Mapping, modeling, and characterization of protein–protein interactions on a proteomic scale

TM Cafarelli[1,2,3,*], A Desbuleux[1,2,3,4,*], Y Wang[1,2,3], SG Choi[1,2,3], D De Ridder[1,2,3] and M Vidal[1,3]

Proteins effect a number of biological functions, from cellular signaling, organization, mobility, and transport to catalyzing biochemical reactions and coordinating an immune response. These varied functions are often dependent upon macromolecular interactions, particularly with other proteins. Small-scale studies in the scientific literature report protein–protein interactions (PPIs), but slowly and with bias towards well-studied proteins. In an era where genomic sequence is readily available, deducing genotype–phenotype relationships requires an understanding of protein connectivity at proteome-scale. A proteome-scale map of the protein–protein interaction network provides a global view of cellular organization and function. Here, we discuss a summary of methods for building proteome-scale interactome maps and the current status and implications of mapping achievements. Not only do interactome maps serve as a reference, detailing global cellular function and organization patterns, but they can also reveal the mechanisms altered by disease alleles, highlight the patterns of interaction rewiring across evolution, and help pinpoint biologically and therapeutically relevant proteins. Despite the considerable strides made in proteome-wide mapping, several technical challenges persist. Therefore, future considerations that impact current mapping efforts are also discussed.

Addresses
[1] Center for Cancer Systems Biology (CCSB), Dana-Farber Cancer Institute, Boston, MA, USA
[2] Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA USA
[3] Department of Genetics, Harvard Medical School, Boston, MA, USA
[4] GIGA-R, University of Liège, Liège, Belgium

Corresponding author: Cafarelli, TM (TizianaM_Cafarelli@dfci.harvard.edu)
* These authors contributed equally.

Introduction
As the most important biological building blocks, proteins mainly carry out their functions by interacting with other biological macromolecules, including DNA, RNA, proteins and small molecules such as lipids and metabolites. Protein–protein interactions, in particular, are incredibly diverse, as they execute a myriad of biological functions. Generating a protein–protein interaction network map at proteome-scale reveals the macromolecular connections that underlie the biology of the cell [1]. Indeed, in order to explore the link between genotype and phenotype and deduce how genetic changes result in disease, an understanding of the cellular network of physical and functional interactions involving proteins is critical [1–4]. As we look to generate the richest and most complete network map possible, we rely on the integration of experimentally derived and computationally predicted interactions. Characterization and application of existing networks has proven useful and highlights the need for expanded network information [5].
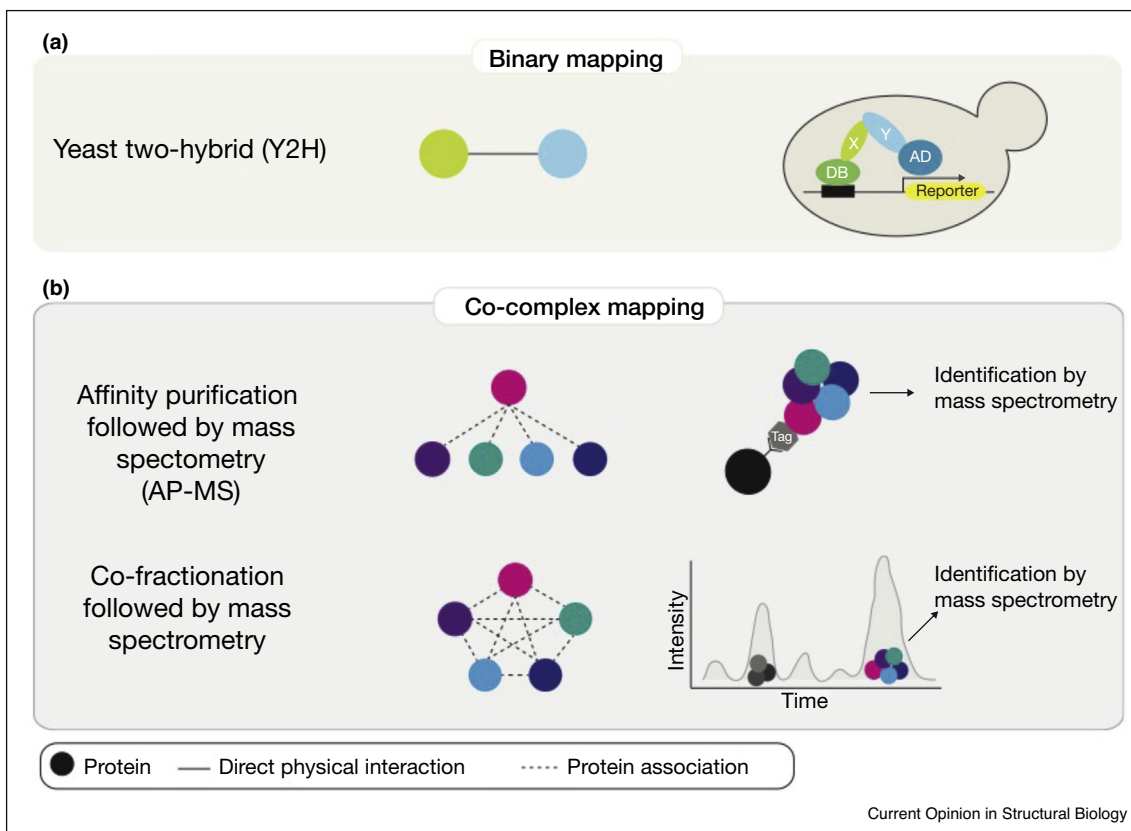
Experimental methods for building a proteome-scale map of the interactome network
There are a number of methods for mapping protein–protein interactions. However, only a few methods are amenable for high-throughput mapping. Recently, proteome-scale interactome maps for human and a number of model organisms have been generated using one of three main techniques. Binary interactome network maps have been generated using yeast two-hybrid (Y2H), and report direct physical interactions. In contrast, co-complex associations can be either direct or indirect protein–protein interactions, and are detected using affinity purification followed by mass spectrometry (AP-MS) or co-fractionation with mass spectrometry (CO-FRAC) (Figure 1). While all of these experimental methods can be adapted to systematically survey the entire proteome, each technique has inherent benefits and limitations.

Binary interaction mapping by yeast two-hybrid (Y2H)
Binary mapping by Y2H detects direct physical interactions between two proteins by the reconstitution of a transcription factor that activates reporter gene expression and promotes yeast cell survival on appropriate selective media (Figure 1a) [6]. Binary interactions identified using Y2H as the primary screening method are validated by a number of orthogonal assays, and data from

Figure 1

Schematic of systematic experimental methods for high-throughput proteome-scale mapping of protein–protein interactions. For each method (left), network representation (middle) and principle (right) are depicted. **(a)** Binary mapping using yeast two-hybrid identifies direct physical interactions between two proteins. **(b)** Co-complex mapping using AP-MS or co-fractionation with mass spectrometry identifies protein associations, which may be either direct or indirect.

such assays indicate that pairs found by this method are of comparable quality as gold standard literature datasets [5,7]. A recent systematic binary mapping study assayed pairs of proteins from a space of ~13 000 × 13 000 human open reading frames (ORFs) and identified ~14 000 protein–protein interactions (PPIs) among ~4300 proteins [8$^{\bullet\bullet}$]. Systematically generated binary maps uniformly identify PPIs in the whole gene space, avoiding sociological bias that may occur in small-scale experiments or literature-curated interactome maps that focus on well-studied genes [8$^{\bullet\bullet}$]. This screening method has therefore proven to be a useful tool, enumerating binary interactions not only for human, but for a number of model organisms as well, including *Saccharomyces cerevisiae* [9–11], *Schizosaccharomyces pombe* [12], *Escherichia coli* [13], *Caenorhabditis elegans* [14,15], and *Arabidopsis thaliana* [16]. While this method is easily scaled and relatively inexpensive, it may fail to capture interactions between proteins which rely on intermediary or scaffold proteins (such as those between protein complex subunits), those involving proteins from specific subcellular compartments (such as membrane proteins), or those which require post-translational modifications [17]. Moreover, this assay requires proteins to be expressed at non-endogenous levels in the yeast nucleus. Such technical requirements may limit the detection of PPIs that require specific protein expression levels (such as protein complexes with strict stoichiometry or stability), or may contribute to the detection of biophysical interactions between proteins that are not endogenously co-expressed or co-localized.

Affinity purification and mass spectrometry (AP-MS)
In interactome mapping by AP-MS, epitope tags are fused to bait proteins, and proteins associated with the bait proteins are purified and identified by mass spectrometry (Figure 1b, top). Two of the latest screening efforts utilizing this method focused on expanding the human interactome network map. The BioPlex dataset reports ~23 700 protein–protein associations (PPAs) among ~7600 proteins, using ~2600 bait proteins over-expressed in HEK293T cells [18$^{\bullet\bullet}$]. An alternative study,

QUBIC, utilized GFP-tagged baits that were incorporated into HeLa cell chromosomes and thus were expressed at near-endogenous expression levels and patterns to identify interactions. The QUBIC dataset includes ∼28 500 PPAs among ∼5500 proteins, resulting from the expression of ∼1100 bait proteins [19]. The investigation of quantitative stoichiometry and protein abundance in the QUBIC dataset suggested that stable complexes are relatively rare, accounting for only 10% of human interactome. Weak, substoichiometric interactions were more critical for maintaining the connectivity of the overall PPI network [19].

AP-MS is efficient in the identification of stably associated and relatively abundant proteins. In particular, PPAs involved in large protein complexes are more likely to be found by AP-MS, as cooperation among multiple proteins may provide the needed structural stability for complexes [11,20]. This method would thereby generate a dataset that is complementary to those constructed using binary PPI mapping methods. However, AP-MS is limited in its ability to detect transient associations, low-abundance complexes, and interactions occurring in particular cell types [21]. In addition, PPAs identified by AP-MS indicate only co-complex associations. For protein complexes in which a mechanistic understanding of assembly is critical, further testing must be done to determine which of these are direct physical interactions.

Co-fractionation and mass spectrometry
In contrast to Y2H and AP-MS, interactome mapping by co-fractionation does not require exogenously introduced ORFs or protein tags (Figure 1b, bottom). Instead, chromatographic and other biochemical separation methods of cell extracts are carried out to generate hundreds to thousands of fractions that are analyzed by mass spectrometry. Protein associations are inferred by co-elution profiles of proteins, as well as weighting from various functional criteria such as co-expression, genetic interactions, domain co-occurrence, and co-evolutionary rates. Havugimana *et al.* were the first to use this method at proteome-scale, and identified ∼14 000 associations among ∼3000 human proteins for ∼600 protein complexes [22•]. An additional effort published in 2016 from the same research group, increased the size of the network map, now composed of ∼28 000 PPAs between ∼4500 proteins [23••]. The human interactome network map derived from co-fractionation showed little evidence for tissue specificity, as few cell type-specific proteins were detected. Moreover, fewer interactions for low abundance proteins were identified [23••], and rigorous computational algorithms are required to reconstruct high confidence co-complex associations. This inference of co-complex association may not be free of sociological bias, as it relies on the quality of previously generated large-scale datasets to train the algorithm and infer the interactions.

Regardless of the mapping strategy employed, protein–protein interactome maps have relatively poor overlap. This is true even when comparing maps generated with the same assay. However, this can be put in perspective when considering the many different technical parameters that influence the detection of PPIs. The genes queried, the strains or cell types used, and the presence and orientation of proteins tags are all examples of the many variables that impact the detection of PPIs [7,17]. Ultimately, the combination of maps generated with different methods provides a more complete view of interactome networks, since each method highlights a different subset of the interactome.

In silico prediction of protein–protein interactions
To fill in the gap between the low coverage of experimentally determined interactome network maps and the urgent need for more complete and detailed interactome information, *in silico* prediction methods of protein–protein interactions have been developed. These facilitate the prioritization of experimental efforts and further our understanding of current biological questions.

Many types of information, such as phylogenetic profiling, gene fusion, co-expression, conservation of gene neighborhood, and gene ontology, can be used individually or together to predict protein–protein interactions [24–29]. Rhodes *et al.* used a naive Bayes model that combines information for ortholog interaction, co-expression, shared gene ontology terms, and enriched domain–domain interaction pairs to predict novel interactions in human [30]. Zhang *et al.* showed that 3D structural information for protein complexes and protein monomers can be very helpful in the prediction of PPIs, when combined with non-structural-based methods [31]. With the use of only protein sequence evolutionary coupling information derived from carefully generated multiple sequence alignments, interactions can be predicted with high accuracy and at the resolution of single residues [32••]. With the combination of 3D docking methods such as HADCOCK, it is possible to acquire protein complex models with even atomic resolution. Finally, new interactions can also be predicted by an algorithm that is purely based on the topology of current network itself, without knowing any additional biological information [33].

While all computational methods rely on different kinds of experimental data, such as sequencing, structural, gene expression, and PPI data, issues such as errors, biases, incompleteness, and misinterpretation of experimental data will influence computational prediction results. The underlying assumptions of computational methods dictate the power of prediction. For example, sequence or structural similarities may not always reflect preservation of biophysical interactions across species. Moreover,

function-based clues may not necessarily lead to biophysical interactions. General network patterns may not be correct for PPI networks. Even if computational predictions can enrich positive interactions many folds better than random, they still suffer from a relatively high false positive rate. The reliability of each method and its predictions must ultimately be assessed by subsequent validation experiments.

Both experimental and predicted PPIs are catalogued in a number of databases and are available to the scientific community. Some major repositories, and their characteristics, are listed in Table 1. A more detailed description of selected repositories has been provided by Zahiri *et al.* [28•].

Modelling PPIs

Since the functions of proteins rely on their unique and dynamic three-dimensional structures in a physiological environment, finer mapping of protein–protein interactions is critical; incorporating structural information to flesh out the molecular details of PPIs will allow us to better understand fundamental cellular processes.

Experimentally derived structures of interacting proteins provide the most detailed description of how PPIs are formed, at times with atomic resolution. For example, they can reveal detailed binding interface residue information or can be used to calculate the affinity of the interaction [34–37]. Such data can also further our understanding of how changes in a protein (*e.g.*, mutations and post-translational modifications) affect interaction or protein stability [38•,39,40]. Notably, detailed structures are critical for drug design and virtual screening of small molecule inhibitors, and serve as a basis for studying the dynamic features of protein complexes, such as conformational changes and allosteric effects [41].

The structures of interacting proteins can be deduced by experimental techniques such as X-ray crystallography, NMR spectroscopy, and cryo-electron microscopy. However, obtaining a high quality complex structure is still very time consuming and labor intensive, and it is rarely used in early stage PPI studies as a result. Considering the size of even small interactomes (*e.g.*, ~18 000 predicted total interactions in *S. cerevisiae* [11]), it would be nearly impossible to experimentally solve all structures. Alternatively, PPI structures can be predicted by methods that incorporate data such as existing PPIs, structure templates, and single protein structures [42••,43–45]. The curation and integration of such data provides a vital resource; an example of such an integral database is Interactome3D, which provides detailed structural information for ~23 500 PPIs in eighteen organisms [42••].

Uses and implications of proteome-scale interactome maps

Once constructed, macromolecular interaction network maps are characterized to determine how reported interactions affect the organization, dynamics, and functions of a given biological system.

Exploration of disease mechanisms

In the era of genomic sequencing, more genetic mutations have been identified than have been functionally characterized. Although over 100 000 mutant alleles associated with Mendelian disorders, complex diseases, and cancer have been catalogued, causal relationships have yet to be determined in many cases [2••,46,47]. It is critical then to discriminate between disease-causing and natural genomic variants, and to determine their respective PPI patterns [2••,8••].

A proteome-scale map serves as a reference, allowing for the enumeration of interactions for wild-type proteins.

Table 1

Most widely used protein–protein interaction databases. The major features of each database, as well as a link to each resource, is provided

Name	Characteristics	Website
BioGrid	Curated protein and genetic interactions, chemical associations, and post-translational modifications for major model organisms	http://thebiogrid.org/
IntAct	PPIs derived from literature curation and from individual user submissions	http://www.ebi.ac.uk/intact/
Interactome3D	Structural annotation of PPIs for eighteen organisms	http://interactome3d.irbbarcelona.org/
STRING	Known and predicted PPIs. Direct and indirect associations with three-dimensional structural annotation	http://string-db.org/
Pathway commons	Provides both pathways and interactions	http://www.pathwaycommons.org/
Mentha	Aggregated data from PPI databases with interaction score that takes supporting evidences into account	http://mentha.uniroma2.it/

Figure 2
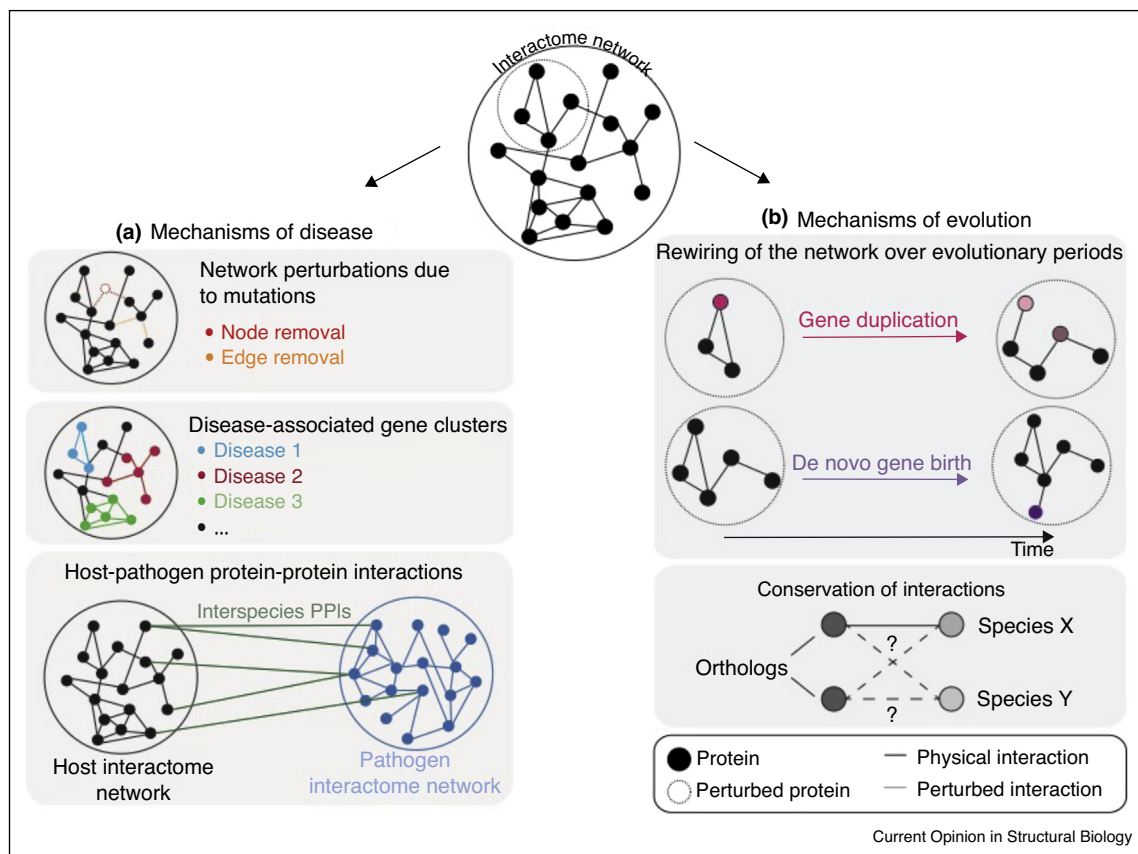
Examples of applications of proteome-scale interactome network maps. A schematic of a network shows proteins (or nodes, indicated as circles) and biophysical interactions (or edges, shown as solid lines). **(a)** Study of disease mechanisms. Top panel: analyses of network perturbations. Mutations can lead to complete loss of the gene product (node removal, illustrated in red) or loss of specific interactions (edge removal, indicated with a dashed orange line). Middle panel: disease-associated genes tend to form clusters, or disease gene modules. Bottom panel: an interspecies interactome network map indicates the interactions between proteins of a host and pathogen. **(b)** Investigation of mechanisms of evolution. Top panel: networks are rewired over evolutionary periods of time. New genes may arise either by the traditional method of gene duplication and diversification (pink) or *de novo* (purple), generating new edges in the PPI network. Bottom panel: possible conservation of protein–protein interactions (represented by the dashed lines) between orthologs of different species.

Determining the loss or gain of key interactions (or edges) of a particular mutant protein, relative to the unperturbed protein, identifies the 'edgotype.' These changes in the interaction profile of the mutant might contribute to a resulting disease state, ultimately allowing for a link between genotype and phenotype [3,4] (Figure 2a, top panel). Analysis of binary interactions from the most recent proteome-scale map of the human interactome network highlights several examples of 'edgetic' alleles (*i.e.*, those which lose some, but not all, edges). For example, Rolland *et al.* determined that R24H and R24C mutations in CDK4, linked to melanoma, disrupt binding to CDKN2C, while the common variants N41S and S52N do not [8••]. Interactions corresponding to functional relationships are more likely to be disrupted by disease-associated mutations than by common variants. An unbiased approach to interactome mapping

has the potential to link uncharacterized genetic mutations to disease phenotypes through altered physical interactions [2••].

A better understanding of disease mechanisms can also be reached through the study of network topology, which allows the identification of patterns in interactome networks. One approach, known as the disease module hypothesis, is based on the observation that disease proteins are not scattered randomly in the interactome, but form topological modules where they tend to interact more with each other than with proteins outside of this neighborhood. These particularly well-connected subgraphs of proteins are called disease modules (Figure 2a, middle panel) [1,48,49]. Based on this hypothesis, Rolland *et al.* expanded the view of the functional cancer landscape by demonstrating that known and

candidate cancer gene products tend to be highly con-nected in the interactome network [8••]. There is growing interest for the disease module hypothesis, as it implies that we can generate lists of genes that are potentially enriched for new candidate disease genes. Those candi-dates are prioritized based on the fact that the number of edges separating the gene products within modules is on average lower than for random sets of gene products [1,50]. Moreover, Kitsak *et al.* recently showed that the tissue specificity of many diseases can be explained by the integrity of expression of the disease module, rather than by the expression of genes carrying the disease-associated mutation alone [51•].

Complementary to intraspecies resource maps, interspe-cies maps of pathogen-host interactomes have been built to study the global landscape of host perturbations by pathogens (Figure 2a, bottom panel). One interactome map published in 2012 illustrated the rewiring of the host cell (*Homo sapiens*) network under viral perturbations from four families of DNA tumor virus proteins [52]. Systematic analysis of host targets of viral proteins can identify cancer genes with a success rate on par with functional genomics and large-scale cataloguing of tumor mutations [52].

Study of protein–protein interaction rewiring in evolution

To date, protein–protein interaction maps have been completed for various model organisms, including *S. cerevisiae* [9–11,53–55], *S. pombe* [12], *A. thaliana* [16], *Drosophila melanogaster* [56] and *C. elegans* [14,15], among others. Comparisons between these interactomes, with a particular focus on orthologous proteins, allows for the study of network evolution (Figure 2b, bottom panel), and allows us to see changes beyond those in the genome. Many studies have examined network evolution at vari-ous scales, in both distantly and closely related species.

To study the conservation of PPI networks, Wan *et al.* generated protein complex maps with a standardized approach for nine species spanning a billion years of evolutionary divergence [23••]. Most of the interacting proteins were conserved across those species, and thus over one million high-confidence co-complex associations for ∼120 eukaryotes with sequenced genomes were pre-dicted. Metazoan complexes, or 'new' complexes, tend to be enriched for roles related to multicellularity, whereas 'old' complexes are involved in core biological processes. Moreover, larger complexes have been shown to be more evolutionarily conserved, while small complexes are more functionally unannotated and recently emerged in verte-brate evolution [22•,23••].

To determine whether proteins retain the ability to interact over evolutionary distances, an interspecies inter-actome map was constructed [57]. Biophysical interac-tions between human and budding yeast proteins were identified and then functionally characterized. This yeast–human inter-interactome map reveals that co-func-tionality and binding ability are preserved. For example, DNA repair proteins, human MLH1 and yeast Mlh1, are both able to bind yeast Ntg2, suggesting that this ances-tral interaction has been preserved despite differentiation of the two species and their proteome. The construction of inter-interactome network maps such as this permit the examination of the relationship between conservation at the level of sequence and three-dimensional structure with functionality, and determine the effect of evolution on biophysical and functional network coordination.

An interactome also evolves as it both gains and loses interactions. A protein interactome map generated for *A. thaliana* provides evidence of dynamic rewiring [16]. Paralogous pairs of proteins arising from gene duplication events and sequence diversification (Figure 2b, top panel) exhibited a range of interaction rewiring, measured as the fraction of shared interactions, and therefore a range of functional divergence. A recent study on a subset of paralogous pairs in budding yeast showed surprising results; instead of only providing robustness, gene dupli-cation may increase mutational fragility in some cases. Half of the paralogous pairs tested required each other's presence to maintain their interaction profile [58].

The birth of new interactions is of particular interest, and a more recent question (Figure 2b, top panel). How do new protein–protein interactions arise? Which mecha-nisms might be used to introduce these new associations in the normal cellular repertoire? It has been proposed that non-functional interactions may represent one reser-voir from which functional interactions may arise [57]. Additionally, the duplication and diversification of exist-ing genes may result in the integration of new interactions (Figure 2b, top panel). Alternatively, new interactions may result from the *de novo* birth of genes. Open reading frames, residing in non-genic regions of the genome and showing evidence of transcription and translation, encode short novel peptides. These so-called proto-genes, which display intermediate features between non-genic sequences and *bona fide* genes, offer adaptive potential if their expression and interactions confer a fitness advan-tage under adverse conditions [59•]. While proto-genes have been identified in a number of organisms, how they participate in and re-shape the cell's protein–protein interactions has yet to be determined at proteome-scale [60,61•,62,63].

Considerations for future interactome mapping efforts
Protein complexity

Protein interactome maps fail to consider the full com-plexity of the human proteome, as generated by alterna-tive splicing, post-translational modifications, somatic hypermutation, and other such mechanisms

Figure 3
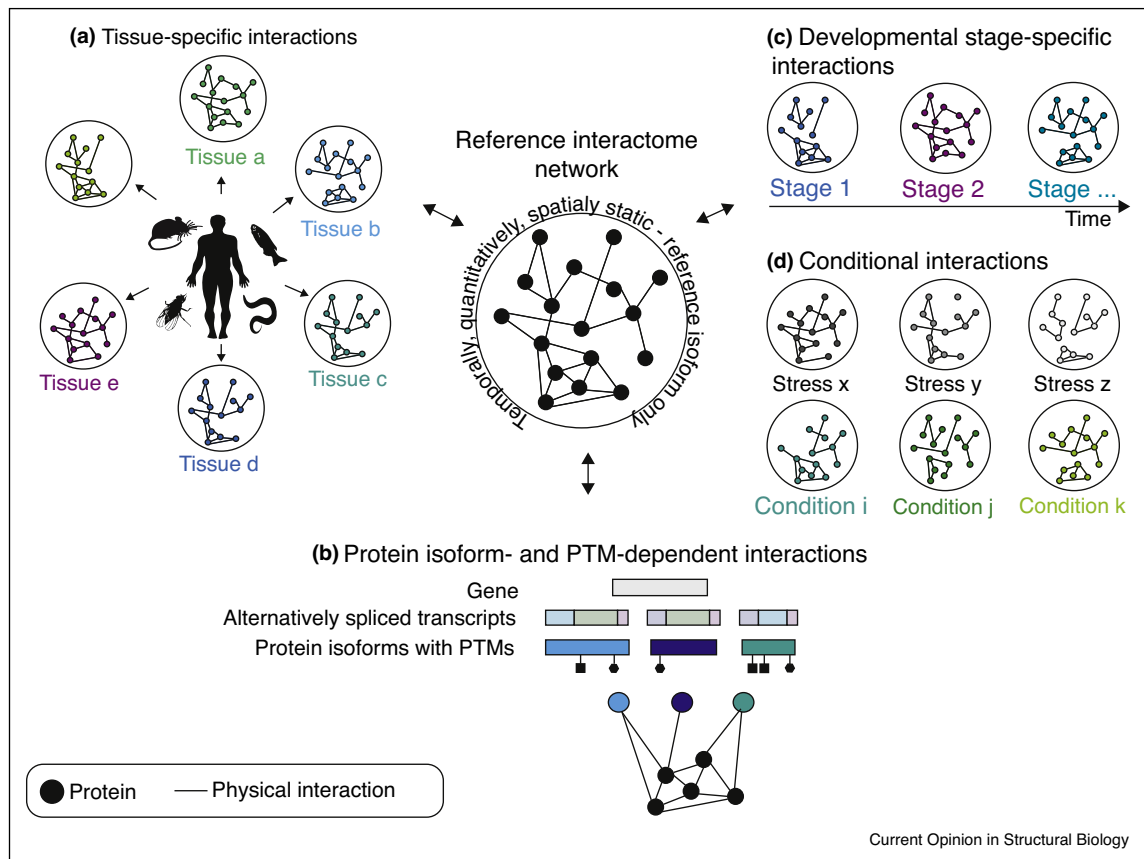
Considerations for future interactome network mapping efforts. **(a)** Tissue-specific interactome network maps. Spatial, quantitative, and temporal information will be integrated. **(b)** Interactome network maps including protein isoforms and post-translational modifications (PTMs). **(c)** Rewiring of interactome networks during developmental stages. **(d)** Interactome network maps generated under specific conditions, such as induction of stress.

(Figure 3b). For example, due to technical challenges and incomplete isoform annotation, human interactome maps utilize the reference isoform; this corresponds to either the longest or most abundantly expressed form of the protein, though this definition varies by database [64]. Only small-scale screening efforts have begun to consider the effects of alternative splicing on protein-binding ability. For example, Corominas *et al.* generated an isoform interactome network map by screening alternatively spliced variants of autism spectrum disorders (ASD) expressed in the brain [65]. Interestingly, half of the interactions in this network involved splicing variants, thereby emphasizing the importance of protein isoforms in interactome networks. An expanded study by Yang *et al.* was the first to examine isoforms across large numbers of genes (∼1500 human genes), including those involved in Mendelian diseases [66[•][•]]. Isoforms for these genes were cloned from five different tissues and screened by Y2H against ∼15 000 human ORF clones to determine protein interaction profiles. While some interactions were shared, others were found to be

isoform-specific. The degree to which interactions are shared may be an approximate measure of shared function, and may allow for the elucidation of redundancy in protein interaction networks [58]. Post-translational modifications (PTMs) also contribute to protein complexity and mediate conditional interactions; recent efforts to identify PPIs that create or rely on PTMs have been reviewed by Woodsmith and Stelzl [67]. More effort will be required to better understand the degree to which protein complexity affects the interactome of various organisms.

Tissue specificity
Current protein interactome network maps provide a vast overview of the biophysical interactions between proteins. However, they often fail to incorporate or fully explore tissue specificity. Indeed, in any given tissue or cell type, only a portion of all possible PPIs will be relevant. Incorporation of gene expression data, such as those collected by the Genotype-Tissue Expression (GTEx) Project, will permit the construction of tissue-

specific subnetworks (Figure 3a). A study of tissue-specific exons by Ellis *et al.* determined that these regions were likely to be flexible and form interaction surfaces [68]. This systematic assessment of neural-regulated exons provides evidence that alternatively spliced exons rewire interactions to generate tissue-specific networks.

Interaction dynamics

Identification of PPIs by biophysical methods reveals only qualitative information. For example, current interactome maps generated by Y2H, AP-MS, or CO-FRAC do not incorporate the dynamics of protein interaction (Figure 3c,d). However, recent interest in quantitative measures of PPI information has grown, spurring the development of new experimental techniques, such as several bimolecular complementation methods, proximity-ligation assays [69] and quantitative AP-MS [70]. Alternative methods such as LUMIER with BACON, BRET, and FRET even allow for the determination of interaction strength [69].

Closing remarks: growth and integration of interactome data

The typical representation of the protein–protein interactome map, that of 'nodes' and 'edges', is a simplified network model. Sophisticated protein molecules become non-differentiable nodes, and the details of protein interaction process degenerate into a simple line. Quantitative and dynamic features, such as protein expression levels and interaction strength, are not yet incorporated. However, even though the network model is simplified, it has great value not only in integrating local PPI information into a global picture, but also in providing an infrastructure where multi-layered information can be linked together to unveil a more realistic description of the protein interactome. Further efforts in interactome mapping, with integration of isoforms and protein variants, as well as quantitative, spatial, and temporal information, will permit a better understanding of a cell's organization and functioning, as well as disease mechanisms.

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

• of special interest
•• of outstanding interest

1. Vidal M, Cusick ME, Barabási A-L: **Interactome networks and human disease**. *Cell* 2011, **144**:986-998.

2. Sahni N, Yi S, Taipale M, Fuxman Bass JI, Coulombe-
•• Huntington J, Yang F, Peng J, Weile J, Karras GI, Wang Y *et al.*: **Widespread macromolecular interaction perturbations in human genetic disorders**. *Cell* 2015, **161**:647-660.
This is the first large-scale systematic study determining an interaction profile for disease-associated alleles and common variants. Disease-associated alleles are more likely to perturb interaction profiles than common variants.

3. Sahni N, Yi S, Zhong Q, Jailkhani N, Charloteaux B, Cusick ME, Vidal M: **Edgotype: a fundamental link between genotype and phenotype**. *Curr. Opin. Genet. Dev.* 2013, **23**:649-657.

4. Zhong Q, Simonis N, Li Q-R, 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.

5. Luck K, Jailkhani N, Cusick ME, Rolland T, Calderwood MA, Charloteaux B, Vidal M: **Interactomes—scaffolds of cellular systems**. *Encycl. Cell Biol.* 2016, **4**:187-198.

6. Fields S, Song O: **A novel genetic system to detect protein–protein interactions**. *Nature* 1989, **340**:245-246.

7. Venkatesan K, Rual J-F, Vazquez A, Stelzl U, Lemmens I, Hirozane-Kishikawa T, Hao T, Zenkner M, Xin X, Goh K-I *et al.*: **An empirical framework for binary interactome mapping**. *Nat. Methods* 2009, **6**:83-90.

8. Rolland T, Tasan M, Charloteaux B, Pevzner SJ, Zhong Q, Sahni N,
•• Yi S, Lemmens I, Fontanillo C, Mosca R *et al.*: **A proteome-scale map of the human interactome network**. *Cell* 2014, **159**:1212-1226.
This study reports the largest systematic map of human binary protein–protein interactions. Comparison with PPIs derived from small-scale studies shows a broader and more homogenous coverage of the interactome network.

9. 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.

10. 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. U. S. A.* 2001, **98**:4569-4574.

11. 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.

12. Vo TV, Das J, Meyer MJ, Cordero NA, Akturk N, Wei X, Fair BJ, Degatano AG, Fragoza R, Liu LG *et al.*: **A proteome-wide fission yeast interactome reveals network evolution principles from yeasts to human**. *Cell* 2016, **164**:310-323.

13. Rajagopala SV, Sikorski P, Kumar A, Mosca R, Vlasblom J, Arnold R, Franca-Koh J, Pakala SB, Phanse S, Ceol A *et al.*: **The binary protein–protein interaction landscape of *Escherichia coli***. *Nat. Biotechnol.* 2014, **32**:285-290.

14. Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, Vidalain P-O, Han J-DJ, Chesneau A, Hao T *et al.*: **A map of the interactome network of the metazoan *C. elegans***. *Science* 2004, **303**:540-543.

15. Simonis N, Rual J, Carvunis A, Tasan M, Lemmens 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.

16. Arabidopsis Interactome Mapping Consortium: **Evidence for network evolution in an Arabidopsis interactome map**. *Science* 2011, **333**:601-607.

17. Braun P, Tasan M, Dreze M, Barrios-Rodiles M, Lemmens I, Yu H, Sahalie JM, Murray RR, Roncari L, de Smet A-S *et al.*: **An experimentally derived confidence score for binary protein–protein interactions**. *Nat. Methods* 2009, **6**:91-97.

18. Huttlin EL, Ting L, Bruckner RJ, Gebreab F, Gygi MP, Szpyt J,
•• Tam S, Zarraga G, Colby G, Baltier K *et al.*: **The BioPlex network: a systematic exploration of the human interactome**. *Cell* 2015, **162**:425-440.
This large-scale systematic study determines protein–protein associations for human proteins using AP-MS. The data permits characterization of proteins of unknown function.

19. Hein MY, Hubner NC, Poser I, Cox J, Nagaraj N, Toyoda Y, Gak IA, Weisswange I, Mansfeld J, Buchholz F *et al.*: **A human interactome in three quantitative dimensions organized by stoichiometries and abundances**. *Cell* 2015, **163**:712-723.

20. Vinayagam A, Stelzl U, Wanker EE: **Repeated two-hybrid screening detects transient protein–protein interactions**. *Theor. Chem. Acc.* 2010, **125**:613-619.

21. Snider J, Kotlyar M, Saraon P, Yao Z, Jurisica I, Stagljar I: **Fundamentals of protein interaction network mapping**. *Mol. Syst. Biol.* 2015, **11** 848.

22. Havugimana PC, Hart GT, Nepusz T, Yang H, Turinsky AL, Li Z,
   • Wang PI, Boutz DR, Fong V, Phanse S *et al.*: **A census of human soluble protein complexes**. *Cell* 2012, **150**:1068-1081.
   The authors present the first large-scale systematic dataset of human protein–protein associations generated by co-fractionation and mass spectrometry. Associations within protein complexes were used to predict unknown protein function and disease association.

23. Wan C, Phanse S, Borgeson B, Tu F, Drew K, Clark G, Xiong X,
   •• Kagan O, Kwan J, Bezginov A *et al.*: **Proteome-wide dataset supporting the study of ancient metazoan macromolecular complexes**. *Nature* 2016, **6**:715-721.
   The authors use co-fractionation and mass spectrometry to explore the evolutionary conservation of protein complexes among various metazoan models.

24. Valencia A, Pazos F: **Computational methods for the prediction of protein interactions**. *Curr. Opin. Struct. Biol.* 2002, **12**:368-373.

25. Shoemaker BA, Panchenko AR: **Deciphering protein–protein interactions. Part II. Computational methods to predict protein and domain interaction partners**. *PLoS Comput. Biol.* 2007, **3**:595-601.

26. Yu J, Fotouhi F: **Computational approaches for predicting protein–protein interactions: a survey**. *J. Med. Syst.* 2006, **30**:39-44.

27. Tuncbag N, Kar G, Keskin O, Gursoy A, Nussinov R: **A survey of available tools and web servers for analysis of protein–protein interactions and interfaces**. *Brief. Bioinform.* 2009, **10**:217-232.

28. Zahiri J, Bozorgmehr JH, Masoudi-Nejad A: **Computational**
   • **prediction of protein–protein interaction networks: algorithms and resources**. *Curr. Genom.* 2013, **14**:397-414.
   The authors provide a comprehensive summary of computational methods for PPI prediction and major PPI databases.

29. Raman K, Eisenberg D, Marcotte E, Xenarios I, Yeates T, Anderson P, Valencia A, Pazos F, Gavin A, Bsche M *et al.*: **Construction and analysis of protein–protein interaction networks**. *Autom. Exp.* 2010, **2**:2.

30. Rhodes DR, Tomlins SA, Tomlins SA, Varambally S, Mahavisno V, Barrette T, Kalyana-sundaram S, Ghosh D, Pandey A, Chinnaiyan AM *et al.*: **Probabilistic model of the human protein–protein interaction network**. *Nat. Biotechnol.* 2005, **23**:951-959.

31. Zhang QC, Petrey D, Deng L, Qiang L, Shi Y, Thu CA, Bisikirska B, Lefebvre C, Accili D, Hunter T *et al.*: **Structure-based prediction of protein–protein interactions on a genome-wide scale**. *Nature* 2012, **490**:556-560.

32. Hopf TA, Schärfe CPI, Rodrigues JPGLM, Green AG,
   •• Kohlbacher O, Sander C, Bonvin AMJJ, Marks DS: **Sequence co-evolution gives 3D contacts and structures of protein complexes**. *Elife* 2014, **3**:e03430.
   The authors demonstrate that evolutionary couplings can be used to predict direct protein interactions at the resolution of amino acid residues.

33. Lei C, Ruan J: **A novel link prediction algorithm for reconstructing protein–protein interaction networks by topological similarity**. *Bioinformatics* 2013, **29**:355-364.

34. Vreven T, Hwang H, Pierce BG, Weng Z: **Prediction of protein–protein binding free energies**. *Protein Sci.* 2012, **21**:396-404.

35. Erijman A, Rosenthal E, Shifman JM: **How structure defines affinity in protein–protein interactions**. *PLoS One* 2014, **9**.

36. Moal IH, Agius R, Bates PA: **Protein–protein binding affinity prediction on a diverse set of structures**. *Bioinformatics* 2011, **27**:3002-3009.

37. Moal IH, Moretti R, Baker D, Fernández-Recio J: **Scoring functions for protein–protein interactions**. *Curr. Opin. Struct. Biol.* 2013, **23**:862-867.

38. Li M, Petukh M, Alexov E, Panchenko AR: **Predicting the impact**
   • **of missense mutations on protein–protein binding affinity**. *J. Chem. Theory Comput.* 2014, **10**:1770-1780.
   By using *in silico* dynamic simulations of protein complex structures, the authors reported a pipeline to predict the PPI affinity changes for single and multiple missense mutations with consideration of protein flexibility.

39. Brender JR, Zhang Y: **Predicting the effect of mutations on protein–protein binding interactions through structure-based interface profiles**. *PLoS Comput. Biol.* 2015, **11**:1-25.

40. Tang H, Thomas PD: **Tools for predicting the functional impact of nonsynonymous genetic variation**. *Genetics* 2016, **203**:635-647.

41. Hassan I: **Editorial recent advances in the structure-based drug design and discovery**. *Curr. Top. Med. Chem.* 2016, **16**:899-900.

42. Mosca R, Céol A, Aloy P: **Interactome3D: adding structural**
   •• **details to protein networks**. *Nat. Methods* 2012, **10**:47-53.
   This resource provides structural information for protein–protein interactions in 18 organisms.

43. Szilagyi A, Zhang Y: **Template-based structure modeling of protein–protein interactions**. *Curr. Opin. Struct. Biol.* 2014, **24**:10-23.

44. Muratcioglu S, Guven-Maiorov E, Keskin O, Gursoy A: **Advances in template-based protein docking by utilizing interfaces towards completing structural interactome**. *Curr. Opin. Struct. Biol.* 2015, **35**:87-92.

45. Vakser IA: **Low-resolution structural modeling of protein interactome**. *Curr. Opin. Struct. Biol.* 2013, **23**:198-205.

46. Stenson PD, Mort M, Ball EV, Shaw K, Phillips AD, Cooper DN: **The human gene mutation database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine**. *Hum. Genet.* 2014, **133**:1-9.

47. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA: **Online Mendelian inheritance in man (OMIM), a knowledge base of human genes and genetic disorders**. *Nucleic Acids Res.* 2005, **33**:514-517.

48. Goh K-I, Cusick ME, Valle D, Childs B, Vidal M, Barabási A-L: **The human disease network**. *Proc. Natl. Acad. Sci. U. S. A.* 2007, **104**:8685-8690.

49. Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, Barabási A-L: **Disease networks. Uncovering disease–disease relationships through the incomplete interactome**. *Science* 2015, **347**:1257601.

50. Lim J, Hao T, Shaw C, Patel AJ, Szabó 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.

51. Kitsak M, Sharma A, Menche J, Guney E, Ghiassian SD,
   • Loscalzo J, Barabási A-L: **Tissue specificity of human disease module**. *Sci. Rep.* 2016, **6**:35241.
   Disease-associated genes tend to cluster in a particular network neighborhoods and show patterns of co-expression in tissues where disease manifests. This study can predict disease-tissue associations.

52. Rozenblatt-Rosen O, Deo RC, Padi M, Adelmant G, Calderwood MA, Rolland T, Grace M, Dricot A, Askenazi M, Tavares M *et al.*: **Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins**. *Nature* 2012, **487**:491-495.

53. Gavin A-C, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dümpelfeld B *et al.*: **Proteome survey reveals modularity of the yeast cell machinery**. *Nature* 2006, **440**:631-636.

54. Collins SR, Kemmeren P, Zhao X-C, Greenblatt JF, Spencer F, Holstege FCP, Weissman JS, Krogan NJ: **Toward a comprehensive atlas of the physical interactome of *Saccharomyces cerevisiae***. *Mol. Cell. Proteom.* 2007, **6**:439-450.

55. 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.

56. Guruharsha KG, Rual J-F, Zhai B, Mintseris J, Vaidya P, Vaidya N, Beekman C, Wong C, Rhee DY, Cenaj O *et al.*: **A protein complex network of *Drosophila melanogaster***. *Cell* 2011, **147**:690-703.

57. Zhong Q, Pevzner SJ, Mosca R, Menche J, Taipale M: **Remnants of co-functionality between two evolutionarily distant proteomes**. *Mol. Syst. Biol.* 2016, **12**:865.

58. Diss G, Gagnon-Arsenault I, Dion-Coté A-M, Vignaud H, Ascencio DI, Berger CM, Landry CR: **Gene duplication can impart fragility, not robustness, in the yeast protein interaction network**. *Science* 2017, **355**:630-634.

59. Carvunis A-R, Rolland T, Wapinski I, Calderwood MA, Yildirim MA,
• Simonis N, Charloteaux B, Hidalgo CA, Barbette J, Santhanam B *et al.*: **Proto-genes and de novo gene birth**. *Nature* 2012, **487**:370-374.
This study computationally identifies candidate proto-genes in the *S. cerevisiae* genome. The authors provide evidence that de novo gene birth may be an important mechanism for generating new protein coding genes.

60. Tautz D, Domazet-Loo T: **The evolutionary origin of orphan genes**. *Nat. Rev. Genet.* 2011, **12**:692-702.

61. McLysaght A, Guerzoni D: **New genes from non-coding**
• **sequence: the role of de novo protein-coding genes in eukaryotic evolutionary innovation**. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 2015, **370**.
This review summarizes various aspects of de novo gene birth, including generation, fixation and functional contribution.

62. Siepel A, Knowles DG, Mclysaght A: **Darwinian alchemy: human genes from noncoding DNA Darwinian alchemy: human genes from noncoding DNA**. *Genome Res.* 2009, **19**:1693-1695.

63. Knowles DG, Mclysaght A: **Recent de novo origin of human protein-coding genes**. *Genome Res.* 2009, **19**:1-9.

64. Rodriguez JM, Maietta P, Ezkurdia I, Pietrelli A, Wesselink JJ, Lopez G, Valencia A, Tress ML: **APPRIS: annotation of principal and alternative splice isoforms**. *Nucleic Acids Res.* 2013, **41**:110-117.

65. Corominas R, Yang X, Lin GN, Kang S, Shen Y, Ghamsari L, Broly M, Rodriguez M, Tam S, Trigg SA *et al.*: **Protein interaction network of alternatively spliced isoforms from brain links genetic risk factors for autism**. *Nat. Commun.* 2014, **5**:3650.

66. Yang X, Coulombe-Huntington J, Kang S, Sheynkman GM, Hao T,
•• Richardson A, Sun S, Yang F, Shen YA, Murray RR *et al.*: **Widespread expansion of protein interaction capabilities by alternative splicing**. *Cell* 2016, **164**:805-817.
This paper presents the first large-scale attempt to identify differences in the interaction profiles of human protein isoforms.

67. Woodsmith J, Stelzl U: **Studying post-translational modifications with protein interaction networks**. *Curr. Opin. Struct. Biol.* 2014, **24**:34-44.

68. Ellis JD, Barrios-Rodiles M, Çolak R, Irimia M, Kim T, Calarco JA, Wang X, Pan Q, O'Hanlon D, Kim PM *et al.*: **Tissue-specific alternative splicing remodels protein–protein interaction networks**. *Mol. Cell* 2012, **46**:884-892.

69. Buntru A, Trepte P, Klockmeier K, Schnoegl S, Wanker EE: **Current approaches toward quantitative mapping of the interactome**. *Front. Genet.* 2016, **7**:1-9.

70. Meyer K, Selbach M: **Quantitative affinity purification mass spectrometry: a versatile technology to study protein–protein interactions**. *Front. Genet.* 2015, **6**:1-7.