Supplementary Materials for 'ppiTrim: constructing non-redundant and up-to-date interactomes'

Aleksandar Stojmirović, and Yi-Kuo Yu

National Center for Biotechnology Information
National Library of Medicine
National Institutes of Health
Bethesda, MD 20894
United States

Supplementary Table 1: Description of ppiTrim MITAB 2.6 columns

Column	Short Name	Description	Example
1	uidA	Smallest Gene ID of the interactor A*†	`entrezgene/locuslink:854647`
2	uidB	Smallest Gene ID of the interactor B*	`entrezgene/locuslink:855136`
3	altA	All gene IDs of the interactor A*	`entrezgene/locuslink:854647`
4	altB	All gene IDs of the interactor B*	`entrezgene/locuslink:855136`
5	aliasA	All canonical gene symbols and integer CROGIDs of interactor A	`entrezgene/locuslink:BNR1|icrogid:2105284`
6	aliasB	All canonical gene symbols and integer CROGIDs of interactor B	`entrezgene/locuslink:MYO5|icrogid:3144798`
7	method	PSI-MI term for interaction detection method	`MI:0018(two hybrid)`
8	author	First author name(s) of the publication in which this interaction has been shown‡	`Tong AH [2002]|tong-2002a-3`
9	pmids	Pubmed ID(s) of the publication in which this interaction has been shown	`pubmed:11743162`
10	taxA	NCBI Taxonomy identifier for interactor A	`taxid:4932(Saccharomyces cerevisiae)`
11	taxB	NCBI Taxonomy identifier for interactor B	`taxid:4932(Saccharomyces cerevisiae)`
12	interactionType	PSI-MI term for interaction type	`MI:0407(direct interaction)`
13	sourcedb	PSI-MI terms for source databases‡	`MI:0000(MPACT)|MI:0463(grid)|MI:0465(dip)|MI:0469(intact)`
14	interactionIdentifier	A list of interaction identifiers⋆	`ppiTrim:tyuGkSOK231dh3YnSi6GbczJCFE=|MPACT:8233|dip:DIP-11198E|grid:147506|intact:EBI-601565|intact:EBI-601728|irigid:288990|edgetype:X`
15	confidence	A list of ppiTrim confidence scores•	`maxsources:2|dmconsistency:full|conflicts:S3oaiXt5tA4vVrUsO1rc1TA9krk=`
16	expansion	Either 'none' for binary interactions or 'bipartite' for subunits of complexes	`none`
17	biologicalRoleA	PSI-MI term(s) for the biological role of interactor A‡	`MI:0499(unspecified role)`
18	biologicalRoleB	PSI-MI term(s) for the biological role of interactor B ‡	`MI:0499(unspecified role)`
19	experimentalRoleA	PSI-MI term(s) for the experimental role of interactor A‡	`MI:0496(bait)|MI:0498(prey)|MI:0499(unspecified role)`
20	experimentalRoleB	PSI-MI term(s) for the experimental role of interactor B‡	`MI:0496(bait)|MI:0498(prey)|MI:0499(unspecified role)`
21	interactorTypeA	PSI-MI term for the type of interactor A (either 'protein' or 'protein complex')	`MI:0326(protein)`
22	interactorTypeB	PSI-MI term for the type of interactor B (always 'protein')	`MI:0326(protein)`
29	hostOrganismTaxid	NCBI Taxonomy identifier for the host organism	`taxid:4932(Saccharomyces cerevisiae)`
31	creationDate	Date when ppiTrim was run	`2011/05/11`
32	updateDate	Date when ppiTrim was run	`2011/05/11`
35	checksumInteraction	ppiTrim ID for an interaction	`ppiTrim:tyuGkSOK231dh3YnSi6GbczJCFE=`
36	negative	Always 'false'	`false`

The above table shows short descriptions for the columns of lines output by ppiTrim with examples. The columns that are not used by ppiTrim (– output) are omitted. List of items are always separated by the | character (without any intervening spaces). This description only applies to ppiTrim output; the full PSI-MI 2.6 TAB format description can be found at `http://code.google.com/p/psimi/wiki/PsimiTab26Format` Notes: *An interactor may be associated with several Gene IDs. In that case the smallest one is written in uid columns while the entire list is shown in alt columns. †Interactor A may be used to denote a protein complex. In that case the uidA is of the form `complex:<ppiTrim ID>`, while altA and aliasA are left empty. ‡Multiple items are possible, originating from all source records contributing to the consolidated interaction. ⋆First ID is always the ppiTrim ID for the consolidated interaction, followed by the original IDs for all contributing interactions and their integer RIGIDs from iRefIndex. The final item is the edge type code. •`maxsources`: an estimate of the maximal number of independent experiments contributing to the consolidated interaction; `dmconsistency`: consistency of contributing detection method terms. Values are one of *invalid* (no method terms present), *single* (only one method term), *min* (minimum term found but not maximum), *max* (maximum term found but not minimum), and *full* (both minimum and maximum term present in subcluster); `conflicts`: ppiTrim IDs of consolidated interactions with detection method term in conflict with the current one.

Supplementary Table 2: Remapping of obsolete PSI-MI terms

Original Term		Mapped Term		Notes
MI:0021	colocalization by fluorescent probes cloning	MI:0428	imaging technique	
MI:0022	colocalization by immunostaining	MI:0428	imaging technique	∗
MI:0023	colocalization/visualisation technologies	MI:0428	imaging technique	∗
MI:0025	copurification	MI:0401	biochemical	
MI:0059	gst pull down	MI:0096	pull down	
MI:0061	his pull down	MI:0096	pull down	
MI:0079	other biochemical technologies	MI:0401	biochemical	
MI:0109	tap tag coimmunoprecipitation	MI:0676	tandem affinity purification	
MI:0045	experimental interaction detection	MI:0492	in vitro	†
MI:0493	in vivo	MI:0493	in vivo	†
MI:0000	coip coimmunoprecipitation	MI:0019	coimmunoprecipitation	⋆
MI:0000	elisa enzyme-linked immunosorbent assay	MI:0411	enzyme linked immunosorbent assay	⋆

∗ Interaction type is also adjusted to MI:0403 as recommended in `psi-mi.obo`; † HPRD terms are treated as a special case, see main text; ⋆ MPPI interactions in the human dataset.

Supplementary Table 3: Mapping PTM labels from BioGRID into PSI-MI terms

Original Term	Mapped Term	
Acetylation	MI:0192	acetylation reaction
Deacetylation	MI:0197	deacetylation reaction
Demethylation	MI:0871	demethylation reaction
Dephosphorylation	MI:0203	dephosphorylation reaction
Deubiquitination	MI:0204	deubiquitination reaction
Glucosylation	MI:0559	glycosylation reaction
Methylation	MI:0213	methylation reaction
Nedd(Rub1)ylation	MI:0567	neddylation reaction
No Modification	MI:0414	enzymatic reaction
Phosphorylation	MI:0217	phosphorylation reaction
Prenylation	MI:0211	lipid addition
Proteolytic Processing	MI:0570	protein cleavage
Ribosylation	MI:0557	adp ribosylation reaction
Sumoylation	MI:0566	sumoylation reaction
Ubiquitination	MI:0220	ubiquitination reaction

Supplementary Table 4: Processing source interactions (RIGIDs)

Species	Initial	Without Gene ID	Retained	With Mapped Gene ID
S. cerevisiae	186530	1272	79931	591
H. sapiens	138570	1917	84860	7158
D. melanogaster	46925	4988	39200	2176

Statistics of initial processing of raw interactions from in terms of iRefIndex RIGIDs. A RIGID for an interaction is a unique hash derived from its interactants' sequences (with order not significant). Thus, multiple interactions with the same interactants share the same RIGID. Shown are the initial number, number removed due to missing Gene ID, total number of retained and the number retained containing at least one interactant with mapped Gene ID. Compared to Table 2 in the main text, this table does not contain a column showing the number of removed RIGIDs due to filtering criteria. This is becuase the ppiTrim filtering routine operates on raw interactions (corresponding to a single record from a source database) and some RIGIDs would be associated with both accepted and removed raw interactions.

Supplementary Table 5: Mapping CROGID identifiers from iRefIndex into Gene IDs: details

Species	I	V	O	R	P	T	M	G	S	B
S. cerevisiae	5552	0	0	607	95	461	0	26	21	386
H. sapiens	11428	11	0	2615	155	2017	71	754	429	0
D. melanogaster	7780	0	30	1569	18	814	2	124	440	0

Detailed statistics of mapping CROGIDs into Gene IDs. All numbers denote CROGIDs: directly mapped to valid Gene IDs in the iRefIndex file (I); directly mapped to Gene IDs but the Gene IDs were updated during validation (V); directly mapped to obsolete Gene IDs (O); not directly mapped to Gene IDs – total orphans (R); orphans with PDB accession as a primary ID (P); orphans with Uniprot accession as a primary ID (T); additionally mapped to a valid Gene ID using mapping.txt file from iRefIndex (M); additionally mapped to a valid Gene ID using a direct reference from Uniprot record (G); additionally mapped to a valid Gene ID using a gene name from Uniprot record (S); additionally mapped to a Gene ID that was not valid (B).

Supplementary Table 6: Randomly sampled deflated complexes from high throughput publications

ppiTrim Complex ID	Sources	Pubmed ID	Members	Comments
8AVRUHG76vkiFn2cZGICNZzr00Y=	grid	14759368	CFT2, YSH1, PTA1, MPE1	Part of mRNA cleavage/polyadenylation complex (4/10 proteins).
9yS57j/gbRbOlNmmimsVeonoraA=	grid	14759368	NUT1, MED7, MED4, SIN4, SRB4	Part of mediator complex.
JU+EOkq6ipLh9DJKRtGRLUvT7vM=	grid,mint	14759368	UBP6, RPT3, RPN9, RPT1, RPN8, RPN2, RPN7, RPN1	Part of proteasome. MINT does not contain complexes from the original paper.
HtTmhGiPyfIT2vFtRZ94uWw0rsY=	grid	16429126	IOC3, HTB1, HTA2, HHF2, ISW1, KAP114, ITC1, RPS4A, VPS1, NAP1, RPO31, ISW2, TBF1, BRO1, MOT1	Part of Complex # 99.
LnNzfyPGShcG7zkKynU6+fsK2eU=	grid	16429126	PSK1, NTH1, BMH2, RTG2, BMH1	Part of complex # 147 (two core proteins plus three attachments).
S2I6VRjFMWC6rkkM+oYXwKCg9YQ=	grid	16429126	RPL4B, MNN10, MNN11, HOC1, MNN9, ANP1	Core complex (# 111 – mannan polymerase II) + one attachment protein (RPL4B).
1fRmAapl2ruoQq202YUJg55maFo=	grid,mint	16554755	RSM24, RSM28, MRPS5, MRP13, MRPS35, RSM27, RSM7, RSM25, MRPS17, MRPS12, RSM19, MRP4	Part of complex # 1.
5tBkYOmK/G1h3vaQmiOnUoBHHMQ=	grid,mint	16554755	CFT2, YSH1, MPE1, PAP1	Part of complex # 18.
9f2DVj2rDGeCP53LHOnWRMwq14A=	grid,mint	16554755	KAP95, RTT103, VMA2, RAI1, RAT1, RPB2, SRP1	True experimental association but not part of any derived complex.
AVawv51+6Fqe3DquygD/XfyrXxE=	grid,mint	16554755	RRP42, RRP45, RRP6, CSL4, MPP6, RRP4, LRP1, DDI1	Part of complex # 19.
NOLEwovavMsFrQEdkSUt/mldeMc=	grid,mint	16554755	CDC3, SHS1, CDC11, CDC12	Part of complex # 121.
WA51i87Lj1wGp/EeF1OV/YvbW1Y=	grid,mint	16554755	GTT2, TRX1, CRN1, SSA3, IPP1, CMD1, TRX2, TDH1, RPL40B, CDC21, OYE2	True experimental association but not part of any derived complex.
YN/hQXQvzoB5HqrgPzVth28mGsY=	grid,mint	16554755	RRP43, RRP42, RRP45, RRP40, DIS3, RRP6, RRP4, LRP1	Part of complex # 19.
1LRk+AgI8HpGOSAgkhDzNJWSvtI=	grid	20489023	RTG3, RTG2, TOR1, TOR2, CKA2, MYO2, MKS1, KOG1	True experimental association.
xWzvxeJFGqjkCihjmQVf5gZhJjQ=	dip,grid,mint	20489023	PUF3, SAM1, GCD6, SPT16, MTC1, YGK3, LSM12	True experimental association.

To partially investigate the fidelity of deflated complexes of type A and N, we randomly sampled 25 such complexes from the final ppiTrim yeast dataset and examined the original publications associated with them. This table contains 15 deflated complexes from high-throughput publications, while Supplemenary Table 7 contains the complexes from low-throughput publications. Most of high-throughput papers referred to in this table present both the lists of bait-prey associations and of derived complexes. The complexes delated by ppiTrim are often derived from the former and form only parts of the latter. In the last column of this table, the complex numbers referred to are labels used by the publication's authors.

Supplementary Table 7: Randomly sampled deflated complexes from low-throughput publications

ppiTrim Complex ID	Sources	Pubmed ID	Members	Comments
15VfQtoe5gxGNwPSY3AG0sq6A2U=	grid	9891041	CCR4, HPR1, PAF1, SRB5, GAL11	NOT a true complex. This is because of bad annotation of PAF1–SRB5 interaction by the BioGRID. Completely opposite interpretation was given in the paper.
d79IdtwfTAENrH8CQ+c8CpS389Y=	grid	10329679	YPT1, VPS21, YPT7, GDI1	True complex. This is the only experiment in the paper.
EtS4cgphEpTqJb/FS5qxyzf0ke8=	grid	11733989	CDC39, CCR4, CDC36, CAF130, CAF40, CAF120, POP2, NOT5, MOT2	True complex. CAF120 is an unusual member that could almost be left out.
2kOyGdwzWywSpN5mhK26gCcC6LQ=	grid	14769921	GBP2, IMD3, TEF1, KEM1, CTK2, CTK1, CTK3	True complex, except that TEF1 should be TEF2. This is an error in the iRefIndex source file; the BioGRID website has the correct assignment.
Kd07BBUF07Sqy9NP3D0lixsS/TY=	grid	15303280	BUD31, RPL2B, PRP19, CDC13, ATP1, RPS4A, SNU114, MDH1, MAM33, MRPL3, MRPL17, PRP8, PRP22, PAB1, BRR2	True association
ZAGz/IZqkEr3/NTDLzPEDAD9cKo=	grid	16179952	CDC40, UFD1, SSM4, UBX2	NOT a true complex, probably due to a typo in annotation. CDC40 cannot be found anywhere in the paper and should most likely be CDC48.
RDu0dsPAN0QEadfSU5sv05Ifihw=	grid	16286007	SIN3, RCO1, RPD3, UME1, EAF3	True complex.
Vqbn3dDwTPgyE9DzbatFNqzdFe0=	grid	16615894	VPS36, VPS25, VPS28, SNF8	Vps28 binds the other three, which form a complex.
lmdypAN9kaHBdasLWS19x8K7KkE=	grid	20159987	UBI4, UFD2, PEX29, SSM4	Biological association but indicated as 'NOT a stable complex' in the paper.
aakRh6qVahGxGvqHe399+faxPvA=	grid	20655618	PEX13, PEX10, PEX8, PEX12	Association is correct, although mutant strain was used to obtain this particular complex.

To partially investigate the fidelity of deflated complexes of type A and N, we randomly sampled 25 such complexes from the final ppiTrim yeast dataset and examined the original publications associated with them. This table contains 10 deflated complexes from low-throughput publications, while Supplemenary Table 6 contains the complexes from high-throughput publications.

Supplementary Table 8: Summary of resolvable conflicts

Consolidated terms	Count
MI:0018 (two hybrid), MI:0045 (experimental interaction detection), MI:0398 (two hybrid pooling approach), MI:0399 (two hybrid fragment pooling approach)	3959
MI:0090 (protein complementation assay), MI:0111 (dihydrofolate reductase reconstruction)	2612
MI:0090 (protein complementation assay), MI:0112 (ubiquitin reconstruction)	2077
MI:0004 (affinity chromatography technology), MI:0676 (tandem affinity purification)	1840
MI:0004 (affinity chromatography technology), MI:0007 (anti tag coimmunoprecipitation)	1408
MI:0018 (two hybrid), MI:0045 (experimental interaction detection), MI:0397 (two hybrid array)	1231
MI:0018 (two hybrid), MI:0045 (experimental interaction detection)	954
MI:0018 (two hybrid), MI:0397 (two hybrid array)	914
MI:0045 (experimental interaction detection), MI:0686 (unspecified method)	628
MI:0004 (affinity chromatography technology), MI:0019 (coimmunoprecipitation)	598
MI:0018 (two hybrid), MI:0398 (two hybrid pooling approach)	506
MI:0004 (affinity chromatography technology), MI:0007 (anti tag coimmunoprecipitation), MI:0676 (tandem affinity purification)	444
MI:0018 (two hybrid), MI:0045 (experimental interaction detection), MI:0686 (unspecified method)	320
MI:0004 (affinity chromatography technology), MI:0096 (pull down)	217
MI:0415 (enzymatic study), MI:0424 (protein kinase assay)	192
MI:0045 (experimental interaction detection), MI:0081 (peptide array)	150
MI:0045 (experimental interaction detection), MI:0676 (tandem affinity purification)	120
MI:0492 (in vitro), MI:0493 (in vivo)	5739
MI:0018 (two hybrid), MI:0398 (two hybrid pooling approach)	5394
MI:0018 (two hybrid), MI:0492 (in vitro), MI:0493 (in vivo)	2796
MI:0096 (pull down), MI:0492 (in vitro), MI:0493 (in vivo)	2760
MI:0096 (pull down), MI:0492 (in vitro)	2134
MI:0018 (two hybrid), MI:0492 (in vitro)	1658
MI:0018 (two hybrid), MI:0493 (in vivo)	1193
MI:0018 (two hybrid), MI:0397 (two hybrid array)	1045
MI:0096 (pull down), MI:0493 (in vivo)	513
MI:0004 (affinity chromatography technology), MI:0006 (anti bait coimmunoprecipitation)	384
MI:0004 (affinity chromatography technology), MI:0019 (coimmunoprecipitation)	309
MI:0004 (affinity chromatography technology), MI:0007 (anti tag coimmunoprecipitation)	195
MI:0114 (x-ray crystallography), MI:0492 (in vitro)	166
MI:0004 (affinity chromatography technology), MI:0096 (pull down)	161
MI:0047 (far western blotting), MI:0492 (in vitro), MI:0493 (in vivo)	106
MI:0018 (two hybrid), MI:0398 (two hybrid pooling approach)	17738
MI:0018 (two hybrid), MI:0399 (two hybrid fragment pooling approach)	1426

All resolvable conflicts with counts of more than 100 for yeast (top), human (middle) and fruitfly (bottom) datasets are shown.