

# Protein–protein interaction networks: unraveling the wiring of molecular machines within the cell

Javier De Las Rivas and Celia Fontanillo

Advance Access publication date 20 August 2012

## Abstract

Mapping and understanding of the protein interaction networks with their key modules and hubs can provide deeper insights into the molecular machinery underlying complex phenotypes. In this article, we present the basic characteristics and definitions of protein networks, starting with a distinction of the different types of associations between proteins. We focus the review on protein–protein interactions (PPIs), a subset of associations defined as physical contacts between proteins that occur by selective molecular docking in a particular biological context. We present such definition as opposed to other types of protein associations derived from regulatory, genetic, structural or functional relations. To determine PPIs, a variety of binary and co-complex methods exist; however, not all the technologies provide the same information and data quality. A way of increasing confidence in a given protein interaction is to integrate orthogonal experimental evidences. The use of several complementary methods testing each single interaction assesses the accuracy of PPI data and tries to minimize the occurrence of false interactions. Following this approach there have been important efforts to unify primary databases of experimentally proven PPIs into integrated databases. These meta-databases provide a measure of the confidence of interactions based on the number of experimental proofs that report them. As a conclusion, we can state that integrated information allows the building of more reliable interaction networks. Identification of communities, cliques, modules and hubs by analysing the topological parameters and graph properties of the protein networks allows the discovery of central/critical nodes, which are candidates to regulate cellular flux and dynamics.

**Keywords:** *Interactome; proteomics; protein interaction; network; protein hub; bioinformatics*

## HOLISTIC APPROACH TO BIOLOGICAL SYSTEMS: FROM BIOMOLECULAR ENTITIES TO NETWORK BIOLOGY

Many large-scale and high-throughput experimental techniques—mostly applied in the last decade—are producing an outstanding advance in molecular and cell biology, moving biological research into a new global scenario. Genomics, transcriptomics, proteomics and all the new ‘omic’ technologies prove

that we are in a new research era that comprehends global biological systems.

To understand a biological system at molecular level, we need to identify and characterize all biomolecular entities—e.g. genes, proteins—that play a role in the particular system. However, it is not enough to obtain the complete list of elements that define a living system (e.g. identify the whole genome and the whole proteome), but we need to build biomolecular maps to show the relative

Corresponding author. Javier De Las Rivas, Bioinformatics and Functional Genomics Research Group, Cancer Research Center (IBMCC, CSIC/USAL), Salamanca, Spain. Tel: +34 923 294819; Fax: +34 923 294743; E-mail: [jrivas@usal.es](mailto:jrivas@usal.es)

**Javier De Las Rivas** MSc, PhD, is CSIC Research Scientist and PI of the Bioinformatics and Functional Genomics group at the Cancer Research Center. He is biochemist and after postdoctoral stays in London (IC) and New York (MSSM), he set up his current group in 2003 focusing his studies on cancer omics and development of bioinformatic methods applied to this field.

**Celia Fontanillo** MsEng, is a young Scientist, degree in Computer Sciences, who has been working for 5 years at the Bioinformatics and Functional Genomics group of the Cancer Research Center. Her expertise includes several programming languages applied to the development of algorithms and methods in bioinformatics and genomics.

location and movement, the paths and ways, the links and crosstalks between the constitutive entities. The aim of achieving such ‘relational maps’ defines the new research field called ‘network biology’ [1]. Moreover, most biological processes arise from complex interactions between the cell’s numerous constituents, such as proteins, DNA, RNA and small molecules. Therefore, a key challenge for biology in the 21st century is to understand the structure and the dynamics of the complex intercellular web of interactions that contribute to the structure and function of a living cell [1].

### **THE PROTEIN INTERACTOME, BITACORA TO UNRAVEL THE COMPLEXITY OF THE BIOMOLECULAR NETWORKS**

Proteins are macromolecular structures that build the nanoscopic working machinery of a living system. Biochemical and biomolecular research for over a century have produced a remarkable compendium of knowledge about the function and properties of many individual proteins. But proteins do not act alone, they team up into molecular machines and complex structures with intricate physicochemical connections to undertake specific functions. The complete map of protein interactions that take place in a living organism is the ‘interactome’ [2].

The collection, verification and validation of the interactions among molecules inside a cell pose considerable challenges and together form an active field in bioinformatics research. Certainly, interactions will not occur all the time and under all conditions. Nevertheless, understanding which proteins interact with one another will give us deeper insights into the molecular machinery underlying complex phenotypes [3].

To draw a comprehensive atlas of all possible protein interactions within a living system is a first-step needed to building its interaction network and to identifying its ‘central nodes’. Complete interactome maps can be most relevant for current biomolecular research, because it is clear that the location of the proteins in their interaction network will allow the evaluation of their centrality and the definition of their role in a relational context. In the case of the human interactome, the identification of protein ‘hubs’ can be a key step to find potential targets, which can be activated or inhibited using drugs to modulate certain pathways altered in specific diseases.

Finally, in the study of the interactomes, we have to consider the dynamic nature of living systems. Each cellular function requires the precise coordination of a large number of events, and the identification of temporal and contextual signals underlying specific protein interactions is a crucial step to understand such functions [4]. Network dynamics can describe, e.g. how cells respond to environmental cues or how a protein network evolves during development or differentiation [4]. Measuring interactome dynamics is much more complicated than obtaining static snapshots of the protein interactions at different times and conditions. However, the construction of reliable protein networks derived from comprehensive mappings is a required step before unraveling the interactome dynamics.

### **TYPES OF PROTEIN ASSOCIATIONS: PHYSICAL, REGULATORY, GENETIC, STRUCTURAL, FUNCTIONAL**

Before analysing the protein interactome, we need to describe the types of relations between proteins that can be found in a biological system [5]. Cellular complexity and cellular dynamics obey many different internal forces and links between the biomolecular entities acting inside an organism. The most common relationships and associations can be organized into the following categories: (i) physical interactions: direct or indirect physical contact between biomolecules; for instance, protein–protein interactions (PPIs) present in processes such as macromolecular protein complex assemblies, protein ligand–receptor activation, signal transduction phosphorylation cascades, etc.; (ii) regulatory associations: activation or inhibition events between biomolecules mediated by intermediate cellular processes; for instance, gene–expression regulation mediated by transcription factors (TFs), regulatory links between extracellular signals and gene response, or transcriptional regulation denoted by gene-to-gene co-expression correlation [6]; (iii) genetic interactions: connection between gene-pairs whose concurrent genetic perturbation leads to a phenotypic result different than that expected from a combination of single gene effects; for instance, synthetic lethal interactions which connect genes that weakly affect an organism viability when are individually deleted, but provoke lethality when are both deleted; (iv) structural similarity: links between two

biomolecular elements that are similar according to a structural attribute; for instance, protein/gene sequence similarity, protein 3D structural similarity, etc. and (v) functional associations: links between two biomolecular elements that have a functional connection because they are involved in the same signalling/metabolic pathway or in the same biomolecular process; for instance, two enzymes that work in the glycolysis pathway or in the Krebs cycle, two proteins enrolled in the WNT signalling pathway, co-location in the same organelle or macrostructure of the cell (e.g. endoplasmic reticulum).

Note that the categories described above are not exclusive. For example, some regulatory associations can include physical interactions as it is the case for allosteric regulation of enzymes and some functional associations can also include in some cases physical proximity and interaction. Therefore, different types of association can be assigned to the same protein-protein pair.

All the described relations and associations can be used to decipher the function of genes and proteins and to identify groups of proteins that work together controlling specific biological processes [7]. Different types of links are sometimes difficult to combine because, usually they have different biological meanings [8]. The strength of each type of protein-protein link depends very much on the experimental data and biological information that support it, but it is clear that the determination of the global map of physical PPIs present in a given biological system will provide a good view of the molecular network that drives the behaviour of such living system.

### **PPIS: SPECIFIC PHYSICAL CONTACTS BETWEEN PROTEINS THAT OCCUR BY SELECTIVE MOLECULAR DOCKING**

PPIs are commonly defined as physical contacts involving molecular docking between proteins that occur in a living organism *in vivo*. Such physical contacts are specific but they can be 'direct', embracing a molecular interface between two proteins, or 'indirect', when the protein-protein contact is mediated by other or others intermediate protein molecules building a complex. The question of whether two proteins share a 'functional association' is quite different from the question of whether two proteins have 'physical contact' with each other [9]. Any protein in the basal transcriptional regulatory apparatus

shares a functional association with the other proteins in these large structures, but certainly not all the proteins involved in the function of a particular cellular system have physical interactions. As indicated in the previous section, it is interesting to explore all types of 'links' between proteins in living organisms, but these associations should not be confused with protein physical interactions. Moreover, identification of different types of protein physical interactions that involve contact with other molecules (i.e. protein-DNA, protein-RNA, protein-cofactor, protein-ligand) is also important for a comprehensive study of the interactome, but again these types of data should not be confused or mixed if we want to build an atlas of PPIs. In conclusion, considering the ideas exposed, we provide a definition of PPIs as: specific, direct or indirect physical contacts between proteins that occur by selective molecular docking in a particular biological context [9].

### **EXPERIMENTAL DETERMINATION OF PHYSICAL INTERACTIONS BETWEEN PROTEINS: BINARY METHODS AND CO-COMPLEX METHODS**

The experimental determination of a given PPI in a biological system is not always easy. Several research groups have indicated that it is not acceptable to conclude that two proteins interact directly, provided only that their interaction is demonstrated by pulldown or co-immunoprecipitation (co-IP) experiments [10, 11]. A positive result with these methods does not imply a direct interaction between two proteins, since the binding can occur by intermediate hidden partners. In addition, there is a widespread misconception that co-IPs from cellular extracts provide '*in vivo* evidence' of the existence of an interaction. This is not accurate, particularly when the experiments are carried out using over-expressed proteins in cell lines. Pull-down assays that rely on glutathione-S-transferase (GST) or other affinity tags can also give rise to problems. For example, it has been reported that interactions found with GST pull-downs using bacterial-expressed protein domains could not be detected using other biophysical techniques [10]. These observations bring about the need to use adequate experimental methods in PPI studies taking into account that not all the methods provide the same information.

The experimental methods to determine PPIs can be divided in two major classes: (i) binary methods: methods that interrogate direct pair-wise PPIs, designed to test each specific interaction between a pair of proteins, (ii) co-complex methods: methods that tag one specific protein (bait-protein) and interrogate its interaction with a group of proteins (prey-proteins) finding direct and indirect physical associations. These methods are designed to find interactions between the tagged protein and a group of proteins without a clear dissection of the pair-wise interactions that occur between each protein pair [9].

In large-scale high-throughput studies, the most common binary methods are the two-hybrid systems, yeast two-hybrid (Y2H) being the most widely and successfully used methodology [12, 13]. Currently two-hybrid (2H)-based methods include a large series of different technologies to be used not only in yeast-cells but also in mammalian-cell systems and in bacterial systems [14, 15]. Also new variants of 2H methodologies have been developed regarding the compartment and the cell type to overcome the limitations of the classic ‘nuclear’ Y2H [14]. A review focused on benchmarking binary interaction assays have been published recently [16].

Large-scale automated 2H approaches have been crucial to achieve global interactome studies that try to cover whole organisms’ proteomes. Matrixes with thousands of open reading frames (ORFs) cloned into bait and prey vectors were used to generate the first overviews of the yeast *Saccharomyces cerevisiae* protein interactome network [17, 18]. Since then, similar comprehensive 2H screens have been undertaken on two metazoan organisms: *Drosophila melanogaster* [19] and *Caenorhabditis elegans* [20]. Later on, several landmark studies addressed the initial mapping of the human interactome [21, 22]. These studies are still partial, but have identified thousands of PPIs.

The most common co-complex method, which has produce large-scale datasets, is tandem affinity purification followed by mass spectrometry (TAP-MS) that was first applied to systematic analysis of multi-protein complexes in yeast *S. cerevisiae* [23, 24]. In this technique a protein mixture—usually a lysate from the cell or tissue of interest—is passed through the matrix where a single protein (bait) is affinity captured, and interacting partners (preys) are retained by interaction with the bait. Proteins that do

not interact, pass through the matrix and are discarded. The captured protein complexes, composed of bait and preys, are analysed by mass spectrometry, identifying interaction participants from their peptide signatures [25]. Mass spectrometry is capable of identifying hundreds of potential interactors simultaneously at subpicomole concentrations [25]. Some recent reviews discussing the capabilities and limitations of AP-MS technology, describe improvements achieved combining multiple biological replicates, and dealing with data generated using different tagging strategies [26, 27]. There are several alternative methods to the affinity purification (AP) step, the most common ones being protein immunoprecipitation (IP) and pull-down of epitope-tagged molecules [28]. The final result of all these approaches is the identification of interactions between multiple proteins, i.e. ‘n-ary interactions’. For this reason they can be called co-complex methods. In these results, each binary PPI between bait and prey cannot be directly deduced without producing some false positive estimations. This is a disadvantage of the co-complex methods. A review about the strengths and weaknesses of mass spectrometry applied to map PPIs can be found in reference [27].

An advantage of AP versus 2H technique is that isolated prey proteins can be in concentrations more similar to the *in vivo* status and can keep the folding native state better than the proteins expressed from cDNAs in the 2H systems. However, it is important to underline that none of the two approaches is able to interrogate the PPIs in their natural *in vivo* cellular context. Both types of techniques require *in vitro* assays where proteins are tested separately, since it is the only way to prove specific interactions.

## PPI DATABASES AND RESOURCES: WHERE AND HOW TO QUERY FOR INTERACTIONS TO BUILD SPECIFIC PROTEIN NETWORKS

An analysis and comparison of public PPI data resources according to the types of interactions included allows them to be divided into three major types: (i) primary databases, which include experimentally proven PPIs coming from either small-scale or large-scale studies, that have been published and are manually curated by experts of the database; (ii) meta-databases, which include only experimentally proven PPIs obtained by consistent

integration and unification of several primary databases (sometimes including small sets of original PPI data); (iii) prediction databases, which include mostly predicted PPIs obtained using different bioinformatic analysis or combine many predicted PPIs with experimentally detected PPIs [9].

Some well known and highly used PPI primary databases are: BioGRID [29], DIP [30], HPRD [31], IntAct [32] and MINT [33]. PPI meta-databases—developed due to the lack of overlap between primary databases and due to the need of using unified non-redundant datasets—are also resources in demand, e.g.: APID [34], iRefWeb [35] and the work done by the IMEx international consortium [36]. These databases provide integrated web access where unified experimental protein interactions can be easily queried and explored. With respect to the third type of PPI databases, one of the most used resources developed by experts in computational prediction methods is STRING, which also includes experimental data on different types of protein associations [37].

In 2010, a challenging project from the HUPO initiative was promoted to unify the access to the main PPI databases [38]. This project, called PSICQUIC, relies on the previous establishment of a controlled vocabulary and a common representation standard developed by the Molecular Interactions group of the HUPO Proteomics Standard Initiative (PSI-MI) [39]. This standardized access allows simultaneous interrogation of a series of associated databases and the search of many types of interactions. All the above described PPI resources and services are designed to facilitate the construction of specific interaction networks for any given protein set of interest.

### **IMPROVING RELIABILITY OF THE PROTEIN NETWORKS: INCREASING CONFIDENCE AND COVERAGE**

Despite the fact that all PPIs included in the primary resources mentioned above come from experimental data, they are sometimes noisy and still incomplete. There are multiple reasons that can bring about errors in the determination of protein interactions. For example, the isolation of the proteins from its natural native environment to test the interactions can give rise to multiple types of artefacts and mistakes in the experimental detection. Other common

reasons provoking error are failures in the consideration of the specific cellular location of the proteins, or lack of specific biomolecular partners needed for the interactions that are lost during isolation. These difficulties are present in all types of techniques though the bias and error propensity are different for each type of experimental approach. However, it has been shown that the error levels are similar in high- or low-throughput systems [11]. We need ways to estimate the error rates in a given PPI network or, at least, ways to assign a confidence level to each interaction present in a network obtained for a given study. It is still a challenge to minimize the occurrence of false positives (FP), resulting in the improvement of confidence in the detected interactions and the minimization of the occurrence of false negatives (FN), while increasing the coverage of the PPI networks built.

Since no single experimental approach has optimal sensitivity (i.e. no FN) and optimal specificity (i.e. no FP), probably one of the best ways to increase the confidence in a given protein interaction is to integrate orthogonal experimental evidences. Several studies have demonstrated a confidence improvement by considering the use of complementary experimental methods applied to test each single interaction [34, 40, 41]. Some strategies based on distances and weights calculated according to the number of experiments have proven that each interaction has been applied quite successfully [42]. Also an empirical framework for assessing completeness of binary interactome mappings has been proposed based on this type of strategy [43]. These efforts to increase the number of experimental methods that validate the interactions are leading to the construction of more accurate interactome networks, which provide more complete and reliable PPI maps.

In these strategies, only experimental detection interaction methods are taken into account, although other simple criteria like the number of supporting publications, the co-expression of the participant genes, co-occurrence in the same biological process or pathway can be used to increase the confidence of the interactions [37]. These types of information about the interaction partners have to be always used as a complementary approach to the experimental PPI data, because none of them is a direct proof of a physical interaction.

Finally, in many studies it can be useful to compare the interactome networks obtained for a set

of query proteins in different organisms by integrating information about orthologous partners (i.e. interologs). Comparative analysis of the conservation of interactions among different species can introduce evolutionary insights about the architecture of the PPI networks, helping to identify essential interactions that are maintained during evolution.

### **PPI NETWORKS: FINDING PROTEIN ‘COMMUNITIES’, PROTEIN ‘CLIQUES’ AND PROTEIN ‘HUBS’ IN THE CELLULAR LANDSCAPE**

Network representation has been widely used in many scientific disciplines (sociology, physics, telecommunications, biology, etc.) where it is necessary to explore and compare large complex datasets that include relationships between elements. A great advantage of networks is that they can be studied by applying graph theory and other powerful analytical techniques. Protein interaction data can be represented as a network diagram where nodes correspond to proteins and edges to interactions between protein pairs. These networks are undirected when there is no experimental information about the source or destination nodes. Directed graphs can be produced including, e.g. identification of the bait and prey proteins or information about the enzyme and target relationship. The networks are unweighted by default, although weights can be assigned to the edges according to the confidence of the interactions or to other properties scored.

As mentioned before, there are still technical problems that need to be solved in order to reduce the FPs and FNs in PPI datasets. However, several studies on interactome networks, even though incomplete, have led to a consensus on several characteristics common for these interactome networks based on their topology [17–20]. According to Barabasi *et al.* [44], it seems that PPI networks are ‘small world’ networks characterized by a low connectivity [1, 44]. This means that the average distance between each pair of nodes is small and that the major part of the nodes are not directly linked, but the length of the shortest path between them is small. These observations lead to the proposal that the protein interactomes are ‘scale free’ networks with a degree distribution that follows a power-law function [1, 44]. This model is still open

to discussion since there are other authors that consider it critically [45, 46]. Despite the lack of clear model identification, the PPI networks show the existence of vertices with a degree that greatly exceeds the average (called ‘hubs’). Several authors have distinguished between two types of hubs in the PPI networks: ‘party’ hubs, which interact with most of their partners simultaneously, and ‘date’ hubs, which bind their partners at different times or locations [47]. Some studies have linked hubs with proteins that are essential for the biological system. That is because they observed that the likelihood that a protein is essential correlates with its connectivity degree [48]. This means that the cells are more vulnerable to the loss of hubs than non-hubs, because the disruption of hubs—especially ‘date’ hubs—causes the breakdown of the network into isolated clusters. In contrast, random node deletion does not lead to a major loss of connectivity in scale-free networks, and this confirms the robustness of cellular networks against random disruptions [44].

Another common practice in the analyses of networks is to find node ‘communities’, ‘cliques’ and ‘modules’. Communities are sets of nodes that have a dense connectivity between them and can be separated from the rest of the network using some topological criteria. Cliques and modules are smaller groups of nodes that have similar characteristics and are closely located in the network. The cliques are defined as a subset of nodes in a network such that every two nodes in the subset are connected by an edge. In this way the cliques are specified by the parameter  $k$  ( $k$ -clique) that indicates the number of nodes that includes, e.g. 5-cliques are groups of 5 nodes, where all are interconnected by edges (i.e. each one is connected with all the others). The definition of module is more open and different in several research forums, but it always try to indicate a group of nodes that are heavily interconnected, often following some specific graph pattern. It is interesting to look for communities, cliques and modules in the PPI networks, because the forming nodes tend to have related biological functions and many times is a good way to predict functional association. Since protein interaction networks are highly connected, the communities, cliques and modules should not be understood only as sets of nodes disconnected from other sets, but rather as nodes that have dense intra-modular connectivity and sparse inter-modular connectivity.

In conclusion, a proper analysis of the cellular landscape requires the deciphering of interaction patterns between all elements of its biomolecular machinery and how such interactions build the complex networks that operate inside cells. Empirical determination and mapping cellular protein networks for a few model organisms and for human is providing the necessary scaffold toward understanding the functional, logical and dynamical aspects of cellular systems. 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 [49].

### Key Points

- Holistic approach to biological systems: from biomolecular entities to network biology.
- The protein interactome, bitacora to unravel the complexity of the biomolecular networks.
- Types of protein associations: physical, regulatory, genetic, structural, functional.
- PPIs: specific physical contacts between proteins that occur by selective molecular docking.
- Experimental determination of physical interactions between proteins: binary methods and co-complex methods.
- PPI databases and resources: where and how to query for interactions to build specific protein networks.
- Improving reliability of the protein networks: increasing confidence and coverage.
- PPI networks: finding protein 'communities', protein 'cliques' and protein 'hubs' in the cellular landscape.

### FUNDING

This work was supported by the Consejo Superior de Investigaciones Científicas (CSIC) [project i-LINK0398]; the Spanish Government (ISCiii) [project PS09/00843]; and the European Commission [project FP7-HEALTH-2007-223411]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### References

1. Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet* 2004;**5**:111–3.
2. Cusick ME, Klitgord N, Vidal M, *et al.* Interactome: gateway into systems biology. *Hum Mol Genet* 2005;**14**:R171–81.
3. Stumpf MPH, Kelly WP, Thorne T, *et al.* Evolution at the system level: the natural history of protein interaction networks. *Trends Ecol Evol* 2007;**22**:366–73.
4. Przytycka TM, Singh M, Slonim DK. Toward the dynamic interactome: it's about time. *Brief Bioinform* 2010;**11**:15–29.
5. Merico D, Gfeller D, Bader GD. How to visually interpret biological data using networks. *Nat Biotechnol* 2009;**27**:921–4.
6. Prieto C, Risueño A, Fontanillo C, *et al.* Human gene coexpression landscape: confident network derived from tissue transcriptomic profiles. *PLoS One* 2008;**3**:e3911.
7. Boone C, Bussey H, Andrews BJ. Exploring genetic interactions and networks with yeast. *Nat Rev Genet* 2007;**8**:437–49.
8. De Las Rivas J, De Luis A. Interactome data and databases: different types of protein interaction. *Comp Funct Genomics* 2004;**5**:173–8.
9. De Las Rivas J, Fontanillo C. Protein-protein interactions essentials: key concepts to building and analyzing interactome networks. *PLoS Comp Biol* 2010;**6**:e1000807.
10. Mackay JP, Sunde M, Lowry JA, *et al.* Protein interactions: is seeing believing? *Trends Biochem Sci* 2007;**32**:530–1.
11. Cusick ME, Yu H, Smolyar A, *et al.* Literature-curated protein interaction datasets. *Nat Methods* 2009;**6**:39–46.
12. Fields S, Song O. A novel genetic system to detect protein-protein interactions. *Nature* 1989;**340**:245–6.
13. Dreze M, Monachello D, Lurin C, *et al.* High-quality binary interactome mapping. *Methods Enzymol* 2010;**470**:281–315.
14. Suter B, Kittanakom S, Stagljar I. Two-hybrid technologies in proteomics research. *Curr Opin Biotechnol* 2008;**19**:316–23.
15. Kanno A, Ozawa T, Umezawa Y. Detection of protein-protein interactions in bacteria by GFP-fragment reconstitution. *Methods Mol Biol* 2011;**705**:251–8.
16. Braun P. Interactome mapping for analysis of complex phenotypes: Insights from benchmarking binary interaction assays. *Proteomics* 2012;**12**:1499–518.
17. Uetz P, Giot L, Cagney G, *et al.* A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 2000;**403**:623–7.
18. Ito T, Chiba T, Ozawa R, *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci USA* 2001;**98**:4569–74.
19. Giot L, Bader JS, Brouwer C, *et al.* A protein interaction map of *Drosophila melanogaster*. *Science* 2003;**302**:1727–36.
20. Li S, Armstrong CM, Bertin N, *et al.* A map of the interactome network of the metazoan *C. elegans*. *Science* 2004;**303**:540–3.
21. Stelzl U, Worm U, Lalowski M, *et al.* A human protein-protein interaction network: a resource for annotating the proteome. *Cell* 2005;**122**:957–68.
22. Rual JF, Venkatesan K, Hao T, *et al.* Towards a proteome-scale map of the human protein-protein interaction network. *Nature* 2005;**437**:1173–8.
23. Ho Y, Gruhler A, Heilbut A, *et al.* Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 2002;**415**:180–3.
24. Gavin AC, Bosche M, Krause R, *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 2002;**415**:141–7.

25. Koh GCKW, Porras P, Aranda B, *et al.* Analyzing Protein–Protein Interaction Networks. *J Proteome Res* 2012;**11**:2014–31.
26. Gavin AC, Maeda K, Kühner S. Recent advances in charting protein–protein interaction: mass spectrometry-based approaches. *Curr Opin Biotechnol* 2011;**22**:42–9.
27. Nesvizhskii AI. Computational and informatics strategies for identification of specific protein interaction partners in affinity purification mass spectrometry experiments. *Proteomics* 2012;**12**:1639–55.
28. Dunham WH, Mullin M, Gingras AC. Affinity-purification coupled to mass spectrometry: Basic principles and strategies. *Proteomics* 2012;**12**:1576–90.
29. Stark C, Breitkreutz BJ, Chatr-Aryamontri A, *et al.* The BioGRID Interaction Database: 2011 update. *Nucleic Acids Res* 2011;**39**:D698–704.
30. Salwinski L, Miller CS, Smith AJ, *et al.* The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res* 2004;**32**:D449–51.
31. Mishra GR, Suresh M, Kumaran K, *et al.* Human protein reference database: 2006 update. *Nucleic Acids Res* 2006;**34**:D411–44.
32. Kerrien S, Aranda B, Breuza L, *et al.* The IntAct molecular interaction database in 2012. *Nucleic Acids Res* 2012;**40**:D841–6.
33. Licata L, Briganti L, Peluso D, *et al.* MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res* 2012;**40**:D557–61.
34. Prieto C, De Las Rivas J. APID: Agile Protein Interaction DataAnalyzer. *Nucleic Acids Res* 2006;**34**:W298–302.
35. Turner B, Razick S, Turinsky AL, *et al.* iRefWeb: interactive analysis of consolidated protein interaction data and their supporting evidence. *Database* 2010;**2010**:baq023.
36. Orchard S, Kerrien S, Abbani S, *et al.* Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. *Nat Methods* 2012;**9**:345–50.
37. Szklarczyk D, Franceschini A, Kuhn M, *et al.* The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* 2011;**39**:D561–8.
38. Aranda B, Blankenburg H, Kerrien S, *et al.* PSICQUIC and PSISCORE: accessing and scoring molecular interactions. *Nat Methods* 2011;**8**:528–9.
39. Orchard S, Kerrien S. Molecular interactions and data standardisation. *Methods Mol Biol* 2010;**604**:309–18.
40. Prieto C, De Las Rivas J. Structural domain–domain interactions: Assessment and comparison with protein–protein interaction data to improve the interactome. *Proteins* 2010;**78**:109–17.
41. Yu H, Braun P, Yildirim MA, *et al.* High-quality binary protein interaction map of the yeast interactome network. *Science* 2008;**322**:104–10.
42. Perez-Fernandez J, Roman A, De Las Rivas J, *et al.* The 90S preribosome is a multimodular structure that is assembled through a hierarchical mechanism. *Mol Cell Biol* 2007;**27**:5414–29.
43. Venkatesan K, Rual JF, Vazquez A, *et al.* An empirical framework for binary interactome mapping. *Nat Methods* 2009;**6**:83–90.
44. Albert R, Barabasi AL. Statistical mechanics of complex networks. *Rev Mod Phys* 2002;**74**:47.
45. Przulj N. Protein–protein interactions: making sense of networks via graph-theoretic modeling. *Bioessays* 2011;**33**:115–23.
46. Kelly WP, Ingram PJ, Stumpf MP. The degree distribution of networks: statistical model selection. *Methods Mol Biol* 2012;**804**:245–62.
47. Han JD, Bertin N, Hao T, *et al.* Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature* 2004;**430**:88–93.
48. Jeong H, Mason SP, Barabasi AL, *et al.* Lethality and centrality in protein networks. *Nature* 2001;**411**:41–2.
49. Vidal M, Cusick ME, Barabasi AL. Interactome networks and human disease. *Cell* 2011;**144**:986–98.