMed: 19913537]

von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, Bork P. Comparative assessment of large-scale data sets of protein-protein interactions. Nature. 2002; 417:399–403. [PubMed: 12000970]

Wachi S, Yoneda K, Wu R. Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. Bioinformatics. 2005; 21:4205–4208. [PubMed: 16188928]

Walhout AJ. Unraveling transcription regulatory networks by protein-DNA and protein-protein interaction mapping. Genome Res. 2006; 16:1445–1454. [PubMed: 17053092]

Walhout AJ, Reboul J, Shtanko O, Bertin N, Vaglio P, Ge H, Lee H, Doucette-Stamm L, Gunsalus KC, Schetter AJ, et al. Integrating interactome, phenome, and transcriptome mapping data for the *C. elegans* germline. Curr Biol. 2002; 12:1952–1958. [PubMed: 12445390]

Walhout AJ, Sordella R, Lu X, Hartley JL, Temple GF, Brasch MA, Thierry-Mieg N, Vidal M. Protein interaction mapping in *C. elegans* using proteins involved in vulval development. Science. 2000a; 287:116–122. [PubMed: 10615043]

Walhout AJ, Temple GF, Brasch MA, Hartley JL, Lorson MA, van den Heuvel S, Vidal M. GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. Methods Enzymol. 2000b; 328:575–592. [PubMed: 11075367]

Walhout AJ, Vidal M. Protein interaction maps for model organisms. Nat Rev Mol Cell Biol. 2001; 2:55–62. [PubMed: 11413466]

Wu X, Jiang R, Zhang MQ, Li S. Network-based global inference of human disease genes. Mol Syst Biol. 2008; 4:189. [PubMed: 18463613]

Wuchty S, Oltvai ZN, Barabasi AL. Evolutionary conservation of motif constituents in the yeast protein interaction network. Nat Genet. 2003; 35:176–179. [PubMed: 12973352]

Yeger-Lotem E, Sattath S, Kashtan N, Itzkovitz S, Milo R, Pinter RY, Alon U, Margalit H. Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction. Proc Natl Acad Sci USA. 2004; 101:5934–5939. [PubMed: 15079056]

Yu H, Braun P, Yildirim MA, Lemmens I, Venkatesan K, Sahalie J, Hirozane-Kishikawa T, Gebreab F, Li N, Simonis N, et al. High-quality binary protein interaction map of the yeast interactome network. Science. 2008; 322:104–110. [PubMed: 18719252]

Zhang LV, King OD, Wong SL, Goldberg DS, Tong AHY, Lesage G, Andrews B, Bussey H, Boone C, Roth FP. Motifs, themes and thematic maps of an integrated *Saccharomyces cerevisiae* interaction network. J Biol. 2005; 4:6. [PubMed: 15982408]

Zhong Q, Simonis N, Li QR, Charloteaux B, Heuze F, Klitgord N, Tam S, Yu H, Venkatesan K, Mou D, et al. Edgetic perturbation models of human inherited disorders. Mol Syst Biol. 2009; 5:321. [PubMed: 19888216]
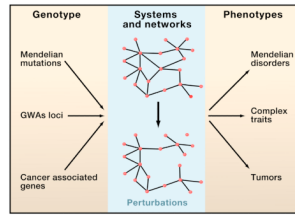
Figure 1. Perturbations in biological systems and cellular networks may underlie genotype-phenotype relationships

By interacting with each other, genes and their products form complex cellular networks. The link between perturbations in network and systems properties and phenotypes, such as Mendelian disorders, complex traits, and cancer, might be as important as that between genotypes and phenotypes.
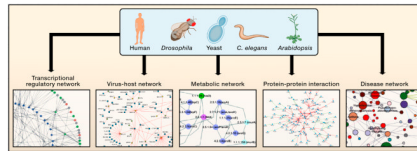
Figure 2. Networks in cellular systems
To date cellular networks are most available for the "super-model" organisms (Davis, 2004) yeast, worm, fly, and plant. High-throughput interactome mapping relies upon genome-scale resources such as ORFeome resources, a segment of which is shown in the background. Several types of interactome networks discussed are depicted around the periphery. In a protein interaction network nodes represent proteins and edges represent physical interactions. In a transcriptional regulatory network nodes represent transcription factors (circular nodes) or putative DNA regulatory elements (diamond nodes) and edges represent physical binding between the two. In a gene-disease network, nodes represent disease genes and edges represent genes mutation of which is associated with disease. In a virus-host network nodes represent viral proteins (square nodes) or host proteins (round nodes) and edges represent physical interactions between the two. In a metabolic network nodes represent enzymes and edges represent metabolites that are products or substrates of the enzymes. The network depictions seem dense, but they represent only small portions of available interactome network maps, which themselves constitute only a few percent of the complete interactomes within cells.
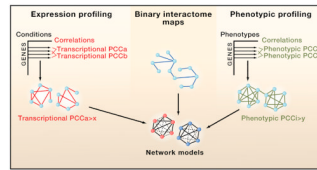
Figure 3. Integrated networks
Co-expression and phenotypic profiling can be thought of as matrices comprising all genes of an organism against all conditions that this organism has been exposed to within a given expression compendium and all phenotypes tested, respectively. For any correlation measurement, Pearson Correlation Coefficients (PCCs) being one of the most widely used, the threshold between what is considered co-expressed and non-co-expressed needs to be set using appropriate titration procedures. The resulting integrated networks have powerful predictive value.