Methods 58 (2012) 317–324

Contents lists available at SciVerse ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

A comparison and optimization of yeast two-hybrid systems

J.H. Caufield, Neha Sakhawalkar, Peter Uetz *

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

ARTICLE INFO

Article history:
Available online 8 December 2012
Communicated by Uetz

Keywords:
Yeast two-hybrid
Protein interactions
Interactome
pDEST
pGBKT7g
pGADT7g
pGBGT7g
pGADCg
pGBKCg

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.

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 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

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.

* Corresponding author. Fax: +1 804 628 6094.
E-mail address: peter@uetz.us (P. Uetz).

1046-2023/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ymeth.2012.12.001

Table 1
Datasets used in this study.

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

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 PPIs reported by Braun et al. were assigned a value of 1, as the exact number of replicates performed in these assays is unclear. All PPIs 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, 3, 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

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]
pGBKCg	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").