REGULAR ARTICLE

Repeated two-hybrid screening detects transient protein–protein interactions

Arunachalam Vinayagam · Ulrich Stelzl ·
Erich E. Wanker

Abstract Yeast two-hybrid (Y2H) screening is a powerful method to detect protein–protein interactions (PPI) at the genomic-scale. A recently proposed framework for binary interactome mapping recommends the repeated screening approach to improve the quality of PPI data. Such repeated screening reveals Y2H interactions ranging from highly sampled to singleton interactions. The quality and the biological significance of interactions from distinguished sampling classes remain unknown. In order to systematically characterize such interactions, we have chosen a dataset of 1,262 interactions that were screened repeatedly four-times. The interactions were classified as highly sampled, weakly sampled, and singleton interactions. We assessed the quality of interactions in different sampling classes using features such as protein structural properties, conservation in yeast and presence of known domain–domain interactions that are previously associated with false positive rates. Our analysis reveals that the quality of singleton interactions is comparable to that of highly sampled interactions. Interestingly, singletons encompass a higher fraction of known domain–domain interactions than highly sampled ones. Furthermore, we observed that the singleton interactions are transient in nature, while the highly sampled interactions are predominantly part of stable complexes. Hence, the repeated Y2H screening method is ideal for detecting transient PPIs that are crucial in cellular signaling pathways.

Keywords Protein–protein interaction network ·
Repeated Y2H screening · Y2H sampling ·
Transient interactions

1 Introduction

Protein–protein interactions (PPIs) are involved in almost every process of a living cell. Mapping interactions therefore provides the basis to understand how proteins function and communicate with each other. A number of human diseases arise because of aberrant PPIs. Hence, PPI maps are crucial to understand the molecular mechanisms behind disease pathogenesis and important for finding new therapeutic targets [1, 2]. Recently, genome-wide interactome maps of proteins from human and other model organisms have been created using a variety of high-throughput technologies such as yeast two-hybrid assays (Y2H) [3–9] and affinity purification chromatography followed by mass spectrometry (AP/MS) [10–12]. Although Y2H and AP/MS data are of equally high quality, they interrogate different subspaces within the whole interactome and are complementary to each other [13].

The low overlap between the interactions identified by different Y2H experiments raises concern about the suitability of the assay to build interactome maps [14]. Even the high-confidence datasets from two independent Y2H experiments have only a small fraction of interactions in

A. Vinayagam (✉) · E. E. Wanker (✉)
Max Delbrueck Center for Molecular Medicine (MDC),
AG Neuroproteomics, Robert-Rössle-Str. 10,
13125, Berlin-Buch, Germany
e-mail: vinu@mdc-berlin.de

E. E. Wanker
e-mail: erich.w@mdc-berlin.de

U. Stelzl
Max Planck Institute for Molecular Genetics (MPI-MG),
Otto-Warburg Laboratory, Ihnestraße 63-73,
14195 Berlin, Germany

common [4, 9]. It has been argued that the large number of false-positives and false-negatives within the PPI data sets are responsible for the low overlap [15]. However, a recent study has shown that the fraction of false-negatives is the primary cause for this phenomenon [13]. To overcome this problem, a framework for Y2H interactome mapping has been proposed [16]. This strategy recommends performing multiple Y2H screens to report all detectable interactions. Such repeated interaction screening improves the overlap between different datasets and increases the interactome coverage [13]. For instance, to identify 90% of all Y2H detectable interactions, at least six repeated Y2H screens are necessary [16]. Importantly, in such an approach, one repeat screen refers to a procedure involving screening and retesting of interaction pairs, a requirement to remove various artifactual and spurious hits.

At low repeat numbers, where each repeat detects still a substantial number of novel interactions, the repeated Y2H screening yields different types of interactions such as highly sampled interactions (that are found in almost all screens), weakly sampled interactions (found in two experiments) and singletons (interactions detected only in a single experiment). The quality and the biological properties of these differently sampled interactions remain unknown. Previous studies have estimated the precision of existing interactome maps and associated the false discovery rates to different sequence and domain features [14, 17, 18]. In this study, we have systematically compared the interactions from different sampling classes with features such as protein foldability, conservation in yeast, and presence of interacting domains (known domain–domain interactions). We found that there is no obvious bias of false-positive features toward any sampling class, suggesting that the quality of the singleton interactions is comparable to that of highly sampled interactions. Furthermore, we observed that singleton interactions are transient in nature compared to the highly sampled interactions, which are often part of proteins complexes.

2 Materials and methods

2.1 Creating a dataset of repeatedly screened Y2H interactions

We have collected 1,262 human PPIs from two recent studies that are tested 4-times with repeated Y2H screening protocols. We refer to this dataset as repeat-Y2H dataset. It contains PPI data from two independent studies: (1) a dataset by Venkatesan et al. [16], derived by retesting 5% subset of space searched for CCSB_HI1 PPI network [6] with four repeat Y2H screens (1,822 baits against 1,796 prey proteins). (2) A subset of the MDC_MAPK PPI

network (a PPI network created for MAPK signaling pathways) is derived by selecting the interactions that are screened four times from the MDC_MAPK network (unpublished data). In the first dataset by Venkatesan et al. [16], a single screen refers to the pooled testing of 188 preys against arrayed single baits, with subsequent retest. For the MDC_MAPK dataset, a single screen refers to the pooled screening of eight baits against individually arrayed preys, followed by a subsequent retest of the potential interacting pairs in a one-to-one manner. These two studies resulted in 239 and 1,023 PPIs, respectively. Combining the two datasets shows there is no overlap between the datasets, thus repeat-Y2H dataset resulted in 1,262 PPIs that are tested four times in Y2H matrix screening.

2.2 Predicting protein features associated with false-positive interactions

2.2.1 A literature-based PPI dataset

The Human Protein Reference Database (HPRD) is a literature-curated database of the human proteome that contains information about PPIs, domain architecture, post-translational modifications, and disease associations [19]. HPRD is considered as a reference data set for literature-based PPIs. A recent version (release 7) of HPRD is downloaded for this study (http://www.hprd.org/). As we used a fraction of the CCSB_HI1 data for our repeat-Y2H dataset [6], we removed these interactions from the HPRD dataset. We found 91 PPIs in the repeat-Y2H dataset that are already reported in HPRD database.

2.2.2 Predicting protein foldability

We have used FoldIndex to predict whether a given protein sequence adopts a defined fold or is intrinsically unfolded, based on the average residue hydrophobicity (Kyte–Doolittle scale) and the net charge of the sequence [20]. We have used a perl script that automatically predicts the protein foldability using the FoldIndex web service (http://bioportal.weizmann.ac.il/fldbin/findex). The program outputs an unfoldability score, where the positive values represent proteins likely to be folded, and the negative values represent proteins likely to be intrinsically unfolded. Unfoldability scores are predicted for 791 proteins that are part of repeat-Y2H dataset. Out of 791 proteins, 191 are predicted as intrinsically unfolded proteins.

2.2.3 Predicting conservation in yeast

To detect the yeast orthologues of human proteins we have used the HomoloGene database (release 63) [21]. The HomoloGene system automatically detects homologs

proteins/genes among the annotated genes of several completely sequenced eukaryotic genomes. We have downloaded the recent version of HomoloGene from ftp://ftp.ncbi.nih.gov/pub/HomoloGene/. Out of 791 proteins in repeat-Y2H dataset, we mapped yeast orthologues for 103 proteins.

2.2.4 Domain annotation and domain–domain interaction dataset

For each protein in the repeat-Y2H dataset, we have defined its domains based on the InterPro database version 20.0 [22]. InterPro is an integrative database of protein families, domains, repeats, and sequence motifs. We have downloaded the InterPro database from ftp://ftp.ebi.ac.uk/pub/databases/interpro/. The domain-domain interactions have been extracted from the DOMINE database version 1.1 [23]. DOMINE is a comprehensive database of known and predicted protein domain–domain interactions. It contains interactions inferred from protein structure data bank (PDB) [24] and those that are predicted by eight different computational approaches using Pfam domain [25] definitions. DOMINE is downloaded from http://domine.utdallas.edu/cgi-bin/Domine. The Pfam based domain interactions are mapped to the InterPro domain identifiers using an InterPro to Pfam domain relation table available within the InterPro database. We have mapped 360 domain–domain interactions to 208 PPIs in repeat-Y2H dataset.

2.3 Dataset of protein complexes and human kinases

We have compiled the human protein-complex data from the recently published large-scale studies using a mass spectrometry-based approach [26]. We further complemented this dataset with the protein-complexes reported in the HPRD database. This resulted in a comprehensive dataset of human protein-complexes with 25,595 binary interactions between 3,598 proteins. We mapped ten interactions in the repeat-Y2H dataset to protein-complex dataset and annotated them as part of stable complex. To define transient interactions, we have mapped 51 known human kinases [27] to 200 PPIs in the repeat-Y2H dataset and defined these interactions as transient interactions.

3 Results and discussions

3.1 Classification of interactions based on Y2H sampling

The interactions in the repeat-Y2H dataset have been classified into three groups according to their sampling in
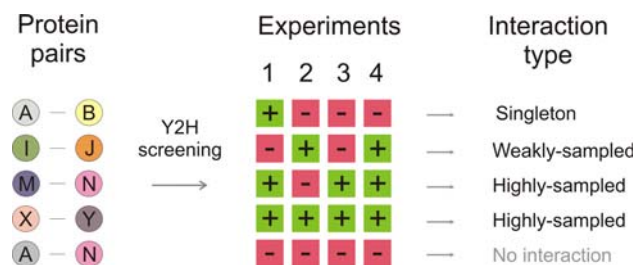

Fig. 1 Schematic representation of the classification of repeated-Y2H interactions. Each protein pair is tested with four independent Y2H experiments and based on the outcome they are classified as singleton, weakly sampled, highly sampled and non-interacting pairs

four repeated Y2H screens. Highly sampled interactions were found in at least three screens; weakly sampled interactions in two out of four screens and singletons were detected only in one of the four experiments (Fig. 1). Table 1 shows the number of interactions in each class. About 56% of the interactions in the dataset are singletons, weakly sampled and highly sampled interactions account for 31 and 13% of the PPIs, respectively.

3.2 Highly sampled interactions show a better overlap with literature-based interactions than singletons

We have analyzed the overlap of interactions in each sampling class with the literature interactions from the HPRD dataset. We found 7% overlap between the repeat-Y2H dataset and HPRD. Strikingly, the overlap increased to 15% when highly sampled interactions were compared. However, an overlap of only 5% was observed by comparing HPRD with singletons (Fig. 2a). The differences between the sampling classes were statistically significant (P value < 0.0001; Chi-square).
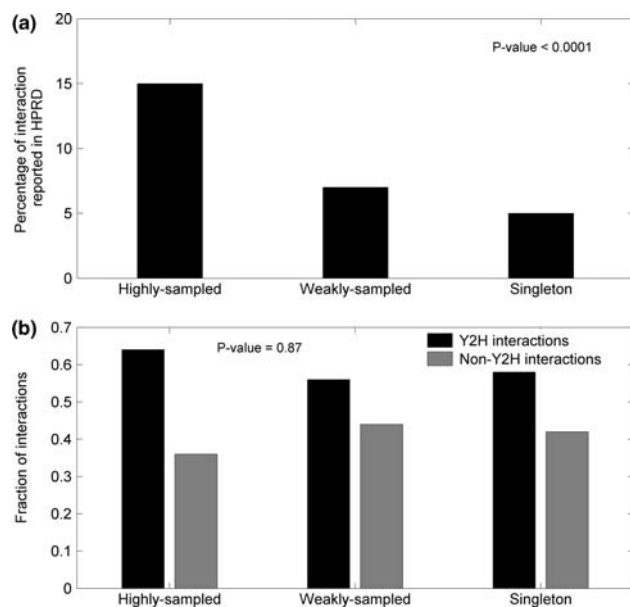
Next, the literature-overlapping interactions were further grouped into Y2H and non-Y2H detected interactions. We found that 60% of the literature-overlapping interactions were found with Y2H screening and 40% with other in vivo and in vitro assays. We did not observe a significant difference in the ratio of Y2H and non-Y2H detected interactions when singletons and highly sampled interactions were analyzed (P value $= 0.87$; Chi-square) (Fig. 2b). Although the singletons show low overlap with the literature, the probability of detecting singletons by a non-Y2H assay is as good as for the highly sampled interactions. Thus, we conclude that singletons are not likely to be false-positive interactions.

3.3 Foldability and conservation of proteins in yeast have no influence on the sampling of interactions

We have performed a systematic analysis to understand the influence of sequence properties of proteins on the Y2H sampling results. Previous studies have shown that the

Table 1 Classification of interaction pairs based on the sampling in the repeat-Y2H dataset

Interaction type	Venkatesan et al.		MDC_MAPK dataset		Total number of interactions	
	Interactions	%	Interactions	%	Interactions	%
Highly sampled	52	21.7	114	11.1	166	13.1
Weakly sampled	62	25.9	331	32.3	393	31.1
Singleton	125	52.3	578	56.5	703	55.7
Total	239	100	1,023	100	1,262	100

Fig. 2 a Overlap between the interactions in different sampling classes with the literature-based interactions. For each sampling class, we computed the percentage of interactions overlapping with HPRD database. **b** Fraction of interactions detected with Y2H and non-Y2H assays for the overlapping literature-interactions

hydrophobicity of proteins influences the false discovery rates of PPIs [15]. Here, we have therefore investigated the impact of protein foldability on the Y2H PPI sampling using the FoldIndex. We observed that 24% of the proteins in the Y2H data set are intrinsically unfolded. Based on the foldability score, we then grouped the interacting pairs into three classes (1) interactions where both partners are intrinsically unfolded; (2) interactions where one of the proteins is unfolded; and (3) interactions where both proteins are likely to be folded according to FoldIndex score. Table 2 shows that there is no obvious bias of protein foldability with respect to the interaction sampling results (P value = 0.32; Chi-square). This suggests that the interaction sampling in Y2H is independent of the protein foldability.

Furthermore, we have tested whether conservation of proteins in yeast has any influence on the interaction sampling. Yeast orthologs of repeat-Y2H dataset proteins were predicted using the HomoloGene database. We found

that ∼13% of the human proteins in the repeat-Y2H dataset have yeast orthologues, indicating that these proteins are conserved. The PPI pairs were grouped into three classes: (1) both the interacting pairs have orthologs in yeast; (2) at least one of the interacting proteins has a yeast ortholog; and (3) none of the interacting proteins is conserved. Table 2 shows that there is no bias of evolutionary conservation with respect to interaction sampling (P value = 0.47; Chi-square).

3.4 Singletons are enriched with proteins containing interaction domains

We have investigated the impact of protein domains and domain–domain interactions on the results of Y2H interaction sampling. We assigned InterPro domain annotation to 89% of the proteins in the repeat-Y2H dataset (703 out of 791 proteins). For the 703 proteins, 2,528 annotated domains were identified, indicating that on average each protein has 3.6 domains. Table 3 shows the ten most frequently found domains in the different sampling classes. Previous studies have shown that certain protein domains like the Homeobox domain might be responsible for the identification false-positive interactions [15]. However, we failed to observe such specific associations with any sampling class, suggesting there is no bias of false-positive associated domains with respect to sampling class (Table 3).
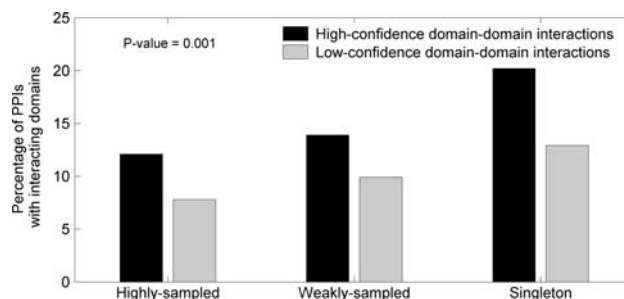
However, the different sampling classes correlated with the appearance of domain–domain interactions. Using the DOMINE database, we analyzed whether annotated domain–domain interactions are overrepresented in the different Y2H sampling classes. To do so, we grouped the interacting pairs into three classes (1) interactions with potentially high-confidence domain–domain interactions; (2) interactions with low-confidence domain–domain interactions and (3) interaction without any known domain–domain interactions. Overall, the repeat-Y2H dataset possess 17% high-confidence and 11% low-confidence domain–domain interactions. Figure 3 shows that singleton dataset contains a significantly higher fraction of known domain–domain interactions (both high- and low-confidence interactions) than the weakly sampled and highly sampled interaction datasets (P value = 0.001; Chi-square). This is surprising, as one would expect the

Table 2 Comparison of different sampling classes with respect to their protein foldability, and conservation in yeast

Features	Interaction class	Sampling class			Significance of difference (Chi-square) (*P* values)
		Highly sampled (%)	Weakly sampled	Singleton	
Conservation in yeast	Between conserved proteins	0.7	2.2	2.9	0.466
	Between conserved and non-conserved protein	21.2	19.8	21.2	
	Between non-conserved proteins	78.1	78	75.9	
Protein foldability	Between folded proteins	6.7	6.4	6.9	0.317
	Between folded and unfolded protein	38.7	41.8	35.6	
	Between intrinsically unfolded proteins	54.6	51.8	57.8	

Table 3 The ten most frequent InterPro domains found in different sampling classes

Interpro ID	Domain name
Highly sampled	
IPR005225	Small GTP-binding protein
IPR016024	Armadillo-type fold
IPR007087	Zinc finger, C2H2-type
IPR001849	Pleckstrin homology
IPR000884	Thrombospondin, type 1 repeat
IPR006941	Ribonuclease CAF1
IPR010909	PLAC
IPR000953	Chromo domain
IPR001806	Ras GTPase
IPR000719	Protein kinase, core
Weakly sampled	
IPR011009	Protein kinase-like
IPR000719	Protein kinase, core
IPR002290	Serine/threonine protein kinase
IPR000980	SH2 motif
IPR001452	Src homology-3 domain
IPR001781	Zinc finger, LIM-type
IPR001680	WD40 repeat
IPR013026	Tetratricopeptide region
IPR016024	Armadillo-type fold
IPR013783	Immunoglobulin-like fold
Singleton	
IPR011009	Protein kinase-like
IPR000719	Protein kinase, core
IPR002290	Serine/threonine protein kinase
IPR011993	Pleckstrin homology-type
IPR013087	Zinc finger, C2H2-type/integrase, DNA-binding
IPR001452	Src homology-3 domain
IPR001680	WD40 repeat
IPR000980	SH2 motif
IPR005225	Small GTP-binding protein
IPR001909	Krueppel-associated box

Fig. 3 Known domain–domain interactions in the PPIs are compared against the different sampling classes. The domain-domain interactions are further distinguished as high- and low-confidence domain-domain interactions

proteins in the highly sampled interactions to have more known interaction domains than the singletons. Although singletons are not well sampled in Y2H screens, they contain proteins with well-characterized domain interactions. This provides additional evidence that singleton interactions are not likely to be false-positives and instead are interactions with biological meaning.

3.5 Singletons are transient protein–protein associations

Physical interactions between proteins are characterized either as stable complexes or transient interactions based on their affinity and lifetime [28, 29]. In a protein complex, the proteins form stable associations to perform their functions in the cell (e.g. basic transcriptional machinery). In contrast, a protein may interact transiently with another protein to modify its function (for example, a protein kinase will add a phospho-group to a substrate and thereby change the activity of a protein). We therefore have investigated the sampling nature of potentially stable and transient interactions in the repeat-Y2H dataset.

In order to study the sampling nature of protein-complex interactions, we mapped the repeat-Y2H dataset to human

protein-complexes reported by Ewing et al. [26] and HPRD. We only detected ten interactions (0.8%) that overlapped between repeat-Y2H dataset and protein-complexes. However, within these interactions, we found a significant difference between the singletons and highly sampled interactions. Figure 4a shows that 3% (5 out of 166 interactions) of the highly sampled interactions are found in the dataset containing protein complexes. In contrast, only 0.1% (1 out of 703 interactions) of singletons was found in the protein-complex data. This difference is statistically significant with the P value of 0.005 (Chi-square). This shows that even when the overlap between Y2H interaction data and protein-complex data are low, the interactions that are part of complexes are enriched with highly sampled Y2H interactions.

In order to analyze the sampling behavior of transient interactions, we have considered kinase interactions as transient interactions. We defined 200 interactions (15%) as kinase-interactions in the repeat-Y2H dataset that contain 51 different protein kinases. Then, these kinase-interactions were analyzed in different Y2H sampling classes. We found that 20% of the singleton interactions are kinase-interactions, while only 5% of highly sampled interactions contain kinases (Fig. 4b). This difference is statistically significant with the P value < 0.0001 (Chi-square). Thus, our results indicate that singleton interactions are of transient nature, and repeated Y2H screens are necessary to detect such low-affinity interactions.
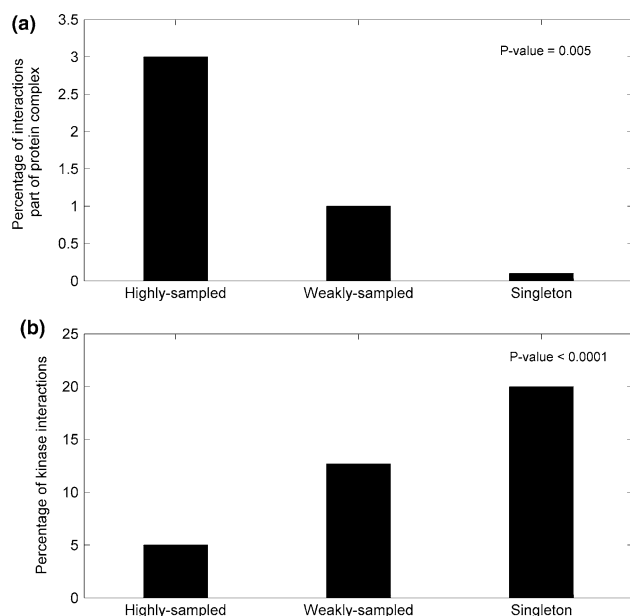
4 Conclusions

Our systematic analysis addresses the question whether Y2H interactions found in different sampling classes have different quality. A simple comparison of the overlap between different sampling classes with literature-based interactions might suggest that the singleton interactions are of poor quality. However, our detailed analysis revealed that both singletons and highly sampled interactions have an equal probability of being recaptured in independent, non-Y2H assays. This shows that singletons consist of true interactions that could even be validated with an independent assay.

Next, the features of proteins within singletons, weakly and highly sampled interactions are similar in their foldability, conservation in yeast and domain occurrence. This suggests that they are similar in quality and none of the previously associated false-positive features is linked to any particular sampling class. Hence, the quality of singletons is comparable to the highly sampled interactions. Furthermore, the singletons contain a high fraction of known domain–domain interactions. This implies that these interactions are feasible and most likely biologically meaningful interactions.

We show that repeated Y2H screening is advantageous as it captures the weak and transient interactions. Such transient interactions are much more difficult to study, because the conditions required for their identification have to be established individually. However, we found that repeated Y2H matrix screening efficiently allows the identification of transient interactions that are particular important in cell signaling.

Fig. 4 a Overlap between the interactions in different sampling classes with protein-complexes. **b** Overlap between the interactions in different sampling classes with kinase-interactions

References

1. Calvano SE, Xiao W, Richards DR, Felciano RM, Baker HV, Cho RJ, Chen RO, Brownstein BH, Cobb JP, Tschoeke SK, Miller-Graziano C, Moldawer LL, Mindrinos MN, Davis RW, Tompkins RG, Lowry SF (2005) Nature 437:1032
2. Chuang HY, Lee E, Liu YT, Lee D, Ideker T (2007) Mol Syst Biol 3:140
3. Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao YL, Ooi CE, Godwin B, Vitols E, Vijayadamodar G, Pochart P, Machineni H, Welsh M, Kong Y, Zerhusen B, Malcolm R, Varrone Z, Collis A, Minto M, Burgess S, McDaniel L, Stimpson E, Spriggs F, Williams J, Neurath K, Ioime N, Agee M, Voss E, Furtak K, Renzulli R, Aanensen N, Carrolla S, Bickelhaupt E, Lazovatsky Y, DaSilva A, Zhong J, Stanyon CA, Finley RL Jr, White KP, Braverman M, Jarvie T, Gold S, Leach M, Knight J, Shimkets RA, McKenna MP, Chant J, Rothberg JM (2003) Science 302:1727

4. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y (2001) Proc Natl Acad Sci USA 98:4569
5. Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, Vidalain PO, Han JD, Chesneau A, Hao T, Goldberg DS, Li N, Martinez M, Rual JF, Lamesch P, Xu L, Tewari M, Wong SL, Zhang LV, Berriz GF, Jacotot L, Vaglio P, Reboul J, Hirozane-Kishikawa T, Li Q, Gabel HW, Elewa A, Baumgartner B, Rose DJ, Yu H, Bosak S, Sequerra R, Fraser A, Mango SE, Saxton WM, Strome S, Van Den Heuvel S, Piano F, Vandenhaute J, Sardet C, Gerstein M, Doucette-Stamm L, Gunsalus KC, Harper JW, Cusick ME, Roth FP, Hill DE, Vidal M (2004) Science 303:540
6. Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, Klitgord N, Simon C, Boxem M, Milstein S, Rosenberg J, Goldberg DS, Zhang LV, Wong SL, Franklin G, Li S, Albala JS, Lim J, Fraughton C, Llamosas E, Cevik S, Bex C, Lamesch P, Sikorski RS, Vandenhaute J, Zoghbi HY, Smolyar A, Bosak S, Sequerra R, Doucette-Stamm L, Cusick ME, Hill DE, Roth FP, Vidal M (2005) Nature 437:1173
7. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S, Timm J, Mintzlaff S, Abraham C, Bock N, Kietzmann S, Goedde A, Toksoz E, Droege A, Krobitsch S, Korn B, Birchmeier W, Lehrach H, Wanker EE (2005) Cell 122:957
8. Suzuki H, Fukunishi Y, Kagawa I, Saito R, Oda H, Endo T, Kondo S, Bono H, Okazaki Y, Hayashizaki Y (2001) Genome Res 11:1758
9. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamodar G, Yang M, Johnston M, Fields S, Rothberg JM (2000) Nature 403:623
10. Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dumpelfeld B, Edelmann A, Heurtier MA, Hoffman V, Hoefert C, Klein K, Hudak M, Michon AM, Schelder M, Schirle M, Remor M, Rudi T, Hooper S, Bauer A, Bouwmeester T, Casari G, Drewes G, Neubauer G, Rick JM, Kuster B, Bork P, Russell RB, Superti-Furga G (2006) Nature 440:631
11. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, Remor M, Hofert C, Schelder M, Brajenovic M, Ruffner H, Merino A, Klein K, Hudak M, Dickson D, Rudi T, Gnau V, Bauch A, Bastuck S, Huhse B, Leutwein C, Heurtier MA, Copley RR, Edelmann A, Querfurth E, Rybin V, Drewes G, Raida M, Bouwmeester T, Bork P, Seraphin B, Kuster B, Neubauer G, Superti-Furga G (2002) Nature 415:141
12. Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis AP, Punna T, Peregrin-Alvarez JM, Shales M, Zhang X, Davey M, Robinson MD, Paccanaro A, Bray JE, Sheung A, Beattie B, Richards DP, Canadien V, Lalev A, Mena F, Wong P, Starostine A, Canete MM, Vlasblom J, Wu S, Orsi C, Collins SR, Chandran S, Haw R, Rilstone JJ, Gandi K, Thompson NJ, Musso G, St Onge P, Ghanny S, Lam MH, Butland G, Altaf-Ul AM, Kanaya S, Shilatifard A, O'Shea E, Weissman JS, Ingles CJ, Hughes TR, Parkinson J, Gerstein M, Wodak SJ, Emili A, Greenblatt JF (2006) Nature 440:637
13. Yu H, Braun P, Yildirim MA, Lemmens I, Venkatesan K, Sahalie J, Hirozane-Kishikawa T, Gebreab F, Li N, Simonis N, Hao T, Rual JF, Dricot A, Vazquez A, Murray RR, Simon C, Tardivo L, Tam S, Svrzikapa N, Fan C, de Smet AS, Motyl A, Hudson ME, Park J, Xin X, Cusick ME, Moore T, Boone C, Snyder M, Roth FP, Barabasi AL, Tavernier J, Hill DE, Vidal M (2008) Science 322:104
14. Bader JS, Chant J (2006) Science 311:187
15. Huang H, Bader JS (2009) Bioinformatics 25:372
16. Venkatesan K, Rual JF, Vazquez A, Stelzl U, Lemmens I, Hirozane-Kishikawa T, Hao T, Zenkner M, Xin X, Goh KI, Yildirim MA, Simonis N, Heinzmann K, Gebreab F, Sahalie JM, Cevik S, Simon C, de Smet AS, Dann E, Smolyar A, Vinayagam A, Yu H, Szeto D, Borick H, Dricot A, Klitgord N, Murray RR, Lin C, Lalowski M, Timm J, Rau K, Boone C, Braun P, Cusick ME, Roth FP, Hill DE, Tavernier J, Wanker EE, Barabasi AL, Vidal M (2009) Nat Methods 6:83
17. Scholtens D, Chiang T, Huber W, Gentleman R (2008) Bioinformatics 24:218
18. Hart GT, Ramani AK, Marcotte EM (2006) Genome Biol 7:120
19. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, Balakrishnan L, Marimuthu A, Banerjee S, Somanathan DS, Sebastian A, Rani S, Ray S, Harrys Kishore CJ, Kanth S, Ahmed M, Kashyap MK, Mohmood R, Ramachandra YL, Krishna V, Rahiman BA, Mohan S, Ranganathan P, Ramabadran S, Chaerkady R, Pandey A (2009) Nucleic Acids Res 37:D767
20. Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg EH, Man O, Beckmann JS, Silman I, Sussman JL (2005) Bioinformatics 21:3435
21. Sayers EW, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S, Feolo M, Geer LY, Helmberg W, Kapustin Y, Landsman D, Lipman DJ, Madden TL, Maglott DR, Miller V, Mizrachi I, Ostell J, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Shumway M, Sirotkin K, Souvorov A, Starchenko G, Tatusova TA, Wagner L, Yaschenko E, Ye J (2009) Nucleic Acids Res 37:D5
22. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, Finn RD, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Laugraud A, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, Mistry J, Mitchell A, Mulder N, Natale D, Orengo C, Quinn AF, Selengut JD, Sigrist CJ, Thimma M, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C (2009) Nucleic Acids Res 37:D211
23. Raghavachari B, Tasneem A, Przytycka TM, Jothi R (2008) Nucleic Acids Res 36:D656
24. Berman H, Henrick K, Nakamura H (2003) Nat Struct Biol 10:980
25. Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer EL, Bateman A (2008) Nucleic Acids Res 36:D281
26. Ewing RM, Chu P, Elisma F, Li H, Taylor P, Climie S, McBroom-Cerajewski L, Robinson MD, O'Connor L, Li M, Taylor R, Dharsee M, Ho Y, Heilbut A, Moore L, Zhang S, Ornatsky O, Bukhman YV, Ethier M, Sheng Y, Vasilescu J, Abu-Farha M, Lambert JP, Duewel HS, Stewart II, Kuehl B, Hogue K, Colwill K, Gladwish K, Muskat B, Kinach R, Adams SL, Moran MF, Morin GB, Topaloglou T, Figeys D (2007) Mol Syst Biol 3:89
27. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S (2002) Science 298:1912
28. Nooren IM, Thornton JM (2003) EMBO J 22:3486
29. Nooren IM, Thornton JM (2003) J Mol Biol 325:991