Cell Research (2008) 18:716-724.

www.nature.com/cr

REVIEW

Mapping the human protein interactome

Daniel Figeys[1]

[1]*The Ottawa Institute of Systems Biology, The Department of Biochemistry, Microbiology and Immunology, University of Ottawa, Ottawa, ON, K1H 8M5, Canada*

Interactions are the essence of all biomolecules because they cannot fulfill their roles without interacting with other molecules. Hence, mapping the interactions of biomolecules can be useful for understanding their roles and functions. Furthermore, the development of molecular based systems biology requires an understanding of the biomolecular interactions. In recent years, the mapping of protein-protein interactions in different species has been reported, but few reports have focused on the large-scale mapping of protein-protein interactions in human. Here, we review the developments in protein interaction mapping and we discuss issues and strategies for the mapping of the human protein interactome.

***Keywords*:** interactome, protein interaction, yeast two hybrid, immunopurification, mass spectrometry, LUMIER, co-localization

Cell Research (2008) **18**:716-724. doi: 10.1038/cr.2008.72; published online 24 June 2008

The latest craze in science is *systems biology*. Scientists in different scientific fields, however, define and view "systems biology" in different ways. Systems biology, for example, can mean mathematical modeling of a system for accurate predictions of the behavior of the system. It can also mean the systematic application of different high-throughput approaches to gather information on the system and to combine different information spaces for a more comprehensible view of the biology at hand. To date, the best modeling and predictions have been on systems for which good biological information is available [1]. Hence, gathering information on the system is required to develop predictive models. Genomics has been the primary source of large-scale information on systems through genome sequencing, large-scale SNPs analyses, and gene expression studies.

The success of genomics has snow-balled into other areas, such as the large-scale data gathering at the protein level (proteomics). One of the achievements of proteomics is the mapping of protein interactions (interactome; the ensemble of the protein interactions related to a proteome or a genome). We have seen a growing number of large-scale studies of the interactome (or part of) for different species including yeast [2-7], Drosophilia [8], *Helicobacter pylori* [9], *Caenorhabditis elegans* [10, 11], cyanobacterium *Synechocystis* [12], and human [13-17]. Furthermore, the mapping of individual protein interactions and studies focused on subgroups of proteins, for example, bacteria Flagella [18], the Smad signaling system [19], and *C. elegans* 26S proteasome [20], are ongoing in many laboratories around the world where mass spectrometry [21] and techniques, such as yeast two hybrid (Y2H) [22], are readily available. Already in yeast we have a semi complete map of the interactome. Unfortunately, in human, we only have a partial map of the interactome, which is complicated by the different cell types and cellular localizations, thus limiting the broader application of systems modeling.

Correspondence: Daniel Figeys
E-mail: dfigeys@uottawa.ca

Should the human protein interactome be mapped?

The first question can be easily answered and is reminiscent of early discussions about sequencing the human genome. Many scientists were doubtful that sequencing the human genome would be worth the effort. It is now clear that today's research in human diseases and biological processes benefits from the human genome project. We already have a glimpse of the potential outcomes of mapping the human interactome by looking at the yeast interactome. The importance of mapping the yeast interactome can be measured in different ways; however, the number of references to the original papers by Fields [22] (over 3 000 citations) for interaction mapping by Y2H; and Ho [4] (over 1 300 citations) and Gavin [6] (over 1 400 citations) for protein interaction mapping by mass spectrometry clearly

illustrate the changes in understanding yeast biology provided by the yeast interactome. We could reasonably expect a similar or greater impact in human biology by the human interactome.

Are techinques with sufficient throughput available?

From a technical point of view, the broadly accepted techniques of Y2H [22] and affinity purification combined to mass spectrometry [21] can potentially be scaled up. Assuming that 24 000 genes are present in human, testing all the possible interactions using Y2H would require 3×10^8 tests, not including the repeats and all the splice variants and mutants. This number might well be within reach by the Y2H technique as recent studies have performed over 10^7 Y2H assays [14, 23]. For affinity purification/mass spectrometry, it would require 24 000 experiments, not including the repeats and all the splice variants and mutants. Recently, we published the interaction map for 400 human genes obtained in 293T cells by affinity purification/mass spectrometry [16]. Hence, it is conceivable that the human protein interactome could be mapped by Y2H and affinity purification/mass spectrometry.

Will the data quality and coverage be sufficient?

Neither Y2H nor affinity purification/mass spectrometry provide the complete list of interactions [24]. Futschik *et al*. [25] compared eight large-scale maps with a total of over 10 000 unique proteins and 57 000 interactions based either on literature search, orthology, or by Y2H assays. Their comparison reveals a small, but statistically significant overlap. More importantly, their analysis gives clear indications that all interaction maps imply considerable selection and detection biases. Our studies of the overlap between Y2H and affinity purification/mass spectrometry in human indicate at best a 11% overlap [16]. A closer assessment of the results indicates that the techniques are complementary because they each provide a partial view of known protein complexes. Furthermore, our current methods to represent and compare Y2H and affinity purification/mass spectrometry do not take into account the differences in the data, i.e. binary interaction data (Y2H) and affinity purification/mass spectrometry data (direct and indirect interactions). We have shown a greater overlap (2.6-fold increase) when taking these differences in consideration [16]. Finally, a significant contribution to the low overlap between the two techniques is false positives. False positives can be greatly reduced by taking into account the known localization of proteins and their functions. Reguly *et al*. [26] demonstrated that of the interaction data obtained in yeast by high-throughput techniques 20% are between proteins that are involved in the same biological process. Furthermore, 27% of the protein interactions are between proteins that have the same cellular localization.

Choosing only one approach to map the human interactome, although appealing from a cost point of view, would likely provide a poor interaction map. We are proposing as a discussion point a process for studying the human interactome that will provide an increasing level of refinement through a multiple pass process with each pass leading to a more confident interaction map (Figure 1). Hence, it is likely

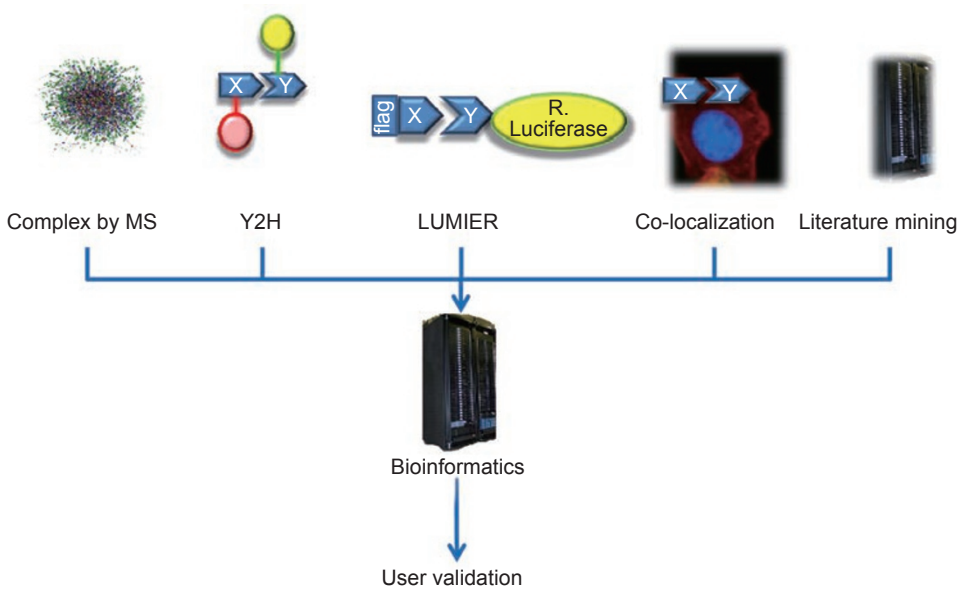

Figure 1 Combination of different protein interaction techniques to increase the level of confidence in the human interactome while minimizing the number of experiments to perform.

that the human interactome will need to be mapped using the combination of different technical approaches.

Are high-throughput techniques standardized?

The mapping of the human protein interactome will require a coordinated international effort. It is not possible to only use one scientific approach; therefore, the coordination of various approaches, such as Y2H and affinity purification/mass spectrometry, needs to be considered. Standardization of methods will be very important for studying the human interactome.

Reagents

Y2H and affinity purification/mass spectrometry both require access to clones. Different collections of clones are available. The latest version of the human ORFeome contains over 12 000 full length clones of human genes [27]. Invitrogen has over 35 000 full length clones in their collection while the FLJ-DB database (www.nedo.go.jp/bioiryo/bio-e/index.html) reports over 20 000 clones (although not necessarily full length). Although the availability of full length clones has improved drastically, the annotation and quality of clones remains an important issue. It is likely that clones representing most human genes will be readily available over the next few years.

High-throughput Y2H

The Y2H system is currently one of the most standardized protein interaction mapping techniques. Over the past years, several large-scale protein-protein interaction datasets have been published [2, 3, 9, 11, 28, 29].

In a Y2H assay, the two proteins to be tested for interactions are expressed with amino-terminal fusion moieties in the yeast Saccharomyces cerevisiae. One protein is fused to a DNA-binding domain (BD) and the other one is fused to a transcription activation domain (AD). An interaction between the two proteins results in the activation of reporter genes that have upstream binding sites for the BD. Large arrays of AD and BD strains representing most of the proteins encoded by a genome have been constructed and used to systematically detect binary interactions. Most large-scale screens have used arrays in a library-screening approach in which the BD strains are individually mated with a library containing all of the AD strains pooled together. Detailed protocols for large-scale two-hybrid analyses have been described in references [30] and are also available on a website (http://vidal.dfci.harvard.edu).

Li *et al.* [31] recently reported a study on how protein-protein interaction (or "interactome") networks relate to multicellular functions. They have mapped a large fraction of the *C. elegans* interactome network. Starting with a subset of metazoan-specific proteins, more than 4 000 interactions were identified from HT-Y2H screens.

Y2H screens are often criticized for generating high rates of false-positives. Vidalain *et al*. [32] described false-positives as either being biological false-positives (artificially occurring in yeast cell) or technical false-positives (due to limitation in the techniques). They developed various approaches to reduced false-positives due to technical limitations of HT-Y2H. von Mering *et al*. [24] also demonstrated that the biological false-positives can be reduced by combining results obtained from different studies including results obtained from mass spectrometry based approach [24]. As well, Jin *et al*. [33] recently described a small pool array that increases the screening efficiency by one order of magnitude and reduces the false positives. However, it still remains that the study of human genes using the Y2H approach cannot reproduce the proper cell types, the localizations of the proteins, the basal levels of expression, and it might not provide the necessary post-translational modifications and processing.

Affinity purification/mass spectrometry

Affinity purification has been a basic methodology of modern biology. Its recent coupling with mass spectrometry has increased the speed and ease of identifying protein interactors [34, 35]. The combination of affinity purification of protein complexes with mass spectrometry allows the rapid identification of different proteins involved in a complex. This approach is promising and is likely to be very useful for generating part of the human interactome map. The standardization of methodologies for protein interaction mapping by affinity purification/mass spectrometry is lacking and many protocols still need to be developed. "One-size-fit-all" protocols will unlikely emerge. It is more likely, however, that a suite of standardized protocols that are applicable to different protein classes, protein localizations, and potentially cell types will be developed.

At this point, we will discuss some of the technical hurdles that affinity purification faces. The major drawback of this approach is that it cannot be done without disturbing the system. This disturbance of protein interactions can occur at different levels such as: at the level of tagging of the bait protein; at the expression level;when lysis of the cells occurs; and also when the protein complexes are purified using different protocols (Figure 2).

To tag or not to tag

The best situation occurs when monoclonal antibodies against the wild-type proteins are available to perform immunopurification [36]. Monoclonal antibodies allow the study of the protein interactions with very limited disturbance of the cells other than the lysis. Unfortunately, good

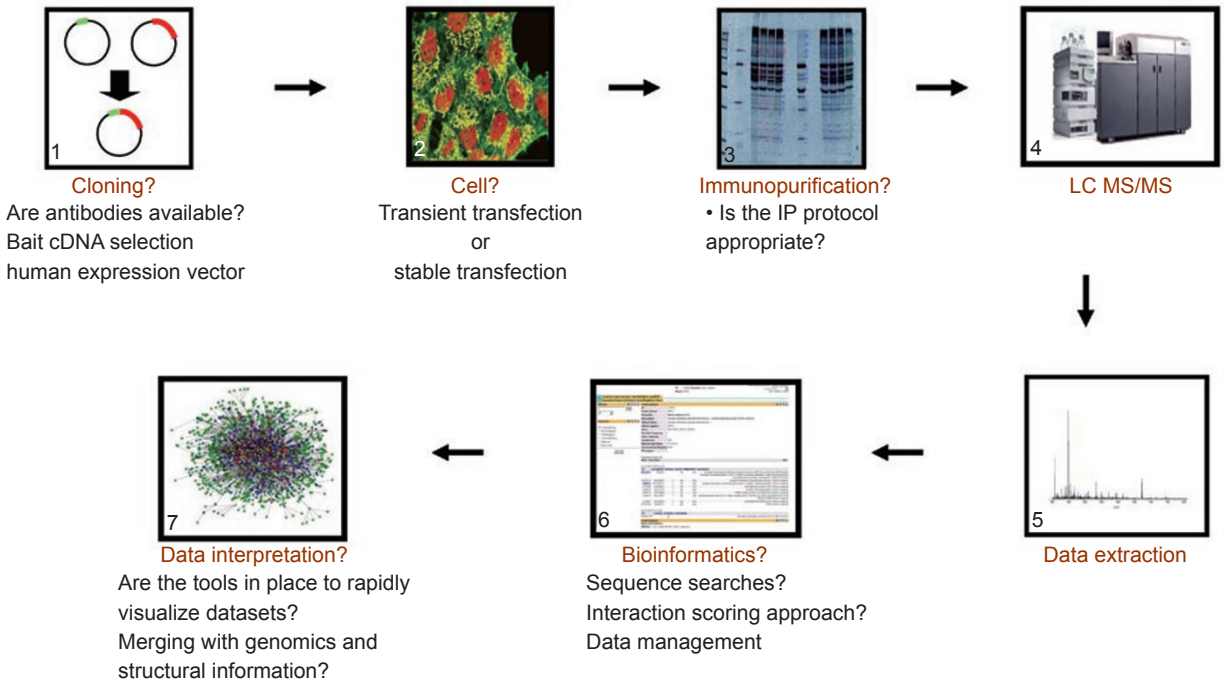

Figure 2 Steps involved in protein interaction mapping by affinity purification coupled to mass spectrometry. Possible issues are highlighted.

quality antibodies that are sufficient for immunopurification of wild-type proteins are often not available and alternative strategies need to be employed. However, this insufficiency of antibiodies could change with projects such as the Human Protein Atlas [37] that generate a suite of high quality monospecific polyclonal antibodies.

The next strategy requires the tagging of specific bait proteins using molecular biology approaches. To date, this has been the primary approach for high-throughput mapping of protein-protein interactions by affinity purification coupled to mass spectrometry [34]. The systematic addition of the tag to proteins provides a universal handle that can be used to affinity purify different complexes. Most current methods for purification and identification of protein complexes use an endogenous expression of an affinity-tagged bait [38]. Once the complex has been allowed to form *in vivo*, the cells are lysed, and the complex is purified using the tag present on the bait protein after which the complexes are denatured and separated by 1D or 2D electrophoresis gel. The protein lanes/spots are excised, digested with trypsin, and analyzed by mass spectrometry (MS). Although the schema is generally the same, different tagging approaches have been reported for the purification of protein complexes.

The tandem affinity purification (TAP) method developed by Seraphin is based on two successive affinity chromatography steps [39, 40]. Originally it was applied in yeast by in vivo recombination of the TAP-tag. The tag fused to a target protein is composed of protein A having a very high affinity for IgG, a TEV protease cleavage site, and a calmodulin binding peptide which has a high affinity for calmodulin. The TAP-tagged target protein can be purified using protein A affinity resin followed by incubation with TEV protease which releases the target protein. The protein complex is further purified through a second affinity step based on calmodulin binding in the presence of calcium. The complex is then released using EGTA which depletes the calcium ions that are essential for the bait-calmodulin binding. These two different affinity purification steps enhance the specificity of the purification procedure. The initial high-throughput example of this approach was the mapping of protein complexes in yeast by combining TAP purification and MS as reported by Gavin [6]. Additional tagging approaches are also used to purify protein complexes by affinity chromatography. Among those, the FLAG tag has been extensively used [41]. This small acidic peptide tag (often coded as a triple FLAG) is selectively recognized by a monoclonal antibody. This tag was used by Ho *et al*. [4] in combination with recombinant-based cloning to tag 725 yeast genes, which were then transfected in yeast. Other tags such as Glutathione S-transferase (GST), His6, biotinylation substrates, and others were used to purify recombinant proteins over-expressed in *Escherichia coli* [42]. However, these tags were

not extensively used, probably due to their limited affinity and/or high background levels [39].

Attempts have been done to compare the results from different tagging approaches. For example, Ito *et al*. [43] compared two large-scale biochemical protein interaction studies performed by the TAP purification method [6], and by the single-step affinity purification with FLAG tag [4] mentioned above. Unfortunately, the tag was not the only difference found in the experiments. The TAP-tagged proteins, for example, were expressed at their natural levels from the endogenous promoters, and the FLAG tag analysis used proteins over-expressed from plasmids which generated very high levels of expression. In both cases, purification was followed by SDS-PAGE and mass spectrometry. The two studies shared 115 targets and showed only about 10% overlap in the proteins recovered.

Tissue, animal models and cell lines

In human, protein interaction mapping needs to take into account i) the diversity of cells, ii) that proteins are not expressed in all cells, and iii) that protein interactions can change between different cell types. Mapping protein interactions in all of cell types in which proteins are expressed is currently too expensive. Although doing immunopurification from cells isolated from human tissue would provide the most relevant results, it remains an expensive proposition. Instead, approaches that rely on expressions in animal models have been proposed. This could be useful for small sets of proteins, but again, it is likely to be too expensive for all proteins. The more realistic options for the first pass mapping of the human interactome are to create stable cell lines or to perform high-throughput transient transfections. Many reports have already used high-throughput transient transfections [44, 45]. The major drawback with transient transfection, however, is that the level of protein expression ends up being one or two orders of magnitude above the endogenous level of expression (based on our experience). This can impact the localization of the protein(s) and cause interactions that are not physiologically relevant. As well, the position and the type of tag employed can drastically affect the localization. For example, GFP as a tag has been shown to lead to miss-localization depending on its N or C termini position [46]. Unfortunately, the localization, the basal level of expression, and the effects of tagging are unknown for the vast majority of proteins in human.

On the other hand, although stable cell lines are more expensive to create and reportedly slow to generate, the cells that express different levels of the bait protein can be selected. We have not found any reports of the large scale generation of stable cell lines other than papers reporting the promise of lentivirus in rapidly generating stable cell lines [47] and cell microarrays [48]. New systems for regulating expression levels based on the modulation of RNA self-cleavage [49] and modulation of translational termination [50] have been proposed to better control the expression levels during transient and stable transfections.

Immunopurification protocols

The majority of the protocols for immunopurification of proteins were developed for soluble proteins. Unfortunately, these proteins only represent a fraction of the human protein interactome. Membrane proteins often do not work with protocols for soluble proteins, and therefore, remain problematic. Protocols based on cross-linking and different stringencies have been developed, but they need to be assessed on a case by case basis. As well, soluble proteins attached to polymeric macromolecules (DNA, microtubules, and microfilaments) have more complex interactions than what is observed by only looking at their interactions while they are free floating. For example, a few novel approaches have been proposed for the study of proteins associated with DNA [51-53]. It is likely that protein complexes anchored on major polymeric macromolecules are poorly represented by current protocols. Hence, the development of novel affinity purification protocols should be a priority in any global effort to map the human interactome.

How will the protein interaction be accessed?

Public repositories of protein interactions are available such as: The MIPS Mammalian Protein-Protein Interaction Database [54] (mips.gsf.de/proj/ppi/), the Human Protein Reference Database [55, 56] (www.hprd.org), the Biomolecular Interaction Network Database (BIND) [57], The Database of Interacting Proteins (DIP) [58], IntAct [59, 60], Molecular INTeraction database (MINT) [61], and BioGRID [62, 63]. Furthermore, HUPO, through the HUPO Proteomics Standard initiative, has recently proposed standards required for the submission of interaction datasets [64]. These repositories provide access to the interaction datasets and some levels of annotation.

The protein interaction data submission standard and the minimal requirements for mass spectrometric information from peptide identification that have been defined by the HUPO Proteomics Standard Initiative [65-67] will also facilitate retrospective studies of protein interaction datasets and the development of new tools. For example, greater attention needs to be focused on the development of confidence scoring approaches for protein-protein interaction. We have developed such a scoring algorithm for our recently released large set of human protein interactions [16]. As well, Krogan developed, through machine learning algorithm, a scoring scheme for yeast protein interactions [7]. The development of new methods to score protein-

protein interactions requires that sufficient information be available in public databases to judge the quality of the information and properly compare different results.

It is clear that the bioinformatic analysis of protein interactions is still evolving. New tools for rapidly analyzing protein-protein interactions need to be developed, particularly for most users who are not interested in seeing the whole interactome, but instead are interested in performing an interaction walk starting from one protein of interest. Software, such as Cytoscape [68], Osprey [69, 70], Biomolecular Interaction Network Database (BIND) [57, 71] and others are good graphical interfaces for interaction datasets. The next step needed is the development of tools that incorporate graphical representations of the interactions with the wealth of annotation available from the protein interaction databases, NCBI, and others. Many plug-ins to Cytoscape have been reported to solve these issues [72-78]. We can also foresee that the "balls and edges" approach for displaying interactions is insufficient because it does not represent the dynamic and the structural aspect of the protein interactions. Already, experiments that study the dynamics of protein interactions are underway (http://dynactome.mshri.on.ca/). As well, high-throughput projects are underway for the elucidation of protein structures; but, the incorporation of structural information in models of protein interactions remains a computational challenge even though Morris *et al*. have reported a plug-in to Cytoscape to link interaction to structure [72].

Novel approaches

LUMIER

Barrios-Rodiles *et al*. [13] developed the high-throughput LUMIER (for luminescence-based mammalian interactome mapping) to systematically map protein-protein interactions in mammalian cells. This strategy uses Renilla luciferase enzyme (RL) fused to proteins of interest which are then coexpressed with individual FLAG-tagged partners in mammalian cells. This group determined protein interactions by performing an RL enzymatic assay on immunoprecipitates using an antibody against FLAG. This approach has not yet been as widely used as the Y2H approach.

co-localization

Protein interactions require co-localization; therefore, the study of co-localizations can be used to reinforce the first pass interactions map. Computational approaches have been proposed for the prediction of protein localizations, and there are other tools that can be used to reinforce the confidence in the results [79-82]. However, laboratory validation of co-localizations is also important because many proteins can have multiple localizations that are hardly predictable through current models.

Recently, a high-throughput approach to determine the localization of GFP labeled proteins has been described and applied in yeast [83]. In this approach, yeast strains were created for 4 156 proteins and microscopic imaging was used to determine the localization of the GFP labeled proteins. 1 839 of these proteins indicated localizations other than in the nucleus or the cytoplasm. Matsuyama *et al*. [84] studied the localization of 4 431 proteins in the yeast *Schizosaccharomyces pombe* by cloning its ORFeome and by tagging each ORF with the yellow fluorescent protein. An automated image analysis software that has over 80% accuracy in evaluating the localization of tagged proteins in yeast has been developed [85]. Although the large-scale mapping of protein localizations in human cells using these types of approaches have not been reported, it is a plausible strategy to enhance the confidence of protein-protein interactions.

Conclusions

The mapping of the human protein interactome is one of the key challenges in the post genome era. Mapping the human protein interactome will require a coordinated international effort. Current technologies such as Y2H and affinity purification/mass spectrometry can be scaled up to map the human protein interactome. It is likely that multiple standard methods will be developed for the affinity purification/mass spectrometry studies of proteins to take into account different protein functions and localizations. The development of novel technologies, methodologies, algorithms, and software should also be part of such an international effort. Here again, one can look back at the history of sequencing the human genome which succeeded rapidly because of a systematic effort to continuously improve technologies. Even today, these improvements have led to the recent reports of individual genome sequencing [86]. Hence, a combination of current technologies and efforts to improve and develop new technologies should also enable us to reach the goal of mapping the human interactome. The mapping of the human interactome would create an invaluable source of information to better understand human biology.

Acknowledgments

Daniel Figeys would like to acknowledge a Canada Research Chair in Proteomics and Systems Biology.

References

1 Kaern M, Blake WJ, Collins JJ. The engineering of gene regulatory networks. *Annu Rev Biomed Eng* 2003; **5**:179-206.

2 Uetz P, Giot L, Cagney G, *et al*. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 2000; **403**:623-627.
3 Schwikowski B, Uetz P, Fields S. A network of protein-protein interactions in yeast. *Nature Biotechnology* 2000; **18**:1257-1261.
4 Ho Y, Gruhler A, Heilbut A, *et al*. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 2002; **415**:180-183.
5 Gavin AC, Aloy P, Grandi P, *et al*. Proteome survey reveals modularity of the yeast cell machinery. *Nature* 2006; **440**:631-636.
6 Gavin AC, Bosche M, Krause R, *et al*. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 2002; **415**:141-147.
7 Krogan NJ, Cagney G, Yu H, *et al*. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* 2006; **440**:637-643.
8 Stuart LM, Boulais J, Charriere GM, *et al*. A systems biology analysis of the *Drosophila* phagosome. *Nature* 2007; **445**:95-101.
9 Rain JC, Selig L, De Reuse H, *et al*. The protein-protein interaction map of *Helicobacter pylori*. *Nature* 2001; **409**:211-215.
10 Walhout AJ, Reboul J, Shtanko O, *et al*. Integrating interactome, phenome, and transcriptome mapping data for the *C. elegans* germline. *Curr Biol* 2002; **12**:1952-1958.
11 Walhout AJ, Sordella R, Lu X, *et al*. Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* 2000; **287**:116-122.
12 Sato S, Shimoda Y, Muraki A, *et al*. A Large-scale protein protein interaction analysis in *Synechocystis* sp. PCC6803. *DNA Res* 2007; **14**:207-216.
13 Barrios-Rodiles M, Brown KR, Ozdamar B, *et al*. High-throughput mapping of a dynamic signaling network in mammalian cells. *Science* 2005; **307**:1621-1625.
14 Rual JF, Venkatesan K, Hao T, *et al*. Towards a proteome-scale map of the human protein-protein interaction network. *Nature* 2005; **437**:1173-1178.
15 Bouwmeester T, Bauch A, Ruffner H, *et al*. A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway. *Nat Cell Biol* 2004 **6**:97-105.
16 Ewing RM, Chu P, Elisma F, *et al*. Large-scale mapping of human protein-protein interactions by mass spectrometry. *Mol Syst Biol* 2007; **3**:89.
17 Lim J, Hao T, Shaw C, *et al*. A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. *Cell* 2006; **125**:801-814.
18 Rajagopala SV, Titz B, Goll J, *et al*. The protein network of bacterial motility. *Mol Syst Biol* 2007; **3**:128.
19 Colland F, Jacq X, Trouplin V, *et al*. Functional proteomics mapping of a human signaling pathway. *Genome Res* 2004; **14**:1324-1332.
20 Davy A, Bello P, Thierry-Mieg N, *et al*. A protein-protein interaction map of the Caenorhabditis elegans 26S proteasome. *EMBO Rep* 2001; **2**:821-828.
21 Vasilescu J, Figeys D. Mapping protein-protein interactions by mass spectrometry. *Curr Opin Biotechnol* 2006; **17**:394-399.
22 Fields S, Song O. A Novel Genetic System to Detect Protein-Protein Interactions. *Nature* 1989; **340**:245-246.
23 Stelzl U, Worm U, Lalowski M, *et al*. A human protein-protein interaction network: a resource for annotating the proteome. *Cell* 2005; **122**:957-968.
24 von Mering C, Krause R, Snel B, *et al*. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* 2002; **417**:399-403.
25 Futschik ME, Chaurasia G, Herzel H. Comparison of human protein-protein interaction maps. *Bioinformatics* 2007; **23**:605-611.
26 Reguly T, Breitkreutz A, Boucher L, *et al*. Comprehensive curation and analysis of global interaction networks in Saccharomyces cerevisiae. *J Biol* 2006; **5**:11.,
27 Lamesch P, Li N, Milstein S, *et al*. hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes. *Genomics* 2007; **89**:307-315.
28 Ito T, Chiba T, Yoshida M. Exploring the protein interactome using comprehensive two-hybrid projects. *Trends Biotechnol* 2001; **19**:S23-S27.
29 Ito T, Tashiro K, Muta S, *et al.* 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.
30 Walhout AJ, Vidal M. High-throughput yeast two-hybrid assays for large-scale protein interaction mapping. *Methods* 2001; **24**:297-306.
31 Li S, Armstrong CM, Bertin N, *et al*. A map of the interactome network of the metazoan C. elegans. *Science* 2004; **303**:540-543.
32 Vidalain PO, Boxem M, Ge H, Li S, Vidal M. Increasing specificity in high-throughput yeast two-hybrid experiments. *Methods* 2004; **32**:363-370.
33 Jin F, Avramova L, Huang J, Hazbun T. A yeast two-hybrid smart-pool-array system for protein-interaction mapping. *Nat Methods* 2007; **4**:405-407.
34 Figeys D, McBroom LD, Moran MF. Mass spectrometry for the study of protein-protein interactions. *Methods Mol Biol* 2001; **24**:230-239.
35 Shevchenko A, Zachariae W. A strategy for the characterization of protein interaction networks by mass spectrometry. *Biochem Soc Trans* 1999; **27**:549-554.
36 Lipman NS, Jackson LR, Trudel LJ, Weis-Garcia F. Monoclonal versus polyclonal antibodies: distinguishing characteristics, applications, and information resources. *ILAR J* 2005; **46**:258-268.
37 Uhlen M, Bjorling E, Agaton C, *et al*. A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol Cell Proteomics* 2005; **4**:1920-1932.
38 Markillie LM, Lin CT, Adkins JN, *et al*. Simple protein complex purification and identification method for high-throughput mapping of protein interaction networks. *J Proteome Res* 2005; **4**:268-274.
39 Rigaut G, Shevchenko A, Rutz B, *et al*. A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* 1999; **17**:1030-1032.
40 Puig O, Caspary F, Rigaut G, *et al*. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* 2001; **24**:218-229.
41 Brizzard BL, Chubet RG, Vizard DL. Immunoaffinity purification

of FLAG epitope-tagged bacterial alkaline phosphatase using a novel monoclonal antibody and peptide elution. *Biotechniques* 1994; **16**:730-735.
42 Terpe K. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* 2003; **60**:523-533.
43 Ito T, Ota K, Kubota H, *et al*. Roles for the two-hybrid system in exploration of the yeast protein interactome. *Mol Cell Proteomics* 2002; **1**:561-566.
44 Chen J, Lake MR, Sabet RS, *et al*. Utility of large-scale transiently transfected cells for cell-based high-throughput screens to identify transient receptor potential channel A1 (TRPA1) antagonists. *J Biomol Screen* 2007; **12**:61-69.
45 Davies A, Greene A, Lullau E, Abbott WM. Optimisation and evaluation of a high-throughput mammalian protein expression system. *Protein Expr Purif* 2005; **42**:111-121.
46 Beitz E, Liu K, Ikeda M, *et al*. Determinants of AQP6 trafficking to intracellular sites versus the plasma membrane in transfected mammalian cells. *Biol Cell* 2006; **98**:101-109.
47 Bianchi BR, Moreland RB, Faltynek CR, Chen J. Application of large-scale transiently transfected cells to functional assays of ion channels: different targets and assay formats. *Assay Drug Dev Technol* 2007; **5**:417-424.
48 Bailey SN, Ali SM, Carpenter AE, Higgins CO, Sabatini DM. Microarrays of lentiviruses for gene function screens in immortalized and primary cells. *Nat Methods* 2006; **3**:117-122.
49 Yen L, Svendsen J, Lee JS, *et al*. Exogenous control of mammalian gene expression through modulation of RNA self-cleavage. *Nature* 2004; **431**:471-476.
50 Murphy GJ, Mostoslavsky G, Kotton DN, Mulligan RC. Exogenous control of mammalian gene expression via modulation of translational termination. *Nat Med* 2006; **12**:1093-1099.
51 Du YC, Gu S, Zhou J, *et al*. The dynamic alterations of H2AX complex during DNA repair detected by a proteomic approach reveal the critical roles of Ca(2+)/calmodulin in the ionizing radiation-induced cell cycle arrest. *Mol Cell Proteomics* 2006; **5**:1033-1044.
52 Foltz DR, Jansen LE, Black BE, *et al*. The human CENP-A centromeric nucleosome-associated complex. *Nat Cell Biol* 2006; **8**:458-469.
53 Mizuguchi G, Shen X, Landry J, *et al*. ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science* 2004; **303**:343-348.
54 Pagel P, Kovac S, Oesterheld M, *et al*. The MIPS mammalian protein-protein interaction database. *Bioinformatics* 2005; **21**:832-834.
55 Peri S, Navarro JD, Amanchy R, *et al*. Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res* 2003; **13**:2363-2371.
56 Mathivanan S, Ahmed M, Ahn NG, *et al*. Human Proteinpedia enables sharing of human protein data. *Nat Biotechnol* 2008; **26**:164-167.
57 Bader GD, Betel D, Hogue CW. BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Res* 2003 **31**:248-250.
58 Salwinski L, Miller CS, Smith AJ, *et al*. The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res* 2004; **32**: D449-D451.
59 Kerrien S, Alam-Faruque Y, Aranda B, *et al*. IntAct--open source resource for molecular interaction data. *Nucleic Acids Res* 2007; **35**:D561-D565.
60 Hermjakob H, Montecchi-Palazzi L, Lewington C, *et al*. IntAct: an open source molecular interaction database. *Nucleic Acids Res* 2004; **32**:D452-D455.
61 Zanzoni A, Montecchi-Palazzi L, Quondam M, *et al*. MINT: a Molecular INTeraction database. *FEBS Lett* 2002; **513**:135-140.
62 Breitkreutz BJ, Stark C, Reguly T, *et al*. The BioGRID Interaction Database: 2008 update. *Nucleic Acids Res* 2008; **36**:D637-D640.
63 Stark C, Breitkreutz BJ, Reguly T, *et al*. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* 2006; **34**: D535-D539.
64 Orchard S, Salwinski L, Kerrien S, *et al*. The minimum information required for reporting a molecular interaction experiment (MIMIx). *Nat Biotechnol* 2007; **25**:894-898.
65 Hermjakob H, Montecchi-Palazzi L, Bader G, *et al*. The HUPO PSI's molecular interaction format--a community standard for the representation of protein interaction data. *Nat Biotechnol* 2004; **22**:177-183.
66 Martens L, Orchard S, Apweiler R, Hermjakob H. Human Proteome Organization Proteomics Standards Initiative: data standardization, a view on developments and policy. *Mol Cell Proteomics* 2007; **6**:1666-1667.
67 Orchard S, Hermjakob H, Apweiler R. The proteomics standards initiative. *Proteomics* 2003; **3**:1374-1376.
68 Shannon P, Markiel A, Ozier O, *et al*. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003; **13**:2498-2504.
69 Breitkreutz BJ, Stark C, Tyers M. The GRID: the General Repository for Interaction Datasets. *Genome Biol* 2003; **4**:R23.
70 Breitkreutz BJ, Stark C, Tyers M. Osprey: a network visualization system. *Genome Biol* 2003; **4**:R22.
71 Bader GD, Donaldson I, Wolting C, *et al*. BIND--The Biomolecular Interaction Network Database. *Nucleic Acids Res* 2001; **29**:242-245.
72 Morris JH, Huang CC, Babbitt PC, Ferrin TE. structureViz: linking Cytoscape and UCSF Chimera. *Bioinformatics* 2007; **23**:2345-2347.
73 Salwinski L, Eisenberg D. The MiSink Plugin: Cytoscape as a graphical interface to the Database of Interacting Proteins. *Bioinformatics* 2007; **23**:2193-2195.
74 Barsky A, Gardy JL, Hancock RE, Munzner T. Cerebral: a Cytoscape plugin for layout of and interaction with biological networks using subcellular localization annotation. *Bioinformatics* 2007; **23**:1040-1042.
75 Ferro A, Giugno R, Pigola G, *et al*. NetMatch: a Cytoscape plugin for searching biological networks. *Bioinformatics* 2007; **23**:910-912.
76 Garcia O, Saveanu C, Cline M, *et al*. GOlorize: a Cytoscape plug-in for network visualization with Gene Ontology-based layout and coloring. *Bioinformatics* 2007; **23**:394-396.
77 Vlasblom J, Wu S, Pu S, *et al*. GenePro: a Cytoscape plug-in for advanced visualization and analysis of interaction networks. *Bioinformatics* 2006; **22**:2178-2179.
78 Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 2005; **21**:3448-3449.
79 Cai YD, Chou KC. Predicting 22 protein localizations in budding

yeast. *Biochem Biophys Res Commun* 2004; **323**:425-428.
80 Chou KC, Cai YD. Predicting protein localization in budding yeast. *Bioinformatics* 2005; **21**:944-950.
81 Chou KC, Shen HB. Euk-mPLoc: a fusion classifier for large-scale eukaryotic protein subcellular location prediction by incorporating multiple sites. *J Proteome Res* 2007; **6**:1728-1734.
82 Su EC, Chiu HS, Lo A, *et al*. Protein subcellular localization prediction based on compartment-specific features and structure conservation. *BMC Bioinformatics* 2007; **8**:330.
83 Huh WK, Falvo JV, Gerke LC, *et al*. Global analysis of protein localization in budding yeast. *Nature* 2003; **425**:686-691.
84 Matsuyama A, Arai R, Yashiroda Y, *et al*. ORFeome cloning and global analysis of protein localization in the fission yeast Schizosaccharomyces pombe. *Nat Biotechnol* 2006; **24**:841-847.
85 Chen SC, Zhao T, Gordon GJ, Murphy RF. Automated image analysis of protein localization in budding yeast. *Bioinformatics* 2007; **23**:i66-i71.
86 Levy S, Sutton G, Ng PC, *et al*. The diploid genome sequence of an individual human. *PLoS Biol* 2007; **5**:e254.