

Protein interaction: same network, different hubs

Robert Hoffmann and Alfonso Valencia.

Supplementary information to *Trends in Genetics* article

Data sources

Table 1 gives an overview of the analysed data sets. To compare our results with previous analyses [1,2] we used the interaction data published by Mering *et al.* [1] for all methods, including the reference set. Please refer to the supplementary information of Mering *et al.* [3] for the exact preparation and filtering of data and the parameter choice for the prediction methods: <http://www.nature.com/nature/journal/v417/n6887/supinfo/nature750.html>. We distinguish experimental high-throughput methods (experimental) from prediction methods (*in silico*).

Table 1 Experimental and in silico methods for large-scale prediction of protein interactions

Method	Description	Preparation and filtering ^a	Type
HMS	High-throughput mass spectrometric protein complex identification [4]	Filtered datasets (removed 'sticky' proteins and components of the ribosome); connections assigned between all proteins present in a purification; no additional manual curation	Experimental
TAP	Tandem affinity purification [5]	Filtered datasets (removed 'sticky' proteins and components of the ribosome); connections assigned between all proteins present in a purification; no additional manual curation	Experimental
Y2H	Yeast two-hybrid [6,7]	Overlapping interactions were counted only once; homotypic interactions were not counted	Experimental
CE	Correlated mRNA expression [8,9]	Pearson correlation coefficient to measure the similarity of their expression profiles	Experimental
SL	Synthetic lethal interactions [10]	High-throughput study on genetic interactions	Experimental
GN	Conserved gene neighbourhood [11,12]	42 completely sequenced genomes; two or more genes required to have the same orientation on the chromosome and to be in a 'run' with intergenic regions of no more than 300 bp	<i>In silico</i>
CO	Co-occurrence of genes [13,14]	Pattern of occurrence among 42 completely sequenced genomes	<i>In silico</i>
GF	Gene fusion events [15]	Presence of a gene in more than one COG cluster	<i>In silico</i>
ANNOT	Reference set (MIPS [16], YPD [17])	Reference assembled from a set of known interactions from two catalogs of protein complexes in Yeast (MIPS and YPD); YPD-data was taken from http://www.incyte.com/	Manual

^aFor details on preparation and filtering, see Ref. [3].

Computation of connectivity

The connectivity (number of interactions) for all proteins was computed from previously published and described interaction data [1]: <http://www.nature.com/nature/journal/v417/n6887/supinfo/nature750.html>. A reformatted version of the output data from Mering *et al.* is available at http://www.pdg.cnb.uam.es/supplement/protein_network/interaction.txt. The connectivity of an individual protein was calculated as the sum of all its interactions for a given method. This calculation was carried out separately for all methods. The result is shown in part in Table 2. All in all, connectivity data for 5295 proteins were used for further correlation analysis.

Table 2^{a,b} Protein connectivities in different methods

PROTEIN (ORF)	ANNOT	HMS	Y2H	TAP	GN	CE	SL	CO	GF
Q0032		3							
Q0045	2				14			4	1
Q0050		16							
Q0085	6				10			5	
Q0092		2							

^aList is truncated. The complete list (5296 lines) is available at http://www.pdg.cnb.uam.es/supplement/protein_network/doi.txt (tab-delimited flat file).

^bAbbreviations: ANNOT, reference set (MIPS and YPD); HMS, high-throughput mass spectrometric protein complex identification; TAP, tandem affinity purification; Y2H, yeast two-hybrid; CE, correlated mRNA expression; SL, synthetic lethal interaction; GN, conserved gene neighbourhood; CO, co-occurrence of genes; GF gene fusion events.

The complete connectivity data are available at http://www.pdg.cnb.uam.es/supplement/protein_network/doi.txt.

Correlation analysis of protein connectivity

To detect and quantify a common tendency between two methods, we assessed the correlation of protein connectivity (Table 2) between them. All methods were compared in pairs, taking into account only those proteins that had been covered by both methods. Because the distribution of connectivity follows approximately a power law (i.e. assumption of normal distribution not valid), a nonparametric correlation analysis according to Spearman's rho was applied and correlation coefficients were tested for significance at level 0.01 and 0.05 (two-tailed) [18]. Results of all correlations are shown as a matrix in Table 3.

A weak though significant correlation was found, for instance, between the two complex purification methods (TAP and HMS). Interestingly, a significant correlation between TAP and HMS even remains when overlapping interactions are excluded from the preceding calculation of connectivity. In other words, both methods assign, independently of each other, similar connectivities for the same proteins: both predict many interactions for hub proteins and fewer interactions for specifically interacting proteins.

Table 3. Spearman's rho^{a,b}

		ANNOT	HMS	TAP	Y2H	CE	SL	GN	CO	GF
ANNOT	Correlation coefficient	1	0.168 ^c	0.421 ^c	0.021	0.035	0.114	0.593 ^c	0.045	-0.338 ^d
	Significance		0.002	0	0.657	0.581	0.075	0	0.756	0.025
	N	777	345	497	465	251	247	230	50	44
HMS	Correlation coefficient	0.168 ^c	1	0.253 ^c	-0.018	-0.068	0.009	-0.042	0.103	-0.053
	Significance	0.002		0	0.583	0.130	0.876	0.478	0.266	0.600
	N	345	1577	659	957	503	304	292	119	102
TAP	Correlation coefficient	0.421 ^c	0.253 ^c	1	-0.048	0.003	0.015	0.214 ^c	-0.077	-0.041
	Significance	0	0		0.170	0.953	0.797	0.001	0.484	0.741
	N	497	659	1375	804	420	288	260	85	67
Y2H	Correlation coefficient	0.021	-0.018	-0.048	1	-0.038	0.112 ^d	-0.092 ^d	-0.073	0.065
	Significance	0.657	0.583	0.170		0.207	0.026	0.033	0.300	0.424
	N	465	957	804	3571	1114	397	540	201	154
CE	Correlation coefficient	0.035	-0.068	0.003	-0.038	1	0.180 ^d	-0.028	0.148	0.029
	Significance	0.581	0.130	0.953	0.207		0.024	0.577	0.069	0.771
	N	251	503	420	1114	1941	157	387	152	103
SL	Correlation coefficient	0.114	0.009	0.015	0.112 ^d	0.180 ^d	1	-0.097	-0.431	0.214
	Significance	0.075	0.876	0.797	0.026	0.024		0.409	0.109	0.284
	N	247	304	288	397	157	678	74	15	27
GN	Correlation coefficient	0.593 ^c	-0.042	0.214 ^c	-0.092 ^d	-0.028	-0.097	1	0.297 ^c	0.237 ^c
	Significance	0	0.478	0.001	0.033	0.577	0.409		0	0
	N	230	292	260	540	387	74	998	311	260
CO	Correlation coefficient	0.045	0.103	-0.077	-0.073	0.148	-0.431	0.297 ^c	1	0.255 ^c
	Significance	0.756	0.266	0.484	0.300	0.069	0.109	0		0.006
	N	50	119	85	201	152	15	311	378	114

GF	Correlation coefficient	-0.338 ^d	-0.053	-0.041	0.065	0.029	0.214	0.237 ^c	0.255 ^c	1
	Significance	0.025	0.600	0.741	0.424	0.771	0.284	0	0.006	
	N	44	102	67	154	103	27	260	114	293

^a Abbreviations: ANNOT, reference set (MIPS and YPD); HMS, high-throughput mass spectrometric protein complex identification; TAP, tandem affinity purification; Y2H, yeast two-hybrid; CE, correlated mRNA expression; SL, synthetic lethal interaction; GN, conserved gene neighbourhood; CO, co-occurrence of genes; GF gene fusion events; N, number of cases.

^b For details on Spearman's rho, see Ref. [18].

^c Correlation significant at 0.01 level (two-tailed) [19].

^d Correlation significant at 0.05 level (two-tailed) [19].

Results of the correlation analysis according to Kendall [18] are shown for comparison purposes in Table 4.

Table 4. Kendall's tau-b^{a,b}

		ANNOT	HMS	TAP	Y2H	CE	SL	GN	CO	GF
ANNOT	Correlation coefficient	1	0.125 ^c	0.329 ^c	0.017	0.027	0.092	0.448 ^c	0.058	-0.261 ^d
	Significance		0.001	0	0.644	0.550	0.066	0	0.597	0.031
	N	777	345	497	465	251	247	230	50	44
HMS	Correlation coefficient	0.125 ^c	1	0.175 ^c	-0.013	-0.047	0.006	-0.030	0.069	-0.037
	Significance	0.001		0	0.593	0.126	0.889	0.456	0.297	0.614
	N	345	1577	659	957	503	304	292	119	102
TAP	Correlation coefficient	0.329 ^c	0.175 ^c	1	-0.037	0.002	0.011	0.145 ^c	-0.053	-0.028
	Significance	0	0		0.165	0.942	0.800	0.001	0.500	0.764
	N	497	659	1375	804	420	288	260	85	67
Y2H	Correlation coefficient	0.017	-0.013	-0.037	1	-0.029	0.092 ^d	-0.071 ^d	-0.059	0.054
	Significance	0.644	0.593	0.165		0.205	0.026	0.034	0.299	0.421
	N	465	957	804	3571	1114	397	540	201	154
CE	Correlation coefficient	0.027	-0.047	0.002	-0.029	1	0.139 ^d	-0.022	0.109	0.023
	Significance	0.550	0.126	0.942	0.205		0.027	0.550	0.066	0.761
	N	251	503	420	1114	1941	157	387	152	103
SL	Correlation coefficient	0.092	0.006	0.011	0.092 ^d	0.139 ^d	1	-0.083	-0.354	0.195
	Significance	0.066	0.889	0.800	0.026	0.027		0.381	0.108	0.257
	N	247	304	288	397	157	678	74	15	27
GN	Correlation coefficient	0.448 ^c	-0.030	0.145 ^c	-0.071 ^d	-0.022	-0.083	1	0.225 ^c	0.181 ^c
	Significance	0	0.456	0.001	0.034	0.550	0.381		0	0
	N	230	292	260	540	387	74	998	311	260
CO	Correlation coefficient	0.058	0.069	-0.053	-0.059	0.109	-0.354	0.225 ^c	1	0.193 ^c
	Significance	0.597	0.297	0.500	0.299	0.066	0.108	0		0.008
	N	50	119	85	201	152	15	311	378	114
GF	Correlation coefficient	-0.261 ^d	-0.037	-0.028	0.054	0.023	0.195	0.181 ^c	0.193 ^c	1
	Significance	0.031	0.614	0.764	0.421	0.761	0.257	0	0.008	
	N	44	102	67	154	103	27	260	114	293

^aAbbreviations: ANNOT, reference set (MIPS and YPD); HMS, high-throughput mass spectrometric protein complex identification; TAP, tandem affinity purification; Y2H, yeast two-hybrid; CE, correlated mRNA expression; SL, synthetic lethal interaction; GN, conserved gene neighbourhood; CO, co-occurrence of genes; GF gene fusion events, N, number of cases.

^bFor more details on Kendall's tau-b, see Ref. [18]

^cCorrelation is significant at 0.01 level (two-tailed) [19].

^dCorrelation is significant at 0.05 level (two-tailed) [19].

Correlation between yeast two-hybrid data from different laboratories

Previous analyses have shown that the number of common protein pairs between yeast two-hybrid (Y2-H) and other methods is extremely low. This is also reflected at the level of network organization, by comparing the connectivities of individual proteins in Y2-H and other methods. However, at the level of network organization, reasonable consistency between yeast two-hybrid datasets from different laboratories can be found, something that was not detected by previous assessments [7] (Table 5).

Table 5. Spearman's rho^a

	UETZ [6]	ITO [7]
UETZ Correlation coefficient	1	0.184 ^b
Significance		0
N ^c	871	871
ITO Correlation coefficient	0.184 ^b	1
Significance	0	
N ^c	871	871

^aFor more details on Spearman's rho, see Ref. [18].

^bCorrelation significant at 0.01 level (two-tailed) [19].

^cN is the number of cases.

Functional preferences of methods

Interestingly, the *in silico* gene neighbourhood method (GN), which is based on evolutionary constraints, correlates with most other methods, linking experimental (i.e. TAP, HMS) and *in silico* predictions (e.g. CO).

We find that, from all common proteins in TAP and GN, the main part is classified as translation, transcription, genome maintenance and protein fate. This is complementary to the proteins that GN has in common with CO (co-occurrence of genes), which are mainly involved in energy production and amino acid metabolism.

This picture suggests that the correlation found between TAP and GN is based mainly on proteins involved in translation and transcription, whereas the correlation between GN and CO depends on proteins mainly involved in metabolism. Indeed, we find that the correlation between TAP and GN increases further when restricted to proteins involved in translation and transcription. The corresponding is true for the correlation between GN and CO (Table 6 and Table 7).

To be sure, this restriction was not carried out by chance, but following the complementary division of functional categories between $GN \cap TAP$ and $GN \cap CO$ (Figure 1b in *TIG* article).

Correlation between conserved gene neighbourhood and tandem affinity purification

Proteins are restricted to those functional categories that were found to be dominant in the proteins common to GN and TAP (see Figure 1b in article): genome maintenance, transcription, translation, protein fate. Adapted connectivity data available at http://www.pdg.cnb.uam.es/supplement/protein_network/doi_not_EGMU.txt.

Table 6. Spearman's rho^{a,b}

	ANNOT	TAP	HMS	GN
ANNOT Correlation coefficient	1	0.477 ^c	0.231 ^c	0.650 ^c
Significance		0	0	0
N	700	458	300	165
TAP Correlation coefficient	0.477 ^c	1	0.258 ^c	0.400 ^c
Significance	0		0	0

	N	458	937	450	127
	Correlation coefficient	0.231 ^c	0.258 ^c	1	0.104
HMS	Significance	0	0		0.281
	N	300	450	914	110
GN	Correlation coefficient	0.650 ^c	0.400 ^c	0.104	1
	Significance	0	0	0.281	
	N	165	127	110	371

^aAbbreviations: ANNOT, reference set (MIPS and YPD); HMS, high-throughput mass spectrometric protein complex identification; TAP, tandem affinity purification; GN, conserved gene neighbourhood; N, number of cases.

^bFor more details on Spearman's rho, see Ref. [18].

^cCorrelation significant at 0.01 level (two-tailed) [19].

Correlation between conserved gene neighbourhood and co-occurrence of genes

Proteins are restricted to those functional categories that were found to be dominant in the proteins common to GN and CO (Figure 1b in *TIG* article): energy production, amino acid metabolism, transport, stress/ defense. Adapted connectivity data available at http://www.pdg.cnb.uam.es/supplement/protein_network/doi_not_PTDFU.txt.

Table 7. Spearman's rho^{a,b}

	GN	CO	GF	
	Correlation coefficient	1	0.457 ^c	0.269 ^c
GN	Significance		0	0.001
	N	459	204	139
	Correlation coefficient	0.457 ^c	1	0.244 ^d
CO	Significance		0	0.035
	N	204	244	75
	Correlation coefficient	0.269 ^c	0.244 ^d	1
GF	Significance	0.001	0.035	
	N	139	75	165

^aAbbreviations: GN, conserved gene neighbourhood; CO, co-occurrence of genes; GF gene fusion events, N, number of cases.

^bFor more details on Spearman's rho, see Ref. [18].

^cCorrelation significant at 0.01 level (two-tailed) [19].

^dCorrelation significant at 0.05 level (two-tailed) [19].

For the complete assignment of each protein to a functional category please refer to the supplement material in Mering *et al.* [1]. <http://www.nature.com/nature/journal/v417/n6887/supinfo/nature750.html>.

Comparison of individual interactions

The overlap of individual interactions between pairs of methods is shown for comparison purposes (analogous to Mering *et al.* [1]). Please note that the number of proteins covered by both methods generally exceeds the number of overlapping interactions. Therefore, the comparison of protein connectivities can be based on more data than the comparison of overlapping interactions. Interestingly, although complex purification methods (TAP and HMS) share only 5–10% of all interactions, we found a correlation regarding protein connectivity (Table 3).

Table 8 Overlap of individual interactions between pairs of methods

TAP / HMS				
Number of proteins covered by both methods				Overlap of interactions
659				1728
Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)

HMS	1578	33 014	9005	19.19%
-----	------	--------	------	--------

TAP / GN				
Number of proteins covered by both methods			Overlap of interactions	
260			121	
Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
TAP	1379	18 027	885	13.67%
GN	998	6387	673	17.98%

TAP / Y2H				
Number of proteins covered by both methods			Overlap of interactions	
804			156	
Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
TAP	1379	18 027	5641	2.77%
Y2H	3575	5125	506	30.83%

TAP / SL				
Number of proteins covered by both methods			Overlap of interactions	
288			55	
Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
TAP	1379	18 027	876	6.28%
SL	678	886	224	24.55%

TAP / CE				
Number of proteins covered by both methods			Overlap of interactions	
420			192	
Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
TAP	1379	18 027	2778	6.91%
CE	1958	16 496	1073	17.89%

TAP / GF				
Number of proteins covered by both methods			Overlap of interactions	
67			11	
Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
TAP	1379	18 027	70	15.71%
GF	293	358	25	44.00%

TAP / CO				
Number of proteins covered by both methods			Overlap of interactions	
85			18	
Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
TAP	1379	18 027	93	19.35%
CO	378	997	84	21.43%

TAP / ANNOT				
-------------	--	--	--	--

Number of proteins covered by both methods			Overlap of interactions	
498			1428	
Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
TAP	1379	18 027	3253	43.90%
ANNOT	778	2301	1504	94.95%

HMS / GN				
Number of proteins covered by both methods			Overlap of interactions	
292			45	
Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
HMS	1578	33 014	2777	1.62%
GN	998	6387	429	10.49%

HMS / Y2H				
Number of proteins covered by both methods			Overlap of interactions	
957			146	
Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
HMS	1578	33 014	11695	1.25%
Y2H	3575	5125	586	24.91%

HMS / SL				
Number of proteins covered by both methods			Overlap of interactions	
304			37	
Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
HMS	1578	33 014	1101	3.36%
SL	678	886	239	15.48%

HMS / CE				
Number of proteins covered by both methods			Overlap of interactions	
503			124	
Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
HMS	1578	33 014	4818	2.57%
CE	1958	16 496	1371	9.04%

HMS / GF				
Number of proteins covered by both methods			Overlap of interactions	
102			9	
Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
HMS	1578	33 014	364	2.47%
GF	293	358	61	14.75%

HMS / CO				
Number of proteins covered by both methods			Overlap of interactions	
119			12	

Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
HMS	1578	33 014	497	2.41%
CO	378	997	113	10.62%

HMS / ANNOT				
Number of proteins covered by both methods			Overlap of interactions	
345			528	
Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
HMS	1578	33 014	2673	19.75%
ANNOT	778	2301	645	81.86%

GN / Y2H				
Number of proteins covered by both methods			Overlap of interactions	
540			5	
Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
GN	998	6387	1517	0.33%
Y2H	3575	5125	137	3.65%

GN / SL				
Number of proteins covered by both methods			Overlap of interactions	
74			4	
Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
GN	998	6387	28	14.29%
SL	678	886	14	28.57%

GN / CE				
Number of proteins covered by both methods			Overlap of interactions	
387			80	
Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
GN	998	6387	1181	6.77%
CE	1958	16496	1096	7.30%

GN / GF				
Number of proteins covered by both methods			Overlap of interactions	
260			110	
Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
GN	998	6387	529	20.79%
GF	293	358	304	36.18%

GN / CO				
Number of proteins covered by both methods			Overlap of interactions	
311			183	

CO	378	997	829	22.07%
----	-----	-----	-----	--------

GN / ANNOT

Number of proteins covered by both methods	Overlap of interactions
230	550

Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
GN	998	6387	1874	29.35%
ANNOT	778	2301	750	73.33%

Y2H / SL

Number of proteins covered by both methods	Overlap of interactions
397	17

Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
Y2H	3575	5125	151	11.26%
SL	678	886	293	5.80%

Y2H / CE

Number of proteins covered by both methods	Overlap of interactions
1114	8

Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
Y2H	3575	5125	465	1.72%
CE	1958	16 496	5359	0.15%

Y2H / GF

Number of proteins covered by both methods	Overlap of interactions
154	2

Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
Y2H	3575	5125	23	8.70%
GF	293	358	80	2.50%

Y2H / CO

Number of proteins covered by both methods	Overlap of interactions
201	1

Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
Y2H	3575	5125	27	3.70%
CO	378	997	244	0.41%

Y2H / ANNOT

Number of proteins covered by both methods	Overlap of interactions
465	125

Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
Y2H	3575	5125	226	55.31%
ANNOT	778	2301	865	14.45%

SL / CE

Number of proteins covered by both methods				Overlap of interactions
157				2
Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
SL	678	886	41	4.88%
CE	1958	16 496	166	1.20%

SL / GF				
Number of proteins covered by both methods				Overlap of interactions
27				1
Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
SL	678	886	2	50.00%
GF	293	358	2	50.00%

SL / CO				
Number of proteins covered by both methods				Overlap of interactions
15				0
Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
SL	678	886	3	0.00%
CO	378	997	4	0.00%

SL / ANNOT				
Number of proteins covered by both methods				Overlap of interactions
247				129
Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
SL	678	886	227	56.83%
ANNOT	778	2301	264	48.86%

CE / GF				
Number of proteins covered by both methods				Overlap of interactions
103				6
Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
CE	1958	16 496	95	6.32%
GF	293	358	41	14.63%

CE / CO				
Number of proteins covered by both methods				Overlap of interactions
152				37
Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
CE	1958	16 496	443	8.35%
CO	378	997	222	16.67%

CE / ANNOT				
Number of proteins covered by both methods				Overlap of interactions
251				172

Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
CE	1958	16 496	500	34.40%
ANNOT	778	2301	396	43.43%

GF / CO				
Number of proteins covered by both methods			Overlap of interactions	
114			29	
Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
GF	293	358	80	36.25%
CO	378	997	167	17.37%

GF / ANNOT				
Number of proteins covered by both methods			Overlap of interactions	
44			14	
Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
GF	293	358	18	77.78%
ANNOT	778	2301	30	46.67%

CO / ANNOT				
Number of proteins covered by both methods			Overlap of interactions	
50			52	
Method	Toll number of proteins	Toll number of interactions	Number of interactions among common proteins	Overlap of interactions (%)
CO	378	997	60	86.67%
ANNOT	778	2301	61	85.25%

References

- 1 von Mering, C. *et al.* (2002) Comparative assessment of large-scale data sets of protein–protein interactions *Nature* 417, 399–403
- 2 Bader, G.D. and Hogue, C.W. (2002) Analyzing yeast protein–protein interaction data obtained from different sources *Nat. Biotechnol.* 20, 991–997
- 3 von Mering, C. *et al.* (2002) Supplementary information: comparative assessment of large-scale data sets of protein–protein interactions <http://www.nature.com/nature/journal/v417/n6887/supinfo/nature750.html>
- 4 Ho, Y. *et al.* (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry *Nature* 415, 180–183
- 5 Gavin, A.C. *et al.* (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes *Nature* 415, 141–147
- 6 Uetz, P. *et al.* (2000) A comprehensive analysis of protei–protein interactions in *Saccharomyces cerevisiae* *Nature* 403, 623–627
- 7 Ito, T. *et al.* (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome *Proc. Natl. Acad. Sci. U. S. A.* 98, 4569–4574
- 8 Cho, R.J. *et al.* (1998) A genome-wide transcriptional analysis of the mitotic cell cycle *Mol. Cell.* 2, 65–73
- 9 Hughes, T.R. *et al.* (2000) Functional discovery via a compendium of expression profiles *Cell* 102, 109–126
- 10 Tong, A.H. *et al.* (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants *Science* 294, 2364–2368
- 11 Overbeek, R. *et al.* (1999) The use of gene clusters to infer functional coupling *Proc. Natl. Acad. Sci. U. S. A.* 96, 2896–2901
- 12 Huynen, M., *et al.* (2000) Predicting protein function by genomic context: quantitative evaluation and qualitative inferences *Genome Res.* 10, 1204–1210
- 13 Pellegrini, M. *et al.* (1999) Assigning protein functions by comparative genome analysis: protein phylogenetic profiles *Proc. Natl. Acad. Sci. U. S. A.* 96, 4285–4288
- 14 Huynen, M.A. and Bork, P. (1998) Measuring genome evolution *Proc. Natl. Acad. Sci. U. S. A.* 95, 5840–5856
- 15 Marcotte, E.M. *et al.* (1999) A combined algorithm for genome-wide prediction of protein function *Nature* 402, 83–86
- 16 Mewes, H.W. *et al.* (2002) MIPS: a database for genomes and protein sequences *Nucleic Acids Res.* 30, 31–34
- 17 Costanzo, M.C. *et al.* (2001) YPD, PombePD and WormPD: model organism volumes of the BioKnowledge library, an integrated resource for protein information *Nucleic Acids Res.* 29, 75–79
- 18 Kendall, M.G. and Gibbons, J.D. (1990) *Rank correlation methods*, Oxford University Press, London
- 19 Gibbons, J.D. and Chakraborti, S. (1992) *Nonparametric statistical inference*, M. Dekker