Published in final edited form as:
Biotechniques. 2008 April ; 44(5): 655–662. doi:10.2144/000112797.

Resolving the network of cell signaling pathways using the evolving yeast two-hybrid system

Vladimir Ratushny[1,2] and **Erica A. Golemis**[1*]

1*Fox Chase Cancer Center, 333 Cottman Ave., Philadelphia, PA 19111*

2*Drexel University College of Medicine, 2900 W. Queen Lane, Philadelphia, PA 19129.*

Abstract

In 1983, while investigators had identified a few human proteins as important regulators of specific biological outcomes, how these proteins acted in the cell was essentially unknown in almost all cases. 25 years on, our knowledge of the mechanistic basis of protein action has been transformed by our increasingly detailed understanding of protein-protein interactions, which have allowed us to define cellular machines. The advent of the yeast two-hybrid (Y2H) system in 1989 marked a milestone in the field of proteomics. Exploiting the modular nature of transcription factors, the Y2H system allows facile measurement of the activation of reporter genes based on interactions between two chimeric or "hybrid" proteins of interest. After a decade of service as a leading platforms for individual investigators to use in exploring the interaction properties of interesting target proteins, the Y2H system has increasing been applied in high throughput applications intended to map genome-scale protein-protein interactions for model organisms and humans. Although some significant technical limitations apply, the Y2H has made a great contribution to our general understanding of the topology of cellular signaling networks.

Introduction

Within the last twenty-five years, the yeast two-hybrid (Y2H) system has helped transform out understanding of the molecular landscape of a cell. By 1983, application of increasingly powerful molecular biological techniques to mammals had begun to identify novel genes encoding proteins nominated as important for study based on functional criteria. For example, mutation of the oncogene Ras (1), or amplification of the oncogene Myc (2) was known to cause cancer in humans. The proteins encoded by these and other important genes were then further characterized primarily by simple overexpression of intact, truncated, or mutated derivatives of the protein of interest, followed by measurements of changes in gross cellular processes such as proliferation or morphological change. This "isolationist", single protein-centered approach was marked by the very limited nature of mechanistic insights that could be gained into the relations between a primary protein of interest and its protein partner(s). For some proteins such as Ras, that had known orthologs in lower eukaryotes (3), some insights into relevant signaling pathways could be gleaned by inference from genetic studies in those organisms (4). For proteins such as Myc that did not, options were more restricted. In 1983, the primary means of identifying interacting proteins was through biochemical co-purification and protein sequencing: difficult, laborious, and expensive.

*corresponding author: Erica A. Golemis, Fox Chase Cancer Center, W406, 333 Cottman Ave., Philadelphia, PA 19111 USA, (215) 728-2860 ph, -3616 fax, EA_Golemis@fccc.edu.

As databases of potentially interesting proteins burgeoned through the 1980s, this pattern continued. While most signaling proteins were thought or known to have at least one effector, such signaling pathways as existed were sparsely populated and drawn to indicate linear connections between interacting proteins. At its initial description in 1989 (5), it was not apparent that the Y2H assay was a technological advance that would transform our understanding of cell biology (6–11). Initially described as a simple, yet powerful, tool to detect protein-protein interactions (PPIs), the Y2H technology evolved in order to address PPIs of increasing complexity, and ultimately has been applied on a genome-wide scale. As most revolutionary technologies are judged by the physical or intellectual products they create, we believe that the Y2H system can be judged by the changing paradigm of protein interactions it helped usher in: from binary protein interactions of twenty-five years ago, to canonical pathways, and, most recently, towards interconnected protein interaction networks.

Y2H basics

The original Y2H system was developed as an assay to study binary protein interactions between proteins already known or strongly suspected to associate (5). The system capitalizes on the modular nature of transcription factors, whose split domains can reconstitute protein function upon physical interaction (6). The protein of interest (X), or bait, is fused to the DNA-binding domain (DBD) of a transcription factor such as Gal4 or LexA. (Figure 1). The potential interactor protein (Y), or prey, is fused to a transcriptional activating domain (AD). DBD-X and AD-Y are co-expressed in an appropriate yeast strain that has been engineered to contain reporter genes whose transcription depends on the association of DBD-X and AD-Y at the promoter. While the first described Y2H system used only a lacZ colorimetric reporter, which was not amenable to high-throughput usage using 1980s technology, the subsequent addition (8,10) of secondary auxotrophic reporters opened the door for the use of the Y2H for library screening. Some of the very earliest uses of the Y2H for such screening involved oncoproteins such as Ras and Myc (11,13). The instructive case of screening with Ras set forth what was to become a typical paradigm: although for a number of years there had been accumulating evidence suggesting functional relationship between Ras and a second oncoprotein, Raf, Y2H screening made clear the observation that a direct physical association underlay the functional relation (7). Suddenly, proteins could quickly find partners.

The establishment of Y2H library screening capabilities allowed many investigators to try to identify other proteins with which their protein of interest interacted. In the approach standard in the early 1990s, a yeast strain expressing bait and reporter that had passed certain optimization steps (below) would then be super-transformed with a cDNA library expressing AD-fused proteins, and $1–3 \times 10^6$ transformants would then be scored for activation of the *lacZ* and auxotrophic reporters. The basic technical manipulations required little more than low-cost microbiological growth media and minimal laboratory equipment, and were readily accomplishable by any laboratory with basic competence in molecular biology. The requisite plasmids and yeast strains required were essentially free, distributed from the laboratories of the inventors of the technique. From start to finish, a screen could be performed in less than six weeks. These advantages underlay the essentially democratic nature of Y2H screening: with such a low investment cost, many could afford to attempt the approach.

Y2H limitations

The older biochemical techniques available to study protein interactions before the development of the Y2H had specific limitations. For example, co-immunoprecipitation experiments required the existence of robust antibodies for target proteins, which were not always available. As a positive, the associations of proteins could be studied in native context, sometimes with complexes of partners; as a negative, binding of the immunoprecipitating

antibody to the target protein had the potential to competitively block the binding of the target to partners of interest. The endpoint of a coimmunoprecipitation or other biochemical purification experiment was analysis of the pool of interacting proteins by protein sequencing or mass spectrometry. This required a very large amount of starting material; even when this was available, in many cases difficulties in purification of sufficient amounts of some proteins caused mass spectrometry to be inadequate for large-scale screening (8), based on technologies of the 1980s and early 1990s.

While the Y2H avoided these limitations by use of a highly engineered, artificial reporting system, the Y2H system possessed its own set of complementary limitations. As a major limitation, bait proteins possessing intrinsic transcriptional activity could not initially be studied, as they would activate the reporter genes in the absence of interaction with an AD-fused partner. To circumvent this fault, a number of approaches were attempted, with varying degrees of success. As one strategy, investigators fused transcriptionally active proteins to ADs and conducted screens of DBD-libraries (15): however, this fix led to the need to develop a strategy to pre-screen the DBD-fused libraries, to eliminate a new set of transcriptionally activating baits. Another creative approach involved use of a RNA-polymerase III transcription system, which was not responsive to classic RNA polymerase II transcriptionally activating proteins, to enable screening of such baits (16,17). Another approach removed transcriptional activating proteins from the nucleus entirely, creating a parallel reporter system based at the cell membrane (9).

An important issue that rapidly emerged for investigators using the Y2H was the fact that most baits yielded at least some false positives, and some baits produced many. This led significant investment of effort towards understanding their cause and reducing their incidence (19–21). False positive results can be broken down into two categories: technical and biological. Technical false positives are faults of the reporter systems in the Y2H assay, wherein reporter activity is induced in the absence of any PPIs. Causes for such artifactual readouts include DBD- or AD-fused proteins that can weakly activate transcription of the reporter gene in absence of mutual interaction, coupled with plasmid copy number changes or rearrangements that result in either weak activators becoming overexpressed or reporter auto-activation (22, 23). Biological false positives are genuine interactions that occur in the context of the Y2H system, where all proteins are forced into the cell nucleus, but are not representative of PPIs in the cellular milieu. Cellular compartmentalization or localization may prevent *in vivo* interaction of proteins that have the appropriate cognate recognition sequences. Arguably, biological false positives are more difficult to circumvent than technical false positives because the former are less amenable to technical optimization. In all, some have estimated that false positives account for 50% of the current high throughput Y2H data sets (24–26). Although the high-throughput methodologies post no better marks, with estimates of spurious results also approximated at 50% (10), grappling with these potentially high error rates was a serious issue as laboratories began to broadly apply Y2H techniques.

Many investigators adapting the Y2H system wanted to ask more complicated questions about PPIs that were not intrinsically amenable to a system based in yeast. This desire was partially addressed by the development of Y2H system variants (11). For example, many eukaryotic protein interactions depend on post-translational modifications. To compensate for absence of the relevant modifying enzymes in yeast species, co-expression of appropriate modifiers, such as tyrosine kinases, along with bait and prey, was used to successfully study signal transduction pathways (28,29).

Others wished to expand the scope of the Y2H system to the study of protein-RNA and protein-drug interactions. For this, investigators employed a known RNA or ligand-binding protein conjugated to the DBD to serve as a ‘hook.’ (30–33). This modification optimized the

presentation of RNA or ligand-binding elements of bait proteins to non-traditional prey. As another example, while most investigators wished to use Y2H techniques to identify protein interactions, others sought to detect specific mutations leading to loss of protein interaction. As such, investigators have used negatively counter-selectable markers to search for mutations that disrupt binding of cell cycle-related proteins (34,35). In addition, a counter-selective Y2H system has been used to search for drugs or peptide ligands that can disrupt specific protein-protein interactions (34,36–38). Finally, the development by Stagljar and coworkers of a Y2H derivative that allowed efficient study of membrane-tethered proteins was an important advance that rendered a notoriously difficult class of protein targets amenable to study (9).

There have been a number of successes (notably in the arena of protein-RNA interactions, or studying interactions of membrane tethered proteins) using these complicating modifications to the Y2H assay. However, as of 2008, the general applicability of some of these Y2H-variations as ‘gold-standard’ techniques is questionable. Many of the "evolved" forms of Y2H system suffer from a very high background of false positives. Others, while deserving enormous commendation for creativity, are technically extremely difficult to set up because of the need to optimize multiple components (11). Pragmatically, widespread use of particular modified Y2H systems has been precluded because of fairly robust alternative technologies available in the field for specific applications: chromatin immunoprecipitation for protein-DNA interactions, or fluorescence polarization or structural model-based screening for protein-targeted drugs are examples of techniques that out-competed the Y2H in these arenas (8).

Genome-wide Y2H: first steps

The large-scale adaptation of the Y2H system by individual investigators was reflected by a surge in publications reporting screening results through the 1990s (Figure 2). Although many of these studies initiated with an interaction predicted by a Y2H "hit", validation of results was extensive, leading to high confidence in the interactions thus obtained. As the Y2H technique was applied to thousands of proteins, most biologically interesting signaling proteins were found to possess multiple interactors. Signaling pathway diagrams became more complex, and research strategies emphasized the analysis of how and under what specific conditions one protein modified the localization and function of each of its known partners, to achieve a final functional endpoint (e.g., increased proliferation).

Since 2000, the publication of complete genome sequences for humans and several important model organisms (39–43) spurred the search for methodologies to characterize previously undescribed open reading frames (ORFs). Clearly, the interaction of a protein product of an uncharacterized ORF with a protein of known function could be used to begin to ascribe functional specificity to the ORF (8). In an attempt to keep the realm of knowledge of protein interactions on par with the rapidly increasing stream of genomic information, investigators co-opted the Y2H system for high-throughput analyses (23,44–53).

The study of the protein interaction map of bacteriophage T7 (12) was the first genomic-wide application of Y2H technology. However, the first truly high-throughput Y2H screens addressed the genome of the Y2H host, *Saccharomyces cerevisiae*, itself, with the work of three competing groups appearing in 2000 and 2001 (24,52,53). The goal of these studies was to begin to elucidate the protein interactions of the approximately 6,000 gene products in the budding yeast. Although some of the work described involved the use of standard library screening, such an undertaking was greatly facilitated by a modification in use of the Y2H system so as to increase its capacity for multiplexing a large number of reactions and to improve its compatibility with robotic automation. These high-throughput analyses were enabled by efficient exploitation of a key biological feature of yeast: the ability of yeast cells of opposite haploid mating types to mate, forming a diploid cell.

As first described by Finley and colleagues in 1994 (13), it was possible to transform a reporter strain of mating type a with a plasmid expressing DBD-X, and a strain of mating type α with a plasmid expressing AD-Y, mix the yeast, and get a positive interaction result scorable in the a/α diploid. Hence, if wishing to test 10 baits against 20 preys, it is only necessary to perform 20 transformations followed by mating (which could easily be performed on 1–2 culture plates, or in microtiter plate format), instead of 200 independent transformations. When considering genome-level screens, the arithmetic versus multiplicative nature of mating-based screening provides a critical enabling tool. The largest of the *S. cerevisiae* screens (53) took a bi-directional approach, employing a mating-based protein array, in addition to a protein library method.

The array method involved the construction of approximately 6,000 yeast transformants, each expressing a unique ORF fused to an AD, arranged in specific positions on microassay plates. Large-scale homologous recombination-based cloning (14) and transformation were performed using PCR-generated products of *S. cerevisiae* ORFs that contained 5' and 3' flanking sequences directly matching those contained within the AD vector. A set of 192 ORF-DBD fusion products was also constructed and transformed into a yeast strain of the opposite mating type. Mating of each of the 192 DBD-fusions to each member of the AD-hybrid array provided information on potential PPIs in *S. cerevisiae*. A parallel approach employed by the investigators involved pooling roughly 6,000 ORFs-AD transformants into a library, while a similar set of roughly 6,000 ORF-DBD transformants was arranged in microassay plates. Each of the DBD-fusion yeast was mated with the AD yeast-library, and potential interacting proteins were identified by growth selection followed by sequencing (15).

Both Y2H screening modalities have some drawbacks, primary of which was reproducibility: Only 20% of identified interactions by the array method were reproduced in a second screen of the array. The library method afforded improved reproducibility, with 68% of potential interactions being validated in independent experiments or multiple times in a single experiment. These numbers were not optimal. The capacity of the library and array methods to detect protein-protein interactions also differed. Only 8% of the total proteins used in the library screen identified interactors, as compared to 45% in the array screen. This discrepancy may be accounted for by the library screen's more limited ability to detect PPIs that result in reduced growth rate or mating ability (8), and may also reflect the lower reproducibility of positives identified by the array method. The library method's redeeming feature, however, is that it is highly amenable high-throughput screening. In comparison to library-based screening, a purely array-based screen is much more labor and time intensive, even with automation.

Over 5600 interactions, involving approximately 70% of yeast proteins, were detected by these initial high-throughput screens of *S. cerevisiae* (24,52,53). As this set of interactions only constitutes a small portion of the 37,800–75,500 estimated interactions in budding yeast, much work remains to be done (57–60). The insights gleaned from these reported interactions can be classified into several categories. First, the screens detected interactions of functionally unclassified proteins with proteins of known function, placing the former into the functional context of the latter. For example, one screen identified a function at the spindle pole body for one previously uncharacterized protein (16). Second, proteins involved in two separate pathways, such as autophagy and cytoplasm-to-vacuole targeting, were shown to interact, urging further experiments to detect pathway cross-talk (53). Finally, novel interconnections discovered between members of the same functional pathway, such as cell cycle progression, opened new avenues of investigation (53).

Genome-wide Y2H: complex eukaryotes

Just as the genomic sequencing of increasingly complex organisms prompted the evolution of the Y2H system to high-throughput scale, the scaled-up Y2H system enabled the PPI-screening of proteomes of increasing size. In addition to *S. cerevisiae*, large-scale protein interaction screens have been conducted for several viruses and bacteria, *Plasmodium falciparum*, *Caenorhabditis elegans*, *Drosophila melanogaster*, and most recently, *Homo sapiens* (44–51).

Focusing on the results addressing humans, Rual *et al.* used a modified mini-library Y2H screen to test for PPIs amongst 7,200 ORF-encoded proteins, or approximately 10% of the comprehensive human proteome (49). This screen detected roughly 2,800 interactions, comprising, the authors posited, ~1% of the human protein interactome (i.e., the set of possible potential interactions). As a testament to the steady improvements in methodology, 78% of the detected interactions were verified by an independent co-affinity purification assay. Stelzl *et al.* also employed a mini-pooling strategy to screen an array comprised of over 5,500 human proteins derived primarily from a fetal brain cDNA library (50). This study identified a network of approximately 3,200 PPIs. Independent pull-down and coimmunoprecipitation assays of a representative sample of the interactions validated over 60% of the findings.

Surprisingly, different Y2H data sets nominally probing the same interaction pool – e.g., the human proteome – have tended to have minimal overlap. Out of the ~2,800 PPIs in the human proteome reported by Rual and the ~3,200 interactions reported by Stelzl, the two studies show concordance in merely 17 interactions (49,50). The explanation for this discrepancy partly lies in the limited clone set used in the two studies, each study encompassing less than roughly 30% of the predicted human genes. Regardless, the two groups screened approximately 1,000 proteins in common (17). Similar lack of overlap is seen when comparing the results of high-throughput Y2H screens of *S. cerevisiae* and *D. melanogaster* to other screens of the same species (24,45,53,61,62). False positive results produced by Y2H screens may partly explain the inconsistencies present in high-throughput data sets. Separately, there is likely to be a high rate of false negatives involved, based on extrapolation of results from many individual investigators. The technical reasons for potential false positives and negatives specific to Y2H have been discussed elsewhere (18): at present, it is likely that existing networks are being greatly under-sampled.

To help establish the biological relevance of their findings, Rual and coworkers demonstrated that interacting pairs of their human interactome had an increased likelihood of: gene co-expression, a common evolutionary-conserved upstream DNA sequence, mouse orthologues with a common phenotype, and annotation with common Gene Ontology (GO) terms. Similarly, Stelzl and colleagues supported the biological relevance of their interactome by demonstrating that their interactor proteins were found not only in human interaction clusters but also in model organism orthologue interaction clusters, and again shared common GO annotations. A major direction of high throughput Y2H screening is the drive to rapidly translate findings to potential insights into disease mechanisms. Both studies cross-referenced their data sets with the disease-associated genes in the Online Mendelian Inheritance in Man (OMIM) database. The data sets produced by the two studies placed over 600 disease-associated proteins into novel context, finding interactions that might offer directions for clinical research. Focusing on one pathway of known relevance to development and disease, Stelzl and colleagues performed functional analysis on two uncharacterized Wnt pathway interactors and identified them as modulators of canonical Wnt signaling.

While the focus of this review is the Y2H system, several other techniques are used to study cellular PPIs: mass spectroscopy of purified complexes (63,64), correlated mRNA expression

profiling (65,66), genetic interaction (synthetic lethality) (67,68), and *in silico* analysis (computed predictions based on gene context) (69–72). If one were to assume, in an ideal scientific world, that each technology is equally adept at detecting PPIs, one might expect overlapping interaction data obtained from studies employing different techniques. However, in a meta-analysis of the ~80,000 interactions currently described amongst yeast proteins, fewer than 5% were found to be corroborated by more than one methodology (10). In this analysis of data sets, von Mering and colleagues reported that different techniques have dissimilar abilities in detecting interactions amongst certain functional classes of proteins (10). For example, mass spectroscopy predicts notably few interactions between proteins involved in transport and sensing; a weakness owning to the difficulty in purification of these transmembrane proteins. Similarly, Y2H methodology seldom detects interactions between proteins involved in translation (10). These findings clearly demonstrate the issue of system bias as a factor in constructing global networks.

In general, the Y2H system detects binary interactions and has the capability to detect relatively weak and transient interactions, based on the overexpression of bait and prey in the nucleus, which can reach micromolar levels (19). In contrast, mass spectroscopy is able to detect more complex interactions in which the energy driving the interaction is distributed among multiple protein partners. However, this technique is biased towards detection of proteins of greater abundance and stability, based on issues with practical detection limits of instrumentation available (14,23,26). For these reasons, these differing techniques can be considered complementary, rather than all-inclusive, in constructing an interactome. Nevertheless, there is good reason to take as "gold standard" interactions that are detected by both Y2H and biochemical techniques (26,73,74).

Interpreting Signaling Networks

Already, the data sets obtained from genome-scale Y2H studies yield a number of insights into network-level cell signaling pathways and the structure of cellular PPI networks. Whilst most proteins in a network have few interactors, certain "hub" proteins interact with a great number of other proteins. As further complication, not all hubs act equivalently. Detailed analyses of protein interactome data sets subclassified hubs into two categories: date and party (75,76). Date hubs contact their interactors at varying times and locations, acting as major convergence points for multiple signaling pathways. Party hubs are defined as proteins that interact with most of their partners concurrently, forming the basis of multiprotein cellular machines. Party hubs have more restricted function than date hubs, guiding the interactions of proteins within individual pathways.

Genetic studies in lower eukaryotes have indicated that most of the proteins in a network can be mutationally inactivated without significant effects on essential signaling pathways (75, 77–79). Notably, essential genes frequently encode hub proteins, and knockout of such genes in yeast confers a three-fold rise in lethality as compared to knockout of non-hub encoding genes (23,76,78,80). In line with their function as convergence points, perturbations in party hubs in particular tend to sensitize the protein interactome to other perturbations (20). It is not surprising to realize that core components of canonical genetically-defined pathways of the pre-genomic era are being recognized today as hubs (21). In considering non-essential proteins, it is clear that natural fail-safes to circumvent a signaling 'blockade' and death due to loss of a hub include the common use of parallel network connections and occupation of key positions by protein paralogues, which buttress the robustness of cellular networks.

It is important to note that many important viral, bacterial and parasitic organisms remain to be subjected to proteome-scale analysis by Y2H and other tools. Increased understanding of the protein interactomes of human pathogens is likely to spawn numerous clinical applications.

The targeting of protein interaction hubs in microorganisms, particularly when identified hub proteins lack mammalian homologues, may prove to be a powerful strategy in intelligent drug design (18). With the currently available technology, exploring these pathogenic organisms is simple by contrast with the exploration of more complex genomes.

Many have noted the importance of caution when attempting to extrapolate the topology of the complete protein interactome from the limited data sets obtained from high-throughput Y2H assays (58,61,81). For example, the PPI maps obtained from the two largest *S. cerevisiae* (24,53) Y2H screens covered less than 10% of the entire yeast protein interactome. Interaction maps obtained from *C. elegans* and *D. melanogaster* screens also largely undersample the potential pool of interactions (45,47). While interaction data derived from other sources helps fill in the gaps, such that some estimate the yeast interactome to be 50% complete, much work remains to be done (22). As a result of such limited sampling, the topologies of Y2H-derived scale-free sub-networks cannot be used to make confident generalizations about the topology of whole cellular networks. Nevertheless, they provide an indispensable base upon which to build truly robust networks in the near future.

The next 25 years

The ultimate goal of researchers is to combine the Y2H system and other technologies to define the comprehensive structural and functional properties of the protein-protein interactome. The continued application of classic two-hybrid screening techniques will continue to yield new candidate interactors. However, description of a full interactome structure will require much more than the assembly-line-style processing of additional ORFs utilizing existing high-throughput Y2H methods, as many proteins remain refractory to use in the classic methods. The accompanying article by Suter *et al.* summarizes a number of more recently developed technologies suitable for high throughput screening that are being applied to build additional protein interaction data sets. As for Y2H and older technologies, there will likely be a significant rate of false positives, false negatives, and lack of overlap between studies with these tools: however, the set of refractory proteins with Y2H is likely to overlap only partially with the refractory proteins in the complementary platforms, allowing us to move asymptotically to a completely defined “potential” interaction landscape.

It is important to stress “potential”, rather than confirmed. First, just as high-throughput methodologies provide us with an abundance of PPI data, they also challenge us to devise high-throughput methodologies to validate and analyze such data. Bioinformatic analyses provide one increasingly important approach to evaluating PPI data, to winnow valid from likely spurious potential interactions. For example, bioinformatic analyis by the Ideker group and others have discovered that interactions that are evolutionary conserved amongst species (i.e. paralogues of interacting proteins are also known to interact) are more likely to be biologically relevant than interactions detected in a single species (22,82–87). Such an insight not only helps rank Y2H-detected interactions based on their detection in multiple species, but also allows researchers with interests in specific pathways to “model forward”, hypothesizing that orthologs of interactor pairs demonstrated in lower organisms are good candidates for interactors in mammals.

A current goal of bioinformatics researchers is to expand and optimize analytic platforms so that they can incorporate PPI data from all high and low-throughput approaches, rank these data on reliability, and integrate these data with current information on protein and gene function, expression regulation and relevance to biological processes. At present, the gold standard remains the independent confirmation and exploration of interaction significance by independent researchers committed to the study of one or more of the interacting proteins. An increasing number of PPI databases, such as the Human Proteome Reference Database (HPRD)

(88), systematically curate information from such sources. However, complementary high throughput screening techniques that do not directly assess PPIs are becoming increasingly informative for PPI screens. One particularly fertile source of such information is provided by medium and high throughput genetic studies in lower organisms. Genetic interaction studies detect synthetic lethal relationships, wherein the deletion two genes in functionally parallel pathways results in a significantly greater loss of viability as compared to the deletion of either individual gene. Progress in the field, such as the development molecular barcode arrays and increasingly comprehensive deletion mutant sets, has broadened the scope of genetic analysis. Genetic interaction studies in yeast and worms have placed many protein complexes into functional pathways, illuminating the biological significance of PPIs in lower eukaryotes and of orthologous PPIs in humans (89–91). As this and other data are merged into databases such as BioGRID (23) and BIND (93), development of apparently prosaic tools such as standardized nomenclature and reporting guidelines for interactions will be critical (24).

As combined high-throughput methods suggest many new interactions between both well-known and minimally annotated proteins, discovering the meaning of these interactions poses a monumental challenge. Currently, high-throughput-derived PPI data can report the number of connections that exist between a protein and its interaction partners, but cannot inform the functional nature of each interaction. There are many potential classes of interaction: inhibition, activation or scaffolding-facilitation, transience or stability, potency, temporal-specificity, subcellular localization-specificity, an others (18). There remain significant knowledge gaps in this area for higher eukaryotes. Hence, accurate modeling and design of networks – the goal of systems biology – is at present most effective in prokaryotes and lower eukaryotes, for which these relations are more readily described. Within 25 years, the potential for the understanding of human systems is unquestionably vast; the potential to improve the human condition, also vast.

Acknowledgments

The authors were supported by NIH RO1s CA63366 and CA113342, and by Tobacco Settlement funding from the State of Pennsylvania (to EAG); by an AACR-Littlefield award; and by NIH core grant CA-06927 and support from the Pew Charitable Fund, to Fox Chase Cancer Center. We are grateful to members of the Golemis laboratory for critical comments on the manuscript.

References

1. McCoy MS, Toole JJ, Cunningham JM, Chang EH, Lowy DR, Weinberg RA. Characterization of a human colon/lung carcinoma oncogene. Nature 1983;302:79–81. [PubMed: 6298638]
2. Schwab M, Alitalo K, Klempnauer KH, Varmus HE, Bishop JM, Gilbert F, Brodeur G, Goldstein M, Trent J. Amplified DNA with limited homology to myc cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumour. Nature 1983;305:245–248. [PubMed: 6888561]
3. Capon DJ, Chen EY, Levinson AD, Seeburg PH, Goeddel DV. Complete nucleotide sequences of the T24 human bladder carcinoma oncogene and its normal homologue. Nature 1983;302:33–37. [PubMed: 6298635]
4. Schwab M, Alitalo K, Varmus HE, Bishop JM, George D. A cellular oncogene (c-Ki-ras) is amplified, overexpressed, and located within karyotypic abnormalities in mouse adrenocortical tumour cells. Nature 1983;303:497–501. [PubMed: 6304530]
5. Fields S, Song O. A novel genetic system to detect protein-protein interactions. Nature 1989;340:245–246. [PubMed: 2547163]
6. Chevray PM, Nathans D. Protein interaction cloning in yeast: identification of mammalian proteins that react with the leucine zipper of Jun. Proc Natl Acad Sci U S A 1992;89:5789–5793. [PubMed: 1631061]

7. Chien CT, 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 U S A 1991;88:9578–9582. [PubMed: 1946372]
8. Durfee T, Becherer K, Chen PL, Yeh SH, Yang Y, Kilburn AE, Lee WH, Elledge SJ. The retinoblastoma protein associates with the protein phosphatase type 1 catalytic subunit. Genes Dev 1993;7:555–569. [PubMed: 8384581]
9. Estojak J, Brent R, Golemis EA. Correlation of two-hybrid affinity data with in vitro measurements. Mol Cell Biol 1995;15:5820–5829. [PubMed: 7565735]
10. Gyuris J, Golemis E, Chertkov H, Brent R. Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2. Cell 1993;75:791–803. [PubMed: 8242750]
11. Vojtek AB, Hollenberg SM, Cooper JA. Mammalian Ras interacts directly with the serine/threonine kinase Raf. Cell 1993;74:205–214. [PubMed: 8334704]
12. Brent R, Ptashne M. Mechanism of action of the lexA gene product. Proc Natl Acad Sci U S A 1981;78:4204–4208. [PubMed: 7027256]
13. Zervos AS, Gyuris J, Brent R. Mxi1, a protein that specifically interacts with Max to bind Myc-Max recognition sites. Cell 1993;72:223–232. [PubMed: 8425219]
14. Golemis, EA., editor. Protein-Protein Interactions. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2005.
15. Du W, Vidal M, Xie JE, Dyson N. RBF, a novel RB-related gene that regulates E2F activity and interacts with cyclin E in Drosophila. Genes Dev 1996;10:1206–1218. [PubMed: 8675008]
16. Marsolier MC, Prioleau MN, Sentenac A. A RNA polymerase III-based two-hybrid system to study RNA polymerase II transcriptional regulators. J Mol Biol 1997;268:243–249. [PubMed: 9159467]
17. Marsolier MC, Sentenac A. RNA polymerase III-based two-hybrid system. Methods Enzymol 1999;303:411–422. [PubMed: 10349657]
18. Stagljar I, Korostensky C, Johnsson N, te Heesen S. A genetic system based on split-ubiquitin for the analysis of interactions between membrane proteins in vivo. Proc Natl Acad Sci U S A 1998;95:5187–5192. [PubMed: 9560251]
19. Serebriiskii I, Estojak J, Berman M, Golemis EA. Approaches to detecting false positives in yeast two-hybrid systems. Biotechniques 2000;28:328–330. 332–326. [PubMed: 10683744]
20. Serebriiskii I, Khazak V, Golemis EA. A two-hybrid dual bait system to discriminate specificity of protein interactions. J Biol Chem 1999;274:17080–17087. [PubMed: 10358061]
21. Serebriiskii IG, Mitina OV, Chernoff J, Golemis EA. Two-hybrid dual bait system to discriminate specificity of protein interactions in small GTPases. Methods Enzymol 2001;332:277–300. [PubMed: 11305104]
22. Fields S. High-throughput two-hybrid analysis. The promise and the peril. FEBS J 2005;272:5391–5399. [PubMed: 16262681]
23. Ito T, Ota K, Kubota H, Yamaguchi Y, Chiba T, Sakuraba K, Yoshida M. Roles for the two-hybrid system in exploration of the yeast protein interactome. Mol Cell Proteomics 2002;1:561–566. [PubMed: 12376571]
24. 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. [PubMed: 11283351]
25. Mrowka R, Patzak A, Herzel H. Is there a bias in proteome research? Genome research 2001;11:1971–1973. [PubMed: 11731485]
26. von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, Bork P. Comparative assessment of large-scale data sets of protein-protein interactions. Nature 2002;417:399–403. [PubMed: 12000970]
27. Fashena SJ, Serebriiskii I, Golemis EA. The continued evolution of two-hybrid screening approaches in yeast: how to outwit different preys with different baits. Gene 2000;250:1–14. [PubMed: 10854774]
28. Osborne MA, Dalton S, Kochan JP. The yeast tribrid system--genetic detection of trans-phosphorylated ITAM-SH2-interactions. Biotechnology (N Y) 1995;13:1474–1478. [PubMed: 9636306]

29. Osborne MA, Zenner G, Lubinus M, Zhang X, Songyang Z, Cantley LC, Majerus P, Burn P, Kochan JP. The inositol 5'-phosphatase SHIP binds to immunoreceptor signaling motifs and responds to high affinity IgE receptor aggregation. J Biol Chem 1996;271:29271–29278. [PubMed: 8910587]
30. Park YW, Wilusz J, Katze MG. Regulation of eukaryotic protein synthesis: selective influenza viral mRNA translation is mediated by the cellular RNA-binding protein GRSF-1. Proc Natl Acad Sci U S A 1999;96:6694–6699. [PubMed: 10359774]
31. Putz U, Skehel P, Kuhl D. A tri-hybrid system for the analysis and detection of RNA--protein interactions. Nucleic acids research 1996;24:4838–4840. [PubMed: 8972875]
32. SenGupta DJ, Zhang B, Kraemer B, Pochart P, Fields S, Wickens M. A three-hybrid system to detect RNA-protein interactions in vivo. Proc Natl Acad Sci U S A 1996;93:8496–8501. [PubMed: 8710898]
33. Wang ZF, Whitfield ML, Ingledue TC 3rd, Dominski Z, Marzluff WF. The protein that binds the 3' end of histone mRNA: a novel RNA-binding protein required for histone pre-mRNA processing. Genes Dev 1996;10:3028–3040. [PubMed: 8957003]
34. Vidal M, Brachmann RK, Fattaey A, Harlow E, Boeke JD. Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions. Proc Natl Acad Sci U S A 1996;93:10315–10320. [PubMed: 8816797]
35. Vidal M, Braun P, Chen E, Boeke JD, Harlow E. Genetic characterization of a mammalian protein-protein interaction domain by using a yeast reverse two-hybrid system. Proc Natl Acad Sci U S A 1996;93:10321–10326. [PubMed: 8816798]
36. Huang J, Schreiber SL. A yeast genetic system for selecting small molecule inhibitors of protein-protein interactions in nanodroplets. Proc Natl Acad Sci U S A 1997;94:13396–13401. [PubMed: 9391035]
37. Vidal M, Endoh H. Prospects for drug screening using the reverse two-hybrid system. Trends Biotechnol 1999;17:374–381. [PubMed: 10461184]
38. Young K, Lin S, Sun L, Lee E, Modi M, Hellings S, Husbands M, Ozenberger B, Franco R. Identification of a calcium channel modulator using a high throughput yeast two-hybrid screen. Nat Biotechnol 1998;16:946–950. [PubMed: 9788351]
39. Consortium. Finishing the euchromatic sequence of the human genome. Nature 2004;431:931–945. [PubMed: 15496913]
40. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, et al. Life with 6000 genes. Science (New York, N.Y 1996;274(546):563–547.
41. McPherson JD, Marra M, Hillier L, Waterston RH, Chinwalla A, Wallis J, Sekhon M, Wylie K, et al. A physical map of the human genome. Nature 2001;409:934–941. [PubMed: 11237014]
42. Consortium. Genome sequence of the nematode C. elegans: a platform for investigating biology. Science (New York, N.Y 1998;282:2012–2018.
43. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, et al. The genome sequence of Drosophila melanogaster. Science (New York, N.Y 2000;287:2185–2195.
44. Bartel PL, Roecklein JA, SenGupta D, Fields S. A protein linkage map of Escherichia coli bacteriophage T7. Nat Genet 1996;12:72–77. [PubMed: 8528255]
45. Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao YL, Ooi CE, et al. A protein interaction map of Drosophila melanogaster. Science (New York, N.Y 2003;302:1727–1736.
46. LaCount DJ, Vignali M, Chettier R, Phansalkar A, Bell R, Hesselberth JR, Schoenfeld LW, Ota I, et al. A protein interaction network of the malaria parasite Plasmodium falciparum. Nature 2005;438:103–107. [PubMed: 16267556]
47. Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, Vidalain PO, Han JD, et al. A map of the interactome network of the metazoan C. elegans. Science (New York, N.Y 2004;303:540–543.
48. Rain JC, Selig L, De Reuse H, Battaglia V, Reverdy C, Simon S, Lenzen G, Petel F, et al. The protein-protein interaction map of Helicobacter pylori. Nature 2001;409:211–215. [PubMed: 11196647]
49. Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, et al. Towards a proteome-scale map of the human protein-protein interaction network. Nature 2005;437:1173–1178. [PubMed: 16189514]

50. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, et al. A human protein-protein interaction network: a resource for annotating the proteome. Cell 2005;122:957–968. [PubMed: 16169070]
51. Uetz P, Dong YA, Zeretzke C, Atzler C, Baiker A, Berger B, Rajagopala SV, Roupelieva M, et al. Herpesviral protein networks and their interaction with the human proteome. Science (New York, N.Y 2006;311:239–242.
52. Fromont-Racine M, Mayes AE, Brunet-Simon A, Rain JC, Colley A, Dix I, Decourty L, Joly N, et al. Genome-wide protein interaction screens reveal functional networks involving Sm-like proteins. Yeast 2000;17:95–110. [PubMed: 10900456]
53. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, et al. A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature 2000;403:623–627. [PubMed: 10688190]
54. Finley RL Jr, Brent R. Interaction mating reveals binary and ternary connections between Drosophila cell cycle regulators. Proc Natl Acad Sci U S A 1994;91:12980–12984. [PubMed: 7809159]
55. Ma H, Kunes S, Schatz PJ, Botstein D. Plasmid construction by homologous recombination in yeast. Gene 1987;58:201–216. [PubMed: 2828185]
56. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 1977;74:5463–5467. [PubMed: 271968]
57. Grigoriev A. On the number of protein-protein interactions in the yeast proteome. Nucleic acids research 2003;31:4157–4161. [PubMed: 12853633]
58. 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. [PubMed: 16003372]
59. Walhout AJ, Boulton SJ, Vidal M. Yeast two-hybrid systems and protein interaction mapping projects for yeast and worm. Yeast 2000;17:88–94. [PubMed: 10900455]
60. Hart GT, Ramani AK, Marcotte EM. How complete are current yeast and human protein-interaction networks? Genome biology 2006;7:120. [PubMed: 17147767]
61. Parrish JR, Gulyas KD, Finley RL Jr. Yeast two-hybrid contributions to interactome mapping. Curr Opin Biotechnol 2006;17:387–393. [PubMed: 16806892]
62. 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 biology 2004;5:R96. [PubMed: 15575970]
63. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 2002;415:141–147. [PubMed: 11805826]
64. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P, et al. Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature 2002;415:180–183. [PubMed: 11805837]
65. Cho RJ, Campbell MJ, Winzeler EA, Steinmetz L, Conway A, Wodicka L, Wolfsberg TG, Gabrielian AE, et al. A genome-wide transcriptional analysis of the mitotic cell cycle. Mol Cell 1998;2:65–73. [PubMed: 9702192]
66. Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA, Coffey E, et al. Functional discovery via a compendium of expression profiles. Cell 2000;102:109–126. [PubMed: 10929718]
67. Mewes HW, Frishman D, Guldener U, Mannhaupt G, Mayer K, Mokrejs M, Morgenstern B, Munsterkotter M, et al. MIPS: a database for genomes and protein sequences. Nucleic acids research 2002;30:31–34. [PubMed: 11752246]
68. Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, Page N, Robinson M, Raghibizadeh S, et al. Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science (New York, N.Y 2001;294:2364–2368.
69. Enright AJ, Iliopoulos I, Kyrpides NC, Ouzounis CA. Protein interaction maps for complete genomes based on gene fusion events. Nature 1999;402:86–90. [PubMed: 10573422]
70. Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, Eisenberg D. Detecting protein function and protein-protein interactions from genome sequences. Science (New York, N.Y 1999;285:751–753.

71. Overbeek R, Fonstein M, D'Souza M, Pusch GD, Maltsev N. The use of gene clusters to infer functional coupling. Proc Natl Acad Sci U S A 1999;96:2896–2901. [PubMed: 10077608]
72. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. Proc Natl Acad Sci U S A 1999;96:4285–4288. [PubMed: 10200254]
73. Huynen M, Snel B, Lathe W 3rd, Bork P. Predicting protein function by genomic context: quantitative evaluation and qualitative inferences. Genome research 2000;10:1204–1210. [PubMed: 10958638]
74. Marcotte EM, Pellegrini M, Thompson MJ, Yeates TO, Eisenberg D. A combined algorithm for genome-wide prediction of protein function. Nature 1999;402:83–86. [PubMed: 10573421]
75. Friedman A, Perrimon N. Genetic screening for signal transduction in the era of network biology. Cell 2007;128:225–231. [PubMed: 17254958]
76. Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJ, et al. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature 2004;430:88–93. [PubMed: 15190252]
77. Friedman A, Perrimon N. A functional RNAi screen for regulators of receptor tyrosine kinase and ERK signalling. Nature 2006;444:230–234. [PubMed: 17086199]
78. Jeong H, Mason SP, Barabasi AL, Oltvai ZN. Lethality and centrality in protein networks. Nature 2001;411:41–42. [PubMed: 11333967]
79. Strogatz SH. Exploring complex networks. Nature 2001;410:268–276. [PubMed: 11258382]
80. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, et al. Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science (New York, N.Y 1999;285:901–906.
81. Gandhi TK, Zhong J, Mathivanan S, Karthick L, Chandrika KN, Mohan SS, Sharma S, Pinkert S, et al. Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. Nat Genet 2006;38:285–293. [PubMed: 16501559]
82. 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. [PubMed: 12118076]
83. Kelley BP, Yuan B, Lewitter F, Sharan R, Stockwell BR, Ideker T. PathBLAST: a tool for alignment of protein interaction networks. Nucleic acids research 2004;32:W83–W88. [PubMed: 15215356]
84. Stuart JM, Segal E, Koller D, Kim SK. A gene-coexpression network for global discovery of conserved genetic modules. Science (New York, N.Y 2003;302:249–255.
85. Yu H, Luscombe NM, Lu HX, Zhu X, Xia Y, Han JD, Bertin N, Chung S, et al. Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs. Genome research 2004;14:1107–1118. [PubMed: 15173116]
86. Bandyopadhyay S, Sharan R, Ideker T. Systematic identification of functional orthologs based on protein network comparison. Genome research 2006;16:428–435. [PubMed: 16510899]
87. Suthram S, Sittler T, Ideker T. The Plasmodium protein network diverges from those of other eukaryotes. Nature 2005;438:108–112. [PubMed: 16267557]
88. Peri S, Navarro JD, Kristiansen TZ, Amanchy R, Surendranath V, Muthusamy B, Gandhi TK, Chandrika KN, et al. Human protein reference database as a discovery resource for proteomics. Nucleic acids research 2004;32:D497–D501. [PubMed: 14681466]
89. Collins SR, Miller KM, Maas NL, Roguev A, Fillingham J, Chu CS, Schuldiner M, Gebbia M, et al. Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. Nature 2007;446:806–810. [PubMed: 17314980]
90. Pierce SE, Fung EL, Jaramillo DF, Chu AM, Davis RW, Nislow C, Giaever G. A unique and universal molecular barcode array. Nature methods 2006;3:601–603. [PubMed: 16862133]
91. Zhong W, Sternberg PW. Genome-wide prediction of C. elegans genetic interactions. Science (New York, N.Y 2006;311:1481–1484.
92. Breitkreutz BJ, Stark C, Reguly T, Boucher L, Breitkreutz A, Livstone M, Oughtred R, Lackner DH, et al. The BioGRID Interaction Database: 2008 update. Nucleic acids research 2008;36:D637–D640. [PubMed: 18000002]

93. Alfarano C, Andrade CE, Anthony K, Bahroos N, Bajec M, Bantoft K, Betel D, Bobechko B, et al. The Biomolecular Interaction Network Database and related tools 2005 update. Nucleic acids research 2005;33:D418–D424. [PubMed: 15608229]

94. Orchard S, Montechi-Palazzi L, Deutsch EW, Binz PA, Jones AR, Paton N, Pizarro A, Creasy DM, et al. Five years of progress in the Standardization of Proteomics Data 4th Annual Spring Workshop of the HUPO-Proteomics Standards Initiative April 23–25, 2007 Ecole Nationale Superieure (ENS), Lyon, France. Proteomics 2007;7:3436–3440. [PubMed: 17907277]

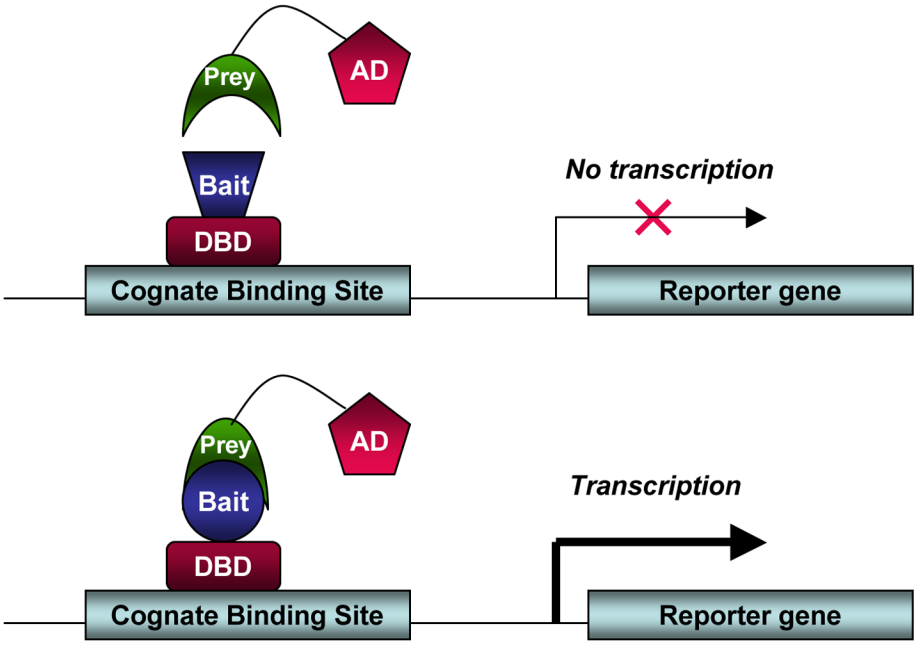

Figure 1. Detecting protein-protein interactions with the Yeast Two-Hybrid system

Protein-protein interactions are detected by the mating of two haploid complementary yeast strains each expressing a distinct expression plasmid. The first strain expresses a protein prey fused to a transcription activation domain (AD). The second strain expresses a protein bait fused to a DNA-binding domain (DBD) that binds to its cognate binding site, usually upstream of a reporter gene. If there is interaction between bait and prey, the AD is brought into the proximity of the DBD and causes transcriptional activation of the reporter gene leading to selection.

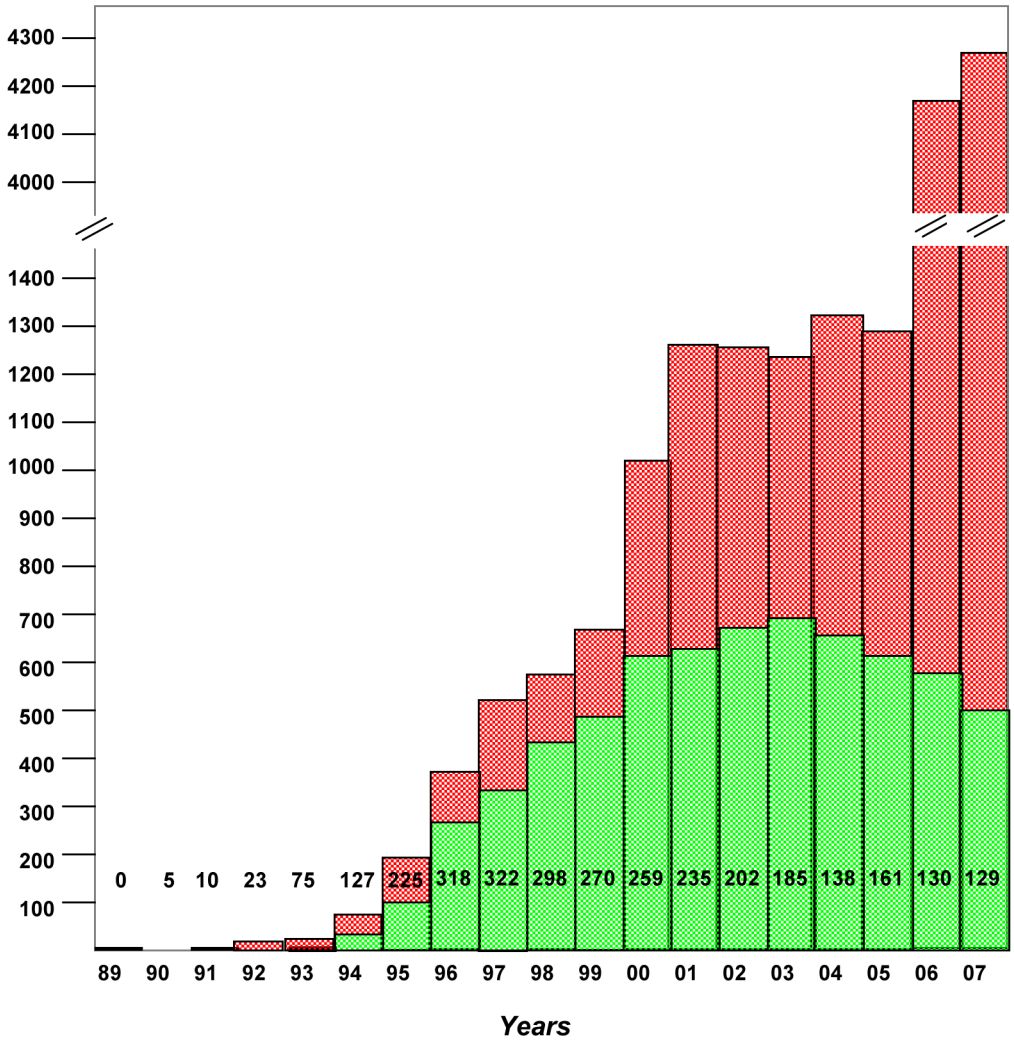

Figure 2. Three assessments of Y2H usage

Numbers appearing across graph represent citations of the seminal 1989 Fields and Song publication of a Y2H approach. Green bars represent the results recovered in each year from the Web of Science for a topic line query “yeast two-hybrid” OR “yeast two hybrid”. Red bars represent results recovered in each year from a PubMed search using the terms, yeast and two-hybrid and YEAR. We hypothesize that the divergence of trends for the last set of data versus the first two in the years 2006 and 2007 reflects the growing utilization of results from high-throughput Y2H screening projects in studies not directly employing the Y2H system.