BMC Bioinformatics

Research article

Triangle network motifs predict complexes by complementing high-error interactomes with structural information

Bill Andreopoulos*[1,2], Christof Winter[1], Dirk Labudde[1,2] and Michael Schroeder[1,2]

Address: [1]Biotechnology Center (BIOTEC), Technische Universität Dresden, 01307 Dresden, Germany and [2]nanometis, Tatzberg 47-49, 01307 Dresden, Germany

Email: Bill Andreopoulos* - williama@biotec.tu-dresden.de; Christof Winter - winter@biotec.tu-dresden.de; Dirk Labudde - dirk.labudde@biotec.tu-dresden.de; Michael Schroeder - ms@biotec.tu-dresden.de

* Corresponding author

Published: 27 June 2009

BMC Bioinformatics 2009, **10**:196 doi:10.1186/1471-2105-10-196

Received: 14 January 2009
Accepted: 27 June 2009

This article is available from: http://www.biomedcentral.com/1471-2105/10/196

Abstract

Background: A lot of high-throughput studies produce protein-protein interaction networks (PPINs) with many errors and missing information. Even for genome-wide approaches, there is often a low overlap between PPINs produced by different studies. Second-level neighbors separated by two protein-protein interactions (PPIs) were previously used for predicting protein function and finding complexes in high-error PPINs. We retrieve second level neighbors in PPINs, and complement these with structural domain-domain interactions (SDDIs) representing binding evidence on proteins, forming PPI-SDDI-PPI triangles.

Results: We find low overlap between PPINs, SDDIs and known complexes, all well below 10%. We evaluate the overlap of PPI-SDDI-PPI triangles with known complexes from Munich Information center for Protein Sequences (MIPS). PPI-SDDI-PPI triangles have ~20 times higher overlap with MIPS complexes than using second-level neighbors in PPINs without SDDIs. The biological interpretation for triangles is that a SDDI causes two proteins to be observed with common interaction partners in high-throughput experiments. The relatively few SDDIs overlapping with PPINs are part of highly connected SDDI components, and are more likely to be detected in experimental studies. We demonstrate the utility of PPI-SDDI-PPI triangles by reconstructing myosin-actin processes in the nucleus, cytoplasm, and cytoskeleton, which were not obvious in the original PPIN. Using other complementary datatypes in place of SDDIs to form triangles, such as PubMed co-occurrences or threading information, results in a similar ability to find protein complexes.

Conclusion: Given high-error PPINs with missing information, triangles of mixed datatypes are a promising direction for finding protein complexes. Integrating PPINs with SDDIs improves finding complexes. Structural SDDIs partially explain the high functional similarity of second-level neighbors in PPINs. We estimate that relatively little structural information would be sufficient for finding complexes involving most of the proteins and interactions in a typical PPIN.

Background

Protein-protein interaction networks (PPINs) derived from high-throughput studies are known to have many errors [1,2]. Data from different studies usually exhibit low overlap; for instance, two large-scale human interactome screens [3,4] share only six interactions, while each has several thousand interactions [5-7]. In some PPINs, more than 50% of reported interactions are estimated to be false positives (FPs) or wrong interactions [8,9]. Moreover, current PPINs are incomplete with an estimated false negative (missing interactions) rate approaching 90% [10-12]. False positives often result when the matrix model, which fully connects the pray and bait proteins, is used for interpreting results of affinity purification followed by mass spectrometry experiments [13].

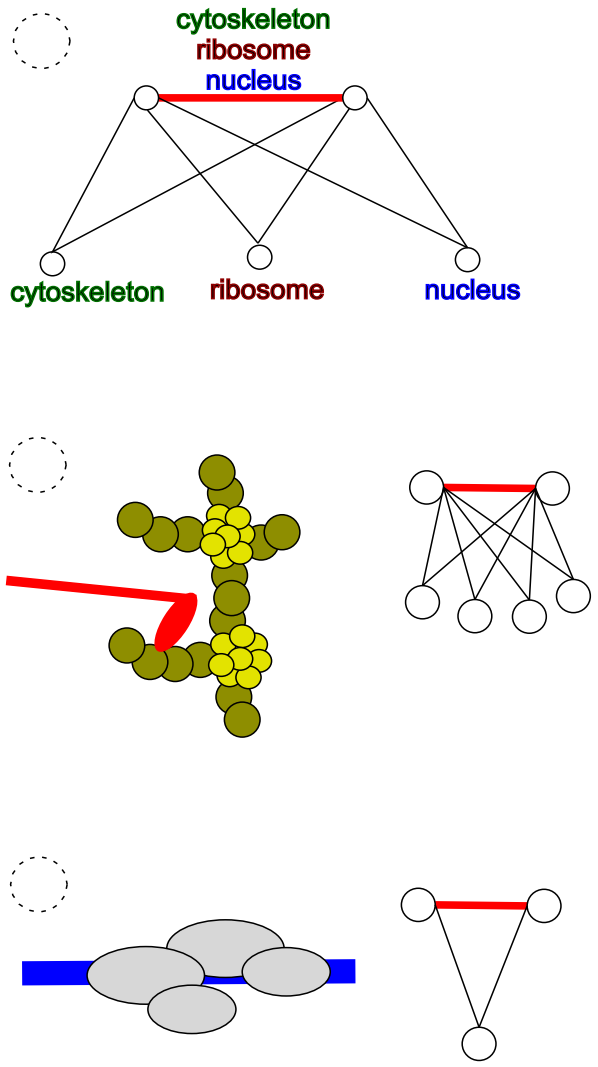

Figure 1

Figure 1
PPIs (black) and structural SDDIs (red). (a) Theme of three PPI-SDDI-PPI triangles sharing the same SDDI. The red SDDI edge is involved in all three triangles. Proteins *D*1 and *D*2 may interact physically with *C* in the cytoskeleton, or *R* in the ribosome, or *N* in the nucleus. The transitive module hypothesis suggests two proteins such as *D*1 and *D*2 that share many common interaction partners are more likely to interact than two proteins that share few common interaction partners [9]. Some PPIs have no common Gene Ontology annotation, hinting to false positives.**(b),(c)** Biological examples of myosin-actin involvement in multiple processes/locations. Their representation as PPI-SDDI-PPI triangle network motifs and themes, as found in integrations of PPINs with SDDIs. **(b)** Myosin in actin cytoskeleton organization and formation. Myosin mediates actin remodelling and vesicular transport. **(c)** Actin and nuclear myosin I (NMI) are required for transcription by RNA polymerases (Pols) I, II, III in the eukaryotic cell nucleus. Actin is directly associated with Pol I, regardless of whether Pol I is engaged in transcription, and NMI interacts with transcription initiation factor TIFIA. TIFIA is phosphorylated. Pol I is then recruited to the DNA promoter through interaction with the phosphorylated TIFIA, which brings actin and NMI into close proximity with each other. Actin, but not NMI, remains associated with Pol I during transcription elongation [121-124].

Not all interactions occur at the same place and time in all cellular states. This implies that representing a PPIN as a set of binary protein-protein interactions (PPIs) is often incomplete [14]. Instead, one wants to restructure protein complexes in PPINs, which are modular units of physical interactions occurring at the same time and cellular component [15,16]. For predicting complexes one wants to include complementary data, such as structural domain-domain interactions (SDDIs) representing binding evidence on proteins [17-22]. At the same time, one wants to leave out of predicted complexes the false positives [22-26].

It was proposed that triangle network motifs represent the basic building blocks of PPINs [27-32]. In this paper, we complement PPIs with SDDIs to form *PPI-SDDI-PPI triangle network motifs*. Triangle network motifs integrate high-throughput PPINs with complementary knowledge, such as structural data, to account for missing edges [25,33-38]. Our proposed paradigm of *PPI-SDDI-PPI triangle network motifs* integrate:

- PPINs from high-throughput experimental studies, which have considerable coverage but also errors, and
- SDDIs that are known to physically mediate PPIs and may be missing in PPINs [39-49].

A *theme* encompasses several PPI-SDDI-PPI triangle network motifs with one SDDI edge as their common organizational principle. Figure 1a shows a theme consisting of three PPI-SDDI-PPI triangle network motifs that share one common SDDI. To demonstrate the biological relevance of triangle network motifs and themes, Figures 1b,c show myosin-actin functions in different cellular locations: cytoskeleton organization and nuclear transcription.

The purpose of PPI-SDDI-PPI triangles is to support revealing biological insights, such as finding complexes of physical interactions occurring at the same time and location [50-55]. Besides complementing PPINs with SDDIs, we additionally form triangle network motifs with other complementary datatypes (CD), such as threading results, and PubMed protein co-occurrence data, thus expanding to other PPI-CD-PPI triangles [56-59]. The complex prediction with other CD is comparable to SDDIs; this supports that the improved complex prediction results are due to a physical relation between proteins and not just coincidence [40,60,61].

A rationale for triangles and themes is the observation that proteins with common interaction partners are likely to have common functions [62-65]. Second-level neighbors in PPINs are functionally similar, and are useful for functional prediction [66-70]. By this "Guilt by Association of Common Interaction Partners" approach, themes can be tied to specific biological phenomena and processes [71-73]. For instance, it was shown for the *E. coli* and *C. elegans* transcriptional network that subgraphs matching two types of transcriptional regulatory circuit triangle – feed-forward and bi-fan – overlap with one another and form large clusters [28,74-76]. Another rationale for triangles and themes is that PPINs are "small-world" implying neighborhood clustering, where neighbors of a given node tend to interact with one another; this results in triangle network motifs of three-node interconnection patterns [77,78]. This led to the "transitive module" hypothesis that is used for predicting missing interactions, as shown in Figure 1a, where proteins with many common interaction partners are likely to interact with one another forming triangles [9].

Extracting triangle network motifs and themes from high-throughput interaction networks

Figure 2 shows the process of extracting triangle network motifs and themes. Given a high-throughput PPIN, we first extract second-level (indirect) neighbors connected by a pair of interactions. Then, we complement them with structural domain-domain interactions (SDDIs), to form PPI-SDDI-PPI triangle network motifs. In the case where a SDDI is involved in more than one triangle, we refer to it as a theme. For evaluation, we examine if the triangles' and themes' overlaps with known MIPS complexes is higher than that of the second-level (indirect) neighbors.

This paper is organized as follows. Next, we present related work on finding errors in PPINs via motifs of interconnection patterns. Then, we present the results on prediction of true positive complexes using triangles. We illustrate this with an example of myosin-actin related activities. Next, we explain the biological basis for triangles: a model for SDDIs that explains the functional similarity of second-level neighbors in PPINs. Finally, we conclude the paper with an outlook of using other data sources to complement interactomes.

Related work

Several papers aim to find errors in PPINs by completing them for missing edges or finding false positives [79-83]. Our approach differs from all of these approaches, since we integrate structural information with PPINs derived from high-throughput studies to find triangle network motifs and themes, which can be used to predict complexes. Moreover, we offer the biological basis for the ability of this structural-PPI hybrid method to predict complexes.

A first category of work involves collecting ensembles of data, such as structural or literature information. Alber et al. (2007) [84] collect diverse high-quality data, and analyse the ensemble to produce a detailed architectural map of the nuclear pore complex. This work translates the data into spatial restraints, instead of using network motifs as in our approach. Ramirez et al. (2007) [22] assessed the quality and value of publicly available human protein network data, by comparing predicted datasets, high-throughput results from yeast two-hybrid screens, and literature-curated protein-protein interactions. This analysis revealed major differences between datasets. Rhodes et al. (2005) [85] demonstrate that a probabilistic analysis integrating model organism protein interactome data, structural domain data, genome-wide gene expression data and functional annotation data predicts nearly 40,000 interactions in humans. Bader et al. (2004) [19] perform an integrated analysis of proteomics data with data from genetics and gene expression. Combining temporal gene expression clustering with proteomics network topology provides an automated method for extracting biological subnetworks. Huang et al. (2004) [86] present POINT, the "prediction of interactome database". POINT integrates several publicly accessible databases, with emphasis placed on mouse, fruit fly, worm and yeast protein-protein interactions datasets from the Database of Interacting Proteins (DIP), followed by converting them into a predicted human interactome. POINT also incorporates correlated mRNA expression clusters obtained from cell cycle microarray databases and subcellular localization from

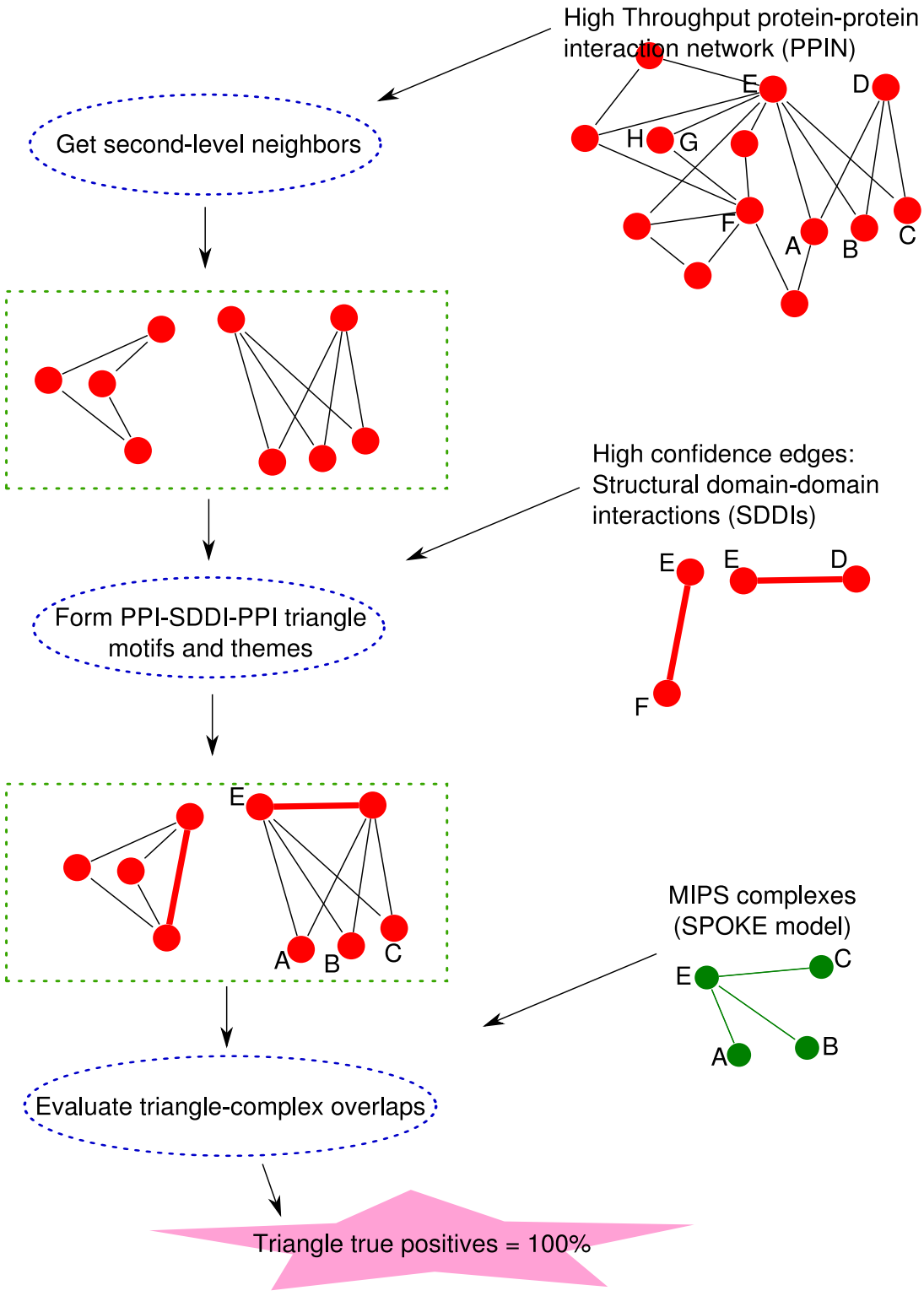

Figure 2
The overall workflow of our process. First, we extract the second-level neighbors from a PPIN. Combining these edges with a complementary data source allows finding triangles and theme motifs. Then, we compare them with known complexes such as MIPS.

Gene Ontology to pinpoint the likelihood of biological relevance of each predicted set of interacting proteins. Patil et al. (2005) [87] find that a combination of sequence, structure and annotation information is a good predictor of true interactions in large and noisy interactomes.

Another large body of work attempted to predict the missing interactions or assign confidences to large noisy interactomes. Some of these use network topology and others use information on SDDIs, while others use Bayesian networks or probabilistic measures. Yu et al. (2006) [68] describe predicting missing PPIs, using only the PPIN topology as observed by a high-throughput experiment. The method searches the interactome for defective cliques, nearly complete complexes of pairwise interacting proteins, and predicts the interactions that complete them. Chen et al. (2008) [88] propose using triplets of observed PPIs to predict and validate interactions. Yeast is the only data set large enough to warrant application of this method. Singhal et al. (2007) [23] present DomainGA, a computational approach that uses information about SDDIs to predict PPIs. This method achieves good prediction for the positive and negative PPIs in yeast. Pitre et al. (2006) [89] present PIPE, a system for predicting PPIs for any target pair of the yeast proteins from their primary structure. Chen et al. (2005) [24] introduce a novel measure called IRAP, "interaction reliability by alternative path", for assessing the reliability of PPIs based on the underlying PPIN topology. IRAP measure is effective for discovering reliable PPIs in large noisy PPIN datasets. Ng et al. (2003) [90] propose an integrative approach that applies SDDIs to predict and validate PPIs. Chen et al. (2005) [24] introduce a SDDI-based random forest of decision trees to infer PPIs. This method is capable of exploring all possible SDDIs and making predictions based on all the protein domains. Wu et al. (2006) [91] propose using the similarity between two Gene Ontology (GO) terms for reconstructing and predicting a yeast PPIN based solely on knowledge of functional associations between the GO annotations.

We have also experimented with using GO similarities in our approach. Chinnasamy et al. (2006) [92] present a probabilistic-based naive Bayesian network to predict PPIs using protein sequence information. This framework provides a confidence level for every predicted PPI. Jansen et al. (2003) [93] also developed an approach using Bayesian networks to predict PPIs in yeast. Han et al. (2004) [94] propose PreSPI, a domain combination based PPI prediction approach. PPIs are interpreted as the result of groups of multiple SDDIs. This approach also provides an interacting probability for PPIs. Recently, Vidal and colleagues [95] used reference sets to calculate the probability that a newly identified PPI is a true biophysical interaction, and assigned confidence scores to all PPIs in interactome networks. Yu et al. (2009) [96] assign confidence scores that reflect the reliability of each PPI, by using multiple independent sets of training positives to reduce the bias inherent in using a single training set.

Another body of work has performed large scale analysis of networks, statistical network motif analysis or error estimation, which is of interest for our work as well. Jin et al. (2007) [32] use network motifs to solve the open question about 'party hubs' and 'date hubs' which was raised by previous studies. At the level of network motifs instead of individual proteins, they found two types of hubs, motif party hubs and motif date hubs, whose network motifs display distinct characteristics on biological functions. Zhang et al. (2005) [28] observed that different types of networks exhibit different triangle profiles, providing a means for network classification. They extended the network triangle concept to an integrated network of many interaction types. Mathivanan et al. (2006) [97] analyzed the major publically available databases that contain literature curated PPI information for human proteins, finding a large difference in their content. This included BIND, DIP, HPRD, IntAct, MINT, MIPS, PDZ-Base and Reactome databases [98]. Chiang et al. (2007) [1] assess the error statistics in all published large-scale datasets for *S. cerevisiae*. Vidal and colleagues [99,100] used an empirically-based approach to assess the quality and coverage of existing human interactomes. They found that high-throughput human interactomes are more precise than literature-curated PPIs from publications.

Several papers used clustering or graph theoretic methods to predict complexes in PPINs. Bader et al. (2003) detected complexes as highly connected subgraphs [101]. Andreopoulos et al. (2007) detected complexes as groups of proteins with similar interaction partners [62]. Cakmak et al. (2007) [102] go beyond complexes to discover unknown pathways in organisms, using Gene Ontology (GO)-based functionalities of enzymes involved in metabolic pathways.

Results and discussion

In our experiments, we employ three high-throughput PPINs, derived by affinity purification followed by mass spectrometry (AP/MS). Krogan06 is based on [103]. Gavin06MATRIX and Gavin06SPOKE are matrix and spoke model interpretations, respectively, of [104]. The matrix model of interpreting pull-down studies connects all prey proteins that were pulled out with a bait, while the spoke model connects only the preys with the bait. We focus on yeast PPINs, since yeast is a well-annotated organism with Gene Ontology terms. The Krogan06 and Gavin06SPOKE yeast PPINs have low overlap. To evaluate the success of our approach, we employ known complexes

Table 1: Overlap of high-throughput PPI networks (Gavin06MATRIX and Krogan06) with the MIPS network (without triangles).

Network	Edge overlap with MIPS[a]	Edges in network but not in MIPS[b]
Gavin06MATRIX	305	3989
Krogan06	359	2225

Symbols below denote *E*, edges; |·|, set cardinality; ∩, intersection; -, set difference.
a $|E_{network} \cap E_{MIPS}|$
b $|E_{network} - E_{MIPS}|$
Only those edges were considered where both proteins were present in the PPI network and in MIPS.

from the MIPS database [105,106]. We evaluate whether known MIPS complexes could be predicted using triangles and theme motifs, consisting of PPINs combined with complementary data such as SDDIs. For illustratory purposes, we use three manually curated networks of myosin-actin involvement in different cellular processes [see Additional files 1, 2, 3, and 4]

Low overlaps of PPINs with complexes

The biological motivation for our work includes low overlap of high-throughput PPINs with known complexes. We compared the overlaps of two high-throughput PPINs, the Gavin06MATRIX and Krogan06 networks, with the MIPS protein complexes dataset. Table 1 shows full results for the overlaps of Gavin06MATRIX and Krogan06 networks to the MIPS complexes. For protein pairs that appear in both PPINs and complexes, we evaluated the number of overlapping edges PPIN ∩ *complexes*. We found *Gavin*06 ∩ MIPS has 305 overlapping edges, *Krogan*06 ∩ MIPS has 359 overlapping edges.

Gavin06MATRIX and Krogan06 had thousands of edges connecting these same proteins, which were not in MIPS. Figure 3 illustrates the overlaps of Gavin06MATRIX and Krogan06 to manually curated myosin-actin networks; the high-throughput PPINs detected disconnected com-

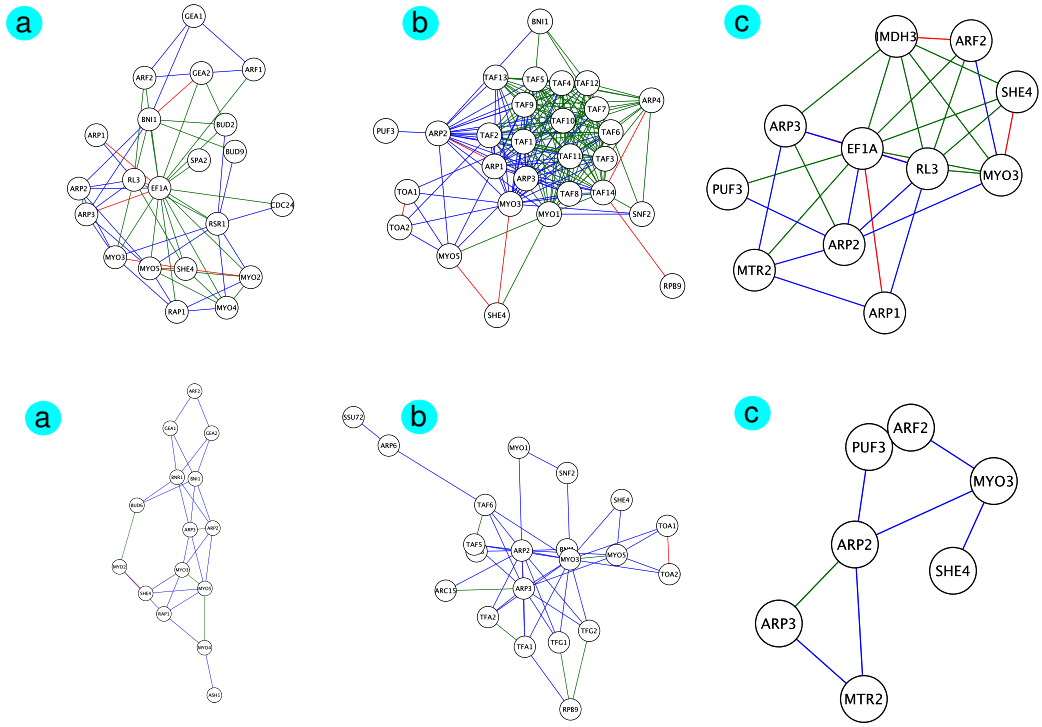

Figure 3
Overlaps of our manually curated myosin-actin networks with high-throughput Gavin06MATRIX (*top*) and Krogan06 (*bottom*) PPINs. The overlap is low. Each row shows the myosin-actin involvement in: *a*. Cytoskeleton organisation, *b*. Nucleus transcription, and *c*. mRNA translocation. Red is both PPINs and myosin-actin; blue is just myosin-actin; green is just PPIN.

Table 2: Success of triangle network motifs and themes in predicting known MIPS complexes.

Complementary datatype	Gavin06MATRIX	Gavin06SPOKE	Krogan06
None [a]	936/166241 = 0.6%	516/10791 = 4.8%	914/33124 = 2.8%
SDDI [b]	254/2832 = 9.0%	143/521 = 27.4%	254/1182 = 21.5%
Literature co-occurrence [c]	710/5592 = 12.7%	416/1340 = 31%	502/1876 = 26.8%
Domain co-occurrence [d]	2004/21876 = 9.2%	892/4268 = 20.9%	1250/4776 = 26.2%
Union of all above	2477/26468 = 9.4%	1446/6129 = 23.6%	1647/6489 = 25.4%

a Second-level indirect relations only
b Structural domain-domain interaction
c PubMed literature co-occurrence of protein mentions
d Pfam domain co-occurrence in IntAct PPIs
Fractions denote *True Positive PPIs/All triangle PPIs* for triangles or second-level neighbors where all three proteins occur in MIPS complexes. Triangle success in MIPS complex prediction is shown as the triangle edges that overlap with complexes. We consistently notice a lower success rate for Gavin06MATRIX than Gavin06SPOKE, which is explained by the higher number of errors in Gavin06MATRIX.

ponents and individual modules, but not the entire connected myosin-actin processes.

PPI-SDDI-PPI triangles predict complexes

Given the many false negatives (missing interactions) and false positives (wrong interactions) in protein-protein interaction networks (PPINs) derived from high-throughput experiments, we evaluated the success of triangle network motifs and themes in finding known MIPS complexes. With structural domain-domain interactions (SDDIs) representing binding evidence on proteins, PPI-SDDI-PPI triangle network motifs are likely to reflect true complexes. To evaluate this, we examined the overlap of triangles from Gavin06 and Krogan06 with MIPS complexes. For the common proteins we evaluated the interactions that are true positives (overlap) or false positives (no overlap) with MIPS.

The first row of table 2 shows the low overlap between PPIN second-level neighbors (without complementary data) and MIPS complexes; where all three proteins in an indirect relation occur in MIPS complexes (denominator), rarely both PPIs occur (numerator). Despite the observed functional similarity of second-level neighbors in PPINs [62-70], second-level neighbors have overlap with MIPS lower than 1%. The other rows show that integrating complementary datatypes (CD) in a PPIN to form PPI-CD-PPI triangle network motifs results in a higher overlap with

Table 3: PPIN triangle success in MIPS complex prediction.

CD = structural SDDI, protein-SCOP domain assignments > confidence threshold			
Confid.thres.	**Gavin06MATRIX**	**Gavin06SPOKE**	**Krogan06**
0	254/2832	143/521	254/1182
40	160/2053	91/367	168/939
50	70/1192	42/215	99/679
60	44/786	29/152	60/467
70	38/704	23/146	55/418
80	36/639	21/130	40/337
90	35/601	21/124	39/306
CD = threading, protein-SCOP domain assignments > confidence threshold			
Confid.thres.	**Gavin06MATRIX**	**Gavin06SPOKE**	**Krogan06**
medium	0/24	0/1	1/12
high	56/296	30/69	68/112
certain	205/4290	123/416	219/1099

Fractions denote *True Positive PPIs /All triangle PPIs*. In triangles where the complementary data (CD) are SDDI structural information or threading, SCOP domain families first need to be assigned to proteins based on confidence. The confidences for protein-SCOP domain assignments are derived, for structural SDDIs based on BLAST sequence alignment similarity, and for threading they are provided by the GTD threading database. Considering different confidence thresholds for protein-SCOP domain assignments affects the MIPS complex prediction success rate.

Table 4: Individual ability of various datatypes to predict MIPS complexes.

Network	Number of nodes [a]	Number of edges [b]	Node overlap with MIPS [c]	MIPS nodes not in network [d]	Network nodes not in MIPS [e]	Nodes in edge overlap [f]	Edges in edge overlap [g]	MIPS nodes not in edge overlap [h]	MIPS edges not in network [i]	Network edges not in MIPS [j]
Gavin06MATRIX	2551	93881	584	791	1967	554	305	821	247	3989
Gavin06SPOKE	2551	22452	584	791	1967	535	232	840	320	950
Krogan06	3670	14291	1031	344	2639	994	359	381	928	2225
SDDI [k]	1551	42222	515	860	1036	500	182	875	207	7375
Threading	4019	95935	1037	338	2982	1017	219	358	1129	11938
Literature co-occurrence [l]	96379	170638	1056	319	95323	1235	491	140	1299	2129
Domain co-occurrence [m]	3560	158704	1042	333	2518	1038	287	337	940	6064
All above combined	100443	504242	1351	24	99092	1358	979	17	1048	24156

Symbols below denote *N*, nodes; *E*, edges; $|\cdot|$, set cardinality; $\cap$, intersection; -, set difference; $\times$, cross product

a $|N_{network}|$

b $|E_{network}|$

c $|N_{MIPS} \cap N_{network}|$

d $|N_{MIPS} - N_{network}|$

e $|N_{network} - N_{MIPS}|$

f |Nodes in $E_{MIPS} \cap E_{network}|$

g $|E_{MIPS} \cap E_{network}|$

h $|N_{MIPS}$ - Nodes in $E_{MIPS} \cap E_{network}|$

i $|(E_{MIPS} \cap (N_{network} \times N_{network})) - E_{network}|$

j $|(E_{network} \cap (N_{MIPS} \times N_{MIPS})) - E_{MIPS}|$

k Structural domain-domain interactions

l PubMed literature co-occurrences of protein mentions

m Pfam domain co-occurrences in IntAct PPIs

This table shows the MIPS overlaps with other network datasets (shown in the first column), indicating the ability of the various networks to predict MIPS. The MIPS network has number of protein nodes $|N_{MIPS}| = 1375$ and number of edges $|E_{MIPS}| = 2099$.

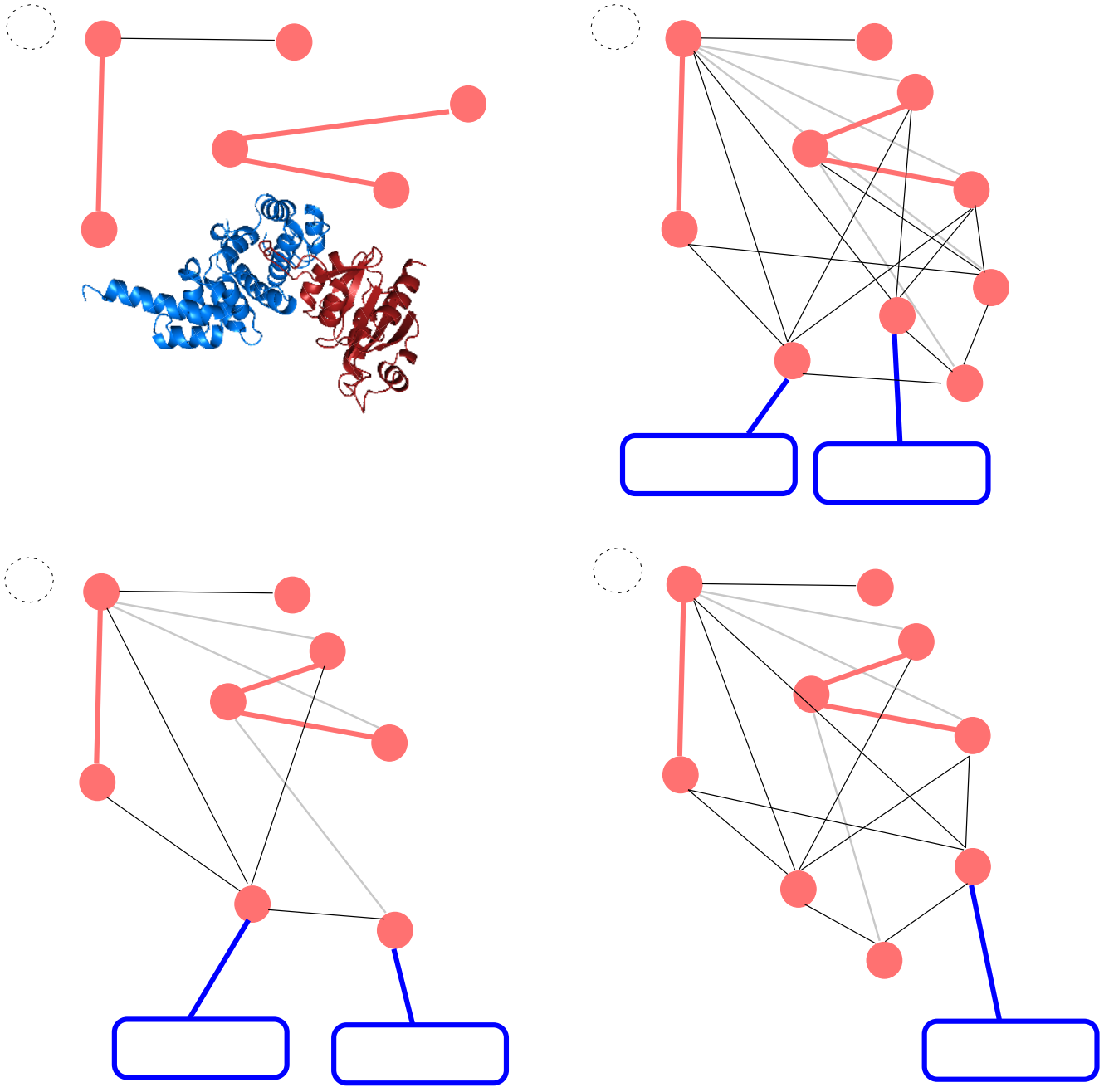

Figure 4
Triangle network motifs and themes from Gavin06MATRIX. Red lines are SDDIs and black lines are PPIs; the SDDIs did not overlap with PPIs from Gavin06MATRIX. The blue boxes show the high number of interaction partners for various proteins in Gavin06MATRIX, supporting that integration with SDDIs can help to find protein complexes. The light gray lines show additional protein connections in the dataset, which resulted in triangles. Each subfigure shows a subnetwork of the original dataset with a specific story. The subfigures represent: *a*. Myosin-actin interactions, mostly SDDIs, which occur in various processes and locations. The blue structure shown is the Sec7 domain (SCOP code a.118.3.1), which was assigned to GEA1. The red structure is the G protein domain (SCOP code c.37.1.8), which was assigned to RSR1 and ARF2. The PDB file that displays these domains as interacting, and from which the image was generated, has the code 1RE0. *b*. These SDDIs are extended with additional PPIs from Gavin06MATRIX, showing specific myosin-actin involvement in cytoskeleton organisation, *c*. nucleus transcription, and *d*. mRNA translocation.

MIPS complexes. In Table 2 the second row shows the PPI-SDDI-PPI triangle overlap with MIPS complexes as a true positive rate as high as 31%; the other triangle interactions are likely false positives. For Gavin06MATRIX the triangle true positive rate is lower than for Krogan06, since Gavin06MATRIX reflects the matrix model interpretation, which resulted in 93, 881 edges including many false positives. Gavin06MATRIX has many errors when overlayed with the MIPS complex dataset. The success rate is higher for Gavin06SPOKE, since there are fewer false positives than Gavin06MATRIX.

Table 3 shows that with varying confidence thresholds for SDDIs, the true positive rate changes. This shows that it is preferable to use the highest-confidence SDDIs. It also shows the significance of using SDDIs for complex prediction.

Triangles with other complementary data

We added to PPINs other complementary datatypes, besides structural SDDIs, to form triangles: PubMed literature co-occurrences of protein mentions, and Interpro Pfam domain co-occurrences in PPIs [107] (see methods section). Table 2 rows 3–4 show the MIPS complex overlaps with triangle network motifs using other complementary datatypes to form triangles. The triangles with other complementary datatypes exhibit little difference in their overlap with MIPS complexes. In the last row 5 where all datatypes are combined, the overlap with MIPS increases. Triangles that include SDDIs or other complementary data to match second-level neighbors have higher overlap with MIPS complexes than second-level neighbors without any complementary data. These results point to the direction of complementing the PPINs with other datatypes as triangle network motifs, rather than simple edges, for improved prediction of MIPS complexes.

Table 4 shows the individual ability of various datatypes to predict the MIPS complexes, showing the edge overlap without forming triangles. As shown under the column "Edges in edge overlap", all datatypes have moderate edge overlap with MIPS. The individual datatypes have little difference in their ability to predict MIPS.

Example: reconstructing distinct myosin-actin biopathways via themes of PPI-SDDI-PPI triangle network motifs

Type I myosin motor proteins (MYO3 or MYO5) have distinct but overlapping functions in multiple cellular processes and locations [108]. Figure 4 shows examples of myosin involvements as PPI-SDDI-PPI triangle network motifs and themes derived from Gavin06MATRIX [104]. Figure 4a shows several core myosin-actin SDDIs that are common to different processes and locations. The SDDIs were validated with the structural interaction network given in [109]. For instance, Myosin type I (MYO3) has SDDIs with the ARP2/3 complex, which plays a major role in the regulation of the actin cytoskeleton, but also plays a role in actin-filament formation during transcription in the nucleus [108]. Figures 4b,c,d extend these core myosin-actin SDDIs with PPIs that are specific to different processes and locations: cytoskeletal actin organization, nuclear transcription, and asymmetric mRNA localization [110].

MYO3 is one of two type I myosins, which utilize the cytoskeleton for movement, moving along microfilaments through interaction with actin. Deletion of MYO3 causes severe defects in growth and actin cytoskeleton organization [111]. Besides myosin, SHE4 is also important for the organization of the actin cytoskeleton. SHE4 is of special interest because it is involved in all of organization of the actin cytoskeleton, asymmetric mRNA localization, and endocytosis [112]. SHE4 has similar Gene Ontology annotations as myosin.

Next, we explore whether triangle network motifs and themes in Gavin06MATRIX can help reconstruct distinct myosin-actin pathways for cellular localization of biomolecules.

Cytoskeletal actin organization

Figure 4b illustrates the relevant triangle network motifs. Yeast cells organize their actin cytoskeleton in a highly polarized manner during vegetative growth. Myosin type I is known to play an important role in moving membranes against actin and membrane-actin interactions. Organization of the actin cytoskeleton requires SHE4. SHE4 is a protein containing a domain that binds to myosin motor domains to regulate myosin function [112].

RSR1, BNI1, GEA1 play a role in cytoskeletal actin localization [113,114]. The correct localization of RSR1 has been shown to be critical for actin cytoskeleton organization. Localization of the Ras-like GTPase RSR1 and its regulators are required for selection of a specific growth site [115]. Regulators direct the correct localization of RSR1 in various organisms. In Figure 4b, while RSR1 interacts with both MYO3 and GEA1, it also interacts with parts of their intersecting neighborhoods. Both GO term similarity and the literature suggest MYO3/GEA1 control of RSR1. The GEA1 RAS superfamily G proteins (small GTPase) has observed SDDIs with both ARF2 and RSR1. GEA1 is a Guanine nucleotide exchange factor for ADP ribosylation factors (ARFs), involved in vesicular transport between the Golgi and ER, Golgi organization, and actin cytoskeleton organization; similar to but not functionally redundant with GEA2. An active Sec7 region in GEA1, which is the

probable catalytic domain for GEF activity, is important for actin cytoskeleton activity. The mechanism by which GEA1 and GEA2 stimulate actin cable formation in a BNI1-dependent manner remains to be determined [116,117].

What is of special interest in this example is the intersection of the neighborhoods of RSR1, ARF2, BNI1 comprising EF1A-RL3, which were previously observed to have a functional significance for F-actin localization [118]. In addition, BNI1 and GEA1 appear to be connected to the ARF2 complex via PYR1 intermediary. Thus, RSR1, GEA1 and BNI1 appear to be linked to one another via EF1A-RL3-PYR1, which are also common partners of ARF2. This suggests a role of EF1A-RL3-PYR1 as the regulators for the RSR1-GEA1-BNI1 complex localization in yeast cytoskeletal actin localization [119].

Overexpression of GEA1 or GEA2 was observed to bypass the requirement for profilin in actin cable formation [116]. Profilin is an actin-binding protein involved in cytoskeleton dynamics. Profilin enhances actin growth as follows: Profilin binds to monomeric actin on the plus end of the filament inducing a shape change of the actin subunit, allowing the G-actin to replace the ADP to which it is bound by ATP and form F-actin. The F-actin then forms a heterodimer which can bind to the plus end of an actin filament. In the process of binding to the actin monomers it also stereochemically inhibits addition to the minus end [120]. On the other hand, in a separate study it was observed that loss of the activity to bind EF1A-RL3 displayed an abnormal phenotype represented by dissociated localizations of F-actin, which were co-localized in wild-type cells [118]. This observation links the two studies, suggesting that the significance of EF1A-RL3 for F-actin localization may help explain why overexpression of GEA1 or GEA2 bypassed the requirement for profilin in actin cable formation.

Nuclear actin and myosin I required for RNA polymerase I, II, III transcription

Figure 4c illustrates the relevant triangle network motifs. The presence of actin and nuclear myosin type I (NMI) in the nucleus suggests a role for these motor proteins in nuclear functions. Both actin and nuclear myosin I (NMI) are associated with ribosomal RNA genes (rDNA) and are required for RNA polymerase I, II, III (Pol I, II, III) transcription [121-124]. Actin and NMI are present in nucleoli as a complex physically associated with RNA polymerase I. This association appears to have a functional relevance in rDNA transcription. Altogether an actin-myosin complex is present on actively transcribing ribosomal genes and, therefore, suggests a direct involvement of actin-myosin in regulating transcription [125].

TBA1/RAP1 play a role in nucleus transciption from RNA polymerase II promoter. TBA1/RAP1 is a DNA-binding protein involved in either activation or repression of transcription, depending on binding site context; it also binds telomere sequences and plays a role in telomeric position effect (silencing) and telomere structure. In Figure 4c, RAP1 is associated with MYO3/SHE4, which transport RAP1 and actin in the nucleus and the cytoplasm. While RAP1 has PPIs to RSR1, BNI1 and ARF2, literature confirms this is an indirect relationship and instead that Myosin type I translocates RAP1 in both the nucleus and cytoplasm (precisely the myosin type I GO annotation) [126,127]. The indirect interaction of RAP1 with RSR1, BNI1 and ARF2 points to the involvement of actin in transciption.

mRNA localization: The SHE protein complex is required for cytoplasmic transport of mRNAs in yeast

Figure 4d illustrates the relevant triangle network motifs. A key feature of eukaryotic cells is their organization into distinct compartments, each with a distinct set of proteins. It has been shown that the sorting of many cytoplasmic proteins involves mRNA localization. Cytoplasmic localization starts in the nucleus where a first set of RNA-binding factors recognize localized mRNAs [124,128]. RNA-protein complexes that are exported to the cytoplasm associate with additional factors, such as molecular motor proteins. Such motors are required to transport their cargo along cytoskeletal filaments to the target site where the mRNA is unloaded and anchored. The SHE protein complex facilitates cytoplasmic localization of ASH1 and other localized mRNAs [129].

ARF2, EF1A, IMDH3 play a role in mRNA localization for translation. ARF2 is an ADP-ribosylation factor involved in regulation of coated formation vesicles in intracellular trafficking within the Golgi [130]. In Figure 4d, ARF2 is likely to interact with subsets of the main cluster; particularly we notice an association of ARF2 with both EF1A and IMDH3:

- **EF1A**: Translation elongation factors are responsible for two main processes during protein synthesis on the ribosome [131]. EF1A (or EF-Tu) is responsible for the selection and binding of the cognate aminoacyl-tRNA to the A-site (acceptor site) of the ribosome. EF2 (or EF-G) is responsible for the translocation of the peptidyl-tRNA from the A-site to the P-site (peptidyl-tRNA site) of the ribosome, thereby freeing the A-site for the next aminoacyl-tRNA to bind. Elongation factors are responsible for achieving accuracy of translation and both EF1A and EF2 are remarkably conserved throughout evolution (InterPro annotation).

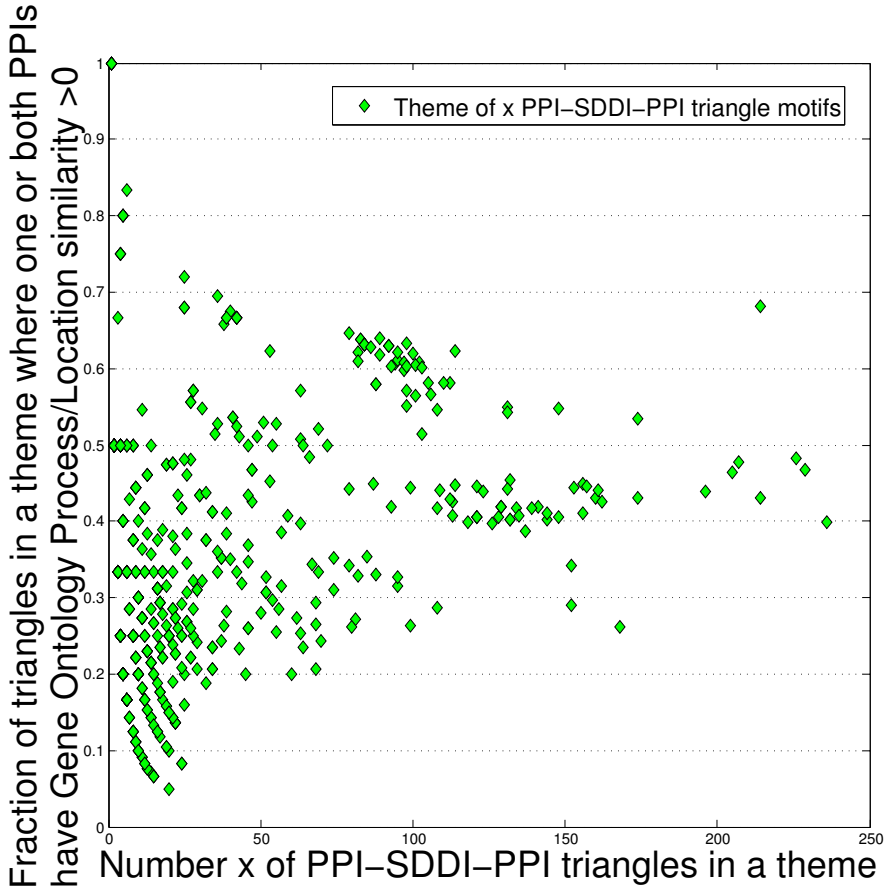

Figure 5

The *x* axis is the number of triangle network motifs in themes of Gavin06MATRIX. The *y* axis is the percentage of those triangles that have non-zero Gene Ontology similarity in their PPI edges. In many triangle network motifs both PPI edges have Gene Ontology annotation similarity equal to zero for the proteins involved.

• **IMDH3**: Involved in the amino acid biosynthesis pathway.

Biological interpretation of PPI-SDDI-PPI triangles: A structural basis for functional similarity of second-level neighbors in PPINs

In this section we propose an explanation for the observation that SDDIs can complement high-error PPINs to improve the finding of complexes. A structural SDDI between two proteins implies that they are likely to be observed with common groups of interaction partners in an experimental study. This especially holds in affinity purification experiments followed by mass spectrometry (AP/MS), since the bait-prey technologies used will cause structurally connected proteins to be detected as prey for similar bait protein(s). Of course this only holds for proteins that are detectable as prey [132]. A SDDI is the likely reason why two proteins are observed with common friends in PPINs from high-throughput AP/MS studies. Then, the SDDI's interaction partners are likely to be observed in different cellular components; Figure 5 shows that many of the SDDI-induced triangles have no common Gene Ontology annotation. Then SDDIs are a partial explanation for the functional similarity of second-level neighbors in PPINs. We propose this *couple-with-common-friends* model as the biological basis for finding complexes via PPI-SDDI-PPI triangle network motifs and themes; subsequently, SDDI edges in triangles can be replaced by other complementary datatypes.

Gene Ontology (GO) similarity in triangle PPI edges

Figure 6 shows an example of a theme from Krogan06, the GO similarities involved, and the evaluated correlations of GO similarities for the PPIs and SDDIs. Table 5 shows that in Gavin06MATRIX and Krogan06 triangle network motifs, SDDIs have significantly higher GO similarities

Table 5: Gavin06MATRIX PPI-SDDI-PPI triangles and Krogan06 PPI-SDDI-PPI triangles: Gene Ontology (GO) similarities and correlations.

	Correlation Gavin06MATRIX	Correlation Krogan06
GO Functional similarity		
Average over protein pairs in SDDI edges	0.48	0.37
Average over protein pairs in PPI edges	0.18 and 0.18	0.35 and 0.36
PPI-PPI similarity correlation coefficient	0.88	0.76
SDDI-PPI similarity correlation coefficient	0.16 and 0.18	0.58 and 0.59
GO Process/Location similarity		
Average over protein pairs in SDDI edges	0.71	0.67
Average over protein pairs in PPI edges	0.28 and 0.27	0.67 and 0.67
PPI-PPI similarity correlation coefficient	0.97	0.6
SDDI-PPI similarity correlation coefficient	0.08 and 0.09	0.46 and 0.45

Often, the two PPIs of a triangle have close GO functional/process/location similarity.

than protein-protein interactions (PPIs). Evaluation of GO similarities in the PPI-SDDI-PPI triangles in Gavin06MATRIX and Krogan06 shows that the PPIs in a triangle represent similar functions and process/location involvements [133]. Table 5 shows a correlation analysis confirming that the PPI-PPI GO similarities (function, process, location) are higher correlated than the SDDI-PPI GO similarities. The correlation of GO similarities of PPI edges in a triangle implies that the SDDI brings together two PPIs involved in similar functions and processes/locations. Figure 5 shows that some triangles' PPIs have GO similarity of zero, hinting at errors. This may also show some promise for finding errors based on GO similarity.

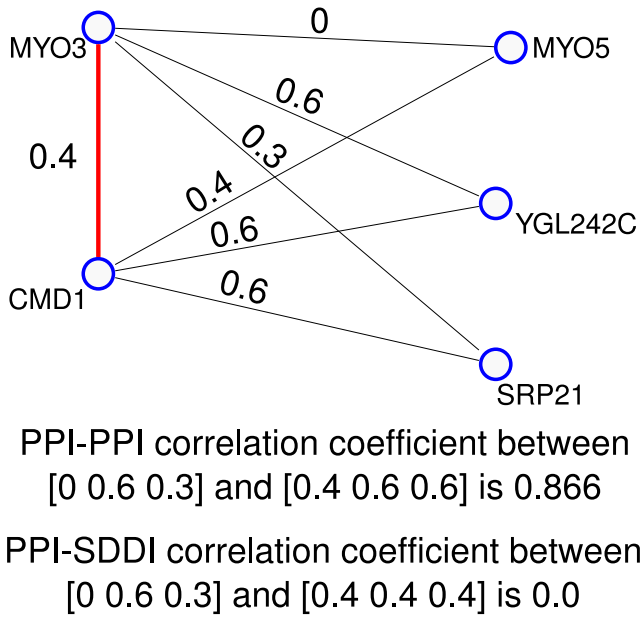

Figure 6
An example of a theme from Krogan06 and the Gene Ontology similarities involved. As shown, computing the correlation coefficients between PPI-PPI vs. DDI-PPI edges gives different correlation values.

Why are few SDDIs detected in high-throughput PPINs experiments?
Table 6 shows that few SDDIs overlap with PPINs, even when considering the highest-confidence SDDIs only. Figure 7 shows a visualization of the SDDIs overlapping with PPINs. The visualization shows that most of these SDDIs are part of *highly connected* components. To assess whether the size of the connected SDDI components that overlap with PPINs is significant, we compared the connected SDDI components to randomly selected SDDIs from the SCOPPI database [134]. We performed 1,000 trials of randomly picking 100 SDDIs from SCOPPI, and we examined how many of these SDDIs were connected each time; on average only 8 SDDIs were connected, a size much smaller than the connected SDDI components that are shown in Figure 7. These results highlight the significance of complementing PPINs via PPI-SDDI-PPI triangles.

Conclusion

How many SDDIs are needed to predict all complexes for an entire PPIN?
Figure 8 is an attempt to predict how many structural SDDIs would be needed for triangles to predict the true positives involving all proteins in a typical PPIN, such as Krogan06. We took all second-level indirect neighbors found in the Krogan06 interactome and, where there was no PPI, added a "hypothetical" SDDI to form PPI-SDDI-PPI triangles. For each SDDI we calculated its theme size, i.e., how many pairs of PPIs the SDDI connected into triangles. Then, we took the theme sizes in decreasing order

Table 6: Few SDDIs overlap with PPINs derived from high-throughput experiments and MIPS complexes.

	SDDI	Gavin06-MATRIX	Gavin06- SPOKE	Krogan06
SDDIs total with both proteins in MIPS and PPIN	SCOPPI [a]	71	71	238
	Threading [b]	3404	3404	9615
SDDIs supported by PPIs in both MIPS and PPIN	SCOPPI	14	9	25
	Threading	61	56	99
SDDIs supported by PPIs in MIPS but not PPIN	SCOPPI	0	5	5
	Threading	20	25	107
SDDIs supported by PPIs in PPIN but not MIPS	SCOPPI	37	30	48
	Threading	144	72	131
SDDIs supported by PPIs in neither PPIN nor MIPS	SCOPPI	20	27	160
	Threading	3179	3251	9278

a protein-SCOP > 90 conf.
b protein-SCOP CERTAIN conf.

from 100 to 1, as shown by blue bars in Figure 8. For each theme size, we indicate on the x-axis how many SDDIs had that theme size, and the red bar shows how many newly encountered proteins were included for that theme size. As the x-axis shows, one could start by finding true positives for the few SDDIs with the largest themes, progressively moving to the many SDDIs with the smallest themes. Somewhere in the middle of the *x*-axis, one would have predicted the true positives for about half of the proteins in the PPIN. However, one would still need to use many SDDIs with the smallest themes, to find all true positives in the Krogan06 PPIN. Therefore, although about half of the true positives could be found with no more than 100 SDDIs, one would need significantly more SDDIs to find all true positives involving all proteins.

SDDIs and the PubMed co-occurrences relate to two different aspects. SDDIs are based on experimental results that are likely to imply a structural interaction. In the case of SDDIs, we can use all information found by mapping structural domains to proteins using BLAST sequence similarity and still get good prediction accuracy. On the other hand, for literature we have to apply a strict filtering, keeping only the top 1% of protein co-occurrences appearing in PubMed as complementary data. We observed that the literature co-occurrences appear to give slightly better results than using SDDIs as complementary data. The main limitation of SDDIs at present is the sparsity of known structural interactions. Since PubMed is expected to grow faster than structural knowledge, using literature co-occurrences might give even better prediction accuracy in the future, as long as a strict cut-off is set.

Conclusion

With the amount of PPINs from high-throughput experiments, structural data and literature-based interactions on the rise, we studied their combined ability to predict known complexes. We found a low overlap of PPINs derived from high-throughput studies with known complexes, as well as low overlap with structural domain-domain interactions.

We proposed PPI-SDDI-PPI triangle network motifs as a model for analysing PPINs and predicting complexes. PPI-SDDI-PPI triangles have higher overlap with MIPS complexes than random second-level neighbors, indicating

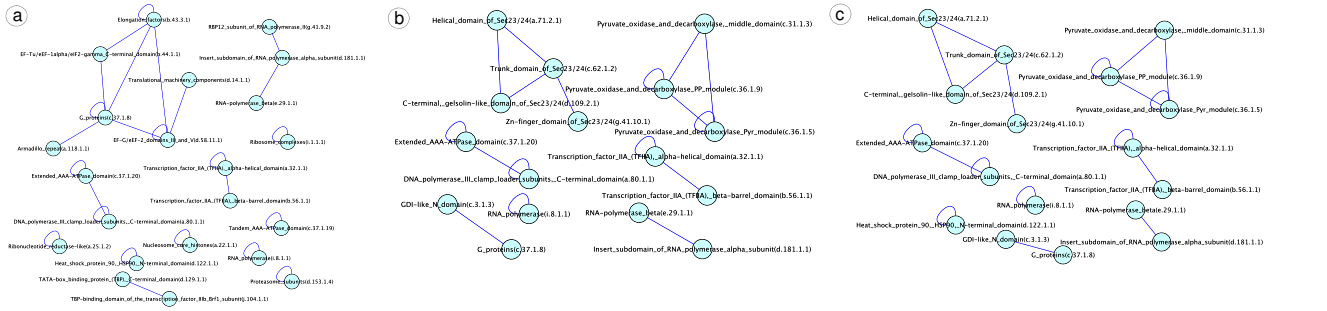

Figure 7
SDDIs that were detected as PPIs in PPINs: (*a*) Krogan06, (*b*) Gavin06SPOKE, (*c*) Gavin06MATRIX. Only one more SDDI was detected in Gavin06MATRIX than in Gavin06SPOKE, pointing to the high number of FPs in the matrix model.

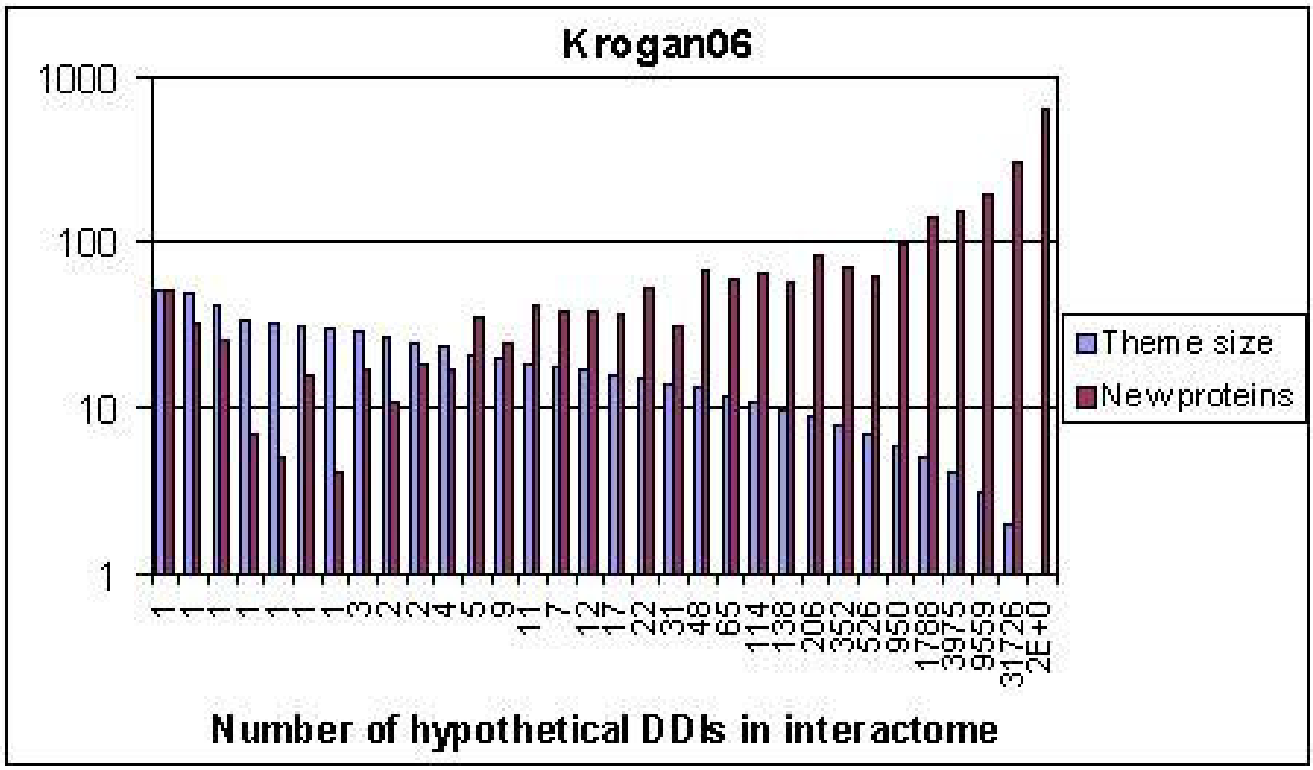

Figure 8
Starting from the few SDDIs (*x*-axis) with the largest theme sizes (blue), and progressively moving to the many SDDIs with the smallest theme sizes. This allows one to eventually find the complexes for all proteins in the PPIN (red).

that structural SDDIs are useful for complementing PPINs in triangles to create a more complete picture of protein cellular involvement. We complemented PPINs with several other datatypes besides SDDIs to create triangle and theme motifs, resulting in similar overlaps with complexes. Themes of PPI-SDDI-PPI triangles helped us to reconstruct complexes in myosin-actin processes that were not detected by PPINs. Our approach is useful for finding true positives in PPINs, as structural knowledge on proteins increases in the future.

SDDIs partially explain the high functional similarity of second-level neighbors in PPINs. A SDDI may cause a structurally connected pair of proteins to be observed with common interaction partners in high-throughput affinity purification experiments followed by mass spectrometry (AP/MS) that use bait-prey technologies. We examined why some SDDIs are detected in PPINs, and we found that SDDIs detected by PPINs are part of highly connected components/complexes, therefore they are more likely to be detected by experimental studies.

Methods

In this section we give an overview of the methods used in this study. Figure 2 illustrates the overall workflow of the process.

PPI-CD-PPI triangle network motifs

PPI-CD-PPI triangles contain three proteins connected by two PPIs and an edge of a complementary datatype (CD), such as a structural SDDI; in this case, we refer to PPI-SDDI-PPI triangles, as Figure 1a shows. Our method can be viewed as finding bicliques in a PPIN, and then connecting second level neighbors via *complementary datatype edges*. For extracting second level neighbors in large networks we used the HIERDENC algorithm, described in [62,135]. Figure 1b and 1c show that PPI-CD-PPI triangles imply that an experiment detected PPIs $A \leftrightarrow B$ and $B \leftrightarrow C$, while a CD edge $A \leftrightarrow C$ exists, such as a structural SDDI. In a PPIN second level neighbors (a pair of PPIs) may be involved across cellular space and time in different processes and locations. Connecting second level neighbors to each other via CD edges gives confidence that the second level neighbors interact at the same cellular space and

time [136-138]. Triangles likely represent a protein complex [139,140].

Let σ_{SDDI} denote the number of PPI-SDDI-PPI triangles a structural SDDI is involved in. A structural SDDI may be involved in $\sigma \geq 1$ triangles, which we refer to as a *theme*. A theme is given by the σ common interaction partners (intersecting neighborhoods) of a SDDI's protein pair, and some PPIs in a theme may be False Positives.

Complementary datatypes

As structural information to complement PPINs, we used the SCOPPI database, which contains SDDIs observed in known protein complex structures [134]. To assign domains, we BLASTed the sequences of all proteins in the "Saccharomyces Genome Database" (which includes yeast PPINs) against all domains sequences of SCOPPI. We considered only BLAST hits with an E-value ≤ 0.01 and a sequence identity percentage $s \geq 30\%$. In addition, we required 75% of the domain to appear in the protein.

Other complementary datatypes (CD) edges we used included The Genomic Threading Database (GTD) [141]. GTD contains yeast protein assignments to SCOP domain structural annotations and interacting structures. An assigned Confidence value gives an indication of the strength of a hit, ranging from "certain" to "guess", which is based on a P-value measure of significance.

The next CD dataset we used was PubMed literature co-occurrences of protein mentions. To extract these, we used the GoPubMed protein mention extraction algorithm to assign proteins to all PubMed documents [142]. Then, we used a version of the Blosum co-occurrence score to find if two proteins p_1 and p_2 co-occur frequently in PubMed documents: $log \frac{Prob(p_1 \ and \ p_2)}{Prob(p_1) \times Prob(p_2)} > 10$. A cutoff of 10 was strict enough to filter out the majority of protein co-occurrences in PubMed, resulting in a network of 170,638 edges. The last CD dataset we used was Interpro Pfam domain co-occurrences in PPIs. For this, we took all IntAct yeast PPIs and assigned to the proteins Pfam domains from InterPro [107]. Then, we used the Blosum co-occurrence score to find which Pfam domains co-occur frequently in the IntAct yeast PPIs. Based on the most co-occurring Pfam domains, we build a network over the yeast PPIs.

High-throughput PPINs and known complexes

We use two yeast PPINs that we denote as Gavin06 [104] and Krogan06 [103]. For Gavin06 we used both the matrix and the spoke model to interpret it, which we refer to as Gavin06MATRIX and Gavin06SPOKE throughout the text. Gavin06MATRIX had 93,881 edges, while Gavin06SPOKE had 22,452 edges. Krogan06 had 14,292 edges, consisting of the binary interactions as provided by the publication. For validation, we used MIPS complexes [105,106]. For MIPS we used the SPOKE model for the interpretation of complexes, since otherwise the result would be biased to give a high overlap with the PPINs [see Additional files 5, 6]. The MIPS complexes had 2,099 edges.

Moreover, for our illustrations we manually curated three network examples from the literature, representing myosin-actin involvement in cytoskeleton organisation, nucleus transcription, and mRNA translocation. Developing these networks involved reading papers from the biomedical literature and recording any interaction(s) described in the articles.

Gene Ontology similarity

It is likely that a PPI is not physical, but a false positive, which may be detected by a GO similarity of zero. PPIs with a GO similarity of zero hint at false positives. For calculating the similarity based on Gene Ontology terms, we searched for GO terms in the current abstract and compared them to the set of GO terms assigned to each gene candidate. For each potential tuple taken from the two sets (text and gene annotation), we calculated a distance of the terms in the ontology tree. These distances yielded a similarity measure for two terms, even if they did not belong to the same sub-branch or were immediate parents/children of each other. The distance took into account the shortest path via the lowest common ancestors, as well as the depth of this lowest common ancestor in the overall hierarchy (comparable to Schlicker et al., 2006 [133]). The distances for the closest terms from each set then defined a similarity between the gene and the text [142].

Correlation

We computed the correlation coefficient between A and B, where A and B are matrices or vectors of the same size. A matrix entry contains a measure of Gene Ontology similarity (0 – 1) for a protein pair involved in a PPI or SDDI. We used the matlab corr2 correlation coefficient:

$$r = \frac{\sum_m \sum_n (A_{mn} - mean(A))(B_{mn} - mean(B))}{\sqrt{(\sum_m \sum_n (A_{mn} - mean(A))^2)(\sum_m \sum_n (B_{mn} - mean(B))^2)}}$$

HIERDENC supplementary material

We implemented the HIERDENC online database, which contains all of the datasets we used. HIERDENC helps a user to visualize and find true positives in PPINs via triangles of high-throughput PPINs and complementary data. http://www.hierdenc.com/ or http://projects.biotec.tu-dresden.de/HIERDENC/

Authors' contributions

All authors read and approved the final manuscript. BA planned the paper, carried out most of the experiments, wrote the software, wrote the python scripts for making the networks, built the manually curated networks, and wrote most of the paper. CW helped in conceptualising the paper with discussions and provided the complementary data. DL helped in formulating the paper with discussions. MS supervised the work and contributed discussions and ideas.

Additional material

Additional file 1

An excel file with Uniprot accession numbers for all protein names used in the text.

Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-10-196-S1.xls]

Additional file 2

Manually curated network example from the literature, representing myosin-actin involvement in cytoskeleton organisation.

Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-10-196-S2.txt]

Additional file 3

Manually curated network example from the literature, representing myosin-actin involvement in mRNA translocation.

Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-10-196-S3.txt]

Additional file 4

Manually curated network example from the literature, representing myosin-actin involvement in nucleus transcription.

Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-10-196-S4.txt]

Additional file 5

MIPS complexes dataset.

Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-10-196-S5.zip]

Additional file 6

A script for converting MIPS complexes to a SPOKE model network.

Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-10-196-S6.zip]

Acknowledgements

This work was funded by the EU Sealife project. Joerg Hakenberg helped with Gene Ontology similarity. Rainer Winnenburg helped with discussions.

References

1. Chiang T, Scholtens D, Sarkar D, Gentleman R, Huber W: **Coverage and error models of protein-protein interaction data by directed graph analysis.** *Genome Biol* 2007, **8(9):**R186.
2. Yip KY, Gerstein M: **Training set expansion: an approach to improving the reconstruction of biological networks from limited and uneven reliable interactions.** *Bioinformatics* 2009, **25(2):**243-50.
3. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck F, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S, Timm J, Mintzlaff S, Abraham C, Bock N, Kietzmann S, Goedde A, Toksöz E, Droege A, Krobitsch S, Korn B, Birchmeier W, Lehrach H, Wanker E: **A human protein-protein interaction network: a resource for annotating the proteome.** *Cell* 2005, **122(6):**957-68.
4. 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: **Towards a proteome-scale map of the human protein-protein interaction network.** *Nature* 2005, **437(7062):**1173-1178.
5. Hart GT, Ramani AK, Marcotte EM: **How complete are current yeast and human protein-interaction networks?** *Genome_Biol* 2006, **7(11):**120.
6. Hoffmann R, Valencia A: **Protein interaction: same network, different hubs.** *Trends Genet* 2003, **19(12):**681-3.
7. 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(5):**349-56.
8. Aloy P: **Shaping the future of interactome networks.** *Genome Biol* 2007, **8(10):**316.
9. Scott MS, Barton GJ: **Probabilistic prediction and ranking of human protein-protein interactions.** *BMC Bioinformatics* 2007, **8:**239.
10. Stumpf MPH, Thorne T, Silva EdS, Stewart R, An HJ, Lappe M, Wiuf C: **Estimating the size of the human interactome.** *Proc Natl Acad Sci USA* 2008, **105(19):**6959-64.
11. D'haeseleer P, Church GM: **Estimating and improving protein interaction error rates.** *Proc IEEE Comput Syst_Bioinform Conf* 2004:216-23.
12. Sprinzak E, Sattath S, Margalit H: **How reliable are experimental protein-protein interaction data?** *J Mol Biol* 2003, **327(5):**919-923.
13. Bader GD, Hogue CWV: **Analyzing yeast protein-protein interaction data obtained from different sources.** *Nat Biotechnol* 2002, **20(10):**991-997.
14. Zhang Y, Xuan J, los Reyes BGdlR, Clarke R, Ressom HW: **Network motif-based identification of transcription factor-target gene relationships by integrating multi-source biological data.** *BMC Bioinformatics* 2008, **9:**203.
15. Sprinzak E, Altuvia Y, Margalit H: **Characterization and prediction of protein-protein interactions within and between complexes.** *Proc Natl Acad Sci USA* 2006, **103(40):**14718-23.
16. Edwards A, Kus B, Jansen R, Greenbaum D, Greenblatt J, Gerstein M: **Bridging structural biology and genomics: assessing protein interaction data with known complexes.** *Trends Genet* 2002, **18(10):**529-36.
17. Singh R, Xu J, Berger B: **Struct2net: integrating structure into protein-protein interaction prediction.** *Pac Symp Biocomput* 2006:403-14.
18. Kim W, Park J, Suh J: **Large scale statistical prediction of protein-protein interaction by potentially interacting domain (PID) pair.** *Genome Inform* 2002, **13:**42-50.
19. Bader J, Chaudhuri A, Rothberg J, Chant J: **Gaining confidence in high-throughput protein interaction networks.** *Nat Biotechnol* 2004, **22:**78-85.
20. Tong A, Drees B, Nardelli G, Bader G, Brannetti B, Castagnoli L, Evangelista M, Ferracuti S, Nelson B, Paoluzi S, Quondam M, Zucconi A, Hogue C, Fields S, Boone C, Cesareni G: **A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules.** *Science* 2002, **295(5553):**321-4.
21. Espadaler J, Romero-Isart O, Jackson R, Oliva B: **Prediction of protein-protein interactions using distant conservation of**

sequence patterns and structure relationships. *Bioinformatics* 2005, **21(16):**3360-8.
22. Ramirez F, Schlicker A, Assenov Y, Lengauer T, Albrecht M: **Computational analysis of human protein interaction networks.** *Proteomics* 2007, **7(15):**2541-2552.
23. Singhal M, Resat H: **A domain-based approach to predict protein-protein interactions.** *BMC Bioinformatics* 2007, **8:**199.
24. Chen X, Liu M: **Prediction of protein-protein interactions using random decision forest framework.** *Bioinformatics* 2005, **21(24):**4394-400.
25. Wojcik J, Schachter V: **Protein-protein interaction map inference using interacting domain profile pairs.** *Bioinformatics* 2001, **17(Suppl 1):**S296-305.
26. Patil A, Nakamura H: **Filtering high-throughput protein-protein interaction data using a combination of genomic features.** *BMC Bioinformatics* 2005, **6:**100.
27. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U: **Network motifs: simple building blocks of complex networks.** *Science* 2002, **298(5594):**824-7.
28. Zhang L, King O, Wong S, Goldberg D, Tong A, Lesage G, Andrews B, Bussey H, Boone C, Roth F: **Motifs, themes and thematic maps of an integrated Saccharomyces cerevisiae interaction network.** *J Biol* 2005, **4(2):**6.
29. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U: **Network motifs: simple building blocks of complex networks.** *Science* 2002, **298(5594):**824-827.
30. Kashtan N, Alon U: **Spontaneous evolution of modularity and network motifs.** *Proc Natl Acad Sci USA* 2005, **102(39):**13773-13778.
31. Albert I, Albert R: **Conserved network motifs allow protein-protein interaction prediction.** *Bioinformatics* 2004, **20(18):**3346-3352.
32. Jin G, Zhang S, Zhang X, Chen L: **Hubs with network motifs organize modularity dynamically in the protein-protein interaction network of yeast.** *PLoS_ONE* 2007, **2(11):**e1207.
33. Kim WK, Park J, Suh JK: **Large scale statistical prediction of protein-protein interaction by potentially interacting domain (PID) pair.** *Genome Inform* 2002, **13:**42-50.
34. Ng SK, Zhang Z, Tan SH: **Integrative approach for computationally inferring protein domain interactions.** *Bioinformatics* 2003, **19(8):**923-929.
35. Aloy P, Russell RB: **InterPreTS: protein interaction prediction through tertiary structure.** *Bioinformatics* 2003, **19:**161-162.
36. Riley R, Lee C, Sabatti C, Eisenberg D: **Inferring protein domain interactions from databases of interacting proteins.** *Genome_Biol* 2005, **6(10):**R89.
37. Guimaraes KS, Jothi R, Zotenko E, Przytycka TM: **Predicting domain-domain interactions using a parsimony approach.** *Genome_Biol* 2006, **7(11):**R104.
38. Deng M, Mehta S, Sun F, Chen T: **Inferring domain-domain interactions from protein-protein interactions.** *Genome_Res* 2002, **12(10):**1540-1548.
39. Jothi R, Cherukuri PF, Tasneem A, Przytycka TM: **Co-evolutionary analysis of domains in interacting proteins reveals insights into domain-domain interactions mediating protein-protein interactions.** *J Mol_Biol* 2006, **362(4):**861-875.
40. Xia K, Fu Z, Hou L, Han JDJ: **Impacts of protein-protein interaction domains on organism and network complexity.** *Genome_Res* 2008, **18(9):**1500-8.
41. Cohen-Gihon I, Nussinov R, Sharan R: **Comprehensive analysis of co-occurring domain sets in yeast proteins.** *BMC_Genomics* 2007, **8:**161.
42. Nye TMW, Berzuini C, Gilks WR, Babu MM, Teichmann SA: **Statistical analysis of domains in interacting protein pairs.** *Bioinformatics* 2005, **21(7):**993-1001.
43. Nye TMW, Berzuini C, Gilks WR, Babu MM, Teichmann S: **Predicting the strongest domain-domain contact in interacting protein pairs.** *Stat_Appl_Genet_Mol_Biol* 2006, **5:**.
44. Liu Y, Liu N, Zhao H: **Inferring protein-protein interactions through high-throughput interaction data from diverse organisms.** *Bioinformatics* 2005, **21(15):**3279-3285.
45. Itzhaki Z, Akiva E, Altuvia Y, Margalit H: **Evolutionary conservation of domain-domain interactions.** *Genome_Biol* 2006, **7(12):**R125.
46. Jothi R, Cherukuri PF, Tasneem A, Przytycka TM: **Co-evolutionary analysis of domains in interacting proteins reveals insights into domain-domain interactions mediating protein-protein interactions.** *J_Mol_Biol* 2006, **362(4):**861-75.
47. Iqbal M, Freitas AA, Johnson CG, Vergassola M: **Message-passing algorithms for the prediction of protein domain interactions from protein-protein interaction data.** *Bioinformatics* 2008, **24(18):**2064-70.
48. Wang RS, Wang Y, Wu LY, Zhang XS, Chen L: **Analysis on multi-domain cooperation for predicting protein-protein interactions.** *BMC Bioinformatics* 2007, **8:**391.
49. Wuchty S: **Topology and weights in a protein domain interaction network-a novel way to predict protein interactions.** *BMC Genomics* 2006, **7:**122.
50. Luo F, Yang Y, Chen C, Chang R, Zhou J, Scheuermann R: **Modular organization of protein interaction networks.** *Bioinformatics* 2007, **23(2):**207-14.
51. Gagneur J, Krause R, Bouwmeester T, Casari G: **Modular decomposition of protein-protein interaction networks.** *Genome_Biol* 2004, **5(8):**R57.
52. Pawson T: **Organization of cell-regulatory systems through modular-protein-interaction domains.** *Philos Transact A Math Phys Eng Sci* 2003, **361(1807):**1251-62.
53. Poyatos J, Hurst L: **How biologically relevant are interaction-based modules in protein networks?** *Genome Biol* 2004, **5(11):**R93.
54. Rives A, Galitski T: **Modular organization of cellular networks.** *Proc Natl Acad Sci USA* 2003, **100(3):**1128-33.
55. Lu H, Shi B, Wu G, Zhang Y, Zhu X, Zhang Z, Liu C, Zhao Y, Wu T, Wang J, Chen R: **Integrated analysis of multiple data sources reveals modular structure of biological networks.** *Biochem Biophys Res Commun* 2006, **345:**302-9.
56. Qi Y, Klein-Seetharaman J, Bar-Joseph Z: **A mixture of feature experts approach for protein-protein interaction prediction.** *BMC Bioinformatics* 2007, **8(Suppl 10):**S6.
57. Qi Y, Bar-Joseph Z, Klein-Seetharaman J: **Evaluation of different biological data and computational classification methods for use in protein interaction prediction.** *Proteins* 2006, **63(3):**490-500.
58. Beyer A, Workman C, Hollunder J, Radke D, Mueller U, Wilhelm T, Ideker T: **Integrated assessment and prediction of transcription factor binding.** *PLoS Comput Biol* 2006, **2(6):**e70.
59. Lin N, Wu B, Jansen R, Gerstein M, Zhao H: **Information assessment on predicting protein-protein interactions.** *BMC Bioinformatics* 2004, **5:**154.
60. Aloy P, Russell R: **Structural systems biology: modelling protein interactions.** *Nat Rev Mol Cell Biol* 2006, **7(3):**188-197.
61. Schlicker A, Huthmacher C, Ramirez F, Lengauer T, Albrecht M: **Functional evaluation of domain-domain interactions and human protein interaction networks.** *Bioinformatics* 2007, **23(7):**859-865.
62. Andreopoulos B, An A, Wang X, Faloutsos M, Schroeder M: **Clustering by common friends finds locally significant proteins mediating modules.** *Bioinformatics* 2007, **23(9):**1124-31.
63. Li H, Li J, Wong L: **Discovering motif pairs at interaction sites from protein sequences on a proteome-wide scale.** *Bioinformatics* 2006, **22(8):**989-996.
64. Okada K, Kanaya S, Asai K: **Accurate extraction of functional associations between proteins based on common interaction partners and common domains.** *Bioinformatics* 2005, **21(9):**2043-8.
65. Goh , Bogan , Joachimiak , Walther , Cohen : **Co-evolution of proteins with their interaction partners.** *JMB* 2000, **299(2):**283-93.
66. Chua HN, Ning K, Sung WK, Leong HW, Wong L: **Using indirect protein-protein interactions for protein complex prediction.** *J Bioinform Comput Biol* 2008, **6(3):**435-66.
67. Chua HN, Sung WK, Wong L: **Using indirect protein interactions for the prediction of Gene Ontology functions.** *BMC Bioinformatics* 2007, **8(Suppl 4):**S8.
68. Yu H, Paccanaro A, Trifonov V, Gerstein M: **Predicting interactions in protein networks by completing defective cliques.** *Bioinformatics* 2006, **22(7):**823-829.
69. Morrison JL, Breitling R, Higham DJ, Gilbert DR: **A lock-and-key model for protein-protein interactions.** *Bioinformatics* 2006, **22(16):**2012-9.
70. Zhang S, Ning X, Zhang X: **Identification of functional modules in a PPI network by clique percolation clustering.** *Comput Biol Chem* 2006, **30(6):**445-51.

71. Chua HN, Sung WK, Wong L: **Exploiting indirect neighbours and topological weight to predict protein function from protein-protein interactions.** *Bioinformatics* 2006, **22(13):**1623-30.
72. Vázquez A, Dobrin R, Sergi D, Eckmann JP, Oltvai ZN, Barab'asi AL: **The topological relationship between the large-scale attributes and local interaction patterns of complex networks.** *Proc Natl Acad Sci USA* 2004, **101(52):**17940-17945.
73. Lo SL, Cai CZ, Chen YZ, Chung MCM: **Effect of training datasets on support vector machine prediction of protein-protein interactions.** *Proteomics* 2005, **5(4):**876-84.
74. Resendis-Antonio O, Freyre-Gonzalez JA, Menchaca-Mandez R, Gutierrez-Rios RM, Martinez-Antonio A, Avila-Sanchez C, Collado-Vides J: **Modular analysis of the transcriptional regulatory network of E. coli.** *Trends Genet* 2005, **21:**16-20.
75. Shen-Orr S, Milo R, Mangan S, Alon U: **Network motifs in the transcriptional regulation network of Escherichia coli.** *Nat Genet* 2002, **31:**64-8.
76. Wuchty S, Oltvai ZN, Barab'asi AL: **Evolutionary conservation of motif constituents in the yeast protein interaction network.** *Nat Genet* 2003, **35(2):**176-179.
77. Clauset A, Moore C, Newman MEJ: **Hierarchical structure and the prediction of missing links in networks.** *Nature* 2008, **453(7191):**98-101.
78. Watts DJ, Strogatz SH: **Collective dynamics of 'small-world' networks.** *Nature* 1998, **393(6684):**440-442.
79. Yu J, Fotouhi F: **Computational approaches for predicting protein-protein interactions: a survey.** *J Med Syst* 2006, **30:**39-44.
80. Valencia A, Pazos F: **Computational methods for the prediction of protein interactions.** *Curr Opin Struct Biol* 2002, **12(3):**368-73.
81. von Mering C, J Jensen L, Kuhn M, Chaffron S, Doerks T, Kruger B, Snel B, Bork P: **STRING 7-recent developments in the integration and prediction of protein interactions.** *Nucleic Acids Res* 2007:D358-62.
82. Ben-Hur A, Noble WS: **Kernel methods for predicting protein-protein interactions.** *Bioinformatics* 2005, **21(Suppl 1):**i38-46.
83. Guo Y, Yu L, Wen Z, Li M: **Using support vector machine combined with auto covariance to predict protein-protein interactions from protein sequences.** *Nucleic Acids Res* 2008, **36(9):**3025-30.
84. Alber F, Dokudovskaya S, Veenhoff L, Zhang W, Kipper J, Devos D, Suprapto A, Karni-Schmidt O, Williams R, Chait B, Rout M, Sali A: **Determining the architectures of macromolecular assemblies.** *Nature* 2007, **450(7170):**683-94.
85. Rhodes D, Tomlins S, Varambally S, Mahavisno V, Barrette T, Kalyana-Sundaram S, Ghosh D, Pandey A, Chinnaiyan A: **Probabilistic model of the human protein-protein interaction network.** *Nat Biotechnol* 2005, **23(8):**951-9.
86. Huang T, Tien A, Huang W, Lee Y, Peng C, Tseng H, Kao C, Huang C: **POINT: a database for the prediction of protein-protein interactions based on the orthologous interactome.** *Bioinformatics* 2004, **20(17):**3273-6.
87. Patil A, Nakamura H: **Filtering high-throughput protein-protein interaction data using a combination of genomic features.** *BMC Bioinformatics* 2005, **6:**100.
88. Chen P, Deane C, Reinert G: **Predicting and Validating Protein Interactions Using Network Structure.** *PLoS Comput Biol* 2008, **4(7):**.
89. Pitre S, Dehne F, Chan A, Cheetham J, Duong A, Emili A, Gebbia M, Greenblatt J, Jessulat M, Krogan N, Luo X, Golshani A: **PIPE: a protein-protein interaction prediction engine based on the re-occurring short polypeptide sequences between known interacting protein pairs.** *BMC Bioinformatics* 2006, **7:**365.
90. Ng S, Zhang Z, Tan S: **Integrative approach for computationally inferring protein domain interactions.** *Bioinformatics* 2003, **19(8):**923-9.
91. Wu X, Zhu L, Guo J, Zhang DY, Lin K: **Prediction of yeast protein-protein interaction network: insights from the Gene Ontology and annotations.** *Nucleic Acids Res* 2006, **34(7):**2137-50.
92. Chinnasamy A, Mittal A, Sung WK: **Probabilistic prediction of protein-protein interactions from the protein sequences.** *Comput Biol Med* 2006, **36(10):**1143-54.
93. Jansen R, Yu H, Greenbaum D, Kluger Y, Krogan N, Chung S, Emili A, Snyder M, Greenblatt J, Gerstein M: **A Bayesian networks approach for predicting protein-protein interactions from genomic data.** *Science* 2003, **302(5644):**449-53.
94. Han DS, Kim HS, Jang WH, Lee SD, Suh JK: **PreSPI: a domain combination based prediction system for protein-protein interaction.** *Nucleic Acids Res* 2004, **32(21):**6312-20.
95. 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: **An experimentally derived confidence score for binary protein-protein interactions.** *Nat Methods* 2009, **6:**91-7.
96. Yu J, Finley RLJ: **Combining multiple positive training sets to generate confidence scores for protein-protein interactions.** *Bioinformatics* 2009, **25:**105-11.
97. Mathivanan S, Periaswamy B, Gandhi T, Kandasamy K, Suresh S, Mohmood R, Ramachandra Y, Pandey A: **An evaluation of human protein-protein interaction data in the public domain.** *BMC Bioinformatics* 2006, **7(Suppl 5):**S19.
98. Galperin MY, Cochrane GR: **Nucleic Acids Research annual Database Issue and the NAR online Molecular Biology Database Collection in 2009.** *Nucleic Acids Res* 2009:D1-4.
99. 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: **An empirical framework for binary interactome mapping.** *Nat Methods* 2009, **6:**83-90.
100. Cusick ME, Yu H, Smolyar A, Venkatesan K, Carvunis AR, Simonis N, Rual JF, Borick H, Braun P, Dreze M, Vandenhaute J, Galli M, Yazaki J, Hill DE, Ecker JR, Roth FP, Vidal M: **Literature-curated protein interaction datasets.** *Nat Methods* 2009, **6:**39-46.
101. Bader GD, Hogue CWV: **An automated method for finding molecular complexes in large protein interaction networks.** *BMC Bioinformatics* 2003, **4:**2.
102. Cakmak A, Ozsoyoglu G: **Mining biological networks for unknown pathways.** *Bioinformatics* 2007, **23(20):**2775-83.
103. Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis AP, Punna T, Peregrín-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, Onge PS, Ghanny S, Lam MHY, 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: **Global landscape of protein complexes in the yeast Saccharomyces cerevisiae.** *Nature* 2006, **440(7084):**637-643.
104. Gavin A, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen L, Bastuck S, Dümpelfeld B, Edelmann A, Heurtier M, Hoffman V, Hoefert C, Klein K, Hudak M, Michon A, Schelder M, Schirle M, Remor M, Rudi T, Hooper S, Bauer A, Bouwmeester T, Casari G, Drewes G, Neubauer G, Rick J, Kuster B, Bork P, Russell R, Superti-Furga G: **Proteome survey reveals modularity of the yeast cell machinery.** *Nature* 2006, **440(7084):**631-6.
105. Mewes H, Frishman D, Mayer K, Muensterkoetter M, Noubibou O, Pagel P, Rattei T, Oesterheld M, Ruepp A, Stuempflen V: **MIPS: analysis and annotation of proteins from whole genomes in 2005.** *Nucleic Acids Res* 2006:D169-72.
106. Pagel P, Kovac S, Oesterheld M, Brauner B, Dunger-Kaltenbach I, Frishman G, Montrone C, Mark P, Stuempflen V, Mewes HW, Ruepp A, Frishman D: **The MIPS mammalian protein-protein interaction database.** *Bioinformatics* 2005, **21(6):**832-4.
107. Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Buillard V, Cerutti L, Copley R, Courcelle E, Das U, Daugherty L, Dibley M, Finn R, Fleischmann W, Gough J, Haft D, Hulo N, Hunter S, Kahn D, Kanapin A, Kejariwal A, Labarga A, Langendijk-Genevaux PS, Lonsdale D, Lopez R, Letunic I, Madera M, Maslen J, McAnulla C, McDowall J, Mistry J, Mitchell A, Nikolskaya AN, Orchard S, Orengo C, Petryszak R, Selengut JD, Sigrist CJA, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C: **New developments in the InterPro database.** *Nucleic Acids Res* 2007:D224-D228.
108. Galletta B, Chuang D, Cooper J: **Distinct Roles for Arp2/3 Regulators in Actin Assembly and Endocytosis.** *PLoS Biol* 2008, **6:**e1.
109. Kim PM, Lu LJ, Xia Y, Gerstein MB: **Relating three-dimensional structures to protein networks provides evolutionary insights.** *Science* 2006, **314(5807):**1938-1941.

110. Lanerolle PdL, Johnson T, Hofmann WA: **Actin and myosin I in the nucleus: what next?** *Nat Struct Mol Biol* 2005, **12(9):**742-6.
111. Evangelista M, Klebl B, Tong A, Webb B, Leeuw T, Leberer E, Whiteway M, Thomas D, Boone C: **A role for myosin-I in actin assembly through interactions with Vrp1p, Bee1p, and the Arp2/3 complex.** *J Cell Biol* 2000, **148(2):**353-62.
112. Toi H, Fujimura-Kamada K, Irie K, Takai Y, Todo S, Tanaka K: **She4p/Dim1p interacts with the motor domain of unconventional myosins in the budding yeast, Saccharomyces cerevisiae.** *Mol Biol Cell* 2003, **14(6):**2237-49.
113. Pruyne D, Evangelista M, Yang C, Bi E, Zigmond S, Bretscher A, Boone C: **Role of formins in actin assembly: nucleation and barbed-end association.** *Science* 2002, **297(5581):**612-5.
114. Evangelista M, Pruyne D, Amberg D, Boone C, Bretscher A: **Formins direct Arp2/3-independent actin filament assembly to polarize cell growth in yeast.** *Nat Cell Biol* 2002, **4(3):**260-9.
115. Park H, Kang P, Rachfal A: **Localization of the Rsr1/Bud1 GTPase involved in selection of a proper growth site in yeast.** *J Biol Chem* 2002, **277(30):**26721-4.
116. Zakrzewska E, Perron M, Laroche A, Pallotta D: **A role for GEA1 and GEA2 in the organization of the actin cytoskeleton in Saccharomyces cerevisiae.** *Genetics* 2003, **165(3):**985-95.
117. Evangelista M, Blundell K, Longtine M, Chow C, Adames N, Pringle J, Peter M, Boone C: **Bni1p, a yeast formin linking cdc42p and the actin cytoskeleton during polarized morphogenesis.** *Science* 1997, **276(5309):**118-22.
118. Yanagihara C, Shinkai M, Kariya K, Yamawaki-Kataoka Y, Hu CD, Masuda T, Kataoka T: **Association of elongation factor 1 alpha and ribosomal protein L3 with the proline-rich region of yeast adenylyl cyclase-associated protein CAP.** *Biochem Biophys Res Commun* 1997, **232(2):**503-7.
119. Nelson WJ: **Adaptation of core mechanisms to generate cell polarity.** *Nature* 2003, **422(6933):**766-74.
120. Lambert AA, Perron MP, Lavoie E, Pallotta D: **The Saccharomyces cerevisiae Arf3 protein is involved in actin cable and cortical patch formation.** *FEMS Yeast Res* 2007, **7(6):**782-95.
121. Bettinger BT, Gilbert DM, Amberg DC: **Actin up in the nucleus.** *Nat Rev Mol Cell Biol* 2004, **5(5):**410-5.
122. Pederson T, Aebi U: **Actin in the nucleus: what form and what for?** *J Struct Biol* 2002, **140(1–3):**3-9.
123. Olave IA, Reck-Peterson SL, Crabtree GR: **Nuclear actin and actin-related proteins in chromatin remodeling.** *Annu Rev Biochem* 2002, **71:**755-81.
124. Franke WW: **Actin's many actions start at the genes.** *Nat Cell Biol* 2004, **6(11):**1013-4.
125. Hofmann WA, Stojiljkovic L, Fuchsova B, Vargas GM, Mavrommatis E, Philimonenko V, Kysela K, Goodrich JA, Lessard JL, Hope TJ, Hozak P, Lanerolle PdL: **Actin is part of pre-initiation complexes and is necessary for transcription by RNA polymerase II.** *Nat Cell Biol* 2004, **6(11):**1094-101.
126. Pina B, Fernandez-Larrea J, Garcia-Reyero N, Idrissi FZ: **The different (sur)faces of Rap1p.** *Mol Genet Genomics* 2003, **268(6):**791-8.
127. Holden JL, Nur-E-Kamal MS, Fabri L, Nice E, Hammacher A, Maruta H: **Rsr1 and Rap1 GTPases are activated by the same GTPase-activating protein and require threonine 65 for their activation.** *J Biol Chem* 1991, **266(26):**16992-5.
128. Lecuyer E, Yoshida H, Parthasarathy N, Alm C, Babak T, Cerovina T, Hughes T, Tomancak P, Krause H: **Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function.** *Cell* 2007, **131:**174-87.
129. Long RM, Singer RH, Meng X, Gonzalez I, Nasmyth K, Jansen RP: **Mating type switching in yeast controlled by asymmetric localization of ASH1 mRNA.** *Science* 1997, **277(5324):**383-7.
130. Stearns T, Kahn RA, Botstein D, Hoyt MA: **ADP ribosylation factor is an essential protein in Saccharomyces cerevisiae and is encoded by two genes.** *Mol Cell Biol* 1990, **10(12):**6690-9.
131. Nilsson J, Nissen P: **Elongation factors on the ribosome.** *Curr Opin Struct Biol* 2005, **15(3):**349-54.
132. Scholtens D, Chiang T, Huber W, Gentleman R: **Estimating node degree in bait-prey graphs.** *Bioinformatics* 2008, **24(2):**218-24.
133. Schlicker A, Domingues FS, Rahnenführer J, Lengauer T: **A new measure for functional similarity of gene products based on Gene Ontology.** *BMC Bioinformatics* 2006, **7:**302.
134. Winter C, Henschel A, Kim W, Schroeder M: **SCOPPI: a structural classification of protein-protein interfaces.** *Nucleic Acids Res* 2006:D310-4.
135. Andreopoulos B, An A, Wang X: **Hierarchical density-based clustering of categorical data and a simplification.** In Proceedings of the 11th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2007), Springer LNCS 4426:11-22. Nanjing, China, May 22-25, 2007
136. Beyer A, Bandyopadhyay S, Ideker T: **Integrating physical and genetic maps: from genomes to interaction networks.** *Nat Rev Genet* 2007, **8(9):**699-710.
137. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, Hanspers K, Isserlin R, Kelley R, Killcoyne S, Lotia S, Maere S, Morris J, Ono K, Pavlovic V, Pico AR, Vailaya A, Wang PL, Adler A, Conklin BR, Hood L, Kuiper M, Sander C, Schmulevich I, Schwikowski B, Warner GJ, Ideker T, Bader GD: **Integration of biological networks and gene expression data using Cytoscape.** *Nat Protoc* 2007, **2(10):**2366-82.
138. Emig D, Cline MS, Lengauer T, Albrecht M: **Integrating expression data with domain interaction networks.** *Bioinformatics* 2008, **24(21):**2546-8.
139. Schuster-Bockler B, Bateman A: **Reuse of structural domain-domain interactions in protein networks.** *BMC Bioinformatics* 2007, **8:**259.
140. Aragues R, Sali A, Bonet J, Marti-Renom MA, Oliva B: **Characterization of protein hubs by inferring interacting motifs from protein interactions.** *PLoS Comput Biol* 2007, **3(9):**1761-71.
141. McGuffin LJ, Street SA, Bryson K, Soerensen SA, Jones DT: **The Genomic Threading Database: a comprehensive resource for structural annotations of the genomes from key organisms.** *Nucleic Acids Res* 2004:D196-9.
142. The GO Consortium: **The Gene Ontology (GO) project in 2006.** *Nucleic Acids Research* 2005:D322-6.