LETTER

Architecture of the human interactome defines protein communities and disease networks

Edward L. Huttlin¹, Raphael J. Bruckner¹, Joao A. Paulo¹, Joe R. Cannon¹, Lily Ting¹, Kurt Baltier¹, Greg Colby¹, Fana Gebreab¹, Melanie P. Gygi¹, Hannah Parzen¹, John Szpyt¹, Stanley Tam¹, Gabriela Zarraga¹, Laura Pontano-Vaites¹, Sharan Swarup¹, Anne E. White¹, Devin K. Schweppe¹, Ramin Rad¹, Brian K. Erickson¹, Robert A. Obar^{1,2}, K. G. Guruharsha², Kejie Li², Spyros Artavanis-Tsakonas^{1,2}, Steven P. Gygi¹ & J. Wade Harper¹

ORFeome-----Proteome-----Interactome

 the proteome can be viewed as constellations of interacting protein modules organized into signal transduction networks, molecular machines, and organelles

 our knowledge of proteome architecture is fragmentary, as is our conception of how protein interconnectivity is influenced by genetic and cellular variation

Challenges....

- Myriad genes, isoforms, and modification states encoded by the human genome
- Low abundance of many proteins, which limits detection
- Many transient interactions that complicate signaling network mapping
- Prevalence of membrane proteins, which often requires specialized methods for purification

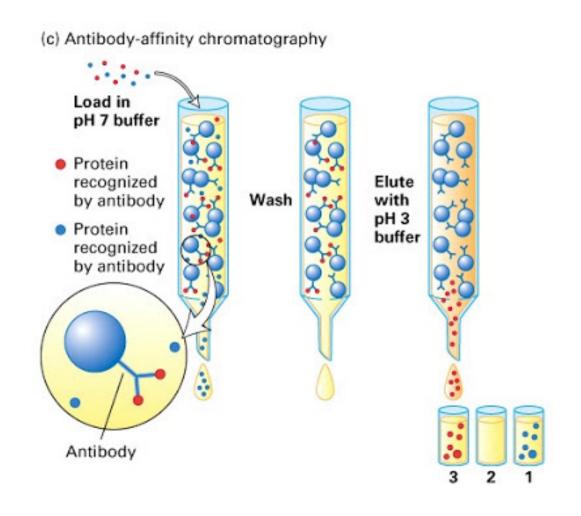
•

•

Main strategies to study mammalian proteome structure

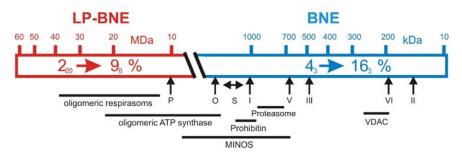
- Biochemical experiments reveal stable macromolecular complexes
- Affinity purification of tagged protein followed by MS (AP-MS)
- Immunoprecipitation followed by MS (IP-MS)
- Protein correlation profiling (Blue native electrophoresis +/- IP) followed by MS
- Yeast 2Hybrid analysis
- Database archive protein interaction from literature (context dependant)

Example of affinity purification-targeted

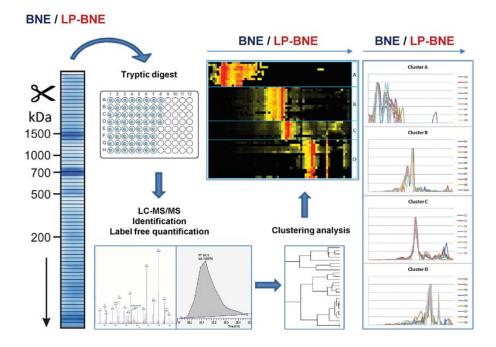


Protein correlation profiling---Non targeted

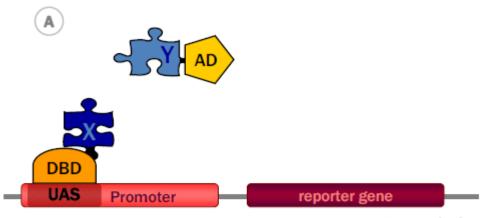
A



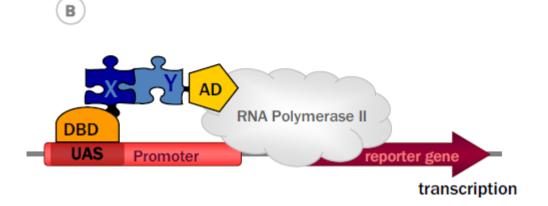
В

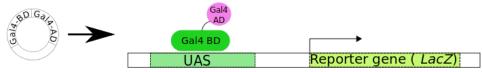


Y2H system



no transcription

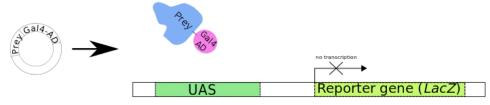




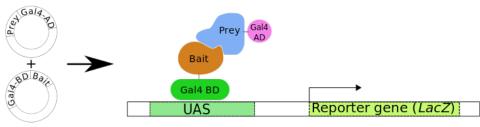
A. Regular transcription of the reporter gene



B. One fusion protein only (Gal4-BD + Bait) - no transcription



C. One fusion protein only (Gal4-AD + Prey) - no transcription



D. Two fusion proteins with interacting Bait and Prey

Databases

- MINT: a Molecular INTeraction database (italian)
- GeneMania
- BioGRID
- STRING (Swiss)
- BioPixie
- IntAct
- CORUM
- •
- •

BioPlex 1.0 ----- BioPlex 2.0

The BioPlex Network: A Systematic Exploration of the Human Interactome

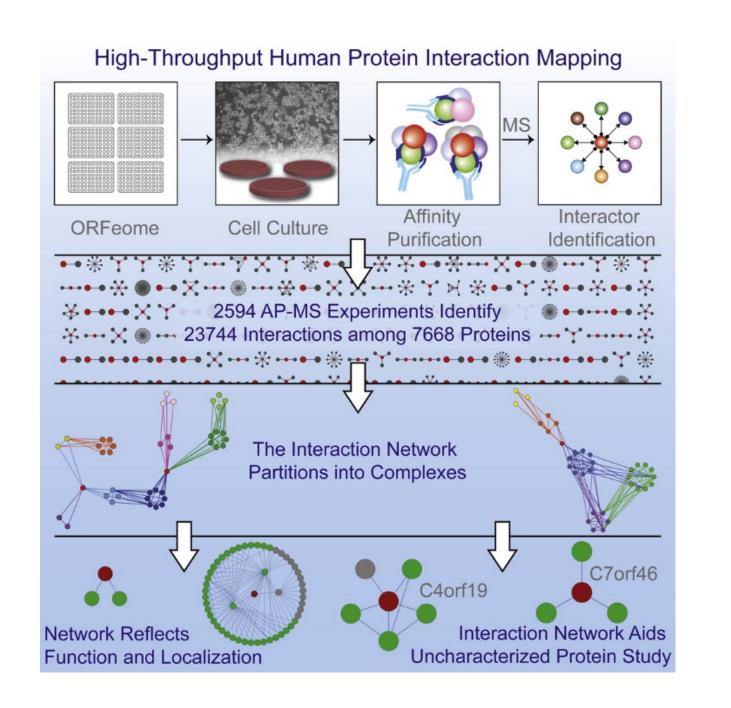
Edward L. Huttlin, Lily Ting, Raphael J. Bruckner, Fana Gebreab, Melanie P. Gygi, John Szpyt, Stanley Tam, Gabriela Zarraga, Greg Colby, Kurt Baltier, Rui Dong, Virginia Guarani, Laura Pontano Vaites, Alban Ordureau, Ramin Rad, Brian K. Erickson, Martin Wühr, Joel Chick, Bo Zhai, Deepak Kolippakkam, Julian Mintseris, Robert A. Obar, 1,3 Tim Harris, 3 Spyros Artavanis-Tsakonas, 1,3 Mathew E. Sowa, 1 Pietro De Camilli, 2 Joao A. Paulo, 1 J. Wade Harper, 1,* and Steven P. Gygi1,*



¹Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA

²Department of Cell Biology and Howard Hughes Medical Institute, Yale School of Medicine, New Haven, CT 06519, USA ³Biogen, Cambridge, MA 02142, USA

^{*}Correspondence: wade_harper@hms.harvard.edu (J.W.H.), steven_gygi@hms.harvard.edu (S.P.G.) http://dx.doi.org/10.1016/j.cell.2015.06.043



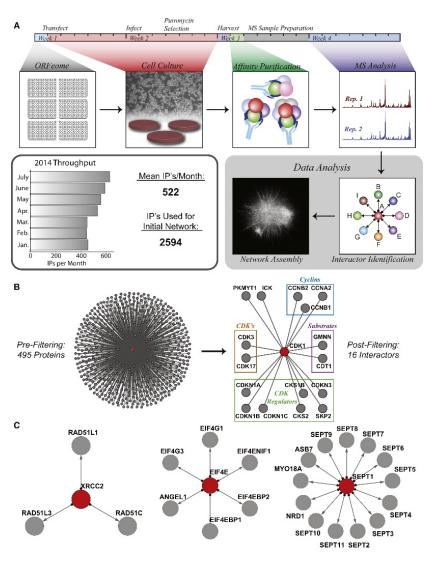
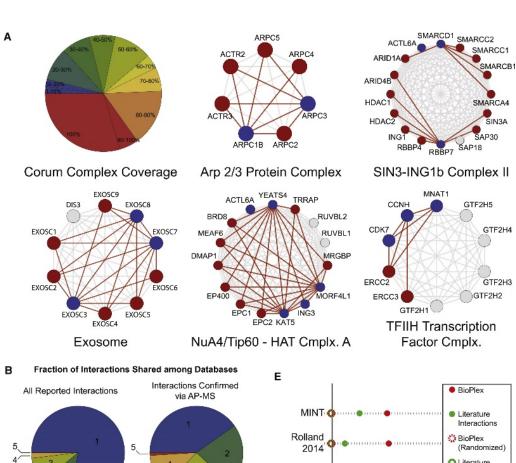
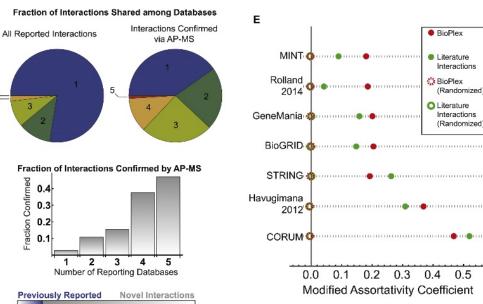


Figure 1. High-Throughput Interaction Mapping via AP-MS

(A) AP-MS platform: (1) a lentiviral library of 13,000 FLAG-HA-tagged ORFs was constructed from the Human ORFEOME; (2) 293T cells were infected and expanded under puromycin selection; (3) baits and preys were immuno-purified; (4) tryptic digests were analyzed in technical duplicate by LC-MS; (5) proteins were identified and specific interactors found; (6) and interactions were assembled to model the human interactome. Up to 600 AP-MS experiments may be completed per month.

- (B) CompPASS-Plus extracts 16 interactors for bait CDK1 from a background of nearly 500 proteins.
- (C) Interaction maps for baits XRCC2, EIF4E, and SEPT1 (red). Nearly all interactions have been previously described. Interactors were identified from backgrounds of 487, 778, and 749 proteins, respectively.





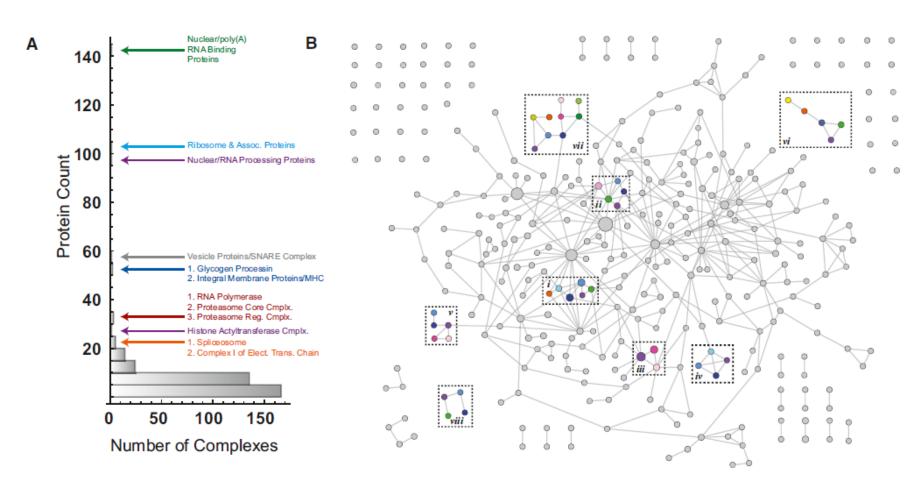
.0 0.2 0.4 0.6 0.8

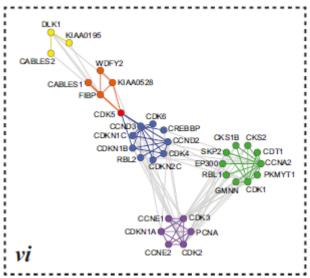
Fraction of Interactions Observed via AP-MS

Figure 3. Evaluation of AP-MS Protein Interactions

- (A) AP-MS interactions superimposed onto CORUM complexes. The pie chart depicts the fraction of complexes achieving the indicated coverage in BioPlex. Only complexes containing two or more baits were considered. Five representative CORUM complexes: baits are colored blue, whereas preys are red and proteins not observed by AP-MS are gray. Interactions among CORUM complex members are gray, whereas interactions confirmed by AP-MS are red.
- (B) Physical protein interactions reported in BioGrid, CORUM, GeneMania, STRING, and MINT were merged. Left: overlap among databases. Right: overlap among databases for interactions confirmed by AP-MS.
- (C) Fraction of database interactions confirmed by AP-MS as a function of the number of supporting database reports. The composite interaction database was filtered to include only interactions connecting one of 2,594 baits with proteins observed as baits or preys in the interaction network.
- (D) 86% of AP-MS interactions have not been reported in the databases listed above.
- (E) Pairwise comparisons of BioPlex with published interaction networks were performed, using graph assortativity to quantify preferential interaction in cases of shared localization among proteins detected in both networks. Literature datasets included BioGRID, CORUM, GeneMania, STRING, and MINT, as well as interactions recently reported via yeast-two-hybrid (Rolland et al., 2014) and LC-MS correlation profiling (Havugimana et al., 2012). Each analysis was repeated with randomized localizations as a control.

Protein "communities" 356





BioPlex 2.0: more proteins, more interactions

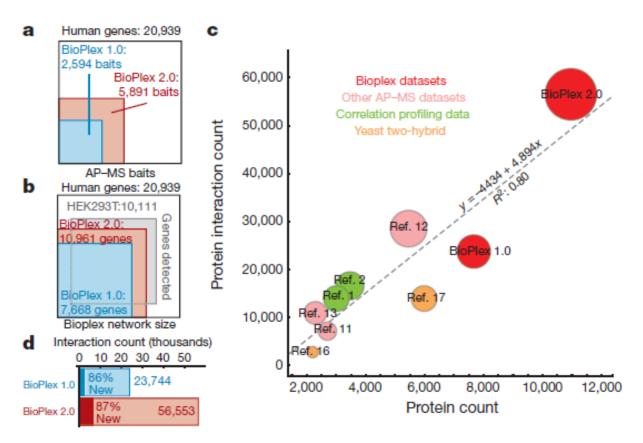


Figure 1 | BioPlex 2.0 substantially increases depth and breadth of interactome coverage. a, Bait proteins targeted for AP–MS analysis. b, Protein-coding genes included in BioPlex 2.0 as baits or preys. c, The BioPlex 2.0 network substantially exceeds previous experimentally derived interaction networks with respect to protein and interaction counts. Circle area is proportional to interaction counts, while shading denotes the experimental strategy used for interaction mapping. d, BioPlex 2.0 doubles the numbers of interactions revealed in BioPlex 1.0.

Better coverage vs. BioPlex 1.0

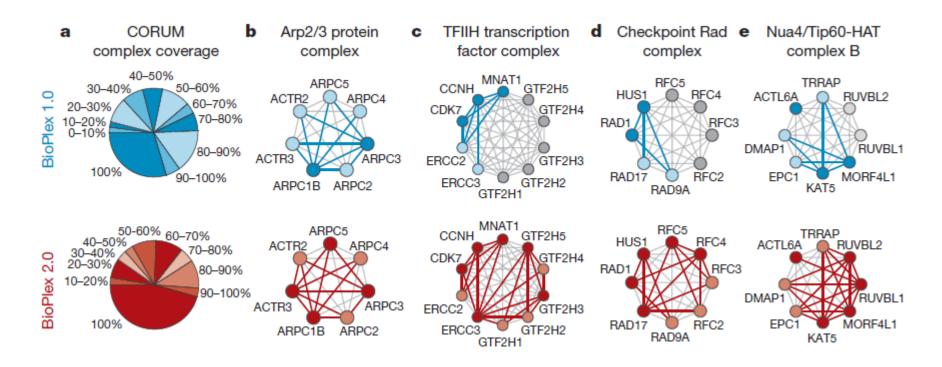


Figure 2 | BioPlex 2.0 maps protein complexes with increased resolution. a, Agreement among BioPlex networks and CORUM complexes. Pie charts indicate the fraction of CORUM complexes that attained the indicated protein coverage. Compared with BioPlex 1.0 (blue), BioPlex 2.0 (red) provides substantially improved coverage. b–e, Network

coverage achieved by BioPlex 1.0 (blue) and BioPlex 2.0 (red) for selected CORUM complexes. Dark and light shades depict bait and prey proteins, respectively, while grey proteins were not observed in the network. Red and blue edges represent detected protein interactions.

BioPlex 2.0 Protein Communities

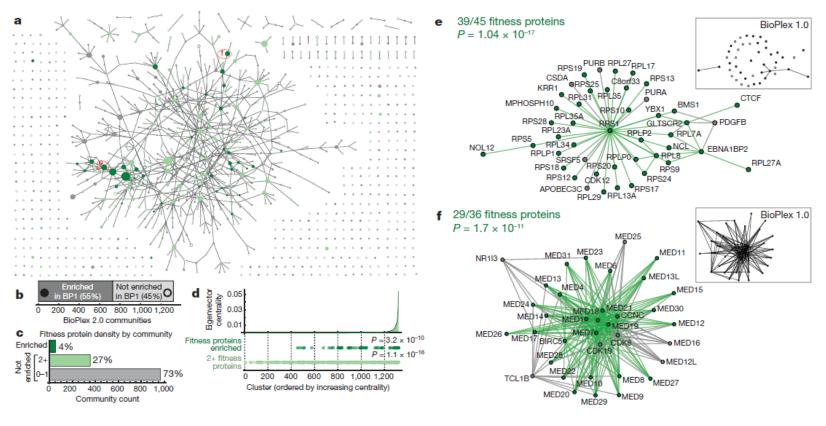


Figure 3 | BioPlex communities subdivide the interaction network according to functional properties and fitness effects. a, Network of communities revealed through MCL clustering of the BioPlex 2.0 network. Nodes represent distinct communities and are scaled to reflect the numbers of proteins in each (3–76 proteins). Nodes are connected by edges when proteins within the respective communities interact with unusually high frequency (see Methods). Filled nodes depict communities that were also found to be interconnected by unusual numbers of interactions in BioPlex 1.0; open circles represent communities of proteins that exhibited only background numbers of interactions in BioPlex 1.0. Communities containing two or more proteins associated with increased cellular fitness are highlighted in light green; communities that are enriched with cellular fitness proteins (1% false discovery rate (FDR)) are highlighted in dark green. Communities circled and marked with red letters 'e' and 'f' refer

to those selected in e and f. b, Mapping BioPlex 2.0 communities onto BioPlex 1.0 reveals lower connectivity, with 45% of complexes showing no significant enrichment of interactions above background levels (binomial test; Benjamini–Hochberg-adjusted P < 0.05). c, Relative fractions of 1,320 communities that contain specified numbers of fitness proteins. d, When BioPlex 2.0 clusters are ranked according to their eigenvector centrality within the BioPlex 2.0 community network (a), clusters that contain multiple fitness proteins (light green) or are enriched for fitness proteins (dark green) tend to have higher centralities (Kolmogorov–Smirnov test). e, f, Selected BioPlex 2.0 communities highlighting proteins associated with cellular fitness (green). Inset maps depict the same communities as observed in BioPlex 1.0. Filled nodes indicate proteins that were in BioPlex 1.0, while black edges indicate interactions that were visible. In contrast, open circles indicate proteins that were not found in BioPlex 1.0.

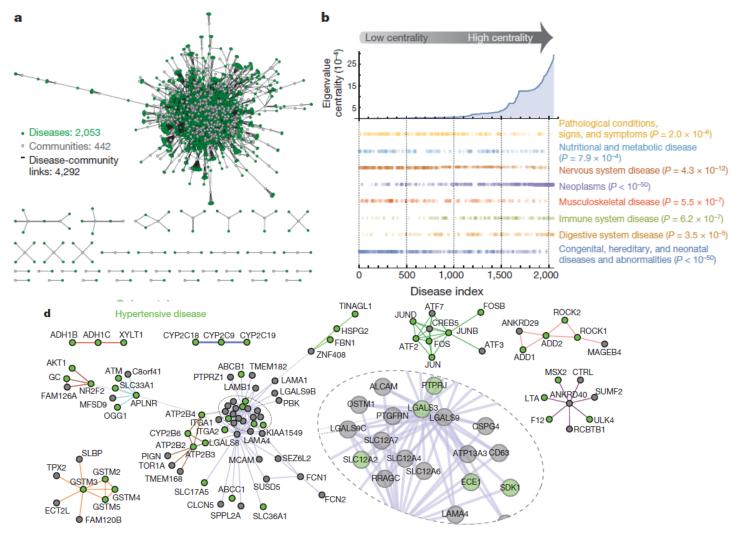
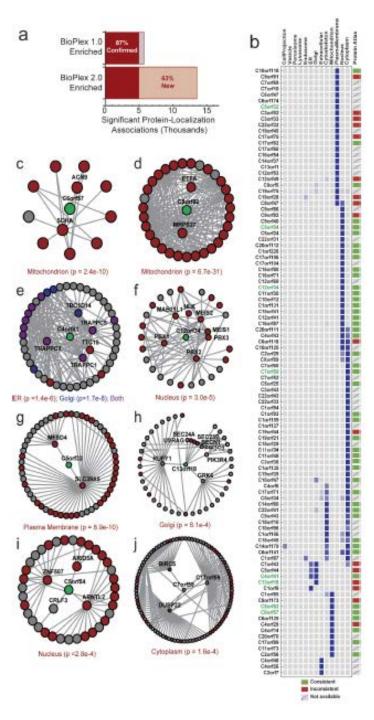


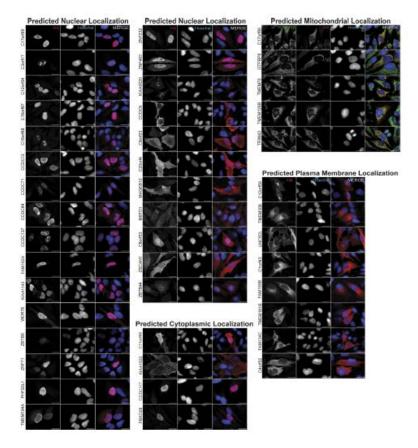
Figure 4 | Integration of BioPlex 2.0 and the DisGeNET network associates protein complexes with disease processes. a, Network of associations among protein interaction communities and disease conditions (see Methods). The network depicts 4,292 associations between 442 protein complexes (grey) and 2,053 disease states (green). b, Ranking of 2,053 disease states on the basis of eigenvalue centrality in the disease–complex network (a). Scatter plots below highlight disease classes that are non-randomly distributed (Kolmogorov–Smirnov test; Benjamini–

Hochberg P < 0.01). c, d, Sub-networks associated with selected disease states: colorectal cancer (BRAF complex: P < 0.05) and hypertensive disease. Nodes associated with the indicated disease are highlighted in green, while other complex members are grey; thick, multi-coloured edges connect proteins belonging to individual communities revealed through MCL clustering; thin, dashed, grey edges connect proteins among adjacent communities.



Extended Data Figure 3 | BioPlex 2.0 enables subcellular localization prediction for additional uncharacterized proteins. a, Increased interaction density expands subcellular localization predictions from BioPlex 2.0. b, Subcellular localization predictions for a selection of uncharacterized human proteins for which no confident prediction could be made in BioPlex 1.0. Where possible, the figure indicates whether predicted localization is consistent with the Human Protein Atlas²¹. c–j, Sub-networks highlighting primary and secondary neighbours for

selected uncharacterized human proteins whose subcellular localization can be predicted using the BioPlex network. Nodes are coloured according to subcellular localization data provided by UniProt. P values were calculated by Fisher's exact test as described in Methods with multiple testing correction. Localizations depicted in c, e, g, and i are consistent with recent characterization as listed in UniProt; The localization given in d is consistent with MitoCarta 2.0 (ref. 41).



Extended Data Figure 4 | Validation of subcellular localization predictions using anti-HA immunofluorescence. The indicated bait proteins fused at their C terminus with an HA tag were expressed after transient infection of lentiviruses at low multiplicity of infection; after 2 days, cells were fixed and subjected to anti-HA-based

immunofluorescence (red). Nuclei were stained with Hoechst. For baits with predicted mitochondrial localization, cells were co-stained with anti-TOMM20 antibodies (green). Z-series optical sections were acquired via spinning disk confocal microscopy; maximum intensity projections are shown. Scale bar, 20 µm. Home Download interactions Browse interactions Download MS data Download CompPASS People Links BioPlex

BioPlex

The BioPlex (biophysical interactions of ORFeome-based complexes) network is the result of creating thousands of cell lines with each expressing a tagged version of a protein from the ORFeome collection. Immunopurification of the tagged protein and detection of associated proteins by mass spectrometry are the building blocks of the network. The overarching project goal is to determine protein interactions for every member of the collection. A first paper in *Cell* reports the first ~2,500 experiments (~23,000 interactions). Our current release with more than 5,000 human proteins as baits (~50,000 interactions) is also now available.



Department of Cell Biology

A Gygi & Harper Lab Collaboration

Funding:

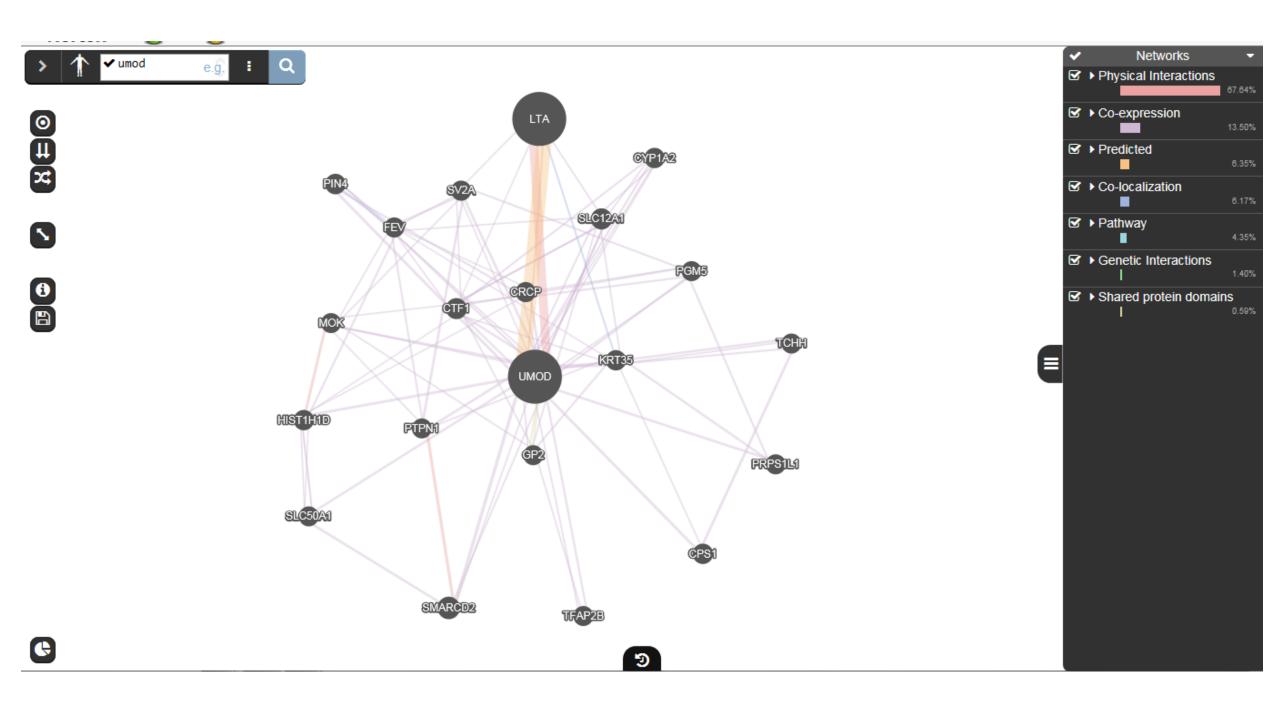
- Biogen
- NHGRI: U41 HG006673

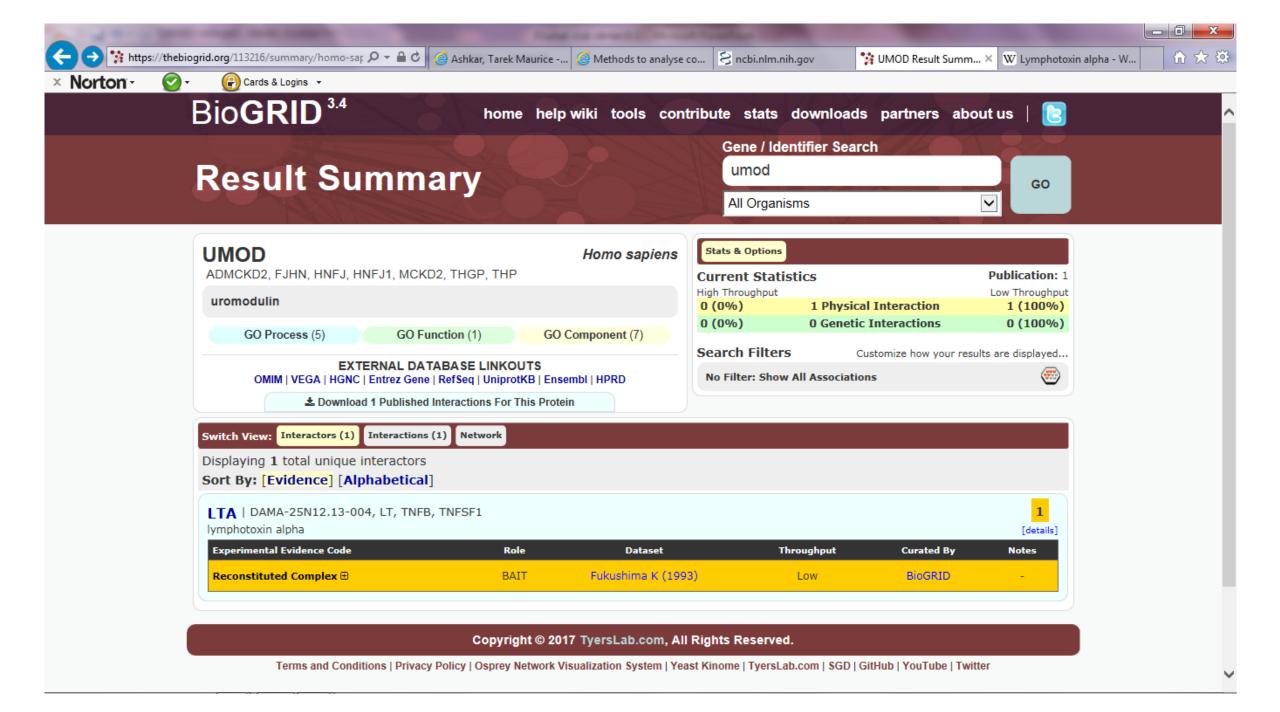
Read more

Huttlin EL, Ting L, Bruckner RJ, Gebreab F, Gygi MP, Szpyt J, Tam S, Zarraga G, Colby G, Baltier K, Dong R, Guarani V, Vaites LP, Ordureau A, Rad R, Erickson BK, Wühr M, Chick J, Zhai B, Kolippakkam D, Mintseris J, Obar RA, Harris T, Artavanis-Tsakonas S, Sowa ME, De Camilli P, Paulo JA, Harper JW, Gygi SP. (2015) The BioPlex Network: A Systematic Exploration of the Human Interactome. Cell 162:425-440.

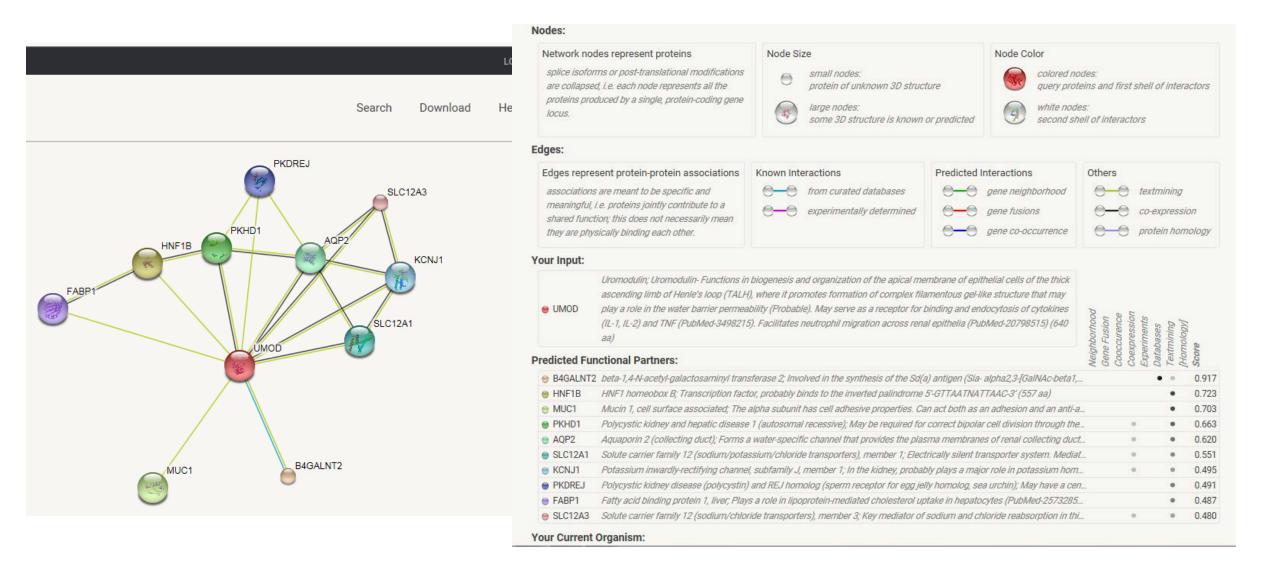
Databases

- MINT: a Molecular INTeraction database (italian)
- GeneMania
- BioGRID
- STRING (Swiss)
- BioPixie
- IntAct
- CORUM
- •
- •





STRING



Network Stats

number of nodes: 11 number of edges: 20 average node degree: 3.64 avg. local clustering coefficient: 0.811 expected number of edges: 10 PPI enrichment p-value: 0.00428

your network has significantly more interactions than expected (what does that mean?)

Functional enrichments in your network

Note: some enrichments may be expected here (why?)

Biological Process (GO)				
pathway ID	pathway description	count in gene set	false discovery rate	
GO:0007588	excretion	4	0.00011	
GO:0001822	kidney development	5	0.000503	
GO:0072001	renal system development	5	0.000503	
GO:0001655	urogenital system development	5	0.00069	

	Molecular Function (GO)		
pathway ID	pathway description	count in gene set	false discovery rate
GO:0015377	cation:chloride symporter activity	2	0.0101
GO:0022892 substrate-specific transporter activity 6		0.0101	
GO:0022891	substrate-specific transmembrane transporter activity	5	0.0279

Cellular Component (GO)			
pathway ID	pathway description	count in gene set	false discovery rate
GO:0045177	apical part of cell	7	1.33e-07
GO:0016324	apical plasma membrane	6	1.39e-06
GO:0044459	plasma membrane part	7	0.00989
GO:0070062	extracellular exosome	7	0.02
GO:0098805	whole membrane	6	0.02
			(more)

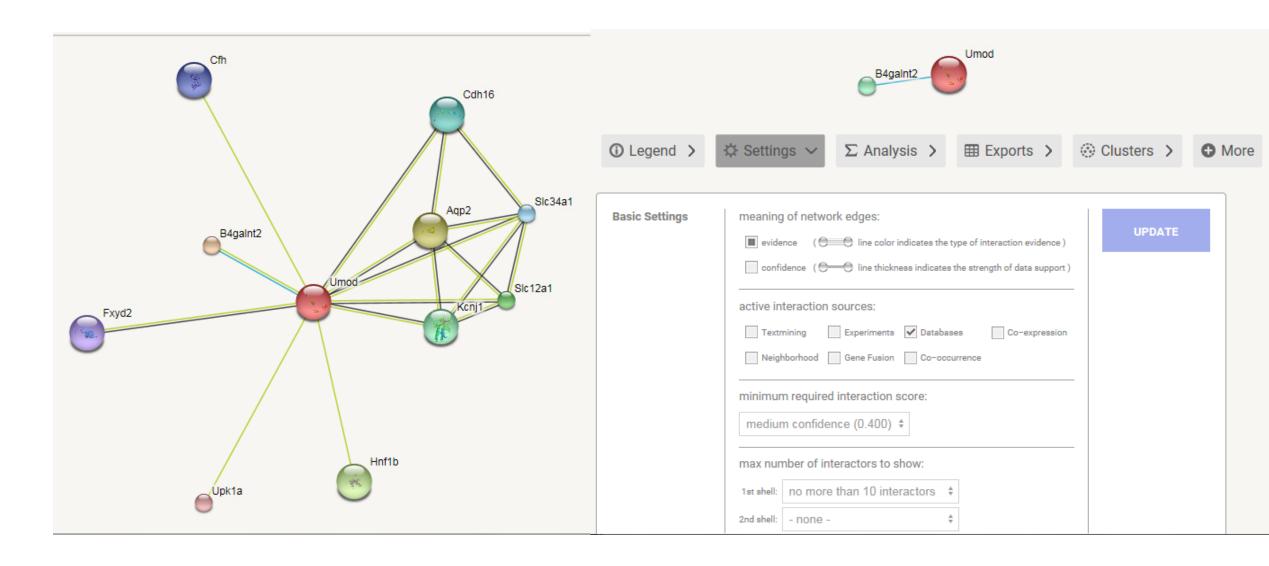
Statistical background

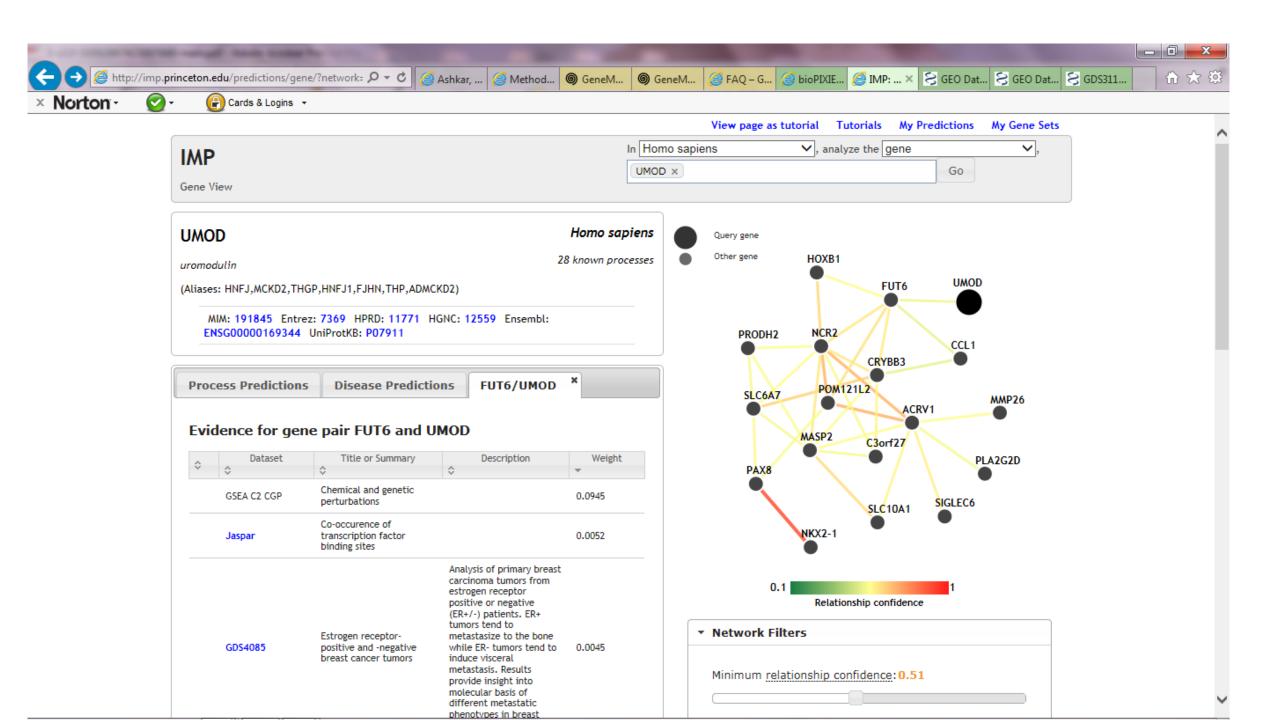
For the above enrichment analysis, the following statistical background is assumed:

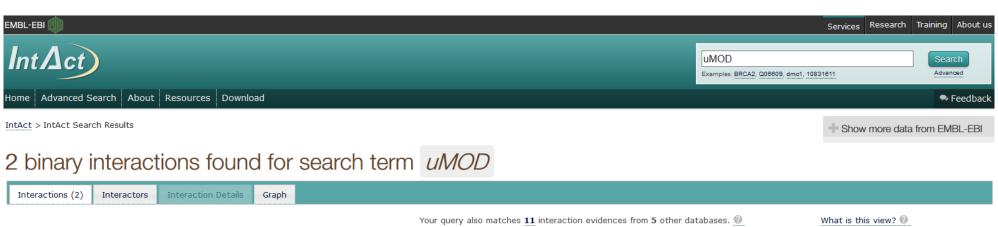
Vhole	Genome	`

UPDATE

STRING- Mouse UMOD

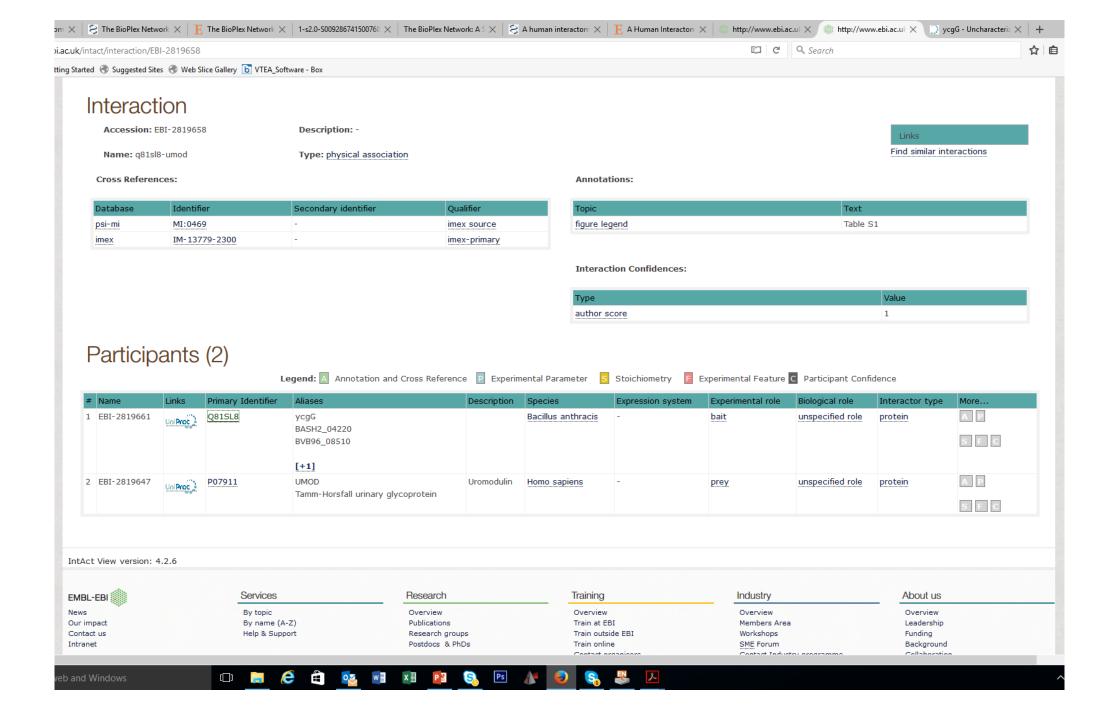




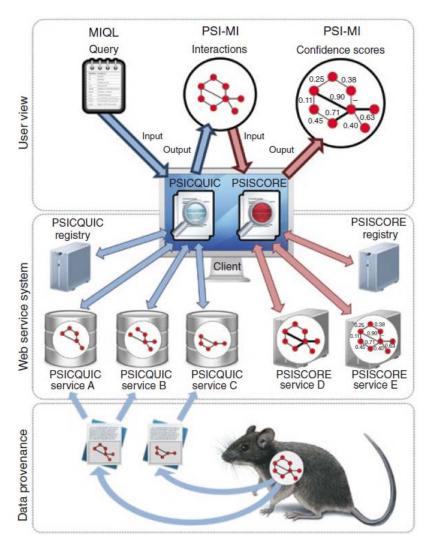




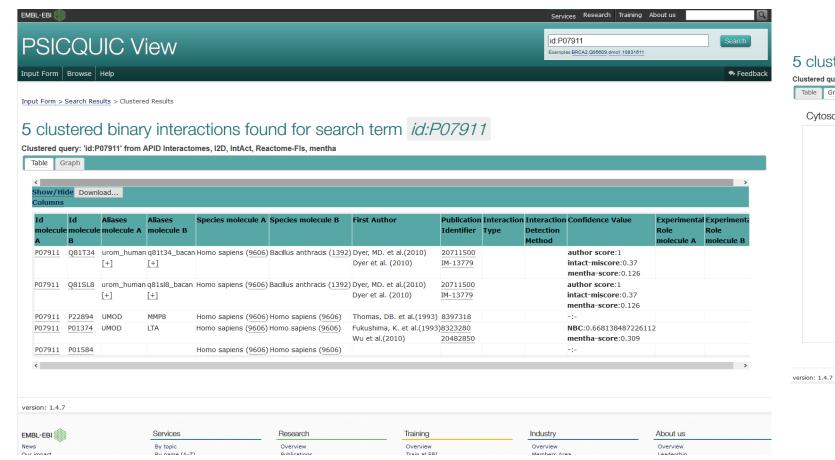
IntAct View version: 4.2.6 Research Services **Training** Industry About us EMBL-EBI News By topic Overview Overview Overview Overview Our impact Train at EBI Members Area By name (A-Z) Publications Leadership Contact us Help & Support Train outside EBI Workshops Funding Research groups Intranet Postdocs & PhDs Train online SME Forum Background Contact organisers Contact Industry programme Collaboration People & groups News Events Visit us Contact us EMBL-EBI, Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK +44 (0)1223 49 44 44



Proteomics Standards Initiative Common Query Interface (PSICQUIC)



PSICQUIC



5 clustered binary interactions found for search term id:P07911 Clustered query: 'id:P07911' from APID Interactomes, I2D, IntAct, Reactome-Fis, mentha Table Graph Cytoscape Graph Q6KV63 P01584