



ELSEVIER

# Yeast two-hybrid contributions to interactome mapping

Jodi R Parrish<sup>1</sup>, Keith D Gulyas<sup>1</sup> and Russell L Finley Jr<sup>1,2</sup>

Interactome mapping, the systematic identification of protein interactions within an organism, promises to facilitate systems-level studies of biological processes. Using *in vitro* technologies that measure specific protein interactions, static maps are being generated that include many of the protein networks that occur *in vivo*. Most of the binary protein interaction data currently available was generated by large-scale yeast two-hybrid screens. Recent efforts to map interactions in model organisms and in humans illustrate the promise and some of the limitations of the two-hybrid approach. Although these maps are incomplete and include false positives, they are proving useful as a framework around which to elaborate and model the *in vivo* interactome.

## Addresses

<sup>1</sup> Center for Molecular Medicine and Genetics, Wayne State University School of Medicine, Detroit, MI, USA

<sup>2</sup> Department of Biochemistry and Molecular Biology, Wayne State University School of Medicine, Detroit, MI, USA

Corresponding author: Finley, Russell L (rfinley@wayne.edu)

**Current Opinion in Biotechnology** 2006, **17**:387–393

This review comes from a themed issue on  
Protein technologies  
Edited by Deb K Chatterjee and Joshua LaBaer

Available online 27th June 2006

0958-1669/\$ – see front matter

© 2006 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.copbio.2006.06.006

## Introduction

A full understanding of any biological system requires an account of the interactions among its constituent molecular parts. Defining the interactions among proteins is essential, because they play a role in virtually every biological process. Early efforts to map all of the protein interactions for a handful of organisms have used two main experimental approaches: yeast two-hybrid screening and the detection of protein complexes by mass spectrometry. These complementary approaches produce static, albeit noisy, snap-shots of the interactions that can occur in an *in vivo* system. Although the resulting interaction maps have lacked sufficient coverage and dynamic information to provide a clear picture of what a complete protein interaction map (or interactome) will look like, they are providing many biological insights and are a useful starting point for systems-level studies. This review will focus on large-scale yeast two-hybrid

screening, which has produced most of the binary protein interaction data currently available. Despite lingering technical issues, including incomplete coverage and the detection of non-biological interactions (false-positives), two-hybrid data are providing a foundation for further interactome studies.

The yeast two-hybrid system is a simple robust assay for protein–protein interactions [1,2] that was developed for high-throughput screening during the early 1990s (Box 1). The first genome-wide interaction map was generated for bacteriophage T7, demonstrating both the value of the screening approach and of the resulting data [3]. Similar efforts with larger genomes, however, have been limited in part by the challenge of generating clone sets with the open reading frames (ORFs) in the two-hybrid vectors for expressing DNA-binding domain (BD) and activation domain (AD) fusions in yeast. A second challenge has been to establish efficient strategies to mate large sets of BD and AD yeast strains to sample all possible combinations of interactions (Box 1). Despite these challenges, large-scale yeast two-hybrid screens have been conducted for several viruses, *Helicobacter pylori*, budding yeast, *Plasmodium falciparum*, *Caenorhabditis elegans*, *Drosophila* and, recently, human [4–10,11<sup>\*</sup>,12,13<sup>\*\*</sup>,14,15,16<sup>\*\*</sup>,17<sup>\*\*</sup>]. The data from these screens are proving to be tremendously useful for individual studies and for interactome modeling. The protein interaction maps generated for human pathogens, for example, have provided clues as to proteins that might function together during pathogenesis, as well as identifying putative protein targets for drug development. For the remainder of this review we will focus on studies with model organisms and recent results with human proteins.

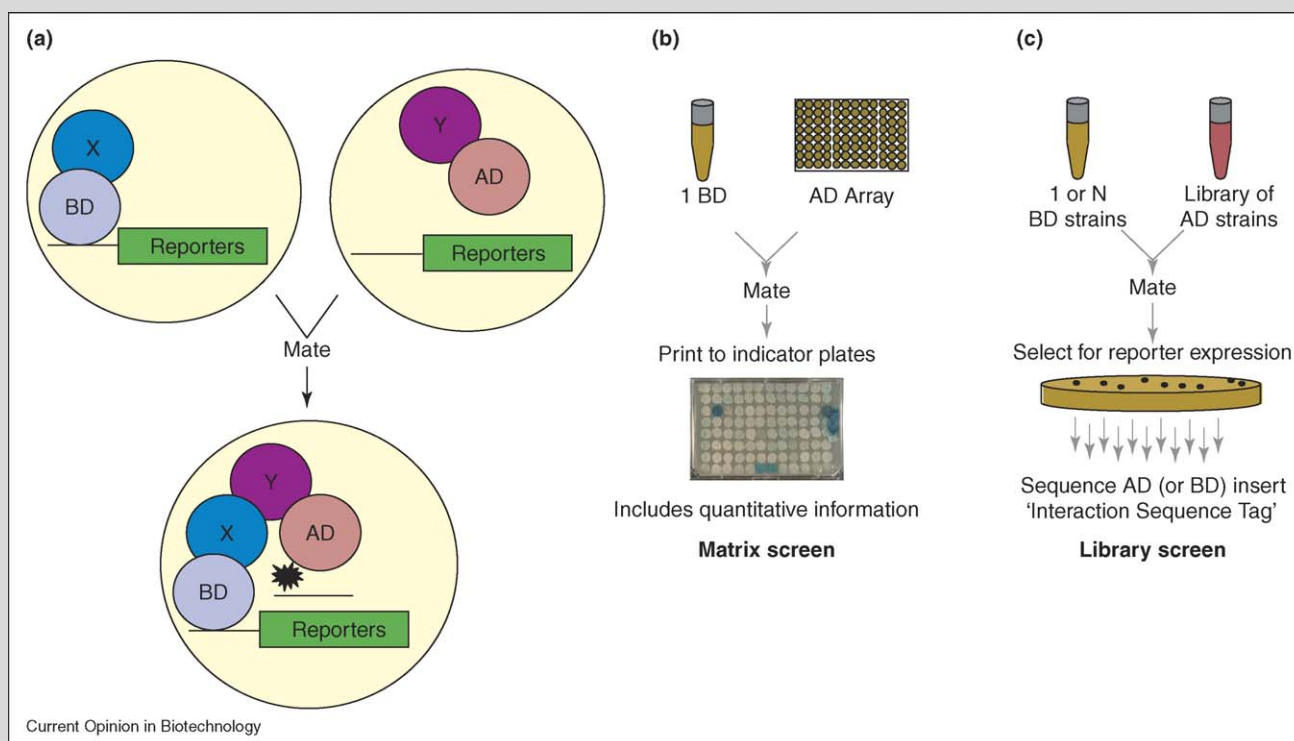
## Yeast two-hybrid interaction maps

The first model organism interaction maps were generated for the budding yeast, *Saccharomyces cerevisiae*. Several large-scale screens generated over 5600 interactions involving 69% of the yeast proteins [9,10,18]. The two papers describing the largest screens [9,10] have been cited over 2200 times, and discussion of that data could form the basis of a lengthy review by itself. Two key points are worth noting here. First, the different screening approaches that were used helped inform all subsequent efforts with larger model organisms (see Box 1). Second, and perhaps more importantly, the yeast two-hybrid data provided a foundation for numerous biological studies, including many that developed generalized methods for analyzing and using interaction data [19<sup>\*</sup>]. The value of this contribution may be difficult to quantify, but probably cannot be overestimated.

**Box 1** High-throughput screening using the yeast two-hybrid approach

High-throughput two-hybrid screens use a mating assay [39,40] as illustrated in Figure 1a. The two proteins to be tested for interaction are expressed in yeast as hybrid fusion proteins. One protein (X) is fused to a DNA-binding domain (BD) from a transcription factor, such as Gal4 or LexA. The BD will bind to upstream sites engineered into reporter genes. The second protein (Y) is fused to a transcription activation domain (AD). If X and Y interact, the AD activates the reporters, leading to the formation of a colony on media where the reporters are required for growth. LacZ is also a common reporter; its activity is detected by blue color on indicator plates. The use of mating enables arrays of BD strains and AD strains to be collected, which can then be screened against each other in high-throughput formats. A BD strain, for example, can be mated to the AD array to identify individual interacting AD fusions (Figure 1b). This 'matrix' approach may be the most sensitive because it effectively tests one BD strain against each AD strain in the array, but for large arrays it is impractical (e.g. thousands of BD strains would be mated with tens or hundreds of AD plates) [9]. A more efficient but less sensitive approach is to mate individual BD strains with libraries of AD strains (Figure 1c). Reporter activation is then selected for and the AD fusion in the surviving clones sequenced [3,41,42]. The AD libraries can be generated from cDNA or from pools of AD clones from an array. The sensitivity of this 'library' approach is limited by sampling efficiency for a large library of AD strains and by the number of clones that researchers can practically pick and sequence. The efficiency of the matrix approach has been improved by strategies that use small pools of clones from one or both arrays (e.g. [10,43,44]). These matrix pooling strategies generally test small pools first, then confirm the interactions by mating individual BD and AD strains. Most recent screens have used some variation of a pooling strategy.

In addition to a variety of screening strategies, other elements of the two-hybrid system have been tweaked in different ways in attempts to maximize the identification of biologically relevant interactions. These include the use of up to three reporters [45], reduced expression vector copy number [42], quantification of reporter activation levels [43], retesting by subcloning open reading frames into fresh yeast, regulated expression of the hybrid proteins [46], and methods to deal with BD fusions that activate transcription [43,47]. Together, these advances have led to a variety of screening systems and strategies in current use, each with its own merits. In fact, the lack of overlap among datasets generated by different two-hybrid screening systems could be due in part to the unique capability of each system to detect only a subset of the interactions, which would argue for the use of multiple systems to maximize coverage. A caveat to this interpretation is that each screen is probably far below saturation, making it difficult to compare two systems by comparing the results of two screens.

**Figure 1**

Yeast two-hybrid high-throughput screening approaches. **(a)** High-throughput two-hybrid screening utilizes yeast mating, in which expression plasmids initially in two different haploid yeast strains are brought together. In the first strain, protein (X) is fused to a DNA-binding domain (BD) and will bind at an engineered site upstream of the reporter gene. In the second strain, protein (Y) is fused to a transcription activation domain (AD). To conduct the assay, the two strains are mated and the reporter activity measured in the resulting diploids. If X and Y interact, AD activates the reporter, leading to selection. **(b)** The matrix approach and **(c)** the library approach for high-throughput screening; both have been superseded by pooling strategies. See text box for details.

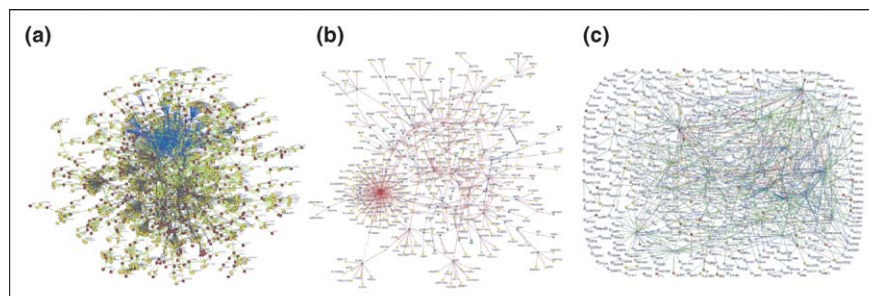
Two-hybrid screens for the multicellular organism *C. elegans* have identified over 5400 interactions covering ~12% of the *C. elegans* genes [12,20]. In the largest screen, Li *et al.* [12] initially focused on multicellular functions by screening *C. elegans* AD libraries with ~1900 BD bait proteins chosen because they have human orthologs or were known to be involved in metazoan-specific functions. Several studies have integrated the interaction data with other functional genomics data to derive models for genetic pathways. In one study, for example, the two-hybrid interactions were combined with correlated expression profiles and loss-of-function phenotype profiles to derive an integrated network that predicted a set of molecular machines functioning during early *C. elegans* embryogenesis [21]. In another example, the two-hybrid map was extended and used to guide a systematic phenotypic analysis of the transforming growth factor  $\beta$  (TGF $\beta$ ) signaling network, leading to a functional map of that pathway [22]. In each of these examples, the binary two-hybrid data were an essential part of the initial framework leading to testable hypotheses about genetic regulatory networks.

Three large two-hybrid screens for *Drosophila* generated the most extensive interaction map yet for any metazoan [13<sup>••</sup>,14,15]. Together, these screens detected over 24 000 interactions involving proteins representing ~54% of the predicted *Drosophila* genes [23]. Each of the screens demonstrated that the data are enriched for biologically relevant interactions, pointing to statistically significant overlap with other datasets including genetic interactions, pairings of specific gene functions and protein domains, and biologically significant network topological properties. Although the map is still incomplete, the broad level of coverage creates opportunities for connecting together more proteins into functional

modules than do maps for other multicellular organisms. This value is illustrated in the potential for placing human disease-associated genes into pathways. Of the 2727 human disease-associated genes currently identified in the Online Mendelian Inheritance in Man (OMIM) database [24], 1716 have obvious *Drosophila* orthologs and 914 (53%) of these are represented in the two-hybrid maps (Figure 1).

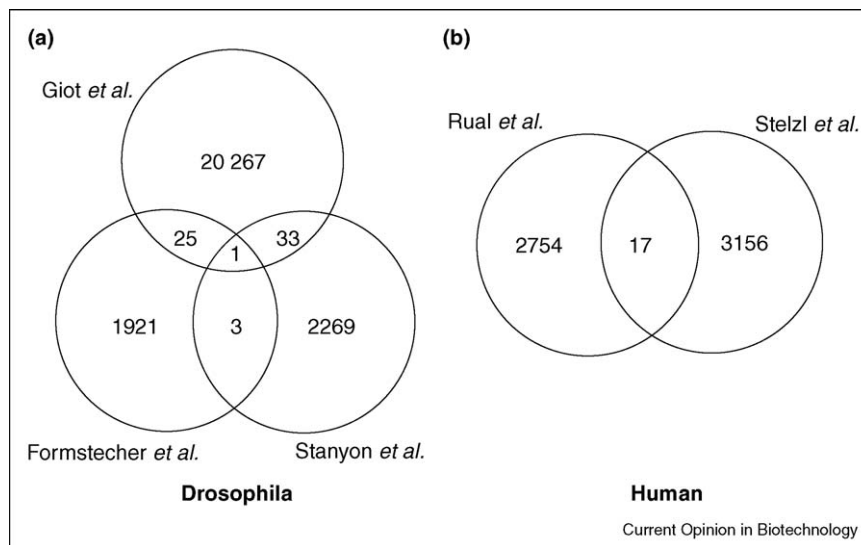
Large two-hybrid screens with human proteins have until recently focused on specific diseases or pathways [25,26] (see also Update). Two new studies, however, have demonstrated the potential for proteome-wide interactome mapping for humans [16<sup>••</sup>,17<sup>••</sup>]. Rual *et al.* [16<sup>••</sup>] screened approximately 7200 human full-length ORFs, randomly chosen on the basis of clone availability, and identified 2754 protein interactions. Stelzl *et al.* [17<sup>••</sup>] screened two-hybrid arrays generated from a human fetal brain cDNA library and from a set of ~2000 full-length ORFs, and identified 3156 interactions. Together, the two datasets identified over 5900 protein interactions (Figure 2), most of which are novel. As with the model organism maps, the human two-hybrid maps showed statistically significant pairing of proteins with the same annotated functions. Rual and colleagues also showed that interacting protein pairs were more likely to have mouse orthologs with a shared phenotype than random pairs, had a higher probability of being encoded by genes with common upstream sequence elements, and were more likely to have correlated gene expression profiles [16]. All of these findings point to the biological significance of the interactions in the two-hybrid data. Both studies give examples by highlighting several subnetworks in their maps that provide biological insights into specific pathways. The maps also connect several OMIM disease-associated proteins into potential pathways and

**Figure 1**



Protein interaction maps place disease-associated proteins into a network context. Shown are portions of yeast two-hybrid protein interaction data from high-throughput screens with *Drosophila* or human proteins. **(a)** The *Drosophila* high-confidence (HC) interactions from three combined studies involving orthologs of human disease-associated proteins (red nodes) [48]. Green edges are from [13], blue edges are from [14], gray edges are from [15]. Orthologs were obtained as described in [48] by BLASTP with disease-associated genes from OMIM at an e-value cut-off of  $10^{-10}$ . Of the 1716 fly orthologs of human disease genes, 914 (53%) are in the combined two-hybrid map. **(b)** Human interactions from the dataset, 'Center for Cancer Systems Biology Human Interactome version 1' (CCSB-HI1) [16], involving OMIM disease-related proteins (green nodes). The CCSB-HI1 two-hybrid dataset included 121 of the disease proteins. (Figure reproduced from [16] with permission.) **(c)** All 911 HC interactions from [17] are shown. The map includes 45 human OMIM disease proteins (orange nodes); 195 disease proteins were found in the entire two-hybrid map. (Figure reproduced from [17] with permission.)

Figure 2



High-throughput two-hybrid screens are subsaturating. Venn diagrams showing overlap among independent high-throughput two-hybrid screens (a) for *Drosophila* proteins or (b) for human proteins. In each case the data represent the entire two-hybrid dataset rather than just the set judged to be high confidence in each study. Numbers indicate unique interactions based on gene locus (i.e. detection of an interaction between protein A and two splice variants of protein B would be counted as one interaction). Data for *Drosophila* were obtained from Giot *et al.* [13<sup>\*\*</sup>], Stanyon *et al.* [14], and Formstecher *et al.* [15], and compiled to remove redundancy in the *Drosophila* Interactions Database [23]. Human data was obtained from Rual *et al.* [16<sup>\*\*</sup>] and Stelzl [17<sup>\*\*</sup>].

functional modules with other proteins (Figure 1), demonstrating the potential of this data for human disease research.

The large amount of interaction data now available from human and model organisms has led to further insight through comparative network analyses. Cross-species comparisons of interaction datasets have identified multiple networks conserved between yeast, *Drosophila* and *C. elegans* [27]. The comparison of interaction maps from multiple organisms has facilitated the prediction of additional interactions missing in one system but found in others [27,28<sup>\*\*</sup>]. The model organism maps have been particularly useful for predicting human interactions [29–31]. Lastly, several studies have noted that interaction maps share certain topological features that correlate with biological properties, including clustering of functionally related proteins and a non-random distribution of interactions per protein [32]. However, more recent studies suggest that the level of interactome coverage could impact the interpretation of topological features (e.g. [28<sup>\*\*</sup>,33]) and, therefore, the full implications of network analyses are likely to continue to emerge as interactome maps become more complete.

### Coverage

One common feature of the large-scale two-hybrid screens is the high frequency of false negatives or missed interactions. These are evident from the minimal overlap

with published low-throughput datasets [12,13<sup>\*\*</sup>,16<sup>\*\*</sup>,34] or between different high-throughput two-hybrid datasets generated for the same organism (Figure 2). One reason for this incomplete coverage is that the clone sets used in the screens are themselves incomplete. For example, the screens of Rual [16<sup>\*\*</sup>] and Stelzl [17<sup>\*\*</sup>] represented only around 23% and 30% of the predicted human genes, respectively, whereas the two-hybrid clone sets for *Drosophila* represent only 78% of the fly genes. A second factor is that two-hybrid screens miss some interactions, even among proteins included in the screening clone set. This occurs in part because some interactions cannot be detected by the yeast two-hybrid approach. Perhaps a larger contributing factor, however, is the subsaturating nature of the high-throughput screens. This is evident from comparing different two-hybrid screens that use the same proteins. Two of the *Drosophila* screens [13<sup>\*\*</sup>,14], for example, used AD arrays generated with the same set of ~12 000 ORFs and screened them with at least 100 of the same BD proteins, yet detected very few interactions in common (Figure 2a). Similarly, in comparing the human screens (Figure 2b) it is clear that there were many more opportunities for overlap than were observed. Of the 1904 genes represented in the interaction map from Stelzl and coworkers, 996 (52%) were among the genes that Rual and colleagues screened. Similar results were found for the extensive two-hybrid screens for yeast proteins [10]. In each case (yeast, *Drosophila* and human), the overlapping data from multiple screens is small and the

non-overlapping data are of the same quality and usefulness typical of two-hybrid data. Together, these studies suggest that screening efficiency remains a significant limitation to achieving complete coverage with this technology.

## Validation

High-throughput two-hybrid data are notoriously cluttered with false positives. Distinguishing these from the biologically relevant interactions is a challenge for biologists studying individual pathways and for computational biologists trying to model the interactome. There are two distinct aspects to validating two-hybrid interactions. One is to determine if the two-hybrid reporter activity is indicative of a true specific binary protein interaction. This can be addressed by demonstrating the interaction with a different assay, such as co-affinity purification. The second, much more important question is whether the interaction takes place *in vivo* where it plays some functional role in the organism. An interaction that can be detected in an orthogonal assay, or even in an independent two-hybrid screen, is less likely to be a two-hybrid artifact and more likely to be biologically significant [14,34,35]. However, at least in the near-term, such cross-validation might be difficult to obtain; for example, of the 43 242 interactions currently recorded in the Database of Interacting Proteins (DIP) [36] for yeast, worm and fly, 94% have been detected in only one experimental system and most two-hybrid interactions have been detected in only one screen (e.g. Figure 2).

An alternative to experimental cross-validation is to annotate every interaction with a computationally derived confidence score that relates to biological significance. For example, computational methods have been described that assign statistical scores to each interaction representing the likelihood that it occurs *in vivo*. In general, the methods use attributes of both the proteins and the interaction data that correlate with biological significance to derive a statistical score for every interaction [37]. Bader and colleagues [13<sup>••</sup>] applied this approach to score the interactions from the *Drosophila* two-hybrid data, showing that interactions receiving higher scores were more likely to be biological true positives. Such scoring systems have several advantages. First, every interaction is scored and maintained within the dataset; this keeps all interactions available for further analysis as new data, computational methods and scoring systems become available. Second, users can view the data at different confidence levels. For sparse networks, for example, a lower confidence level may be tolerated to increase sensitivity. Third, statistical scoring systems facilitate integration of datasets [38]. Finally, rather than indicating surrogate measures of significance, such as *in vitro* reproducibility, statistical scoring systems have the potential to relate directly to *in vivo* biological significance.

## Conclusions

Technological advances will be needed before it is possible to test interactions among all protein isoforms in a setting that captures the *in vivo* dynamics of protein networks. In the meantime, *in vitro* approaches such as the two-hybrid system can map many of the specific interactions that occur *in vivo* and provide a framework for modeling the interactome. Large-scale yeast two-hybrid protein interaction maps are now available for organisms from phage to human. These maps are providing insights into individual protein functions, pathways, molecular machines, functional protein modules, and evolution. Despite the success of the large-scale screens, they have identified only a fraction of the interactions that could be defined by yeast two-hybrid screening. With the continued development of complete clone sets and increasingly efficient screening strategies, the yeast two-hybrid system is likely to continue to make significant contributions to interactome mapping efforts for some time.

## Update

A recent high-throughput two-hybrid screen focused on proteins involved in inherited neurodegenerative disorders [49<sup>\*</sup>]. The screen resulted in a map with 770 mostly novel interactions centered on 20 ataxia-related proteins. The map linked many of the poorly characterized disease proteins to each other and to proteins with known functions, providing new clues about the pathways involved in the ataxia diseases. This study also illustrates the continued value of two-hybrid screens that focus on specific diseases or pathways, while the coverage of random large-scale screens remains incomplete.

## Acknowledgements

We would like to thank Guozhen Liu for help with compiling protein interaction database statistics.

## References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Fields S, Song O: **A novel genetic system to detect protein-protein interactions.** *Nature* 1989, **340**:245-246.
2. Chien C-T, Bartel PL, Sternglanz R, Fields S: **The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest.** *Proc Natl Acad Sci USA* 1991, **88**:9578-9582.
3. Bartel PL, Roecklein JA, SenGupta D, Fields S: **A protein linkage map of *Escherichia coli* bacteriophage T7.** *Nat Genet* 1996, **12**:72-77.
4. McCraith S, Holtzman T, Moss B, Fields S: **Genome-wide analysis of vaccinia virus protein-protein interactions.** *Proc Natl Acad Sci USA* 2000, **97**:4879-4884.
5. Uetz P, Dong YA, Zeretzke C, Atzler C, Baiker A, Berger B, Rajagopala SV, Roupelieva M, Rose D, Fossum E *et al.*: **Herpesviral protein networks and their interaction with the human proteome.** *Science* 2006, **311**:239-242.

6. Flajolet M, Rotondo G, Daviet L, Bergametti F, Inchauspe G, Tiollais P, Transy C, Legrain P: **A genomic approach of the hepatitis C virus generates a protein interaction map.** *Gene* 2000, **242**:369-379.
7. Guo D, Rajamaki ML, Saarna M, Valkonen JP: **Towards a protein interaction map of potyviruses: protein interaction matrixes of two potyviruses based on the yeast two-hybrid system.** *J Gen Virol* 2001, **82**:935-939.
8. Rain JC, Selig L, De Reuse H, Battaglia V, Reverdy C, Simon S, Lenzen G, Petel F, Wojcik J, Schachter V *et al.*: **The protein-protein interaction map of *Helicobacter pylori*.** *Nature* 2001, **409**:211-215.
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 USA* 2001, **98**:4569-4574.
11. LaCount DJ, Vignali M, Chettier R, Phansalkar A, Bell R, Hesselberth JR, Schoenfeld LW, Ota I, Sahasrabudhe S, Kurschner C *et al.*: **A protein interaction network of the malaria parasite *Plasmodium falciparum*.** *Nature* 2005, **438**:103-107.
- The authors used a novel two-hybrid screening strategy to overcome the fact that *P. falciparum* genes are difficult to express owing to their high (80%) AT content. They identified 2846 interactions among 1308 of the parasite's proteins, placing a huge number of previously uncharacterized proteins into putative networks.
12. Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, Vidalain PO, Han JD, Chesneau A, Hao T *et al.*: **A map of the interactome network of the metazoan *C. elegans*.** *Science* 2004, **303**:540-543.
13. Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao YL, Ooi CE, Godwin B, Vitols E *et al.*: **A protein interaction map of *Drosophila melanogaster*.** *Science* 2003, **302**:1727-1736.
- The authors subcloned ~78% of the predicted *Drosophila* ORFs into AD and BD arrays and conducted two-hybrid screens. They detected over 20 000 interactions involving over 7,000 proteins. Combined with more recent screens from Stanyon *et al.* [14] and Formstecher *et al.* [15], this work constitutes the most complete metazoan interaction map yet determined experimentally, covering ~54% of the predicted genes.
14. Stanyon CA, Liu G, Mangiola BA, Patel N, Giot L, Kuang B, Zhang H, Zhong J, Finley RL Jr: **A *Drosophila* protein-interaction map centered on cell-cycle regulators.** *Genome Biol* 2004, **5**:R96.
15. Formstecher E, Aresta S, Collura V, Hamburger A, Meil A, Trehin A, Reverdy C, Betin V, Maire S, Brun C *et al.*: **Protein interaction mapping: a *Drosophila* case study.** *Genome Res* 2005, **15**:376-384.
16. Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N *et al.*: **Towards a proteome-scale map of the human protein-protein interaction network.** *Nature* 2005, **437**:1173-1178.
- Among the two largest two-hybrid screens for human protein interactions (Stelzl *et al.* [17\*\*] is the other). The authors constructed yeast two-hybrid BD and AD arrays from full-length ORFs representing ~7200 human genes and screened all possible combinations using a pooling strategy. Some 2754 interactions were detected among 1549 proteins, including 121 disease-associated proteins.
17. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S *et al.*: **A human protein-protein interaction network: a resource for annotating the proteome.** *Cell* 2005, **122**:957-968.
- Among the two largest two-hybrid screens for human protein interactions (Rual *et al.* [16\*\*] is the other). The authors screened BD and AD arrays containing a mixture of cDNA clones and full-length ORFs using a pooling strategy. Some 3186 interactions were detected among 1705 proteins, including 195 disease-associated proteins.
18. Fromont-Racine M, Mayes AE, Brunet-Simon A, Rain JC, Colley A, Dix I, Decourty L, Joly N, Ricard F, Beggs JD *et al.*: **Genome-wide protein interaction screens reveal functional networks involving Sm-like proteins.** *Yeast* 2000, **17**:95-110.
19. Fields S: **High-throughput two-hybrid analysis. The promise and the peril.** *FEBS J* 2005, **272**:5391-5399.
- Recent review of yeast two-hybrid technology with particular emphasis on recent high-throughput screens with *P. falciparum* proteins and work on detecting interactions with yeast membrane proteins. The review also discusses methods for dealing with false positives.
20. Walhout AJ, Sordella R, Lu X, Hartley JL, Temple GF, Brasch MA, Thierry-Mieg N, Vidal M: **Protein interaction mapping in *C. elegans* using proteins involved in vulval development.** *Science* 2000, **287**:116-122.
21. Gunsalus KC, Ge H, Schetter AJ, Goldberg DS, Han JD, Hao T, Berriz GF, Bertin N, Huang J, Chuang LS *et al.*: **Predictive models of molecular machines involved in *Caenorhabditis elegans* early embryogenesis.** *Nature* 2005, **436**:861-865.
22. Tewari M, Hu PJ, Ahn JS, Ayivi-Guedehoussou N, Vidalain PO, Li S, Milstein S, Armstrong CM, Boxem M, Butler MD *et al.*: **Systematic interactome mapping and genetic perturbation analysis of a *C. elegans* TGF-beta signaling network.** *Mol Cell* 2004, **13**:469-482.
23. Pacifico S, Liu G, Guest S, Parrish JR, Fotouhi F, Finley RL Jr: **A database and tool, IM Browser, for exploring and integrating emerging gene and protein interaction data for *Drosophila*.** *BMC Bioinformatics* 2006, **7**:195.
24. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA: **Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders.** *Nucleic Acids Res* 2005, **33**:D514-D517.
25. Colland F, Jacq X, Trouplin V, Mouglin C, Groizeleau C, Hamburger A, Meil A, Wojcik J, Legrain P, Gauthier JM: **Functional proteomics mapping of a human signaling pathway.** *Genome Res* 2004, **14**:1324-1332.
26. Goehler H, Lalowski M, Stelzl U, Waelter S, Stroedicke M, Worm U, Droege A, Lindenbergs KS, Knoblich M, Haenig C *et al.*: **A protein interaction network links GIT1, an enhancer of huntingtin aggregation, to Huntington's disease.** *Mol Cell* 2004, **15**:853-865.
27. Sharan R, Suthram S, Kelley RM, Kuhn T, McCuine S, Uetz P, Sittler T, Karp RM, Ideker T: **Conserved patterns of protein interaction in multiple species.** *Proc Natl Acad Sci USA* 2005, **102**:1974-1979.
28. Gandhi TK, Zhong J, Mathivanan S, Karthick L, Chandrika KN, Mohan SS, Sharma S, Pinkert S, Nagaraju S, Periaswamy B *et al.*: **Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets.** *Nat Genet* 2006, **38**:285-293.
- An analysis that combines the model organism interaction maps, including the two-hybrid data and interactions obtained from other technologies, with a large database of human protein interactions manually curated from the literature. Several important conclusions are drawn about protein networks, disease pathways, and prediction of interactions.
29. Rhodes DR, Tomlins SA, Varambally S, Mahavisno V, Barrette T, Kalyana-Sundaram S, Ghosh D, Pandey A, Chinnaiyan AM: **Probabilistic model of the human protein-protein interaction network.** *Nat Biotechnol* 2005, **23**:951-959.
30. Brown KR, Jurisica I: **Online predicted human interaction database.** *Bioinformatics* 2005, **21**:2076-2082.
31. Lehner B, Fraser AG: **A first-draft human protein-interaction map.** *Genome Biol* 2004, **5**:R63.
32. Barabasi AL, Oltvai ZN: **Network biology: understanding the cell's functional organization.** *Nat Rev Genet* 2004, **5**:101-113.
33. Han JD, Dupuy D, Bertin N, Cusick ME, Vidal M: **Effect of sampling on topology predictions of protein-protein interaction networks.** *Nat Biotechnol* 2005, **23**:839-844.
34. von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, Bork P: **Comparative assessment of large-scale data sets of protein-protein interactions.** *Nature* 2002, **417**:399-403.
35. Deane CM, Salwinski L, Xenarios I, Eisenberg D: **Protein interactions: two methods for assessment of the reliability of high throughput observations.** *Mol Cell Proteomics* 2002, **1**:349-356.

36. Xenarios I, Rice DW, Salwinski L, Baron MK, Marcotte EM, Eisenberg D: **DIP: the database of interacting proteins.** *Nucleic Acids Res* 2000, **28**:289-291.
37. Bader JS, Chaudhuri A, Rothberg JM, Chant J: **Gaining confidence in high-throughput protein interaction networks.** *Nat Biotechnol* 2004, **22**:78-85.
38. Fraser AG, Marcotte EM: **A probabilistic view of gene function.** *Nat Genet* 2004, **36**:559-564.
39. Finley RL Jr, Brent R: **Interaction mating reveals binary and ternary connections between *Drosophila* cell cycle regulators.** *Proc Natl Acad Sci USA* 1994, **91**:12980-12984.
40. Bendixen C, Gangloff S, Rothstein R: **A yeast mating-selection scheme for detection of protein-protein interactions.** *Nucleic Acids Res* 1994, **22**:1778-1779.
41. Fromont-Racine M, Rain JC, Legrain P: **Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens.** *Nat Genet* 1997, **16**:277-282.
42. Walhout AJ, Vidal M: **High-throughput yeast two-hybrid assays for large-scale protein interaction mapping.** *Methods* 2001, **24**:297-306.
43. Zhong J, Zhang H, Stanyon CA, Tromp G, Finley RL Jr: **A strategy for constructing large protein interaction maps using the yeast two-hybrid system: regulated expression arrays and two-phase mating.** *Genome Res* 2003, **13**:2691-2699.
44. Jin F, Hazbun T, Michaud GA, Salcius M, Predki PF, Fields S, Huang J: **A pooling-deconvolution strategy for biological network elucidation.** *Nat Methods* 2006, **3**:183-189.
45. James P, Halladay J, Craig EA: **Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast.** *Genetics* 1996, **144**:1425-1436.
46. Finley RL Jr, Zhang H, Zhong J, Stanyon CA: **Regulated expression of proteins in yeast using the *MAL61-62* promoter and a mating scheme to increase dynamic range.** *Gene* 2002, **285**:49-57.
47. Walhout AJ, Vidal M: **A genetic strategy to eliminate self-activator baits prior to high-throughput yeast two-hybrid screens.** *Genome Res* 1999, **9**:1128-1134.
48. Reiter LT, Potocki L, Chien S, Gribskov M, Bier E: **A systematic analysis of human disease-associated gene sequences in *Drosophila melanogaster*.** *Genome Res* 2001, **11**:1114-1125.
49. Lim J, Hao T, Shaw C, Patel AJ, Szabo G, Rual JF, Fisk CJ, Li N, Smolyar A, Hill DE *et al.*: **A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration.** *Cell* 2006, **125**:801-814.
- Recent high-throughput two-hybrid screen focusing on proteins involved in inherited neurodegenerative disorders. The screens resulted in a map with 770 interactions centered on 20 ataxia-related proteins. Interestingly, 18 of these proteins interacted either directly or indirectly, and the entire network was enriched for proteins involved in specific biological processes.