Published in final edited form as:

Methods. 2012 December ; 58(4): 317–324. doi:10.1016/j.ymeth.2012.12.001.

A comparison and optimization of yeast two-hybrid systems

J. Harry Caufield, Neha Sakhawalkar, and Peter Uetz*

Center for the Study of Biological Complexity, Virginia Commonwealth University, PO Box 842030, 1015 Floyd Ave., Richmond, VA 23284, USA

Abstract

Two-hybrid (Y2H) assays are available in a variety of different versions, including bacterial, yeast, and mammalian systems. However, even when done exclusively in yeast, multiple different host strains, vectors, reporter genes, or protocols can be used. Here we systematically compare protein-protein interactions (PPIs) from several previously published Y2H datasets. PPIs of a human gold-standard dataset were generated by Y2H assays as well as other methods such as LUMIER or protein fragment complementation assays (PCAs). Different Y2H methods detect substantially different subsets of these PPIs, even when protocols are standardized. In order to maximize the number of interactions found and to minimize the number of false positive interactions we recommend to combine multiple vectors and protocols. While the combined results of all 18 methods detected about 92% of a gold-standard interaction set, a combination of just three Y2H assays detected up to 78% of these protein pairs, or up to 83% when a fourth assay was included. These findings indicate that three or four separate assays may be sufficient to detect the majority of protein-protein interactions in many systems.

Keywords

yeast two-hybrid; protein interactions; interactome; pDEST; pGBKT7g; pGADT7g; pGBGT7g; pGADCg; pGBKCg

1. Introduction

The yeast two-hybrid (Y2H) system has been one of the most successful experimental systems to detect and analyze protein-protein interactions [1]. Importantly, the Y2H system can be applied on a large scale to whole genomes or large sets of proteins, to the point that Y2H results have been major contributors to protein-protein interaction databases (e.g. IntAct [2]). However, the system has also been criticized for producing non-overlapping, non-reproducible results and thus an excess of false positives and false negatives [3]. While this may be true, only recent studies have attempted to benchmark various incarnations of the Y2H system. These studies have used “gold-standard” sets of interactions composed of

*Corresponding author, Fax: 804-628-6094, Tel. 804-827-4573 peter@uetz.us.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

well-studied protein interactions that can serve as “true positives”. More difficult to identify are true negative results (see below). In addition, even if such gold-standards are used, they have rarely been thoroughly investigated using several different variations of the Y2H system concurrently. An exception is the set of human gold-standard interactions described by Braun *et al.* [4] which has been studied by about 10 different Y2H variants [5]. However, there are still dozens of others [6, 7].

In this paper we attempt a comparison of various Y2H systems, primarily based on published interaction data. Many large-scale Y2H studies have been published thus far, including several genome-wide screens (e.g. [2, 8]). It remains difficult to meaningfully compare these datasets as they have been compiled with different Y2H systems, different prey libraries, or under different experimental conditions. This lack of equivalence is a source of confusion and frustration. Many studies have also tried to compare the interactomes of various species, repeatedly raising concerns over the apparent lack of overlap between datasets. This limited overlap may be due to low data quality or actual biological divergence. Alternatively, we show here that these differences may be, in part, the result of methodological differences between the various Y2H systems currently in use.

We examined three protein interaction data sets generated with multiple Y2H systems under nearly identical conditions (Table 1). More extensive analysis was performed with a positive set of human proteins as mentioned above. While this data has been available in the literature, to our knowledge no such comparisons have been attempted. We conclude that differences between datasets primarily stem from technical differences, not from the lack of reliability or reproducibility of the Y2H system *per se*.

2. Materials and Methods

2.1 Data

We used three datasets for our analysis (Table 1). The interactions among human proteins used by Braun *et al.* [4] were originally selected from detailed small-scale studies and subsequently systematically retested [4, 5]. Here we reanalyzed the raw data from the Chen dataset (i.e. images of the Y2H screens in [5]). Re-analysis resulted in slightly different numbers than were originally reported [5]. The interactions of both Varicella Zoster Virus [9] and phage lambda [10] proteins were also included in this analysis as published. Unlike many other sets of published protein-protein interactions, these datasets have been systematically generated by use of four different Y2H vectors. These vectors are listed in Table 2.

2.2 Analysis

For our comparison we counted the Y2H positives (defined as “true” positives of physiological relevance in the case of human proteins) of each dataset and for each vector pair used. Raw data from [5] was re-analyzed using images of the original screens, such that slightly different results were obtained, given the somewhat arbitrary cutoff for positives when background growth was visible in certain screens (raw data is available in the supplement of [5]). In the original Chen assays [5], each set of assays was performed in

duplicate and each interaction screen was grown in quadruplicate per plate. Here we counted all yeast colonies that grew to above background levels in at least two of four colonies per plate and on at least one of the two plates used. Data analysis was performed using the R statistical package.

2.3 Clustering

The aggregate results from each method used by Braun *et al.* and Chen *et al.* were compared by clustering to determine how similar the detected subsets of the reference set are. The results of all assays from both studies were treated as an array of 92 weighted values. Each result for a specific PPI within the PRS and RRS was treated as a single value, with positive results holding a maximum value of 1 and negative results holding a value of 0. All PPI reported by Braun *et al.* were assigned a value of 1, as the exact number of replicates performed in these assays is unclear. All PPI observed in the Chen *et al.* dataset were assigned a weighted value as follows: if a PPI was observed for all replicates at a 3-AT concentration of 0 mM, 3 mM, or 10 mM, they were assigned a value of 0.1, 0.4, or 0.5, respectively. PPI observed in only 1 of 2 replicates at the same 3-AT concentrations were assigned half of the full values, for 0.05, 0.2, or 0.25, respectively. The weighted values for all three 3-AT concentrations were added for each PPI in the PRS and RRS, such that the results for each vector combination could be treated as an aggregate of stringency and replication, with greater values for PPI observed at multiple stringency levels and in multiple replicates.

All results arrays were aligned and clustered using the PermutMatrix graphical data analysis package [11]. A tree was used to visualize the extent to which methods clustered in a pairwise fashion using the unweighted pair group method with arithmetic mean (UPGMA) and Euclidean distance to reflect similarities within the assay data.

3. Results

3.1. Performance of Y2H vectors in independent screens

Despite thousands of successful Y2H screens, it is nearly impossible to compare individual screens, given the different libraries, yeast strains, or screening conditions used. Most commonly, one or a few baits are screened against a random cDNA library, leading to variations in experimental conditions and a generally random selection of positive interactions. More carefully controlled assays are critical when transient interactions are studied, as these interactions are physically weaker and more sensitive to selection criteria. In this work, we analyze screens which used multiple Y2H vectors under very similar conditions to compare the effects of using each vector. We then focus on sets of Y2H data using identical proteins with different interaction detection assays.

3.2. N- vs C-terminal fusions

The vast majority of Y2H screens use the DNA-binding (DBD) and activation domains (AD) of yeast Gal4 fused to the N-terminus of bait and prey proteins. Recent studies incorporating C-terminal fusion vectors [9] have shown markedly different results (Fig. 1). Here we show these differences are characteristic for each series of screens (Fig. 1). For instance, while

Varicella Zoster Virus N-terminal baits and N-terminal preys (NN) produced the highest numbers of interactions, NC screens yielded the lowest number (Fig. 1A). However, with the human gold-standard set used by Braun *et al.*, Chen *et al.* revealed that all bait and prey fusions produced similar results (Fig. 1B). In even stronger contrast, in phage lambda screens [10], NN was the most productive overall but shares little overlap with NC or CN terminal fusions (Fig. 1C). When all three screens were combined, only NN screens produced significantly more PPIs than the other combinations: out of all the PPIs reported in these three studies, 59% were detected by N-N-terminal fusions while NC, CN, or CC arrangements revealed 39%, 45%, and 46% each of all interactions, respectively. The studies working with human and phage lambda proteins also used combinations of pDEST vectors, which in the former case confirmed many interactions but in the latter found very few. The threshold at which pDEST detects interactions remains unclear. Even so, these comparisons strongly suggests that N- and C-terminal vector systems should become the standard in Y2H screening to maximize the number of unique interactions obtained. Single-vector Y2H screens clearly capture only a fraction of all physiological interactions.

3.3. Stringency and the role of 3-AT

Usage of 3-aminotriazole, or 3-AT, in Y2H screens allows for differing levels of stringency when histidine-based growth selection is used, as it competitively inhibits the *HIS3* gene product. Increasing levels of 3-AT raise the minimum amount of gene expression necessary to produce yeast growth. At a 10 mM concentration of 3-AT, for example, only the strongest protein interactions of those visible at 3 mM 3-AT should be visible.

In the Chen results [5], the pGBGT7g-pGADT7g screens produced the highest number of positives (70 at the lowest stringency level, Fig. 2A). At least half of all positive interactions were visible at 10 mM 3-AT (Fig. 2B). The vector combination pGBKCg-pGADT7g produced the highest number of strong interactions but also produced the fewest overall interactions. Interestingly, at least 60% were visible at any concentration of 3-AT above 3 mM (Fig. 2C).

3.4. The role of bait-prey swapping

We have determined the number of interactions among gold-standard interactions (positive reference set, or PRS) and a random reference set (RRS) as defined by Braun *et al.* (2009) [4]. Each vector pair consists of a bait and a prey fusion protein. However, bait and prey proteins can be swapped, such that the “prey” protein is fused to the DBD and the “bait” protein is fused to the AD. Surprisingly, such bait-prey swaps can produce remarkably different interactions ([5], Fig. 3). For instance, within the positive reference set (PRS), the pGBKCg/pGADT7g vectors yielded a total of 71 and 56 interactions, respectively, when 0 and or 3 mM 3AT were used (Fig. 3A). At the higher stringency level (3 mM 3AT), 4 interactions were only visible in the original bait/prey arrangement but another 50 were detectable in both bait/prey combinations (shown as red sections in Fig. 3A). Within the random reference set (RRS) (Fig. 3B) very limited overlap between bait and prey swaps was visible for any vector combination, especially at higher stringency levels, demonstrating the non-reproducible nature of these “interactions”. The nature of the fusion proteins (i.e. N- vs C-terminal) may dramatically affect the outcome of a Y2H assay, even when the same

vectors are used and the “same” proteins are fused (i.e. the same DBD, AD, bait, and prey proteins). Overall, the numbers of interactions seen in both bait/prey arrangements using PRS proteins are not seen with RRS proteins, even when RRS proteins produced multiple interactions in single arrangements. Only one interaction was reproducible when bait and prey swaps were analyzed at 3AT concentrations above 3 mM (Fig. 3B, right panel).

3.5. Signal to noise ratio

Maximizing the signal to noise ratio for a particular assay is critical to maximizing the number of true positive results. We compared the number of interactions among the PRS and the RRS to determine this ratio (Fig. 4). Interactions detected for the PRS are considered to be true interactions while those detected within the RRS are considered to be false positives (FP). No vector set produced more than 10 FPs (or about 11 percent) when XY and YX PPIs were combined (red sections in Fig. 3B). For the same interactions assayed using 3 mM 3-AT, only one false positive was observed in both XY and YX for pDEST. Notably, while pGBGT7g-pGADCg and pGBGT7g-pGADT7g produced the highest number of interactions, they also produced the highest number of false positives. The best overall signal-to-noise ratios were generated by pGBKCg-pGADCg and pGBKCg-pGADT7g which yielded no reproducible false positives at 3 mM 3-AT or higher.

3.6. How do different PPI detection systems compare?

While it has been known for some time that different assays for protein-protein interactions yield different results, few comparisons concern different systems *and* directly comparable datasets (e.g. [4, 5, 9, 12]). Here we go beyond a simple pairwise comparison and use a clustering approach to compare all methods that have been applied to the Braun *et al.* gold-standard dataset (Fig. 5). Two methods may detect similar percentages of the interactions within this dataset yet these methods detect different subsets of the total set of all possible interactions. Clustering provides the benefit of going beyond sums of interaction results in that it compares the patterns of results, revealing further differences between experimental methods.

When the available PRS data is reduced to binary decisions regarding whether an interaction is visible (i.e. without considering 3-AT concentrations), the results are striking: our Y2H results and the Braun results each cluster together very consistently except for the pDEST vectors (which were also used by Braun *et al.*). Not surprisingly, the Braun Y2H assays with the same vectors – but different reporters – clustered together, with the 2-reporter assays simply producing fewer interactions [4]. Our Y2H assays were notable as bait/prey swaps typically clustered together too, e.g. in the pDEST, pGBGT7-pGADT7, and pGBGT7-pGADC assays, but not in the pGBKC-pGADC/pGADT7 cases (Fig. 5). This is insofar surprising, as even bait/prey swaps usually result in distinctly different interaction patterns, so that this result is not immediately obvious when examining the raw data. Overall, these results indicate that each method may detect different subsets of interactions within the same set of protein pairs, especially when multiple sets of Y2H results are compared.

3.7. An optimized strategy for PPI detection and discovery

The data used for Figures 1 and 5 allow us to devise an optimization strategy to detect a maximum number of interaction with a minimal number of assays. For the PRS, combined data from the 10 Y2H assays performed by Chen *et al.* and the 8 assays performed by Braun *et al.* – including 4 Y2H assays – appeared to detect 85 of the total of 92 previously confirmed protein interactions, for about 92% coverage of the dataset. The different subsets of interactions produced by each assay suggested that fewer than 18 independent assays should be necessary to detect this number of unique interactions. On average, six of the assays could detect each PRS interaction (when low stringencies were included), though 9 of the interactions were detected by just one assay and 14 interactions were found with only two assays. Such difficult-to-detect interactions were not consistently detected by the same assays. Within the RRS, all assays detected a total of 21 interactions, more than half of which were visible in only one assay.

We counted the number of interactions each assay pair, triplet, and quartet produced and ranked these combinations according to the number of non-overlapping interactions found. For *pairs* of assays the highest possible number of unique interactions, 63, was produced by the Braun Y2H system (Y strain, 2 μ , 1 reporter; pVV212/pVV213) combined with pGBKCg-Y-pGADT7g-X. The least productive binary pair was pDEST22-X-pDEST32-Y and pDEST32-X-pDEST22-Y (i.e. a bait/prey swap), resulting in only 20 interactions. However, the latter combination may reflect the small total number of interactions found with this more stringent vector pair in the first place; on their own, the pDEST vector pairs only detected 12 and 15 interactions, respectively.

Similarly, optimal combinations of *more than two* assays were determined (Fig. 6). For sets of three methods, the combinations finding the greatest number of interactions within the PRS detected nearly 80% of the dataset (Fig. 6). Sets of four methods detected, at most, slightly more than 80% of the set (Fig. 6). At the lower end of the spectrum, combinations of three or four assays detected 40% of the PRS or less. These assays may detect fewer interactions overall, but like the pDEST vector pairs they may offer greater stringency and may eliminate many weaker interactions.

The specific assays composing the ten highest-coverage and lowest-coverage combinations are also shown in Fig. 6. The combination of three methods including Braun Y2H (Y strain, 2 μ , 1 reporter), pGBGT7g-X-pGADCg-Y, and pGBKCg-Y-pGADT7g-X was the best 3-method combination in terms of the number of PRS interactions – 72 unique interactions were observed with this set, or more than 78% of the PRS. When Braun Y2H (MaV strain, CEN, 2 reporters), pDEST32-X - pDEST22-Y, and pDEST22-X-pDEST32-Y are used, only 31 PPIs are detected. Nearly all of the highest-coverage combinations of four assays include the LUMIER assay as performed by Braun *et al.* On its own, this method detects less than 36% of the PRS but detects many interactions less frequently detected by other assays. Notably, as the PRS is composed of human proteins and LUMIER requires use of mammalian cells, these results may vary for non-mammalian proteins.

It will be more difficult to find an assay to verify interactions generated by an independent protocol, given that the goal here is to maximize *overlaps* rather than to maximize *non-*

overlaps. While different Y2H assays produce almost as different results as entirely different assays, and thus could arguably serve as tools for verifying independent Y2H results, some may argue that independent protocols may be more appropriate as they are “truly independent”. Our results suggest the best “independent” type of verification for each assay (Fig. 6). Researchers can thus optimize their time and resources to produce a maximum number of unique interactions with the fewest assays.

4. Discussion

4.1. Y2H vectors and protein fusions

It has long been known that different Y2H systems yield different results, even when applied to the same proteins or libraries. For instance, Fromont-Racine *et al.* [13] screened the two yeast spliceosomal proteins Lsm2 and Lsm8 in two different bait vectors each – as Gal4 and LexA fusions – against a yeast genomic library, yielding dramatically different results. The two Lsm2 screens yielded 33 and 13 interactions, respectively, of which only 2 were shared among Gal4 and LexA baits. Unfortunately, very few attempts have been undertaken to understand the differences among Y2H systems on a mechanistic level, so the molecular basis for these differences remains unclear.

Interestingly, NN interactions seem to be overrepresented in several datasets (Fig. 1). It is possible that the PPI literature overall is systematically biased towards NN interactions, given that a many studies traditionally use N-terminal fusions not just for Y2H assays but also for other purposes such as protein purification (e.g. using glutathione-S-transferase, GST) or detection (e.g. green fluorescent protein, GFP).

4.2. 3-AT and stringency

Our analysis of the raw data published by Chen *et al.* [5] found a number of discrepancies between Chen *et al.* and this study. Almost all of the discrepancies appeared to be due to arbitrary interpretations of what constitutes a real positive result or a background result. There are often gray zones of weak positives that may be false positives above background but do not stand out as true positives. This problem can be alleviated by increasing the concentration of 3-AT to suppress background activity of the *HIS3* reporter gene, hence increasing the stringency of the assay and reducing its sensitivity to weaker interactions. However, higher 3-AT concentrations may also suppress true but weak interactions. A range of 3-AT concentrations provides for selection of the ideal stringency level. A concentration resulting in at least one or a few interactions but no visible background is generally sufficient, though this threshold needs to be decided on a case-by-case basis.

4.3. The role of bait-prey swapping

Confirming an interaction with bait and prey swapping may be one way to separate true positives from false positives, as a pair of interacting proteins often interacts in bait/prey swaps while false-positives typically do not. If a pair of seemingly interacting proteins appears to fail the bait/prey swap test, there is no guarantee but rather a higher probability that the protein pair represents a false-positive result. Working under the assumption that PRS proteins produce true positive interactions, a majority of these true PPI were visible in

both bait/prey configurations on average across all 3-AT concentrations (Fig. 3A). Within the RRS, about 12% of interactions were visible on average in both configurations across all 3-AT concentrations. Screens of the RRS using 3-AT concentrations of 3 mM or above yielded little to no overlap between swaps (Fig. 3B). When non-related vector pairs are compared (i.e. pairs that do not share any of their main characteristics such as markers), the overlap is not dramatically different, e.g. ranging from 14.3% to 36.8 % for the pDEST32/pDEST22 pair compared to other pairs (Fig. 3, 5). Interestingly, Braun *et al.* [4] did a number of bait/prey swap experiments for most of their assays, including MAPPIT and Y2H. In addition, they conducted experiments where N- and C-terminal swaps were used, including wNAPPA, LUMIER, and PCA experiments (see Suppl. Fig. 5 in [4]). However, they have not published any details regarding which configuration produced which interaction so further dataset comparisons are unavailable.

4.4. Laboratory effects on interaction screens

The failure to produce similar protein interaction results by independent assays may not be fully explained by differences in methodology. Rather, unreported differences between laboratory protocols may contribute to the differences between results. The Y2H system may be sufficiently sensitive to otherwise disregarded differences (i.e., in plasmid copy number, gene expression, or growth conditions) to produce noticeably different sets of predicted protein interactions. Braun *et al.* specify the genotypes of yeast strains used in their analyses but the composition of their Y2H vectors (pVV212/pVV213) remains unclear; the work cited regarding these vectors [14] offers no further detail of their structure or sequence. This may be the result of the all-too-common difficulties with consistent plasmid naming. The Braun *et al.* study and the work by Chen *et al.* [5] also differ in the yeast strains used, the reporters used (multiple reporters were used in tandem rather than individually), and the concentrations of 3-AT used. Final observation thresholds may also differ, as growth phenotypes are commonly determined by a subjective viewer rather than in an objective manner (especially when background growth is present). Systematic methods of monitoring yeast growth or gene expression do exist, including simple reporters such as the *LacZ* gene. Any or all of these experimental differences may contribute to non-identical results for otherwise similar Y2H assays.

The massive sets of data produced by any high-throughput method can magnify otherwise undetectable differences in physical properties or methodology. The background noise contributing to false-positive or false-negative results can become more apparent when large screens are performed, especially when the signal-to-noise ratio remains small. These large batches of results do highlight one of the primary advantages of high-throughput methods: the chances of producing reproducible results are increased. A single protein interaction may not be reproducible, but a large set of interactions may produce useful and reproducible material for other researchers to study further.

The phenomenon of finding different results in otherwise similar studies is certainly not unique to yeast two-hybrid assays. The large-scale mass spectrometry studies performed by Hu *et al.* (2009) [15] and Arifuzzaman *et al.* (2006) [16] serve as just another example. Both of these studies attempted to define physical interactions between nearly the complete

proteome of *E. coli* K-12. Despite using the same protein-coding genes from the same bacterial species and strain, these studies found 5,993 and 2,667 protein interactions respectively, with less than 1% overlap in their inferred PPIs. The lack of similarity may suggest that large studies should not underestimate the complexity of the systems they study.

4.5. Towards a mechanistic understanding of Y2H results

Y2H results suggest interactions between protein pairs, but eventually these interactions need to be explained in structural terms, especially when N- and C-terminal fusion proteins yield different results. For this study, we have not attempted to model all available protein structures and their complexes as this approach is technically demanding. The effect of fusions on protein structure and interactions remains unclear. In addition, we need to take into account how such an approach would require knowledge of the structure of interacting bait and prey proteins bound to the reporter gene promoter as well as the orientation and interaction of the complex with the transcriptional machinery. To our knowledge, modeling of Y2H interactions in structural terms has not been attempted but this will clearly remain a challenge for the future.

5. Conclusions and recommendations

Yeast two-hybrid assays have the potential to reveal most of an organism's protein-protein interactions, no matter whether a full proteome or subset of proteins is investigated. Numerous factors contribute to the final results obtained through Y2H assays, many of which appear to be independent of the proteins in each screen. This study shows that the vectors used in each assay and the configuration of their AD or DBD domains are critical to obtaining comprehensive and consistent results. If Y2H assays are carried out under non-identical conditions with the same plasmids they will produce non-identical results, especially when performed by and interpreted by different laboratories. This phenomenon is true for assays other than Y2H but the structural factors involved in screening for protein interactions emphasize the importance of consistent experimental design in two-hybrid screens.

Here we show how Y2H assays can be combined to maximize detection of positive results while limiting incidence of false positive results. Researchers working with sets of less well-studied proteins should perform three or four Y2H assays with different vectors, ideally concurrent with bait/prey swaps and several levels of stringency. 3-AT concentrations of 3 mM and 10 mM are sufficient to eliminate most false positive results. Further analysis with the LUMIER assay may also help to maximize results and confirm many of those detected by Y2H. As has been recommended previously [17], use of pilot studies with positive and negative controls will save time and resources for any Y2H study. Researchers screening for protein-protein interactions need not perform every assay available – they should simply use a few carefully selected methods to maximize results obtained during their time invested.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Early stages of this project were supported by NIH grant RO1GM79710.

Abbreviations

3-AT	3-aminotriazole
AD	Activation domain
DBD	DNA-binding domain
LUMIER	Luminescence-based mammalian interactome mapping
PCA	protein fragment complementation assays
PPI	protein-protein interaction
PRS	positive reference set
RRS	random reference set
VZV	Varicella Zoster Virus
Y2H	yeast two-hybrid

References

1. Fields S, Song O. Nature. 1989; 340:245–246. [PubMed: 2547163]
2. Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, Duesbury M, Dumousseau M, Feuermann M, Hinz U, Jandrasits C, Jimenez RC, Khadake J, Mahadevan U, Masson P, Pedruzzi I, Pfeifferberger E, Porras P, Raghunath A, Roechert B, Orchard S, Hermjakob H. Nucleic Acids Res. 2012; 40:D841–D846. [PubMed: 22121220]
3. Scott MS, Barton GJ. BMC Bioinformatics. 2007; 8:239. [PubMed: 17615067]
4. Braun P, Tasan M, Dreze M, Barrios-Rodiles M, Lemmens I, Yu H, Sahalie JM, Murray RR, Roncari L, de Smet AS, Venkatesan K, Rual JF, Vandenhaute J, Cusick ME, Pawson T, Hill DE, Tavernier J, Wrana JL, Roth FP, Vidal M. Nat Methods. 2009; 6:91–97. [PubMed: 19060903]
5. Chen YC, Rajagopala SV, Stellberger T, Uetz P. Nature Methods. 2010; 7:667–668. [PubMed: 20805792]
6. Bartel, PL., Chien, C-T., Sternglanz, R., Fields, S. Cellular Interactions in Development: A Practical Approach. Hartley, DA., editor. Oxford: Oxford University Press; 1993. p. 153-179.
7. Fu, H. Protein-Protein Interactions. Methods and Applications. Totowa, NJ: Humana Press; 2004.
8. Bouveret E, Brun C. Methods Mol Biol. 2012; 804:15-3. [PubMed: 22144146]
9. Stellberger T, Häuser R, Baiker A, Pothineni VR, Haas J, Uetz P. Proteome Sci. 2010; 8:8. [PubMed: 20205919]
10. Rajagopala SV, Casjens S, Uetz P. BMC Microbiol. 2011; 11:213. [PubMed: 21943085]
11. Caraux G, Pinloche S. Bioinformatics. 2005; 21:1280–1281. [PubMed: 15546938]
12. Rajagopala SV, Hughes KT, Uetz P. Proteomics. 2009; 9:5296–5302. [PubMed: 19834901]
13. Fromont-Racine M, Mayes AE, Brunet-Simon A, Rain JC, Colley A, Dix I, Decourty L, Joly N, Ricard F, Beggs JD, Legrain P. Yeast. 2000; 17:95–110. [PubMed: 10900456]
14. Vidal M, Brachmann RL, Fattaey A, Harlow E, Boeke JD. PNAS. 1996; 93:10315–10320. [PubMed: 8816797]
15. Hu P, Janga SC, Babu M, Díaz-Mejía JJ, Butland G, Yang W, Pogoutse O, Guo X, Phanse S, Wong P, Chandran S, Christopoulos C, Nazarians-Armavil A, Nasseri NK, Musso G, Ali M, Nazemof N,

- Eroukova V, Golshani A, Paccanaro A, Greenblatt JF, Moreno-Hagelsieb G, Emili A. PLoS Biol. 2009; 7:e96. [PubMed: 19402753]
16. Arifuzzaman M, Maeda M, Itoh A, Nishikata K, Takita C, Saito R, Ara T, Nakahigashi K, Huang HC, Hirai A, Tsuzuki K, Nakamura S, Altaf-Ul-Amin M, Oshima T, Baba T, Yamamoto N, Kawamura T, Ioka-Nakamichi T, Kitagawa M, Tomita M, Kanaya S, Wada C, Mori H. Genome Res. 2006; 16:686–691. [PubMed: 16606699]
17. Häuser R, Stellberger T, Rajagopala SV, Uetz P. Methods Mol Biol. 2012; 812:1–20. [PubMed: 22218851]

Figure 1. N- and C-terminal proteins in Y2H screens

Three studies using the same set of vectors expressing different sets of proteins were compared. For each set of results, absolute numbers of interactions with each bait/prey orientation are plotted against the fusion combination used. For instance, NN denotes constructs with N-terminally fused DBD and AD domains while NC denotes an N-terminal DBD fusion (bait) interacting with a C-terminal AD fusion (prey). The Venn diagram below each histogram shows the overlap among vector pairs. For each plasmid configuration, the total number of PPIs found using that configuration and the percentage of all PPIs found using this configuration are provided in parentheses. Raw data reported in three publications using different data sets were used to determine totals. (A) Varicella-Zoster-Virus [9]. (B) Human gold-standard interactions (the positive reference set, PRS [4]). Raw data is from [5] but differs from published counts due to re-counting in this study. (C) Bacteriophage lambda [10]. (D) Total numbers combined from (A–C). Vectors using the pDEST backbone were also used in the data presented in parts B and C. These results are included in the Venn diagrams in red and within brackets as these vectors differ significantly in composition and sequence from the other Y2H constructs used. For example, part D indicates that 39 PPI were visible in all four vector configurations and an additional 24 PPI were found with pDEST vectors.

Figure 2. The role of 3-AT in Y2H screens of human gold-standard interactions

(A) The number of PPI observed at three different stringency levels is shown. Stringency is indicated as 3-AT concentration, for the same PPI at 0 (yellow), 3 (redorange) or 10 mM 3-AT (blue). Values are stacked to indicate redundancy; for instance, many PPIs visible at 0 mM were not visible at 10 mM. (B) Percentage of interactions detected at the highest stringency level, 10 mM 3-AT. For instance, 77% of all interactions detected with pGBGT7g-X and pGADCg-Y were found at high stringency (10 mM 3-AT) but the remainder were only visible at 3 mM 3-AT or less. (C) As in (A) but including 3-AT concentrations of 3 mM and higher.

Figure 3. Bait and prey swaps in Y2H interactions of human PRS proteins

Bait (“X”) and prey proteins (“Y”) may be swapped among their vectors, such that both interacting proteins are fused to both the DBD and the AD. These bait and prey screening configurations are denoted as XY and YX, respectively. Interactions that are common to both configurations are indicated in red. Interactions found in only one combination (XY or YX) are shown in blue and yellow, respectively. The X-axis shows absolute numbers of interactions. (A) PRS results. The left panel shows all data. The right panel shows only PPIs detected at 3 mM 3-AT or higher. (B) RRS results. The left panel shows all data. The right panel shows only PPIs detected at 3 mM 3-AT or higher. Based on raw data from [5]. For explanation of PRS see [4].

Figure 4. Signal to noise ratios in Y2H assays of human PRS interactions

For each vector pair the absolute number of positive PRS and RRS [4] interactions is shown. Bait/prey swaps were combined. False positive (FP) counts, shown in green denote counts of false-positive interactions (i.e. those detected within the RRS). PPI counts, shown in blue denote PPIs detected within the PRS. The ratio of PPI to false positives is indicated for all assays yielding at least one false positive at more than one stringency level. Based on raw data from [5].

Figure 5. Clustering of PPI detection assays

The assays used to detect each protein pair in the (A) PRS and (B) RRS [4] are clustered by the number and similarity of the interactions detected across the full reference set. For Braun et al. assays, columns denote whether a specific protein interaction was reported. For all other assays, values are weighted values as described in Methods, with increasing brightness indicating greater value. Black spaces indicate that no interaction was detected. The number of each protein pair in each set is visible at the top of each grid. Total set coverage of each assay at all stringency levels, in percent, is shown at the right of the figure. For details of the clustering algorithm see Methods.

Figure 6. Optimizing the combination of multiple PPI detection assays

PPIs detected by single assays or vectors and by combinations of two, three or four assays (or vectors) are shown, ranked by the percentage of PRS PPI found from high (left) to low percentage (right). Single assays are shown in orange and represent the 18 assays shown in Fig. 5. Pairs of assays are shown in green. Blue indicates combinations of three assays while red indicates combinations of four assays. The results of each combination are presented together for clarity; the total number of combinations in each set (that is, the data on the x-axis) is non-equal. Data are from [4] and [5]. The ten highest-coverage and lowest-coverage combinations of three and four PPI detection assays are shown in the lower part of the figure (left and right, respectively). Assays present in combinations of three are shown in blue, assays present in combinations of four are shown in red, and assays present in combinations of three and four are shown in purple. Names of yeast two-hybrid constructs encoding C-terminal protein fusions are highlighted in red. Names of specific assays present in each combination are shown in Supplementary Data.

Table 1

Datasets used in this study.

Species	Interactions	Vector pairs	Ref.
Varicella Zoster virus	348 (nr)	2 (4 vectors total)	[9]
Human	92	5	[5]
Phage lambda	97	2 (4 vectors total)	[10]

nr = non-redundant with respect to open reading frames; Stellberger *et al.* 2010 list more interactions but the published list is partially redundant (e.g. listing PPIs involving full-length proteins and fragments thereof separately).

Table 2

Y2H vectors used in the three compared studies

See Table 1 for details of datasets and sources. Baits contain DNA-binding domains (DBD) and preys contain activation domains (AD) as used in [5, 9, 10]. Yet other vector variants (such as pLP-GADT7, pAS1-LP etc.) have been used and described in [12, 14]

		Gal4-Fusion			Selection			
Vector	Promoter	N / C	AD/DBD	yeast	bacterial	ori	Source	
pDEST22	fl-ADH1	N	AD	Trp1	Ampicillin	CEN	Invitrogen	
pDEST32 *	fl-ADH1	N	DBD	Leu2	Gentamicin	CEN	Invitrogen	
pGBKT7g	t-ADH1	N	DBD	Trp1	Kanamycin	2μ	[15]	
pGADT7g	fl-ADH1	N	AD	Leu2	Ampicillin	2μ	[15]	
pGBGT7g	t-ADH1	N	DBD	Trp1	Gentamicin	2μ		
pGADCg	fl-ADH1	C	AD	Leu	Ampicillin	2μ	[9]	
pGBK Cg	t-ADH1	C	DBD	Trp	Kanamycin	2μ	[9]	

* also encodes CYH2; fl-, t-ADH1 = full length and truncated ADH1 promoters. The bacterial origin in all cases is from pUC (= ColE1). The pDEST, pGBKT7g, and pGADT7g vectors are Gateway-compatible (as indicated by “g”).