Binary Interactome Models of Inner- Versus Outer-Complexome Organization

Luke Lambourne,[1-3,18] Anupama Yadav,[1-3,18] Yang Wang,[1-3,18] Alice Desbuleux,[1-4,18] Dae-Kyum Kim,[1,5-7] Tiziana Cafarelli,[1-3] Carles Pons,[8] István A. Kovács,[1,9,10] Noor Jailkhani,[1-3] Sadie Schlabach,[1-3] David De Ridder,[1-3] Katja Luck,[1-3] Wenting Bian,[1-3] Yun Shen,[1-3] Miles Mee,[5,6,11] Yves Jacob,[1,12,13] Irma Lemmens,[14,15] Thomas Rolland,[1-3] Jan Tavernier,[14,15] Kerstin Spirohn,[1-3] Quan Zhong,[1-3,16] Patrick Aloy,[8,17] Tong Hao,[1-3] Benoit Charloteaux,[1-3] Frederick P. Roth,[1,5-7,11,*] David E. Hill,[1-3,*] Michael A. Calderwood,[1-3,*] Jean-Claude Twizere,[4,*] and Marc Vidal[1,2,*]

[1]Center for Cancer Systems Biology (CCSB), Dana-Farber Cancer Institute, Boston, MA, USA
[2]Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA, USA
[3]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA, USA
[4]Laboratory of Viral Interactomes, Molecular Biology of Diseases Unit, GIGA Institute, University of Liege, Liege, Belgium
[5]The Donnelly Centre, University of Toronto, Toronto, ON, Canada
[6]Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada
[7]Lunenfeld Tanenbaum Research Institute, Sinai Health System, Toronto, ON, Canada
[8]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute for Science and Technology, Barcelona, Catalonia, Spain
[9]Network Science Institute, Northeastern University, Boston, MA, USA
[10]Wigner Research Centre for Physics, Institute for Solid State Physics and Optics, Budapest, Hungary
[11]Department of Computer Science, University of Toronto, Toronto, ON, Canada
[12]Unité de Génétique Moléculaire des Virus à ARN (GMVR), Département de Virologie, Institut Pasteur, UMR3569, Centre National de la Recherche Scientifique (CNRS), Paris, France
[13]Université Paris Diderot, Paris, France
[14]Center for Medical Biotechnology, Vlaams Instituut voor Biotechnologie (VIB), Ghent, Belgium
[15]Cytokine Receptor Laboratory (CRL), Department of Biomolecular Medicine, Faculty of Medicine and Health Sciences, Ghent University, Ghent, Belgium
[16]Department of Biological Sciences, Wright State University, Dayton, OH, USA
[17]Institució Catalana de Recerca I Estudis Avançats (ICREA), Barcelona, Catalonia, Spain
[18]These authors contributed equally
*Correspondence: marc_vidal@dfci.harvard.edu (M.V.), jean-claude.twizere@uliege.be (J.-C.T.), michael_calderwood@dfci.harvard.edu (M.A.C.), david_hill@dfci.harvard.edu (D.E.H.), fritz.roth@utoronto.ca (F.P.R.)

SUMMARY

The interactome is often conceived of as a collection of hundreds of multimeric machines, collectively the “complexome”. However, the large proportion of the proteome that exists outside of the complexome, (the “outer-complexome”) is expected to account for most of the functional plasticity exhibited by cellular systems. To compare features of inner- versus outer-complexome organization, we generated an all-by-all yeast systematic binary interactome map, integrated it with previous binary maps, and compared the resulting binary interactome “atlas” with systematic co-complex association and functional similarity network maps. Direct binary protein-protein interactions in the inner-complexome tend to be readily detected in different assays and exhibit high levels of coherence with functional similarity relationships. In contrast, pairs of proteins connected by relatively transient, harder to detect binary interactions in the outer-complexome appear to exhibit higher levels of functional heterogeneity. Thus a small proportion of the interactome corresponds to a stable, functionally homogeneous inner-complexome, while a much greater proportion consists of mostly transient interactions between pairs of functionally heterogeneous proteins in the outer-complexome.

INTRODUCTION

Cells perform large numbers of biochemical tasks that are tightly controlled in response to complex, ever-changing environmental conditions. As important as lactose permease and β-galactosidase (β-gal) respectively are to transport and cleave lactose when it is present, the role of LacI in repressing these activities in the absence of lactose is also important for short-term cellular physiology and crucial for long-term evolutionary fitness (Jacob and Monod, 1961). Hence the dichotomy between; i) "*structural genes*", such as *lacZ* and *lacY*, which code for enzymes, transporters and other types of polypeptides that perform specific biochemical tasks; and ii) "*regulator genes*", such as *lacI*, which encode regulators that control and coordinate these biochemical activities. How these two fundamental aspects of cellular life interconnect at the scale of the whole proteome remains largely unresolved. What relative proportion of the proteome, in terms of overall mass and number of polypeptides, is involved in processing or manufacturing systems versus regulatory or control systems? Are these two parts of the proteome distinguishable in terms of fundamental principles underlying their global organization? Might the strikingly distinct roles of these systems correspond to distinct types of sub-proteome organization, allowing the "manufacturing proteome" to operate in a relatively constant, robust and persistent manner, and the "regulatory proteome" to exhibit greater flexibility, plasticity, environmental responsiveness and, perhaps, evolvability?

A huge part of the manufacturing proteome produces and catabolizes metabolites through chemical reactions catalyzed by enzymes, which collectively make up ~40% of cellular proteomic mass and ~10% of all polypeptides (Liebermeister et al., 2014). These processes are often organized in a "channel-like" manner effected by protein-protein interactions (PPIs) (Narayanaswamy et al., 2009; Petrovska et al., 2014; Srere, 1987). The manufacturing proteome is also composed of large molecular machines or protein complexes that produce, process, transport or degrade large

macromolecules such as DNA, RNA or proteins. The full complement of complexes (Benschop et al., 2010; Cherry et al., 2012; Meldal et al., 2019; Pu et al., 2009), or "complexome" (Deshaies et al., 2002), contains on the order of 30% of the protein-coding genes and together accounts for ~40% of the proteomic mass of the cell (Liebermeister et al., 2014). Large machines are assembled as sets of intertwined direct, binary PPIs between subunits, often contacting one another via discrete interaction-mediating domains.

Global organizational principles have also been identified for the regulatory proteome. Large numbers of signaling or regulatory "modules" are involved in responding to developmental and environmental changes, cell cycle progression or stress-response (Hartwell et al., 1999). While this part of the proteome only comprises ~20% of the cellular protein mass, it involves more than 60% of all proteins (Liebermeister et al., 2014). Many signal transduction pathways are tethered together as PPI-dependent multimeric components by interactions with scaffold proteins (Choi et al., 1994; Gustin et al., 1998; Hohmann, 2009; Pawson, 1995). Another organizational principle is the co-localization of enzymes and activities impinging on shared biochemical processes within membrane-less organelles formed by liquid-liquid phase separation phenomena (Gall, 2003; Hyman et al., 2014; Mahen and Venkitaraman, 2012). PPIs involved in membrane-less organelles are often characterized by a greater degree of disorder and overall flexibility, and by "multivalent" proteins that can bind multiple proteins simultaneously.

Here, we investigate organizational principles in the interactome by comparing intra-complex interactions taking place "inside" each complex of the complexome, i.e., within the "inner-complexome", to "outer-complexome" interactions. Whether biophysical interactions in the inner- and outer-complexome exhibit different properties, or how such interactions might generally relate to functional relationships is not well understood. We used *S. cerevisiae* as a model system to address these questions by: i) generating for the first time an "all-versus-all" systematic binary interactome map based on

interrogation of nearly all yeast proteins; ii) integrating it with three previously available binary maps, and iii) comparing the resulting systematic binary interactome "atlas" to systematic co-complex association network and functional similarity network maps. We find strong support for an emerging model according to which a relatively small proportion of the interactome corresponds to the stable, functionally coherent, inner-complexome, while a vast majority of the interactome consists of interactions between functionally heterogeneous proteins, forming the outer-complexome.

RESULTS

Current Status of Yeast Binary Interactome Landscape

To distinguish inner- from outer-complexome, we subdivided the yeast proteome-by-proteome search space of 18 million protein-protein pairwise combinations into a "complexogram" organized in four different "zones" (Figure 1A; STAR Methods #1): i) Zone A, the inner-complexome, corresponds to all combinations where both proteins are subunits of the same complex; ii) Zone B corresponds to pairs of complex subunits where each subunit belongs to a different complex; iii) Zone C represents all pairwise combinations between complex subunits and proteins not known to belong to any complex, which we will refer to as "non-complex proteins"; and iv) Zone D corresponds to all pairwise combinations between non-complex proteins. While populated by highly abundant proteins that form high-molecular weight assemblies, the yeast inner-complexome Zone A covers a tiny proportion of the whole interactome search space, representing only ~17,600 co-complex protein pairs that make up only ~0.1% of all 18M possible protein pairs (see STAR Methods #2). By contrast, the outer-complexome, with its three zones, B, C, and D, represents ~99.9% of protein pairs (Figure 1A). Finally, we also organize complexograms according to abundance of non-complex proteins in Zones C and D, which leads to metabolic enzymes, which constitute 44% of the proteomic mass (Liebermeister et al., 2014; Nagaraj et al., 2012), occupying relatively small subzones of the outer-complexome in C and D (Figure 1A).

In order to understand fundamental differences between inner- and outer-complexome, it is crucial to distinguish between co-complex association maps and PPI maps that report directly contacting, or "touching" pairs of proteins, often referred to as "binary" interactions (Figure S1A). To investigate the extent to which currently available high quality binary interactome datasets cover Zones A, B, C, and D, we analyzed three different sources as they stood in August 2020 (Figure 1B; STAR Methods #3-4): i) systematic yeast two-hybrid (Y2H) proteome-scale maps collectively reporting 2,613 heterodimeric pairs from "Y2H-union" (Ito et al., 2001; Uetz et al., 2000; Yu et al., 2008); literature-curated binary pairs with multiple evidence representing 5,056 heterodimeric "Lit-BM-20" PPIs (Calderone et al., 2013; Chatr-Aryamontri et al., 2017; Licata et al., 2012; Orchard et al., 2014; Salwinski et al., 2004); and iii) a set of PPIs with three-dimensional (3D) structural information, which provides definitive proof of physical, direct interaction between 1,761 experimentally derived heterodimeric Interactome3D or "I3D-exp-20" pairs (Mosca et al., 2013) (Figures 1B, S1B-C). Systematic Y2H datasets were released in three publications and produced using a single assay method, although the details of the Y2H versions used were different for each map. In contrast, Lit-BM-20 information was curated from a total of 6,127 publications and produced with a total of 81 different PPI assays and versions thereof, among which Y2H is the most widely used, with ~76% of heterodimeric Lit-BM-20 interactions (3,986/5,056) supported by at least one Y2H piece of evidence (Figure 1B). Lit-BM-20 makes up only 4% of all literature-curated pairs since the vast majority of curated pairs correspond to co-complex associations and/or are supported by a single experiment.

To compare the quality and genuine binary interaction content of literature-curated versus systematic datasets, we used two orthogonal assays that were not used in any of these datasets so far (Figure 1C, STAR Methods #5-6): i) the mammalian protein-protein interaction trap (MAPPIT) assay (Eyckerman et al., 2001); and ii) the mammalian *Gaussia princeps* complementation assay (GPCA) (Cassonnet et al., 2011). We benchmarked MAPPIT and GPCA against a new positive reference set of

108 well-characterized yeast PPIs (scPRS-v2), and, as a negative benchmark, a new random reference set of 198 random protein pairs (scRRS-v2) (see STAR Methods #7). Across a wide range of stringency for both MAPPIT and GPCA, all three systematic Y2H datasets constituting Y2H-union, i.e. "Uetz-screen" (Uetz et al., 2000), "Ito-core" (Ito et al., 2001) and "CCSB-YI1" (Yu et al., 2008) validated *on par* with both scPRS-v2 and "Lit-BM-13", a random sample of literature binary datasets assembled in 2013 (Figure 1C). At the lowest stringency threshold returning 0/198 scRRS-v2 positives, MAPPIT recovered scPRS-v2, Lit-BM-13 and systematic Y2H pairs at rates between 18 and 27% that were statistically indistinguishable from one another. At a similarly stringent threshold, GPCA recovered significantly fewer Lit-BM-13 pairs than Y2H-union pairs (Figure 1C). Interestingly, a proteome-scale dataset generated using a dihydrofolate reductase protein complementation assay (DHFR PCA), which detects physically-proximal, but not necessarily directly contacting, protein pairs (Tarassov et al., 2008), was indistinguishable from the RRS negative control in MAPPIT, but validated *on par* with the Y2H-union datasets in GPCA, (Figure 1C). As shown before for human (Rolland et al., 2014), putative yeast binary PPIs supported by only a single piece of evidence in the literature (Lit-BS) failed to validate with either assay and thus were not used further here (Figure S1D). Finally, combining MAPPIT and GPCA leads to ~30% recovery of Y2H-union, which again, is on par with the recovery of positive control pairs and significantly higher than Lit-BM-20 recovery (Figure 1C bottom panel), indicating that the Y2H-union map is of high quality.

To further investigate possible differences and similarities of direct, binary PPI information content between Interactome3D, literature curation, and systematic binary maps, we retested all binary interactions available in 2017 for the three approaches, altogether representing a total of 11,234 PPI pairs (I3D-exp-17, Lit-BM-17, Y2H-union) (Figure 1D). As an orthogonal Y2H version, we used Y2H v4, a modified Y2H version (see STAR Methods #8) that had not been used for systematic screening. Y2H v4 detects 19% (12-28%, 95% CI) of successfully tested scPRS-v2 pairs under conditions in which 0/198 scRRS-v2 are recovered (Figure 1D). Importantly, the subsets of

scPRS-v2 pairs detected by Y2H v4 or Y2H v1 are as different from each other as they each are from MAPPIT (Figure S1E, STAR Methods #9), further supporting the notion that interactions detectable by different versions of the same assay can show substantial independence (Choi et al., 2019). Together, MAPPIT, Y2H v1 and Y2H v4 detect 37% of scPRS-v2 pairs under conditions that detect 1/198 scRRS-v2 pairs (Figure S1E). Under these conditions, Y2H v4 recovered 11%, 14%, and 15% of Interactome3D, literature curation, and systematic binary pairs respectively. Strikingly, Y2H-union was indistinguishable from scPRS-v2 (P = 0.2, Fisher's exact test), and from I3D-exp-17 (P = 0.6, Fisher's exact test) our benchmark of true direct, heterodimeric binary interactions, and they both performed slightly better than Lit-BM-17 (P = 0.006, Fisher's exact test) (Figure 1D), again demonstrating that the biophysical quality of systematic datasets is at least as good, if not superior to that of literature-curated interactions.

In striking contrast, the binary, direct, proximal interaction content of co-complex association datasets was very close to that of the negative control scRRS-v2 (Figure 1D). Interpretation of co-complex association is complicated by large proportions of indirect, distal associations (Figure S1A), both in literature-curated information (Pu et al., 2009) and in AP-MS derived proteome-scale maps (Gavin et al., 2002, 2006; Ho et al., 2002; Krogan et al., 2006). More specifically, our analysis showed that: i) for CYC2008, the recovery rate by Y2H v4 was only slightly higher than that of our negative control scRRS-v2; ii) pairs curated from the literature with no clear evidence of direct, binary interactions (Lit-NB), which represent 84% of all protein pairs curated (100,102/119,288) were indistinguishable from random (P = 0.27, Fisher's exact test); and iii) the four systematic, non-binary, AP-MS co-complex association maps available heretofore (Gavin et al., 2002, 2006; Ho et al., 2002; Krogan et al., 2006) were recovered by Y2H v4 at rates similar to CYC2008, i.e. with approximately a four-fold lower recovery rate compared to the positive controls scPRS-v2 and Interactome3D (averages of 5% vs 21%, $P = 6 \times 10^{-11}$, Fisher's exact test) (Figure 1D). This observation is consistent with the fact that, wherever structural information is available, protein pairs

involved in PPIs obtained by binary assays are two to five times more likely to be in direct contact than co-complex association literature-curated and systematic pairs as well as Lit-BS pairs (Figure S1F, STAR Methods #10).

Having tested a comprehensive set of Interactome3D pairs allowed us to address a few long-standing issues pertaining to the use of Y2H to systematically map the binary interactome. First, although the number of subunits involved in forming large protein complexes available in Interactome3D appears to have some impact on the rate of interaction recovery by Y2H v4 (Figure 1E), these results demonstrate that binary assays can readily detect pairs of interacting proteins even in large complexes. Second, although all Y2H v4 attempts were generated with full-length yeast proteins, the detection rate appeared unaffected by whether the co-crystal structures of interacting proteins had been solved with full-length proteins or discrete domains thereof (P = 0.14, Kolmogorov-Smirnov test) (Figure S1G). Third, although, we observed a trend towards larger surface areas among Y2H v4 pairs compared to Y2H v4 negatives, Y2H v4 was able to detect interactions with interface areas ranging widely from 100 to 10,000 $Å^2$ (Figure S1H, STAR Methods #11). Y2H v4 was better able to detect PPIs with small interaction interfaces compared to MAPPIT and GPCA (Figure S1I). Finally, Y2H v4 detected interactions with K_d values as high as 10^{-6} M (Figure S1J, STAR Methods #12), consistent with the results of a previous study (Estojak et al., 1995).

To investigate overall homogeneity of coverage, we displayed PPIs in a representation of the search space organized by ranking proteins in both dimensions according to their number of publications (Rolland et al., 2014) (Figure 1F, STAR Methods #13). Most Interactome3D or literature-curated PPI pairs belong to a “dense zone” with 15% or 25% of the whole yeast search space containing 80% of the information, respectively. In striking contrast, Y2H-union interacting pairs are distributed more homogeneously across the yeast search space with 80% of pairs belonging to over 50% of the space (Figure 1F). Likewise, the relative coverage of binary interactions in the four zones defined above, A, B, C and D, is very different between the three

sources of PPIs, with Interactome3D pairs being highly over-represented in Zone A (65% of pairs) while PPIs from systematic maps are more homogeneously distributed (Figure 1G).

In summary, even though Interactome3D, literature-curated information and systematic maps are comparable to each other and vastly superior to co-complex association datasets in terms of overall quality of binary, direct interaction content, systematic maps present a clear advantage over literature-curated information for global interactome network analyses with their greater homogeneity of coverage of the search space.

A First All-Versus-All Reference Binary Map of the Yeast Interactome

Although systematic mapping strategies appear advantageous to inform us about inner- versus outer-complexome organization and other properties of the interactome, the three maps currently available fail to cover the whole yeast interactome with full homogeneity (Figure 1F). This is in part due to limitations of both assay and sampling sensitivity of the Y2H versions used so far and, importantly, "search space incompleteness" of the screens (Yu et al., 2008). For example, CCSB-YI1 and Ito-core were obtained using incomplete open reading frame (ORF) collections, or "ORFeomes" (Walhout et al., 2000), so that only ~70-75% of the search space were screened in these studies and only ~60% screened by both mapping projects.

To improve coverage of the yeast binary interactome, we collected a complete ORFeome of high quality ORFs by verifying an already existing, yet incomplete collection of 4,933 ORFs (Hu et al., 2007) and cloning an additional 921 ORFs (19% increase), together representing 99% of the annotated protein-coding genes (5,854/5,883) (See STAR Methods #14, #15). To maximize independence with the three previous maps, we moved our complete ORFeome into Y2H v4 vectors (see STAR Methods #16) and systematically screened a space of 2.7 x 10^7 bait-prey combinations three times independently (Figure 2A, STAR Methods #17-18). Pairs identified from

these primary screens were then retested in two pairwise tests under conditions maximizing the recovery of scPRS-v2 while minimizing that of scRRS-v2 (see STAR Methods #19). Putative interacting pairs that were positive in both attempts, and for which the identity of both ORFs could be sequence-confirmed (see STAR Methods #20, #21), were validated using both MAPPIT and GPCA (see STAR Methods #5). To maximize statistical power of this validation step, *all* pairs of Y2H v4 positives, rather than a random sample of pairs as we implemented in our original empirical framework (Venkatesan et al., 2009; Yu et al., 2008), were compared under conditions recovering on average 20% of scPRS-v2 and none of the 198 scRRS-v2. Thus, we mapped a new "yeast reference interactome" or "YeRI" dataset of 1,910 PPI pairs between 1,351 proteins, of which ~75% are novel, that exhibited validation at a rate similar to Lit-BM-20 in both MAPPIT and GPCA (Figures 2A, S2A). YeRI displays significant enrichment for interactions between proteins that share a cellular compartment, pathway and protein complex annotation (Figure 2B, STAR Methods #22), demonstrating its overall high levels of biological relevance.

We investigated the extent to which YeRI expands coverage of the interactome within the search space, relative to the three previous maps using the ranking by number of publications as described above (Rolland et al., 2014). Compared to Uetz-screen, Ito-core and CCSB-YI1, YeRI covers the search space substantially more homogeneously (Figure 2C). Since all four maps, Uetz-screen, Ito-core, CCSB-YI1 and YeRI were shown to validate to the same extent by MAPPIT and GPCA, we combined all 4,556 Y2H pairs into a single "atlas of binary biophysical interactions", referred to as "ABBI-21", with YeRI adding 1,723 new PPIs to the preexisting Y2H-union (Ito et al., 2001; Uetz et al., 2000; Yu et al., 2008).

To test whether functionally-related proteins tend to be closely connected in ABBI-21, we performed network-based spatial enrichment analysis for functions and phenotypes (SAFE), to detect functional modules in systematic biophysical network maps (Baryshnikova, 2016) (see STAR Methods #23). By progressively combining

systematic maps based on their chronological order of generation, we show that the numbers of functionally enriched clusters increase substantially with the addition of each map (Figure 2D). Running SAFE on YeRI individually finds 12 functional modules (Figure S2B). ABBI-21, with the addition of YeRI, is of equivalent size to Lit-BM-20 (Figure 2E). ABBI-21, covers 12-25% of the estimated yeast binary interactome (Figure 2F, STAR Methods #24) (Sambourg and Thierry-Mieg, 2010; Stumpf et al., 2008; Yu et al., 2008).

Despite yeast being one of the most well-studied organisms, gene function remains completely unknown for almost one-sixth of genes (Wood et al., 2019) (see STAR Methods #25). We investigated the number of PPIs involving the products of these genes in both literature and systematic maps. Systematic maps identify substantially more PPIs connecting proteins encoded by 'genes of unknown function' than literature-derived maps (Figure S2C). Altogether, 33% of such proteins have at least one interaction in ABBI-21, while 19% have interactions identified only in YeRI (Figure 2G). Given the lack of progress in characterizing the functions of these genes, systematically generated PPIs provide a potential method to infer their cellular roles. We predicted functions of genes of unknown function using a guilt-by-association approach with first and second degree neighbors using their GO term annotations (Luck et al., 2020) (see STAR Methods #26), in both ABBI-21 and Lit-BM networks. Using ABBI-21, the most highly ranked prediction resulting from the addition of YeRI was for *PEX35* (*YGR168C*) to function in the peroxisomal importomer complex. In fact, the function of *PEX35* is known: it has been demonstrated to be a regulator of peroxisomal abundance (Yofe et al., 2017). Correctly predicting the function of this gene, misclassified as being of unknown function, supports the idea that using a guilt-by-association approach with our PPI network is an effective way to predict gene function. Pex35 has 23 PPIs in ABBI-21, all from YeRI, out of which eight are proteins involved in peroxisomal biology (Figure 2H). Pex35 also interacts with another protein encoded by a gene of unknown function, YKL018C-A, also predicted to be involved in the same function (Figure 2H). Another highly ranked prediction is for YJR015W to facilitate endoplasmic reticulum

(ER) transport activity, based on its protein interactions with components of the ER secretory pathway such as Sec11, Spc1 and Sar1. Supporting this prediction, YJR015W was demonstrated to be localized to the ER (Koh et al., 2015).

High Discrepancy Among Alternative Views of Inner- Versus Outer-Complexome Organization

While β-gal homotetramerization and the LacI/operator protein-DNA interaction are equally relevant to the Lac operon system, the former is much more conveniently detected biochemically than the latter, largely owing to the many orders of magnitude difference in abundance between the two molecules (Gilbert and Müller-Hill, 1966; Juers et al., 2012). To investigate similar global differences of relevance versus abundance at the scale of the whole interactome, we compared currently available views of inner- versus outer-complexome organization based on co-complex association versus binary interaction maps using our complexogram strategy (Figure 1A). In each dimension of the yeast search space, we: i) separated all 5,883 proteins between 1,897 complex subunits and 3,986 non-complex proteins; ii) ordered complex subunits in decreasing order of the number of different proteins belonging to their cognate complex, with a range between 81 and 3 subunits; iii) ordered non-complex proteins by increasing abundance, where there is a three-order of magnitude range between low and high abundance; resulting in: iv) a majority of metabolic enzymes located in high abundance subzones in Zones C and D (Figure 3A).

The co-complex association Sys-NB-06 map (Gavin et al., 2002, 2006; Krogan et al., 2006) is, as expected, extremely enriched for associations in the inner-complexome Zone A and to a lesser extent in Zone B. In addition, Sys-NB-06 is enriched in a subzone of Zone C containing complex subunit and highly abundant, mostly metabolic enzymes, and in a small subzone of Zone D mostly containing metabolic enzymes (Figure 3A). The literature-curated Lit-BM-20 map follows similar trends with the exception of the two C and D subzones covering relatively abundant proteins (Figure 3A). In striking contrast, the systematic binary atlas ABBI-21, is less enriched in Zone A,

and shows a highly homogenous distribution across Zones B, C, and D (Figure 3A). These results are most likely explained, for Lit-BM-20, by a general study bias towards highly abundant proteins, due to their ease of use with various molecular biology experimental techniques, and for Sys-NB-06, by the specific technical bias of AP-MS towards highly abundant proteins and relatively stable associations. ABBI-21 is the most uniform in the outer-complexome, for proteins of both high and low abundance, compared to Sys-NB-06 or Lit-BM-20. In short, we observe high discrepancy among alternative views of inner- versus outer-complexome organization obtained from co-complex associations as opposed to binary interaction maps.

To explore other differences between maps representing co-complex associations as opposed to binary interaction maps, we examined five intrinsic protein properties potentially associated with protein complex membership: i) number of publications (see above); ii) protein abundance; iii) evolutionary conservation; iv) essentiality; and v) disorder, by comparing the average values of complex subunits and non-complex proteins (Figure 3B, STAR Methods #27). Complex subunits exhibit significantly higher average values in all five variables ($P < 10^{-83}$, Mann-Whitney *U* test), independent of the complex curation dataset used (Figures S3A, S3B), with strong correlations between these protein properties (Figure S3C). We thus ranked the proteome-by-proteome search space according to these variables and compared the coverage of the different binary PPI and co-complex association maps (Figure 3C). In all cases, ABBI-21 covers the interactome more homogeneously, across multiple biological properties, compared to Lit-BM-20 and Sys-NB-06, even when restricted to Zone D (Figure S3D). In the lowest abundance and/or conservation zones, the number of proteins for which at least one interactor is reported is significantly higher for ABBI-21 than Lit-BM-20 or Sys-NB-06 (Figure S3E).

To understand the extent to which currently available maps might be biased, relative to the real interactome, towards containing pairs from the inner- rather than the outer-complexome, we used gold-standard, reference interactions from the inner- and

outer-complexome, examining potential differences in how well ABBI-21 and Sys-NB-06 might be able to capture these two different types of interactions. We extracted four sets of gold-standard functionally characterized PPIs from different datasets (see STAR Methods #28), two for the inner-complexome: i) complex subunits in direct contact in experimental structural data (Mosca et al., 2013) and ii) co-complex associations in signaling pathways from the KEGG database (Kanehisa et al., 2019); and two for the outer-complexome: i) PPIs from KEGG, not in the same protein complex; and ii) the high-quality subset of literature-curated kinase-substrate interactions from the KID database (Sharifpoor et al., 2011). We observe that, although both co-complex association and binary interaction maps capture the inner-complexome pairs more readily, ABBI-21 is closer to being equally good at capturing both inner- and outer-complexome pairs than Sys-NB-06, which is significantly more biased towards capturing inner-complexome pairs (Figure 3D).

To investigate whether PPIs in ABBI-21 from both the inner- and outer-complexome are of similar biophysical quality, we compared their recovery rates in MAPPIT and GPCA, using Lit-BM-13 as a benchmark (Figures 3E, S3F). The higher positive rate in the inner-complexome of ABBI-21 pairs relative to Lit-BM-13 is likely explained by their higher fraction of direct vs indirect PPIs (Figure 2H). ABBI-21 pairs from the outer-complexome show lower recovery rates ($P = 3 \times 10^{-12}$, Fisher's exact test) than those in the inner-complexome. That the rates of outer-complexome pairs of the Lit-BM-13 benchmark are also lower and indistinguishable from ABBI-21, suggests that the difference in recovery rates stems from differing biophysical factors, for example interaction affinity or requirement of PTMs for the interaction, between the inner- and outer-complexome. This demonstrates that ABBI-21 pairs in the outer-complexome are of good biophysical quality. The striking observation of inner-complexome PPIs being more readily detected by different PPI assays, implies that inner-complexome PPIs tend to be overrepresented in interactome maps relative to their proportion in the real interactome.

Functional Profile Similarity Networks are Enriched in the Inner-Complexome

A comprehensive understanding of cellular biology requires integrating information of multiple types of biological relationships (Vidal et al., 2011). To understand the functional relationships between physically interacting proteins, we used 'functional networks' which link genes/proteins with similar profiles of different molecular and phenotypic readouts. These profile similarity networks (PSNs) came from three different published, comprehensive and systematic sources: genetic interactions (GI-PSN) (Costanzo et al., 2016), condition sensitivity (CS-PSN) (Hillenmeyer et al., 2008), and mRNA expression (GE-PSN) (Obayashi et al., 2019). We constructed each PSN from the 1% highest PCCs of all measured pairs, giving GI-PSN 135,000 edges, CS-PSN 65,000 edges and GE-PSN 99,000 edges, making them much larger than the 5,000 to 13,000 edges of the biophysical networks. These PSNs exhibited a significant but small overlap with each other, after restricting to the genes tested in all three (4% of tested pairs are connected in more than 1 PSN, $P < 0.001$, one-sided empirical test), suggesting that they are capturing complementary distinct functional similarities (Figure S4A).

Since each functional network is capturing distinct information, we investigated how their connections are distributed between the inner- and outer-complexome. We repeated our analysis of the coverage of the space of pairwise protein combinations according to different protein variables, using the PSNs in place of the PPI networks (Figures 4A, S4B). We observed that all three PSNs are highly enriched in the inner-complexome, with the largest protein complexes having the highest density of interactions. GI-PSN and GE-PSN both showed a higher density between highly expressed, more conserved and essential genes (Figure 4B). GI-PSN shows a high density of interactions between the most well-studied proteins (Figure 4B), which was unexpected since this network is systematically generated. Upon investigation we found this was due to a combination of the higher density for essential genes (Costanzo et al., 2016) and the highly studied, highly abundant non-essential ribosomal subunits being

very densely connected in GI-PSN (Figure S4C). Many yeast ribosomal proteins retained paralogs after a whole-genome duplication event (Wolfe and Shields, 1997) often rendering them non-essential where paralogs can functionally compensate for one another's deletions. Their high connectivity illustrates the preference of genetic interactions to capture biological processes with high functional redundancy. Because we used data from the homozygous deletion collection to assemble condition sensitivity profiles, essential genes were not included in CS-PSN, which likely explains the relatively more homogeneous distributions of CS-PSN (Figure 4B).

Since the different functional networks are enriched in protein complexes, we investigated the biophysical nature of interactions between proteins connected in functional PSNs. As an example, the experimentally solved Mediator structures show 33 direct contacts between the 25 proteins, much less than the 300 possible pairwise combinations (Figure 4C, STAR Methods #29). This is unsurprising, as the number of other proteins a complex subunit may be in contact with is fundamentally limited by the surface area of the subunit. The number of direct interactions within a protein complex scales roughly linearly with the size of the complex, whereas the number of indirect interactions scales quadratically. ABBI-21 and Lit-BM-20 find primarily direct interactions, whereas the AP-MS based Sys-NB-06 finds a majority of both direct and indirect pairs of proteins. In addition, GI-PSN connects genes encoding proteins within the complex that are in both direct and indirect contact, which is understandable as all the proteins in the complex collectively contribute to a common function, irrespective of whether they are in direct contact. We look at this trend across the different protein complexes (Figures 4D) and observe that the binary PPI datasets primarily find directly-contacting pairs, whereas both Sys-NB-06 and the PSNs connect both directly-contacting and indirect pairs. ABBI-21 is the most enriched for direct PPIs vs indirect associations.

After observing that within protein complexes, the functional networks connect both direct PPIs and indirect associations, we then tested GI-PSN pairs in binary PPI

assays, obtaining the first direct experimental estimate of the fraction of these functional relationships that are mediated by binary PPIs. We tested random samples of GI-PSN pairs with different PCC cutoffs (Figures 4E, 4F, S4D) in Y2H v4 and GPCA. The positive test rate increases with the increasing PCC threshold, and on average is similar to that of the CYC2008 protein complex dataset. Together this is consistent with protein complexes dominating the high PCC GI-PSN pairs (Costanzo et al., 2016) and GI-PSN linking both directly and indirectly contacting complex subunits (Figures 4C, 4D) and consistent with the reported higher average PCC values for direct than indirect (Meldal et al.).

To further examine how biophysical and functional networks relate to each other at a global-scale, we examined their network topology. The degree distributions of the biophysical networks are heavy-tailed, with most proteins having a low number of interactions and a few having a very large number (Figure S4E, STAR Methods #30). ABBI-21 and Lit-BM-20 show a scale-free degree distribution, previously observed for PPI networks (Albert, 2005). In contrast, the degrees of genes in the functional networks are more uniformly distributed. The degree correlations of the biophysical networks are weak, negative for the binary maps and positive for the non-binary map. The functional networks have strongly positive degree correlations, meaning that genes with similar degree tend to be connected to each other (Figure S4F). This 'assortative' network topology is consistent with the PSNs containing highly-connected modules formed by protein complexes, where each subunit of the complex would have a similar degree of roughly the number of subunits in the complex. Binary PPI maps, on the other hand, do not connect complexes as densely, as they are primarily direct PPIs rather than indirect associations, and they also connect proteins involved in different functional modules.

The Binary Interactome Mainly Consists of Interactions Between Functionally Heterogeneous Proteins in the Outer-Complexome

Systematically generated functional networks can now be used to investigate the functional relationships mediated by PPIs at genome scale. We computed the fraction of protein pairs from Lit-BM-20, Sys-NB-06 and ABBI-21 where the corresponding gene pairs were also connected in functional networks (Figures 5A, STAR Methods #31). Pairs from Lit-BM-20 and Sys-NB-06 show a higher likelihood to connect proteins that are also connected in PSNs, compared to ABBI-21. We observed similar patterns for all three biophysical datasets when directly using the positive and negative genetic interactions, instead of the profile-similarities. In general, genes encoding interacting proteins found in all biophysical maps have a higher likelihood to show negative genetic interactions than positive genetic interactions (VanderSluis et al., 2018) (Figure S5A), and genes encoding protein pairs from the inner-complexome have an increased tendency to show negative GI than pairs in the outer-complexome (Figure S5A). ABBI-21 and Lit-BM-20 have similar biophysical quality as inferred from validation rates (Figure 2A) and functional networks are enriched in highly abundant, conserved inner-complexome direct and indirect pairs (Figures 4A-C). Therefore, the differential likelihood of pairs from different biophysical maps to be connected in the three PSNs suggests varying functional relationships between proteins interacting in the three biophysical maps rather than interaction quality.

Investigating the difference in the fraction of PPIs connected in the functional networks, we note is that ABBI-21 and Lit-BM-20 are not directly comparable as ABBI-21 was produced by uniformly testing a full proteome-by-proteome space, whereas most studies from the literature tend to focus on a particular pathway or process of interest. As an example we looked at a single study (McCann et al., 2015) which, of all the individual publications curated to form Lit-BM-20, contributed the publication that contributes the largest number of PPIs that intersect with the functional networks: 5% and 26% of the subset of Lit-BM-20 pairs that are also connected in GI- and GE-PSN, respectively. McCann et al. tested every pairwise combination of 70 pre-ribosomal proteins using Y2H. The detected PPIs have a very high rate of being connected in the GI- and GE-PSNs of 40% and 91% respectively but not much higher

than the fraction of every tested combination, which is 29% and 88%. For comparison, the corresponding value is only 1% for the proteome-by-proteome space. Thus, the high overlap likely stems more from the choice of protein pairs to test rather than the specific interactions are detected. So the rate of overlap of ABBI-21 with functional networks should more accurately reflect the real binary protein interactome.

The functional relationships between proteins interacting in the inner- and outer-complexome were investigated by comparing the ability of PSNs to connect pairs from the inner- and outer-complexome within the three biophysical maps (Figure 5B). Interactions from the inner-complexome have a uniformly high probability of being connected in the functional networks whether from Lit-BM-20, Sys-NB-06 or ABBI-21 (Figure 5B). By contrast, interactions from all three zones of the outer-complexome, B, C and D, in all three biophysical maps are found at a much lower extent in the functional networks. This demonstrates a higher tendency for functional networks to preferentially connect constitutive, co-complex interactions rather than the likely-transient interactions of the outer-complexome. These three biophysical datasets differ substantially in their composition of inner-complexome pairs, with more than one-quarter of Lit-BM-20 interactions and Sys-NB-06 associations being between proteins that are subunits of the same complex (Zone A) compared to around one-tenth for ABBI-21 interactions (Figure 5B). This difference contributes to the overall reduced tendency of ABBI-21 pairs to be connected in functional networks.

High degree, or hub proteins in biophysical maps can be classified as either ‘date’ or ‘party’ hubs depending on the degree to which interacting partners are also co-expressed (Han et al., 2004). We observed a similar trend for party and date hubs as we did for the inner- and outer-complexome. Party hubs tend to be in complexes, whereas date hubs tend to be outside complexes (Kim et al., 2006). Systematic binary maps have mostly date-hubs whereas literature-curated maps have more party-hubs (Figure S5B, STAR Methods #32), consistent with previous maps (Yu et al., 2008). Party-hubs, which use multiple interfaces to bind multiple partners simultaneously,

overlap with functional networks twice as much as date-hubs, which usually interact with partners one at a time transiently (Figure S5B). There has been debate about the definitions of date and party hubs (Batada et al., 2007; Bertin et al., 2007) and we therefore repeated our analysis using a range of different cutoffs and observed consistent results across all the different cutoffs (Figures S5B, S5C). Our results suggest that while the literature-curated PPIs and indirect associations are biased towards high-affinity constitutive interactions, a systematically generated map of the binary interactome is enriched in transient, context-specific interactions.

We observe a similar bias with essential genes, where interactions between proteins encoded by essential genes show a higher likelihood to be connected in GI-PSN in all three binary datasets. However, Lit-BM-20 and Sys-NB-06 are biased towards proteins encoded by essential genes resulting in overall increased overlap with GI-PSN, whereas ABBI-21 covers the proteome and interactome more uniformly. (Figure S5D).

We next investigated how the degree of the proteins in biophysical networks influences their overlap with functional networks. Fraction of PPIs connected in the functional networks decreases as the degree of the interacting proteins increases across the different biophysical networks (Figure S5E, STAR Methods #33). We would expect high-degree proteins to be less functionally similar to their binding partners, as they are more pleiotropic (Yu et al., 2008) and that the interactions involving hubs would be more likely to be transient domain-motif binding (Kim et al., 2008). Together these results indicate that systematic binary maps have a higher tendency to identify transient, context-specific interactions.

Integrating Biophysical and Genetic Network Maps Aids Understanding Cellular Organization

Because functional networks show a substantially higher tendency to connect protein pairs in the inner-complexome (Figure 5B) than the outer-complexome, it is likely that

PSNs capture more constitutive cellular functions involving highly abundant, essential proteins rather than context-specific functions involving less abundant, non-essential proteins. We investigated our hypothesis by testing the likelihood of PPIs from four well-characterized yeast pathways from the KEGG database: Cell cycle; Meiosis; MAPK signaling; and Autophagy, each highly conserved across the eukarya (Kanehisa et al., 2019) to be connected in functional networks (Figures 6A). Interestingly, genes encoding interacting proteins from different pathways have a different likelihood of being connected in the functional networks. Over half of the interacting pairs in Cell cycle and Meiosis pathways, essential for yeast reproduction and growth, are connected in GI-PSN, whereas less than 20% of the interactions from the context-specific pathways, MAPK signaling and Autophagy, are detected (P = 0.002, Fisher's exact test). GE-PSN and CS-PSN show a similar bias towards cell-cycle and meiosis compared to MARK and Autophagy (Figure 6A).

To validate our results that functional networks have a higher tendency to connect interacting proteins from the inner- as opposed to the outer-complexome, we investigated the ability of PSNs to connect proteins involved in gold-standard, reference PPIs from both inner- and outer-complexome. We used the same set of reference PPIs used in Figure 4F. All three PSNs captured significantly more interactions from the direct co-complex and KEGG co-complex datasets than the outer-complexome regulatory PPIs in pathways (KEGG regulation) and kinase-substrate interactions (P ranges from 5×10^{-9} to 5×10^{-45}, Fisher's exact test, table 4) (Figures 6B). By demonstrating that even PPIs with well-understood functions from the inner- and outer-complexome have different tendencies to be connected in PSNs, these results support our conclusions from the overlap of systematic biophysical maps and PSNs (Figures 5A, 5B) that the inner-complexome tends to consist of functionally similar interacting proteins whereas the outer-complexome (representing the majority of systematic binary maps) tends to consist of interactions between functionally heterogeneous proteins necessary for intracellular crosstalk.

The uniform coverage of interactions in both the inner- and outer-complexome space by ABBI-21 can now be leveraged to elucidate molecular mechanisms. For example, the endosomal sorting complex required for transport (ESCRT) pathway plays a key role in the biogenesis of multivesicular bodies and turnover of membrane proteins (Elia et al., 2011). The main players in the ESCRT pathway are the five ESCRT complexes, supporting auxiliary proteins and the cargo to be sorted. By integrating biophysical and genetic networks, we observe that the core constituents of the five ESCRT complexes interact biophysically in both ABBI-21 and Lit-BM-20 and are highly interconnected in functional networks (Figure 6C). In contrast, outer-complexome ABBI-21 PPIs between subunits of ESCRT complex and non-complex proteins, important for endosomal sorting, are not connected in the functional networks. For example, ABBI-21 contains PPIs between Vfa1, important for vacuolar sorting, and Vps4 and Vta1, subunits of the ESCRT-4 complex. However, despite known functional roles, Vfa1 is not connected with Vps4 and Vta1 in any of the PSNs.

Systematic maps can begin to help us understand how proteins within and outside complexes function together to mediate various biological processes. One such example is that of Snn1, a subunit of the biogenesis of lysosome-related organelles complex 1, BLOC-1, important for endosomal maturation (Hayes et al., 2011; John Peter et al., 2013). In ABBI-21, Snn1 interacts with proteins of the ESCRT complex like Vps28 as well as other non-complex endosomal proteins like Nkp2 (Figure 6C). ABBI-21 interactors of Snn1 are significantly enriched in proteins located in endosomes (13%, vs 2% overall for proteins in ABBI-21, $P = 0.0007$, Fisher's exact test). In fact, five out of six BLOC-1 complex proteins have PPIs primarily in ABBI-21, and none of the interacting protein pairs are connected in any of the functional networks.

The uniform coverage of inner- and outer-complexome by ABBI-21 can also shed light upon potential mechanisms by which previously under-studied complexes act. For example, the oxidant-induced cell-cycle arrest (OCA) complex mediates G1 arrest under stress conditions through a yet unknown mechanism(Alic et al., 2001). The six

components of this complex are connected, both biophysically and in the functional networks, exhibiting similar genetic interaction and condition-sensitivity profiles (Figure 6D). Although inner-complexome interactions with OCA may well be critical for its function, they do not explain the stress-specificity of the complex. Outer-complexome interactions of OCA proteins do not show overlap with the genetic networks but might be instrumental in understanding the mechanism through which the complex mediates its function. Of particular interest is the interaction between Oca1 and Tos4, newly reported in YeRI (Figure 6D). Tos4 is a transcription factor that binds to the promoters of genes involved in the G1/S transition (Horak et al., 2002), offering a hypothesis for the mechanism by which OCA mediates G1 arrest.

Together our results demonstrate the ability of ABBI-21 to uniformly cover both the inner- and outer-complexome, and highlight the need for integrating biophysical and genetic maps to comprehensively understand cellular functional organization across inner- and outer-complexomes.

DISCUSSION

Four decades ago, McConkey suggested that the term "*quinary structure*" be used to describe a "fifth level of organization", referring to macromolecular interactions that, although highly functionally relevant, might be "transient *in vivo"* (McConkey, 1982). According to McConkey, such interactions "will not be evident from the composition of purified proteins" since quaternary structures tend to be more resistant to the "cataclysmic violence of the most gentle homogenization procedure", while quinary interactions, "although stable *in vivo*, might be largely destroyed by cell fractionation".

The observations presented in this paper suggest fundamental differences of organization between the inner-complexome, having mostly quaternary structures that are highly detectable by biochemical pull-down approaches, and the outer-complexome which has a greater tendency towards quinary structures "largely destroyed by cell fractionation" (but detectable by binary assays in living cells that reconstitute functional

proteins like transcription factors, fluorescent proteins or signal transduction receptors as in the Y2H, GPCA and MAPPIT assays respectively). From this perspective, it is easy to see why there should be a substantial discrepancy among alternative views of inner- versus outer-complexome organization revealed by approaches as different as co-complex association detection and binary interaction assays

Our results establish that the biophysical and functional dichotomy between “structural” and “regulatory” genes, as identified by Jacob and Monod, are reflected in the *interactome*. By organizing the interactome into four different zones and integrating biophysical interactions and associations with systematic functional profile similarity maps, we show that whether a protein encoded by a “structural” gene mediates house-keeping, constitutive functions or pleiotropic, regulatory processes is a property of its interactions. For example, Rpl8A, a component of the large ribosomal unit shows interactions with both functionally similar and functionally heterogeneous proteins in Zones A, B and C. A functionally coherent interaction between Rpl8A and Rpl36A in Zone A is important for constituting the ribosome that performs the constitutive function of translating mRNA into proteins. In contrast, a Zone B interaction between Rpl8A and Erb1, a constituent of 66S pre-ribosomal particles, is important for maturation of the ribosome. Finally, a Zone C interaction between functionally diverse Rpl8A and Sba1, a co-chaperone of the Hsp90 family, likely contributes to turnover and maintenance of ribosomal proteins under diverse stresses.

The inner-complexome thus constitutes the “manufacturing machinery” operating in a relatively constant, robust and persistent manner, and the outer-complexome comprises the “regulatory processes" exhibiting greater flexibility, plasticity, environmental responsiveness and evolvability. The outer-complexome mostly consists of quinary interactions between proteins encoded by structural and regulatory genes, mediating dynamic biological processes that allow for cellular communication and create avenues for rapid response to changing environmental conditions.

Despite the early example of *lacZ* and *lacI* illustrating the principle that being a structural gene or accounting for a significant percentage of mass does not equate to higher cellular importance, literature-curated binary maps and systematic non-binary maps provide highly biased views of the interactome that favor the inner-complexome. In contrast, systematically generated binary maps, which are less biased by biophysical and molecular attributes, demonstrate that the outer-complexome is indeed highly populated with interactions between functionally heterogeneous proteins as expected for the mediation of regulatory processes and cross-talk between pathways.

Over twenty years of interactome mapping has produced a wealth of data that has facilitated our understanding of biological processes at cellular, organismal and systems levels. However, these mapping efforts have typically been focused on single conceptual features of biological complexity such as binary protein-protein interactions, co-complex associations, signaling pathways, and genetic interactions. While this plethora of data and integrative analysis has enriched our understanding of the interactome, there has not been a comprehensive integration of network information that describes how all the disparate activities of the cell collectively communicate and coordinate cellular functions. By generating a new proteome-wide binary PPI network and integrating that data with existing protein interaction maps to constitute an “atlas of binary maps”, we have laid out an alternative model of the interactome according to which the interactome is built around inner-complex and outer complex associations that can be merged to provide a more integrated view of cellular function, linking disparate but essential cellular processes. Moreover, we have observed that only a small proportion of the interactome is composed of stable, functionally coherent, inner-complexome interactions, while the vast majority of the interactome consists of transient, context-dependent interactions in the outer-complexome. This is consistent with emerging evidence suggesting the majority of the interactome may be transient and context-specific (Hein et al., 2015; Liu et al., 2020; Tompa et al., 2014). Our analysis suggests that these outer-complexome interactions allow for cellular communication

and create avenues for rapid response to changing environmental conditions. However, mapping and functional characterization of these outer-complexome interactions remains challenging. Development of assays that can efficiently detect transient, context-specific PPIs and approaches that can characterize their functions will be an important next step towards understanding the global organization of the interactome.

DATA AVAILABILITY

YeRI, ABBI-21 and Lit-BM-20 maps are available at http://yeast.interactome-atlas.org/. Analysis code and data are available at https://github.com/ccsb-dfci/YeRI_paper.

ACKNOWLEDGEMENTS

We thank S. de Rouck for help with the MAPPIT experiments. We thank Gary Bader and acknowledge past and current members of the Center for Cancer Systems Biology (CCSB) for helpful discussions and experimental help. This work was supported primarily by National Institutes of Health (NIH) grant R01HG006061 (M.V, D.E.H, M.A.C, M.E.C, P.F.-B.) with additional support from R01GM130885 (M.V.), R01GM133185 (M.V., M.A.C, F.P.R.) and Institute Sponsored Research funds from the Dana-Farber Cancer Institute Strategic Initiative (M.V). A.D. was supported by a Leon Fredericq grant and a FRS-FNRS-Télévie Fellowship #7651317F (J-C.T). M.V. is a Chercheur Qualifié Honoraire from the Fonds de la Recherche Scientifique (FRS-FNRS, Wallonia-Brussels Federation, Belgium). J-C.T is Maitre de Recherche of the FRS-FNRS. F.P.R was supported by a CIHR Foundation grant. D.-K.K. was supported by a Banting Postdoctoral Fellowship through the Natural Sciences and Engineering Research Council (NSERC) of Canada and by the Basic Science Research Program through the National Research Foundation (NRF) of Korea funded by the Ministry of Education (2017R1A6A3A03004385). M.W.M received support from the Google Summer of Code 2015 through the National Resources For Network Biology (NRNB). C.P. was supported by a Ramon y Cajal fellowship (RYC-2017-22959).

AUTHOR CONTRIBUTIONS

Computational analyses were performed by Y.W., L.L., with help from B.C., D.D.R., T.R., and K.L. Interactome mapping experiments were performed by A.D., T.C., with help from K.S., S.S., N.J., and Q.Z. Y2H vectors were designed and generated by Q.Z. with help from N.J. The preparation of Y2H, GPCA and MAPPIT destination clones by *en masse* gateway cloning and yeast transformations were performed by Q.Z., N.J., A.D., and T.C. Experimental results were processed by Y.W., T.H., and K.L. GPCA validation experiments were done by A.D., and T.C., with help from Y.J. MAPPIT validation experiments were done by I.L., supervised by J.T. Functional enrichment analyses were done by D.-K.K., L.L., and Y.W. Extraction of the literature datasets was performed by L.L., and T.H. YeRI web portal was built by M.W.M., supervised by J.R., and M.H. Structural analyses were done by C.P., L.L., and Y.W., supervised by P.A. Topological analyses were done by L.L. Sequencing analyses were done by T.H., W.B., Y.S., B.C. and Y.W. I.K. performed network-based functional prediction. M.V., F.P.R., M.A.C., D.E.H., P.F.-B., Y.W., A.D., L.L., and A.Y. designed and conceptualized the overall research effort. Interactome mapping was supervised by B.C., M.V., M.A.C., D.E.H., and T.H. A.Y., L.L., Y.W., A.D., M.V. wrote the draft manuscript. A.Y., L.L., Y.W., M.V., D.E.H, T.H., J.-C.T., F.P.R. and M.A.C. reviewed and edited the manuscript with contributions from other coauthors. M.V., F.P.R., M.A.C., D.E.H. supervised and/or advised the overall research effort. M.V. conceived of the project. Major funding acquisition was by M.V., D.E.H., M.A.C., F.P.R. P.F.-B., M.E.C. and J.-C.T.

FIGURE LEGENDS

Figure 1. Current Status of the Yeast Binary Interactome Landscape

(A) Illustrative network diagram showing the space of pairwise protein combinations categorized based on protein complex membership and additionally ordered by protein abundance.

(B) Curation of the three binary PPI datasets.

(C) Experimental validation of yeast systematic binary maps and literature-curated pairs in GPCA and MAPPIT, using positive and random reference sets as benchmarks.

(D) The fraction of pairs from different binary, co-complex association and prediction datasets positive in Y2H v4.

(E) Fraction of PPIs identified by Y2H v4 in complex structures of different sizes.

Error bars in (C), (D), (E) indicate 68.3% Bayesian CI.

(F) Heatmaps of PPI counts, ordering the proteome-by-proteome space of pairwise combinations by the number of publications per protein. I3D-exp-20 (yellow), Lit-BM-20 (blue), and Y2H-union (purple). The upper and right bar charts show the median number of publications per gene for each bin. The color intensity of each square reflects the total number of interactions between proteins for the corresponding bins. The red triangle demarcates the fraction of the search space containing 80% of PPIs in the dataset.

(G) Pie charts showing the proportion of pairs in the inner- and outer-complexome Zones within the three biophysical maps.

Figure 2. A First All-Versus-All Reference Binary Map of the Yeast Interactome

(A) Overview of YeRI generation pipeline and experimental validation in GPCA and MAPPIT.

(B) Enrichment of YeRI PPIs between proteins in the same cellular compartments, pathways and protein complexes.

(C) Heatmaps of high-quality binary PPI counts over time, ordered by the number of publications per protein, from all four systematic binary interactome maps.

(D) Composite networks generated by the addition of each systematic map (top row, left). Network regions enriched for GO terms (bottom row, left). Merge of network and enriched regions for the most recent composite network (right).

(E) Increase in the number of high-quality binary PPIs from available experiments in Interactome-3D (yellow), Lit-BM (blue) and systematic studies (purple).

(F) Coverage of the yeast interactome by ABBI-21, based on three reported estimates of the total interactome size.

(G) The number of proteins encoded by genes of unknown function with at least one interaction in ABBI-21 or Lit-BM.

(H) PPI network of PEX35 and its first- and second-degree interactors. Named proteins are those annotated or predicted to have the GO terms related to ‘peroxisome importer complex’.

Figure 3. High Discrepancy Among Alternative Views of Inner- Versus Outer-Complexome Organization

(A) Heatmaps of protein interactions and associations, with the proteome first ordered by size of protein complexes in which the proteins are involved then ordered by protein abundance.

(B) Mean of publication count, abundance, gene conservation, and predicted disordered fraction, and the fraction of essential genes for proteins either in or outside complexes. Error bars denote 95% CI.

(C) Heatmaps of protein interactions and associations in Lit-BM-20, Sys-NB-06 and ABBI-21, ordered by number of publications per protein, protein abundance, gene conservation, essentiality and predicted fraction of residues in disordered regions. Upper limits of the color scales are set to 5x the average number of PPIs per bin, separately for each network.

(D) Enrichments of ABBI-21 and Sys-NB-06 to contain different gold-standard sets of PPIs from the inner- and outer-complexome. Direct PPIs within protein complexes with three or more subunits; co-complex pairs from KEGG pathways; PPIs regulating activation or inhibition from KEGG; and high-quality kinase-substrate pairs from the KID database.

(E) Fraction of pairs in the inner- and outer-complexome from Lit-BM-13 and ABBI-21 that tested positive using GPCA and MAPPIT.

The color code used in all panels: Lit-BM (blue), Sys-NB-06 (red) and ABBI-21 (purple).

Figure 4. Functional Profile Similarity Networks are Enriched in the Inner-Complexome

(A) Heatmaps of number of connected gene pairs in functional networks, with the proteome first ordered by size of protein complexes in which the proteins are involved then ordered by protein abundance.

(B) Heatmaps of number of connected gene pairs in GI-PSN, CS-PSN and GE-PSN, with the proteome ordered by number of publications per gene, protein abundance, gene conservation, essentiality and the predicted fraction of residues in disordered regions. Upper limits of the color scales are set to 5x the average number of PPIs per bin, separately for each network.

(C) Using experimental structures of Mediator, the direct interactions and indirect associations that are contained within the different biophysical and functional networks.

(D) Recovery of protein pairs in direct contact and pairs not in direct contact within the corresponding 3D structures, by different biophysical and functional networks. Each point is a separate protein complex with at least 5 distinct protein subunits.

(E-F) Results of testing samples of pairs from GI-PSNs, across different PCC cutoffs, in Y2H v4 (E) and GPCA (F).

Figure 5. The Binary Interactome Mainly Consists of Interactions Between Functionally Heterogeneous Proteins in the Outer-Complexome

(A) The fraction of pairs from three biophysical maps connected in functional networks.

(B) The fraction of pairs from the inner-complexome (Zone A) and outer-complexome (Zones B, C and D) also connected in functional networks. The pie charts show the proportion of pairs in the inner- and outer-complexome within the three biophysical maps.

Figure 6. Integrating Biophysical and Genetic Network Maps Aids Understanding Cellular Organization

(A) The fraction of PPIs from four different yeast pathways connected in different functional networks.

(B) The fraction of PPIs from the four gold-standard biophysical interaction datasets (see Figure 4E) connected in functional networks.

(C) Integrated network showing interactions among subunits of ESCRT complexes and their interacting partners.

(D) Integrated network showing interactions among subunits of the OCA complex and their interacting partners. Physically interacting proteins also connected in functional networks are connected with thick orange lines in (C, D).

Figure S1. Current Status of the Yeast Binary Interactome Landscape, Related to Figure 1

(A) Left: structure of 20S proteasome (PDB: 1RYP); middle left: direct protein-protein interactions within the structure; middle right: indirect associations derived from experimental structures; right: all possible pairwise combinations between different subunits.

(B) Venn diagram of PPIs in three high-quality binary datasets.

(C) Venn diagram of proteins with interactions in three high-quality binary PPI datasets.

(D) Experimental validation of literature pairs, with single and multiple pieces of evidence, using GPCA

(E) Benchmarking different binary PPI assays, using positive and random reference sets (scPRS-v2 and scRRS-v2). Colored bars indicate positives.

(F) The fraction of directly contactacting interactions, taking all reported pairs where both proteins are in a protein complex structure with at least three subunits, for different datasets.

(G) Y2H v4 recovery of I3D-exp-17 pairs as a function of the fraction of the protein that is contained in the experimental structure. The lower of the two values of coverage of the different proteins in each interaction is used.

(H) Distribution of interface area for I3D-exp-17 PPIs, that tested either positive or negative in Y2H v4.

(I) Distribution of interface area for I3D-exp-17 PPIs, that tested positive in Y2H v4, GPCA and MAPPIT, restricting to pairs that were successfully tested in all three assays.

(I) Interface areas of PPIs with experimental structure positive in Y2H v4, GPCA and MAPPIT.

(J) Dissociation constants of pairs positive or negative in Y2H v4.

Figure S2. A First All-versus-All Reference Binary Map of the Yeast Interactome, Related to Figure 2

(A) The results of different batches of experimental validation of YeRI in GPCA and MAPPIT. In total, all pairs from YeRI were tested in both assays.

(B) Network-based spatial enrichment analysis (SAFE) for YeRI. Clusters of genes enriched for GO terms in YeRI are highlighted.

(C) Number of PPIs involving proteins encoded by genes of unknown function in Lit-BM-20 or ABBI-21.

Figure S3. High Discrepancy Among Alternative Views of Inner- Versus Outer-Complexome Organization, Related to Figure 3

(A) Comparison of two datasets of yeast protein complexes, restricting to those with three or more subunits.

(B) Mean of publication count, abundance, gene conservation, and predicted disordered fraction, and the fraction of essential genes for proteins either in or outside complexes using the collection of complexes from the EBI complex portal. Error bars denote 95% CI.

(C) The Correlation matrix with clustering of different protein-level properties based on Spearman's rank correlation coefficient.

(D) Heatmaps of protein pair counts in biophysical maps in Zone D, ordered by the number of publications per protein, protein abundance, gene conservation, essentiality and predicted fraction of disordered regions. Upper limits of the color scales are set to 5x the average number of PPIs per bin, separately for each network.

(E) The fraction of genes in Lit-BM, Sys-NB-06 and ABBI-21 across sliding windows of 1,000 proteins, ordered by abundance and conservation. Grey dotted lines show the overall fraction.

(F) The fraction of pairs testing positive in GPCA and MAPPIT in Zones A, B, C and D from Lit-BM and systematic binary maps.

Figure S4. Functional Profile Similarity Networks are Enriched in the Inner-Complexome, Related to Figure 4

(A) Overlap of the three functional profile similarity networks (PSNs), genetic interaction (GI-PSN), gene expression at mRNA level across different conditions (GE-PSN) and growth across different conditions PSN (CS-PSN). Restricted to genes tested in all three PSNs with PCC value in the top 1% of tested pairs in each PSN.

(B) Heatmaps of protein pair counts of functional networks in Zone D, ordered by the number of publications per protein, protein abundance, gene conservation, essentiality and predicted fraction of disordered regions. Upper limits of the color scales are set to 5x the average number of PPIs per bin, separately for each network.

(C) Heatmaps of the number of connected gene pairs in GI-PSN, ordered by protein abundance, and further segmented based on essentiality and involvement in the ribosome.

(D) Estimation of the fraction GI-PSN pairs that are also a binary PPI, based on results of testing samples from GI-PSNs at different PCC cutoffs using Y2H v4 and GPCA and comparing to the results of the PRS.

(E) Degree distribution of Lit-BM, Sys-NB-06, ABBI-21 and three functional networks in log-log scale.

(F) Degree correlation plot, indicating network assortativity of Lit-BM, Sys-NB-06 and ABBI-21, and three functional networks in log scale.

Figure S5. The Binary Interactome Mainly Consists of Interactions Between Functionally Heterogeneous Proteins in the Outer-Complexome, Related to Figure 5

(A) Composition of edges from three biophysical maps in inner-complexome (Zone A) and outer-complexome (Zones B, C and D) and the fraction of pairs from biophysical maps also connected in positive or negative genetic interaction networks (top 1%) within the four zones.
(B) Composition of edges from three biophysical maps involving date and party hubs using different degree and PCC thresholds, and the fraction of pairs from biophysical maps of each category connected in functional networks.
(C) Average co-expression PCC of proteins in biophysical maps with different degrees. A cutoff between party and date hubs of 0.3 is shown by the grey horizontal line.
(D) Composition of edges from three biophysical maps between essential genes and involving non-essential genes.
(E) The fraction of edges that are also connected in functional networks, for three biophysical maps, binned by the higher degree of the two proteins for each pair. Logistic regression and 95% CI are shown.

REFERENCES

Albert, R. (2005). Scale-free networks in cell biology. J. Cell Sci. *118*, 4947–4957.

Alic, N., Higgins, V.J., and Dawes, I.W. (2001). Identification of a Saccharomyces cerevisiae gene that is required for G1 arrest in response to the lipid oxidation product linoleic acid hydroperoxide. Mol. Biol. Cell *12*, 1801–1810.

Baryshnikova, A. (2016). Systematic Functional Annotation and Visualization of Biological Networks. Cell Syst *2*, 412–421.

Batada, N.N., Reguly, T., Breitkreutz, A., Boucher, L., Breitkreutz, B.-J., Hurst, L.D., and Tyers, M. (2007). Still stratus not altocumulus: further evidence against the date/party hub distinction. PLoS Biol. *5*, e154.

Benschop, J.J., Brabers, N., van Leenen, D., Bakker, L.V., van Deutekom, H.W.M., van Berkum, N.L., Apweiler, E., Lijnzaad, P., Holstege, F.C.P., and Kemmeren, P. (2010). A consensus of core protein complex compositions for Saccharomyces cerevisiae. Mol. Cell *38*, 916–928.

Bertin, N., Simonis, N., Dupuy, D., Cusick, M.E., Han, J.-D.J., Fraser, H.B., Roth, F.P., and Vidal, M. (2007). Confirmation of organized modularity in the yeast interactome. PLoS Biol. *5*, e153.

Calderone, A., Castagnoli, L., and Cesareni, G. (2013). mentha: a resource for browsing integrated protein-interaction networks. Nat. Methods *10*, 690–691.

Cassonnet, P., Rolloy, C., Neveu, G., Vidalain, P.-O., Chantier, T., Pellet, J., Jones, L., Muller, M., Demeret, C., Gaud, G., et al. (2011). Benchmarking a luciferase complementation assay for detecting protein complexes. Nat. Methods *8*, 990–992.

Chatr-Aryamontri, A., Oughtred, R., Boucher, L., Rust, J., Chang, C., Kolas, N.K., O'Donnell, L., Oster, S., Theesfeld, C., Sellam, A., et al. (2017). The BioGRID interaction database: 2017

update. Nucleic Acids Res. *45*, D369–D379.

Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., et al. (2012). Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res. *40*, D700–D705.

Choi, K.Y., Satterberg, B., Lyons, D.M., and Elion, E.A. (1994). Ste5 tethers multiple protein kinases in the MAP kinase cascade required for mating in S. cerevisiae. Cell *78*, 499–512.

Choi, S.G., Olivet, J., Cassonnet, P., Vidalain, P.-O., Luck, K., Lambourne, L., Spirohn, K., Lemmens, I., Dos Santos, M., Demeret, C., et al. (2019). Maximizing binary interactome mapping with a minimal number of assays. Nat. Commun. *10*, 3907.

Costanzo, M., VanderSluis, B., Koch, E.N., Baryshnikova, A., Pons, C., Tan, G., Wang, W., Usaj, M., Hanchard, J., Lee, S.D., et al. (2016). A global genetic interaction network maps a wiring diagram of cellular function. Science *353*, aaf1420–aaf1420.

Deshaies, R.J., Seol, J.H., McDonald, W.H., Cope, G., Lyapina, S., Shevchenko, A., Shevchenko, A., Verma, R., and Yates, J.R., 3rd (2002). Charting the protein complexome in yeast by mass spectrometry. Mol. Cell. Proteomics *1*, 3–10.

Elia, N., Sougrat, R., Spurlin, T.A., Hurley, J.H., and Lippincott-Schwartz, J. (2011). Dynamics of endosomal sorting complex required for transport (ESCRT) machinery during cytokinesis and its role in abscission. Proc. Natl. Acad. Sci. U. S. A. *108*, 4846–4851.

Estojak, J., Brent, R., and Golemis, E.A. (1995). Correlation of two-hybrid affinity data with in vitro measurements. Mol. Cell. Biol. *15*, 5820–5829.

Eyckerman, S., Verhee, A., der Heyden, J.V., Lemmens, I., Ostade, X.V., Vandekerckhove, J., and Tavernier, J. (2001). Design and application of a cytokine-receptor-based interaction trap. Nat. Cell Biol. *3*, 1114–1119.

Gall, J.G. (2003). The centennial of the Cajal body. Nat. Rev. Mol. Cell Biol. *4*, 975–980.

Gavin, A.-C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.-M., Cruciat, C.-M., et al. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature *415*, 141–147.

Gavin, A.-C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dümpelfeld, B., et al. (2006). Proteome survey reveals modularity of the yeast cell machinery. Nature *440*, 631–636.

Gilbert, W., and Müller-Hill, B. (1966). Isolation of the lac repressor. Proc. Natl. Acad. Sci. U. S. A. *56*, 1891–1898.

Gustin, M.C., Albertyn, J., Alexander, M., and Davenport, K. (1998). MAP kinase pathways in the yeast *Saccharomyces cerevisiae*. Microbiol. Mol. Biol. Rev. *62*, 1264–1300.

Han, J.-D.J., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., Zhang, L.V., Dupuy, D., Walhout, A.J.M., Cusick, M.E., Roth, F.P., et al. (2004). Evidence for dynamically organized modularity in

the yeast protein-protein interaction network. Nature *430*, 88–93.

Hartwell, L.H., Hopfield, J.J., Leibler, S., and Murray, A.W. (1999). From molecular to modular cell biology. Nature *402*, C47–C52.

Hayes, M.J., Bryon, K., Satkurunathan, J., and Levine, T.P. (2011). Yeast homologues of three BLOC-1 subunits highlight KxDL proteins as conserved interactors of BLOC-1. Traffic *12*, 260–268.

Hein, M.Y., Hubner, N.C., Poser, I., Cox, J., Nagaraj, N., Toyoda, Y., Gak, I.A., Weisswange, I., Mansfeld, J., Buchholz, F., et al. (2015). A human interactome in three quantitative dimensions organized by stoichiometries and abundances. Cell *163*, 712–723.

Hillenmeyer, M.E., Fung, E., Wildenhain, J., Pierce, S.E., Hoon, S., Lee, W., Proctor, M., St Onge, R.P., Tyers, M., Koller, D., et al. (2008). The chemical genomic portrait of yeast: uncovering a phenotype for all genes. Science *320*, 362–365.

Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.-L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., et al. (2002). Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature *415*, 180–183.

Hohmann, S. (2009). Control of high osmolarity signalling in the yeast *Saccharomyces cerevisiae*. FEBS Lett. *583*, 4025–4029.

Horak, C.E., Luscombe, N.M., Qian, J., Bertone, P., Piccirrillo, S., Gerstein, M., and Snyder, M. (2002). Complex transcriptional circuitry at the G1/S transition in Saccharomyces cerevisiae. Genes Dev. *16*, 3017–3033.

Hu, Y., Rolfs, A., Bhullar, B., Murthy, T.V.S., Zhu, C., Berger, M.F., Camargo, A.A., Kelley, F., McCarron, S., Jepson, D., et al. (2007). Approaching a complete repository of sequence-verified protein-encoding clones for Saccharomyces cerevisiae. Genome Res. *17*, 536–543.

Hyman, A.A., Weber, C.A., and Jülicher, F. (2014). Liquid-liquid phase separation in biology. Annu. Rev. Cell Dev. Biol. *30*, 39–58.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc. Natl. Acad. Sci. U. S. A. *98*, 4569–4574.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. *3*, 318–356.

John Peter, A.T., Lachmann, J., Rana, M., Bunge, M., Cabrera, M., and Ungermann, C. (2013). The BLOC-1 complex promotes endosomal maturation by recruiting the Rab5 GTPase-activating protein Msb3. J. Cell Biol. *201*, 97–111.

Juers, D.H., Matthews, B.W., and Huber, R.E. (2012). LacZ β-galactosidase: structure and function of an enzyme of historical and molecular biological importance. Protein Sci. *21*, 1792–1807.

Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K., and Tanabe, M. (2019). New approach for

understanding genome variations in KEGG. Nucleic Acids Res. *47*, D590–D595.

Kim, P.M., Lu, L.J., Xia, Y., and Gerstein, M.B. (2006). Relating three-dimensional structures to protein networks provides evolutionary insights. Science *314*, 1938–1941.

Kim, P.M., Sboner, A., Xia, Y., and Gerstein, M. (2008). The role of disorder in interaction networks: a structural analysis. Mol. Syst. Biol. *4*, 179.

Koh, J.L.Y., Chong, Y.T., Friesen, H., Moses, A., Boone, C., Andrews, B.J., and Moffat, J. (2015). CYCLoPs: A Comprehensive Database Constructed from Automated Analysis of Protein Abundance and Subcellular Localization Patterns in Saccharomyces cerevisiae. G3 *5*, 1223–1232.

Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P., et al. (2006). Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature *440*, 637–643.

Licata, L., Briganti, L., Peluso, D., Perfetto, L., Iannuccelli, M., Galeota, E., Sacco, F., Palma, A., Nardozza, A.P., Santonico, E., et al. (2012). MINT, the molecular interaction database: 2012 update. Nucleic Acids Res. *40*, D857–D861.

Liebermeister, W., Noor, E., Flamholz, A., Davidi, D., Bernhardt, J., and Milo, R. (2014). Visual account of protein investment in cellular functions. Proc. Natl. Acad. Sci. U. S. A. *111*, 8488–8493.

Liu, Z., Miller, D., Li, F., Liu, X., and Levy, S.F. (2020). A large accessory protein interactome is rewired across environments. Elife *9*.

Luck, K., Kim, D.-K., Lambourne, L., Spirohn, K., Begg, B.E., Bian, W., Brignall, R., Cafarelli, T., Campos-Laborie, F.J., Charloteaux, B., et al. (2020). A reference map of the human binary protein interactome. Nature *580*, 402–408.

Mahen, R., and Venkitaraman, A.R. (2012). Pattern formation in centrosome assembly. Curr. Opin. Cell Biol. *24*, 14–23.

McCann, K.L., Charette, J.M., Vincent, N.G., and Baserga, S.J. (2015). A protein interaction map of the LSU processome. Genes Dev. *29*, 862–875.

McConkey, E.H. (1982). Molecular evolution, intracellular organization, and the quinary structure of proteins. Proc. Natl. Acad. Sci. U. S. A. *79*, 3236–3240.

Meldal, B.H.M., Bye-A-Jee, H., Gajdoš, L., Hammerová, Z., Horácková, A., Melicher, F., Perfetto, L., Pokorný, D., Lopez, M.R., Türková, A., et al. (2019). Complex Portal 2018: extended content and enhanced visualization tools for macromolecular complexes. Nucleic Acids Res. *47*, D550–D558.

Meldal, B.H.M., Pons, C., Perfetto, L., Del-Toro, N., Wong, E., Aloy, P., Hermjakob, H., Orchard, S., and Porras, P. Analysing the yeast complexome—the Complex Portal rising to the challenge. Nucleic Acids Res.

Mosca, R., Céol, A., and Aloy, P. (2013). Interactome3D: adding structural details to protein

networks. Nat. Methods *10*, 47–53.

Nagaraj, N., Kulak, N.A., Cox, J., Neuhauser, N., Mayr, K., Hoerning, O., Vorm, O., and Mann, M. (2012). System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. Mol. Cell. Proteomics *11*, M111.013722.

Narayanaswamy, R., Levy, M., Tsechansky, M., Stovall, G.M., O'Connell, J.D., Mirrielees, J., Ellington, A.D., and Marcotte, E.M. (2009). Widespread reorganization of metabolic enzymes into reversible assemblies upon nutrient starvation. Proc. Natl. Acad. Sci. U. S. A. *106*, 10147–10152.

Obayashi, T., Kagaya, Y., Aoki, Y., Tadaka, S., and Kinoshita, K. (2019). COXPRESdb v7: a gene coexpression database for 11 animal species supported by 23 coexpression platforms for technical evaluation and evolutionary inference. Nucleic Acids Res. *47*, D55–D62.

Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., Campbell, N.H., Chavali, G., Chen, C., del-Toro, N., et al. (2014). The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Res. *42*, D358–D363.

Pawson, T. (1995). Protein modules and signalling networks. Nature *373*, 573–580.

Petrovska, I., Nüske, E., Munder, M.C., Kulasegaran, G., Malinovska, L., Kroschwald, S., Richter, D., Fahmy, K., Gibson, K., Verbavatz, J.-M., et al. (2014). Filament formation by metabolic enzymes is a specific adaptation to an advanced state of cellular starvation. Elife *3*, e02409.

Pu, S., Wong, J., Turner, B., Cho, E., and Wodak, S.J. (2009). Up-to-date catalogues of yeast protein complexes. Nucleic Acids Res. *37*, 825–831.

Rolland, T., Taşan, M., Charloteaux, B., Pevzner, S.J., Zhong, Q., Sahni, N., Yi, S., Lemmens, I., Fontanillo, C., Mosca, R., et al. (2014). A proteome-scale map of the human interactome network. Cell *159*, 1212–1226.

Salwinski, L., Miller, C.S., Smith, A.J., Pettit, F.K., Bowie, J.U., and Eisenberg, D. (2004). The Database of Interacting Proteins: 2004 update. Nucleic Acids Res. *32*, D449–D451.

Sambourg, L., and Thierry-Mieg, N. (2010). New insights into protein-protein interaction data lead to increased estimates of the S. cerevisiae interactome size. BMC Bioinformatics *11*, 605.

Sharifpoor, S., Nguyen Ba, A.N., Youn, J.-Y., van Dyk, D., Friesen, H., Douglas, A.C., Kurat, C.F., Chong, Y.T., Founk, K., Moses, A.M., et al. (2011). A quantitative literature-curated gold standard for kinase-substrate pairs. Genome Biol. *12*, R39.

Srere, P.A. (1987). Complexes of sequential metabolic enzymes. Annu. Rev. Biochem. *56*, 89–124.

Stumpf, M.P.H., Thorne, T., de Silva, E., Stewart, R., An, H.J., Lappe, M., and Wiuf, C. (2008). Estimating the size of the human interactome. Proc. Natl. Acad. Sci. U. S. A. *105*, 6959–6964.

Tarassov, K., Messier, V., Landry, C.R., Radinovic, S., Serna Molina, M.M., Shames, I.,

Malitskaya, Y., Vogel, J., Bussey, H., and Michnick, S.W. (2008). An in vivo map of the yeast protein interactome. Science *320*, 1465–1470.

Tompa, P., Davey, N.E., Gibson, T.J., and Babu, M.M. (2014). A million peptide motifs for the molecular biologist. Mol. Cell *55*, 161–169.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., et al. (2000). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature *403*, 623–627.

VanderSluis, B., Costanzo, M., Billmann, M., Ward, H.N., Myers, C.L., Andrews, B.J., and Boone, C. (2018). Integrating genetic and protein-protein interaction networks maps a functional wiring diagram of a cell. Curr. Opin. Microbiol. *45*, 170–179.

Venkatesan, K., Rual, J.-F., Vazquez, A., Stelzl, U., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Zenkner, M., Xin, X., Goh, K.-I., et al. (2009). An empirical framework for binary interactome mapping. Nat. Methods *6*, 83–90.

Vidal, M., Cusick, M.E., and Barabási, A.-L. (2011). Interactome networks and human disease. Cell *144*, 986–998.

Walhout, A.J.M., Temple, G.F., Brasch, M.A., Hartley, J.L., Lorson, M.A., van den Heuvel, S., and Vidal, M. (2000). [34] GATEWAY recombinational cloning: Application to the cloning of large numbers of open reading frames or ORFeomes. In Methods in Enzymology, J. Thorner, S.D. Emr, and J.N. Abelson, eds. (Academic Press), pp. 575–IN7.

Wolfe, K.H., and Shields, D.C. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. Nature *387*, 708–713.

Wood, V., Lock, A., Harris, M.A., Rutherford, K., Bähler, J., and Oliver, S.G. (2019). Hidden in plain sight: what remains to be discovered in the eukaryotic proteome? Open Biol. *9*, 180241.

Yofe, I., Soliman, K., Chuartzman, S.G., Morgan, B., Weill, U., Yifrach, E., Dick, T.P., Cooper, S.J., Ejsing, C.S., Schuldiner, M., et al. (2017). Pex35 is a regulator of peroxisome abundance. J. Cell Sci. *130*, 791–804.

Yu, H., Braun, P., Yildirim, M.A., Lemmens, I., Venkatesan, K., Sahalie, J., Hirozane-Kishikawa, T., Gebreab, F., Li, N., Simonis, N., et al. (2008). High-quality binary protein interaction map of the yeast interactome network. Science *322*, 104–110.

STAR Methods

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Marc Vidal.

Data and code availability

YeRI, ABBI-21 and Lit-BM-20 maps are available at http://yeast.interactome-atlas.org/. Analysis code and data is available at https://github.com/CCSB-DFCI/YeRI_paper.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Yeast strains

Yeast haploid strains *MAT*α Y8930 and *MAT***a** Y8800, derived from PJ69-4 (James et al., 1996), were used previously (Yu et al., 2008). Both strains also harbor the following genotype: *leu2-3,112 trp1-901 his3Δ200 ura3-52 gal4Δ gal80Δ GAL2::ADE2 GAL1::HIS3@LYS2 GAL7::lacZ@MET2 cyh2*^{R}. Yeast cells, parental strains or transformants, were cultured either in synthetic drop out media, supplemented as needed, or in non-selective rich YEPD medium (yeast extract, peptone, dextrose) and incubated at 30°C.

Bacterial strains

Chemically competent DH5α *E. coli* cells were used for all bacterial transformations in this study. Transformed cells were cultured in Luria Broth (LB) or Terrific Broth (TB), supplemented with antibiotics (50 µg/ml of ampicillin, spectinomycin or kanamycin) as needed and incubated at 37°C.

Human cell lines

HEK293T cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) (Life Technologies) supplemented with 10% fetal bovine serum, 2mmol/L L-glutamine, 100 I.U./mL penicillin, and 100 µg/mL streptomycin. Cells were incubated at 37°C with 5% CO2 and 95% humidity.

METHOD DETAILS

#15 Yeast genome space

The Yeast ORF annotation file was downloaded from the *Saccharomyces* Genome Database (SGD) (https://www.yeastgenome.org/) on January 14th, 2017. Four ORFs (YCR097W/HMRA1, YCR096C/HMRA2, YCL066W/HMLALPHA1, YCL067C/HMLALPHA2) annotated in SGD as "silenced gene" were removed. Only SGD-annotated "Verified" and "Uncharacterized" ORFs were included whereas ORFs annotated as "Dubious" were excluded, leaving a total of 5,883 ORFs with 5,155 and 728 ORFs classified as Verified and Uncharacterized, respectively, as of 2017. All datasets analysed have been restricted to these 5,883 ORFs and old ORF names that appear as aliases for one of these ORFs have been mapped to the new name.

#2 Complexome – list of protein complexes

Complexes were taken from Data File S12 of (Costanzo et al., 2016) and filtered for those containing 3 or more different protein subunits, resulting in 339 complexes containing 1,897 different proteins. An alternative yeast protein complex dataset was used from the EBI Complex Portal (Meldal et al., 2019), downloaded on May 5th 2020 and also filtered for 3 or more different protein subunits.

#1 Assignment of pairwise combinations to zones

The search space of all possible pairwise combinations of proteins can be classified into four different "zones" based on their relationship to the complexome (Figure 1A). We define Zone A as the pairwise combinations within the inner-complexome. Such pairs would include for example Rpt4 and Rpt5, two interacting subunits of the proteasome (Finley et al., 2012), and Rps1A and Rps14A of the ribosome (Scaiola et al., 2018). Zone B corresponds to pairs of proteins where each protein belongs to a different complex. For example, the RNA polymerase II (RNA Pol-II) Rpb2 subunit is capable of interacting with the Tfg2 subunit of the transcription factor II complex TFIIH (Plaschka et al., 2016). Zone C represents all pairwise combinations where one protein is in a complex and the other is not. For example, Rpl10, a component of the large ribosomal subunit, interacts with Sqt1, a chaperone important for Rpl10 assembly into the ribosome (Eisinger et al., 1997). Another example would be Rbp2, a component of RNA Pol-II

which interacts with Rad26, a nucleotide excision repair protein recruited to DNA lesions by RNA Pol II (Xu et al., 2017). Finally, Zone D corresponds to protein pairs where neither protein belongs to a complex. Examples of Zone D interactions include most PPIs within signal transduction pathways, individual chaperones and their clients or kinase-substrate pairs involved in cellular processes such as autophagy.

While populated by relatively highly abundant proteins and large molecular size machines, the inner-complexome covers only a tiny proportion of the full yeast interactome "search space", *i.e.* all ~18,000,000 pairwise combinations between all ~6,000 proteins. For example, the ribosome, which accounts for a quarter of the proteomic mass, is encoded by only 2% of all genes, representing ~0.04% of the whole search space. Together the 339 complexes in our complexome map represent 17,607 pairwise combinations between their respective subunits, which corresponds to only ~0.1% of the proteome-by-proteome space. This leaves us with ~99.9% of the whole search space for the outer-complexome, with its three zones, B, C, and D, corresponding to 10%, 44%, and 46% of the proteome-by-proteome space, respectively. Finally, we also organize Zones C and D according to relative abundance.

#3 Assembly and description of biophysical and genetic datasets

Y2H-union: Uetz-screen, Ito-core and CCSB-YI1

As described previously (Yu et al., 2008), Uetz-screen is a subset of Y2H PPIs from Uetz et al ((Ito et al., 2001; Uetz et al., 2000) that was obtained from a proteome-scale systematic screen, excluding a smaller-scale, relatively biased, targeted experiment with a smaller number of well-studied bait proteins, and Ito-core is a subset of PPIs found three times or more in Ito et al (Ito et al., 2001), excluding unreliable pairs of proteins found only once. CCSB-Y1 is a proteome-scale dataset of Y2H PPIs validated using the two orthogonal assays MAPPIT and yPCA (Yu et al., 2008). After restricting to PPIs involving the 5,883 ORFs (described above) the dataset sizes are: Uetz-screen: 645 PPIs; Ito-core: 816 PPIs; CCSB-YI: 1,772 PPIs.

Literature curated biophysical datasets (Lit-NB, Lit-BS, Lit-BM)

Literature curated pairs were obtained from the databases MINT (Licata et al., 2012), IntAct (Orchard et al., 2014), DIP (Salwinski et al., 2004), and BioGRID (Chatr-Aryamontri et al., 2017). The data files used were the 2020-07-14 release from IntAct and BioGRID release 3.5.187 (from 2020-06-25). Data was filtered to have valid IDs for UniProt accession numbers, Pubmed

IDs and PSI-MI terms. Each piece of evidence for a protein pair must consist of a Pubmed ID and an interaction detection method code in the PSI-MI controlled vocabulary (http://www.psidev.info/groups/molecular-interactions). Duplicated evidence arises in cases where different source databases curate the same paper. We merged duplicated entries for each pair, as detected by multiple pieces of evidence with the same Pubmed ID and experimental interaction detection codes which are either the same, or have an ancestor-descendent relationship in the PSI-MI ontology. In the latter case, the more specific descendent term was assigned to the merged evidence. In order to select the subset of PPIs corresponding to binary interactions (as opposed to co-complex associations), we developed a manual classification of the PSI-MI interaction detection method terms, which define the experimental technique used (Rolland et al., 2014). Our classification has been updated since previous versions, in order to cover new experimental methods which have been added to the controlled vocabulary in the intervening time. The methods are classified into three categories; 'invalid', 'binary' and 'indirect'. Where 'invalid' corresponds to terms that are not considered valid experimental protein-protein interaction detection methods, 'binary' to terms that detect binary protein-protein interactions and 'indirect' to terms that detect potentially indirect associations. An example term in the "invalid" category is "colocalization". All PPI annotated with "invalid" terms are removed and not considered.

Defining the literature-curated dataset, we excluded evidence corresponding to the eight systematic, proteome-scale datasets (Gavin et al., 2002, 2006; Ho et al., 2002; Ito et al., 2001; Krogan et al., 2006; Tarassov et al., 2008; Uetz et al., 2000; Yu et al., 2008). 'Binary' or 'indirect' evidence is used to divide the protein pairs in the literature-curated dataset into categories, as follows. If a pair has no evidence that corresponds to a binary method it is classified as Lit-NB (100,940 pairs), one piece of evidence with a binary method, Lit-BS (14,477 pairs), two or more pieces of evidence and at least one with a binary method, Lit-BM (5,589 pairs).

Older literature curated datasets generated in 2017 and 2013 were used as a source dataset for pairs tested in GPCA, MAPPIT and Y2H-v4 experiments. These were generated and processed as above with small differences. Lit-17 obtained the data via the mentha resource (Calderone et al., 2013) data file dated August 28th 2017. Lit-13 was constructed similarly as described previously (Rolland et al., 2014). Yeast PPIs annotated through December 2013 from six source databases: BIND (Bader et al., 2003), BioGRID (Chatr-Aryamontri et al., 2017), DIP (Salwinski

et al., 2004), MINT (Licata et al., 2012), IntAct (Kerrien et al., 2012) and PDB (Berman et al., 2000) were extracted and processed using the same protocol.

Direct PPIs with experimental structures

One of the most definitive proofs of physical, direct, “touching” binary interaction between two proteins is the resolution of the three-dimensional (3D) structure of their interface. We used the subset of Interactome3D (Mosca et al., 2013) restricted to experimental structures, excluding homology models. We refer to this dataset from the January 2020 release of Interactome3D as I3D-exp-20, which is used for most analysis, and the dataset from the June 2017 release, I3D-exp-17, was tested in its entirety in Y2H v4. The date of the PPI was taken from the PDB database taking the earliest release date of all PDB structures for that PPI from the ‘complete’ Interactome3D dataset.

Table S3. Summary of biophysical datasets

					Number of pairs attempted		
Category			**Dataset**	**N**	**Y2H (2u-2u)**	**GPCA**	**MAPPIT**
Experimentally detected	Co-complex	Proteome x proteome	Gavin (2002)	3,210	200		
			Gavin (2006)	6,531	200		
			CYC2008[a]	11,136	200	405	
			Krogan	7,059	200		
			Ho	3,584	200		
		Other	Lit-NB-17	71,260	151		
	Binary	Proteome x proteome	Ito-core[b]	816	738	199	199
			Uetz-screen[c]	645	470	193	193
			CCSB-YI1	1,772	1,536	200	200
			Tarassov	2,761	199	199	199
		With structure	I3D-exp[d]	1,787	1,231[#]		
		Literature	Lit-BS-17	13,981	146	149*	
			Lit-BM-17	4,623	4,128	584*	168*
Predicted			Jansen	9,870	200		
			PrePPI[e]	30,184	200	198	
PRS and RRS			scPRS-v2	108	108	108	108
			scRRS-v2	198	198	198	198
YeRI			YeRI	1,910		1,910	1,910

[a] pairwise combinations of proteins within each complex

[b] core dataset

[c] high-throughput screens

[d] experimentally determined subset

[e] high confidence subset

*LIT-BM-13 was tested

[#]Interactome3D June 17 was tested

Genetic and expression profile similarity networks (PSNs)

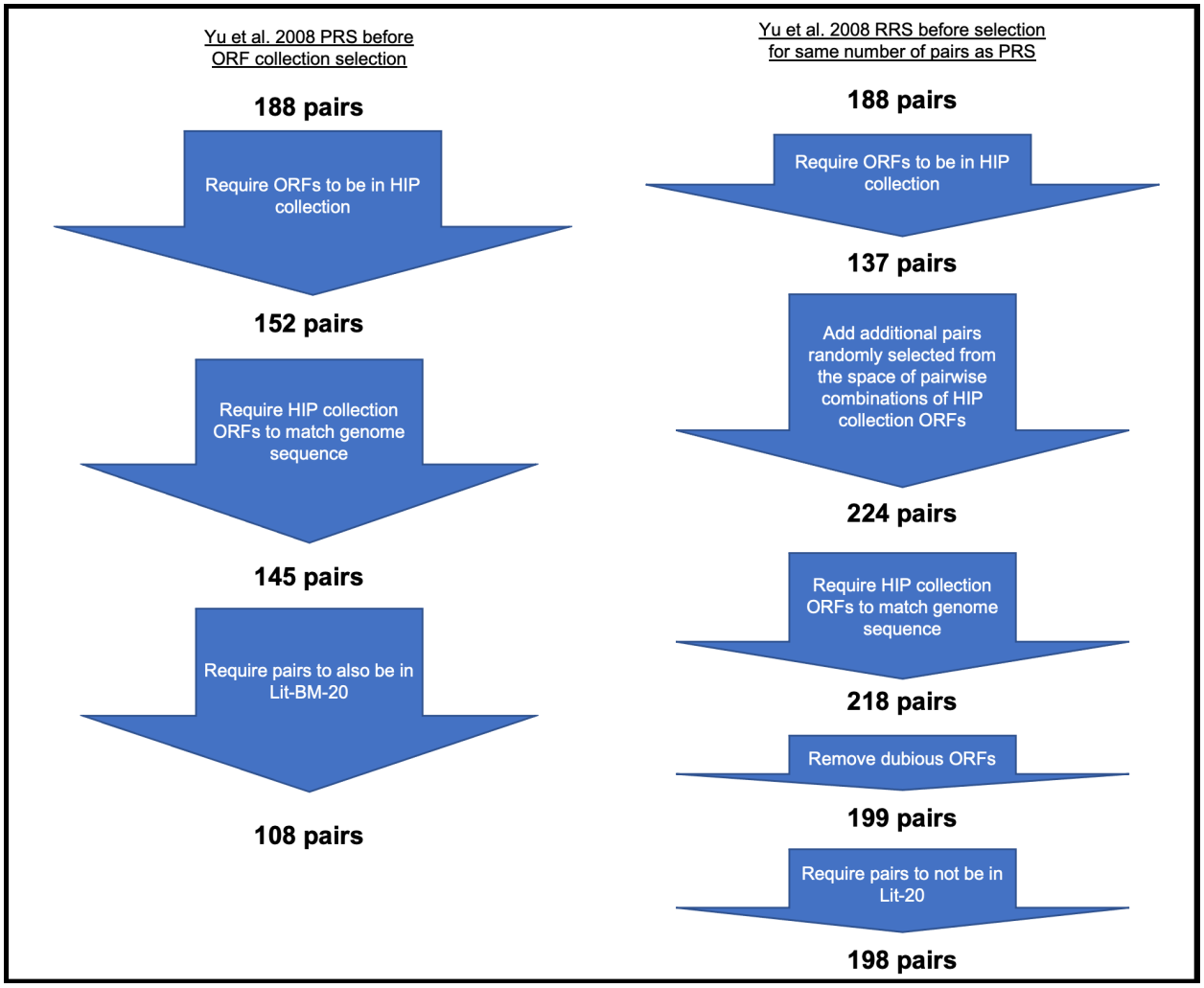
Genetic interaction similarity profile data (GI-PSN) were also extracted from (Costanzo et al., 2016). The average PCC of a pair was used if multiple PCCs were available. Pairs with PCCs ranked in the top 1% were used to generate the GI PSN. Condition-sensitivity data (CS-PSN) was extracted from (Hillenmeyer et al., 2008). Homozygous log ratio results were used to calculate PCC for each pair. Pairs with PCCs ranked in the top 1% were used to generate the condition-sensitivity PSN. Co-expression data (CE-PSN) was downloaded from https://coxpresdb.jp (Obayashi et al., 2019). The union dataset (Sce-m.c3-0 Sce-r.c1-0, 2018.11.07) was used. Pairs with PCCs ranked in the top 1% were used to generate the co-expression PSN.

Table S4. Numbers of genes and interactions at the top 1% percentile of the genetic maps

PSN	PCC threshold (top 1%)	Number of nodes	Number of edges
GI	0.12	5,328	134,972
GE	0.57	3,832	99,454
CS	0.42	3,479	65,147

#7 Generation of scPRS-v2 and scRRS-v2

Due to the change in yeast ORFeome used, we updated our positive reference set (PRS) and random reference set (RRS) from our original set (Yu et al., 2008). We named the updated *Saccharomyces cerevisiae* reference sets scPRS-v2 and scRRS-v2. In Yu et al., 188 PPIs with five or more papers were finalized as PRS candidates of which 116 had both ORFs in the collection at the time. Of those 188 PPIs, 145 had both ORFs in the HIP collection. We further filtered those pairs to also be in Lit-BM-20, resulting in a final scPRS-v2 of 108 PPIs. Of 188 RRS pairs in Yu et al., 137 had both ORFs in the HIP collection. To that we increased the size by adding additional pairs randomly selected from the space of all possible pairwise combinations of ORFs in the HIP collection. Since the RRS is used as a negative control, we then filtered out any pairs that appeared in any of the experimental PPI and co-complex association datasets, which resulted in removing one pair that appeared in Lit-NB-20.



#8 Engineering of new Y2H destination vectors

High-copy 2µ Destination vectors pVV212 and pVV213 (Hallez et al., 2007) were modified to contain *LEU2* or *TRP1* as selectable markers, and *CAN1* or *CYH2* as counterselectable markers, respectively. The pVV212 derivative vector, pDest-DB-QZ212, encodes the Gal4 DNA binding domain (DB) fused to ORFs encoding proteins of interest through the Gateway attL1 site leading to DB-X amino-terminal fusions. The pVV213 derivative vector, Dest-AD-QZ213, encodes Gal4 activation domain (AD) fused to ORFs encoding proteins of interest through the Gateway attL1 site leading to AD-Y amino-terminal fusions. *CYH2* and *CAN1* serve as counterselection markers to facilitate plasmid shuffling in the identification of auto-activators (Walhout and Vidal, 1999). See **Table S1** for detailed information.

Table S1. Yeast destination vectors

Name	pDest-DB	pDest-AD -*CYH2*	pDest-AD -QZ213	pDest-DB -QZ212
Fusion partner (aa)	Gal4-DB (1-147)	Gal4-AD (768-881)	Gal4-AD (768-881)	Gal4-DB (1-147)
Fusion location	N-terminus	N-terminus	N-terminus	N-terminus
Yeast Promoter (nt)	Truncated *ADH1* (-701 to +1)	Truncated *ADH1* (-701 to +1)	Truncated *ADH1* (-410 to +1)	Truncated ADH1 (-410 to +1)
Yeast replication of origin	CEN	CEN	2μ	2μ
Linker sequence 3' of Gal4 element (aa)	SRSNQ	GGSNQ	ICMAYPYDVPD YASLGGHMAM EAPS	PEFPS
Yeast terminator	*ADH1* Term	*ADH1* Term	*ADH1* Term	*ADH1* Term
***E. coli* Selection marker**	Ampicillin	Ampicillin	Ampicillin	Ampicillin

#9 Benchmarking Yeast Two-Hybrid (Y2H) assay versions

Assay versions were benchmarked using scPRS-v2 and scRRS-v2. The new Y2H version with destination clones in vectors pDest-DB-pQZ212 and pDest-AD-QZ213 carrying a yeast 2μ origin was named as Y2H version 4 (Y2H v4). Y2H v1 - v3 can be found in Luck et al, Nature, 2020 (Luck et al., 2020). The positive recovery rate Y2H v4 was compared to one of the CEN-CEN assay versions (Y2H v1), consisting of destination clones in pDest-AD-*CYH2* and pDEST-DB, both centromeric vectors, also used to generate CCSB-YI1 (Yu et al., 2008). Briefly, Y8930:DB-ORF and Y8800:AD-ORF haploid strains were inoculated and mated. After enrichment for diploids in SC-Leu-Trp, diploids were spotted on SC-Leu-Trp-His+3AT solid media, testing for *GAL1::HIS3* activation and on a set of SC-Leu-His+3AT plates supplemented with 10 mg/L cycloheximide (CHX) to identify spontaneous DB-ORF auto-activators (Dreze et

al., 2010). After 3-4 days incubation at 30°C, colonies only growing on SC-Leu-Trp-His+3AT solid media and not on SC-Leu-His+3AT+CHX media were scored as positives.

#14 Generation of an expanded yeast ORFeome collection

The Yeast FLEXGene open reading frame entry clone collection (already cloned in either pDONR201 or pDONR221, both Kan^R) contains 4,933 ORFs (Hu et al., 2007). For the remaining annotated ORFs from SGD (version 2014) (https://www.yeastgenome.org/) not in the FLEXGene ORFeome, entry clones were generated in-house and are referred to as supplemental ORFeome collection. ORF sequences were amplified without native stop codon sequences from either *S. cerevisiae* S288C genomic DNA (ORFs without introns) or cDNA (ORFs containing introns) using KOD high fidelity polymerase (Novagen) and ORF-specific forward and reverse PCR primers tailed with Gateway attB1 and attB2 sequences 5' GGGGACAAGTTTGTACAAAAAAGCAGGCTCCACC and 5' GGGGACCACTTTGTACAAGAAAGCTGGGTCCTA from Hu et al (Hu et al., 2007), respectively, essentially as described (Rual et al., 2004). The CTA sequence in the Gateway tail of the reverse primer provided a synthetic stop codon for all ORFs. Amplified ORFs were transferred to pDONR223 ($Spec^R$) by Gateway BP recombination cloning (Invitrogen) and transformed into chemically competent DH5α *E. coli* cells. Sanger sequencing of PCR products, generated with universal forward and reverse primers, was used to confirm the identity of all cloned ORFs as described (Rual et al., 2004). 921 ORFs were obtained using this approach.

#16 ORFeome cloning in Y2H destination vectors

To generate an arrayed library of DB-ORF and AD-ORF hybrid proteins, the yeast ORFs were transferred into both destination vectors, pDest-DB-pQZ212 and pDest-AD-QZ213, by Gateway LR recombination cloning (Invitrogen). Gateway LR reaction products were transformed into DH5α *E. coli* cells, plasmid DNA was extracted and used to transform yeast strains. pDest-DB-pQZ212 and pDest-AD-QZ213 expression clones were transformed into yeast strains *MAT*α Y8930 and *MAT***a** Y8800, respectively (Dreze et al., 2010).

#17 Auto-activator detection for filtering before Y2H screening

We tested for auto-activation of the *GAL1::HIS3* reporter gene in the absence of any AD-ORF plasmid in both haploid and diploid yeast cells. To identify auto-activator clones in haploid yeast,

Y8930:DB-ORF and Y8800:AD-ORF strains were grown to saturation in SC medium lacking Leucine (Leu) (SC-Leu) or Tryptophan (Trp) (SC-Trp), respectively. After 24 hours of incubation, Y8930:DB-ORF and Y8800:AD-ORF haploids were spotted on SC-Leu-His+3AT or SC-Trp-His+3AT to test for *GAL1::HIS3* activation. Viability of the haploids was confirmed with growth on SC-Leu or SC-Trp, respectively.

To identify auto-activators in diploid yeast, *MAT*α Y8930:DB-ORF and *MAT***a** Y8800:AD-ORF strains were mated against their respective opposite mating type strains carrying the corresponding destination vectors without any fused ORFs. Mating was conducted in rich medium, YEPD, and resulting diploids were enriched following growth in SC-Leu-Trp. Diploids were spotted on SC-Leu-Trp-His+3AT, to test for *GAL1::HIS3* activation, and on SC-Leu-Trp to confirm the viability of the diploids. For both haploids and diploids, after incubation at 30°C for 3-4 days, strains growing in the absence of histidine were considered auto-activators 560 DB-ORFs and 1 AD-ORF were removed from the final screening collection.

The remaining DB-ORF and AD-ORF clones were re-arrayed into four different groups to separate ORFs with similar nucleotide sequences, defined as BLAST scores of 100 and above. Separation of similar ORFs makes the downstream sequence identification of the short NGS reads more accurate, as the reads are aligned to specific groups of ORFs without sequence ambiguity. The final collection used for systematic screening included 4,778 DB-ORF clones and 5,700 AD-ORF clones, covering a total of 5,854 yeast ORFs.

#18 Primary yeast two-hybrid (Y2H) screening

Three replicate Y2H screens were performed. Individual *MAT*α Y8930:DB-ORFs were mated against a pool of ~700 (HIP collection) or ~200 (supplemental collection) *MAT***a** Y8800:AD ORFs. AD-ORF pool size was decreased for the supplemental collection to increase sensitivity. After enrichment in SC-Leu-Trp, 5µl of the culture was spotted on SC-Leu-Trp-His+3AT solid media and on SC-Leu-His+3AT+ 10mg/L CHX to identify spontaneous DB-ORF auto-activators (Dreze et al., 2010). Preliminary investigations using four technical replicates demonstrated that in 97% of the cases, the quadruplicates behaved identically (data not shown). Given the high reproducibility of technical replicates, the SC-Leu-Trp culture was spotted only once per selective media. After incubation at 30°C for 3-4 days, colonies growing on SC-Leu-Trp-His+3AT

but not on SC-Leu-His+3AT+CHX were picked and grown in liquid SC-Leu-Trp. As we used libraries of pools of *MAT***a** Y8800:AD-ORF, it is possible to obtain more than one interaction per mini-library. To account for that, we picked three colonies per growth spot. Cell lysates were prepared from the saturated cultures and used as templates in PCR reactions to amplify and identify the bait and prey sequences.

#20 Yeast colony sequencing

To efficiently and cost-effectively identify both bait and prey proteins for thousands of positive colonies, we used a method called SWIM-seq (Shared-Well Interaction Mapping by sequencing) as described (Luck et al., 2020). Briefly, DB and AD-ORFs were simultaneously amplified from 3µl yeast lysate, using well-specific primers. PCR reactions were performed using Platinum Taq (Life Technologies). After PCR amplification, barcoded PCR reactions from an entire 96 well plate were pooled together and purified (Qiagen, PCR Purification Kit). A final Illumina sequencing library was prepared by adding plate indexes using the i5 and i7 Illumina adapter sequences. Next generation sequencing was performed with Illumina Solexa technology allowing for identification of interacting first pass pairs of proteins (FiPPs). As an exception, interacting pairs from the first screen of the supplemental space were amplified with the universal AD and DB forward and reverse primers and ORF sequences were identified by Sanger sequencing (Genewiz). All SWIM-primers (**Table S2**) were synthesized by Thermo Fisher Scientific, whereas the universal AD, DB and term primers were synthesized by Eurofins Genomics.

Table S2. Primers used (Ns denotes 13-mer well index)

	AD		DB	
	SWIM	**Universal**	**SWIM**	**Universal**
Forward	5'-AGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNCGATGATGAAGATACCCCACCA-3'	5'-CGCGTTTGGAATCACTACAGGG-3'	5'-AGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNGGTCAAAGACAGTTGACTGTATCGT-3'	5'-GGCTTCAGTGGAGACTGATATGCCTC-3'
Reverse	5'-GGAGACTTGACCAAACCTCTGGCG-3'	5'-GGAGACTTGACCAAACCTCTGGCG-3'	5'-GGAGACTTGACCAAACCTCTGGCG-3'	5'-GGAGACTTGACCAAACCTCTGGCG-3'

#19 Pairwise test

To confirm all FiPPs, a pairwise test was performed in the same DB-X/AD-Y orientation they were found in the primary screens. Briefly, glycerol stocks from Y8930:DB-ORF and Y8800:AD-ORF haploid strains were inoculated in SC-Leu or SC-Trp, respectively. Saturated cultures were mated in YEPD. After enrichment for diploids, yeast were spotted on SC-Leu-Trp-His+1 mM 3AT solid media, testing for *GAL1::HIS3* activation. To identify *de novo* auto-activators potentially missed during the first pass screening efforts, Y8930:DB-ORF strains were mated against a Y8800:AD with no ORF fused to the activation domain (Y8800:AD-Empty ORF) and spotted on SC-Leu-Trp-His+1 mM 3AT solid media. Diploids that gave rise to growth on SC-Leu-Trp-His+1 mM 3AT media, but did not grow when the respective Y8930:DB-ORF was mated to Y8800:AD-Empty ORF, were selected as positive interacting pairs of proteins. Positive protein pairs were sequence confirmed as done for the primary screens as described above. As positive and negative controls, the scPRS-v2 and scRRS-v2 pairs were distributed randomly across the respecitvemating plates and tested at the same time.

#5 Validation in orthogonal assays

To assess for the precision of datasets (Venkatesan et al., 2009), PPIs were validated in two orthogonal assays: Mammalian protein-protein interaction trap (MAPPIT) (Eyckerman et al., 2001) and *Gaussia princeps* luciferase protein complementation assay (GPCA) (Cassonnet et al., 2011). As positive and negative controls, we used pairs of scPRS-v2 and scRRS-v2 respectively. For both assays, expression clones were generated by Gateway LR recombination cloning as described above. Expression clones for GPCA were generated by transferring ORFs into pSPICA-N1 and pSPICA-N2 destination vectors (Cassonnet et al., 2011). These two vectors are based on two fragments of the humanized *Gaussia princeps* luciferase (GL1 and GL2). MAPPIT expression clones were generated by LR transfer of ORFs into pMG1-X and pSEL+2L-X destination vectors (Eyckerman et al., 2001). After transformation of all expression clones into DH5α *E. coli* cells, plasmid DNA was extracted and purified using Qiagen 96 Turbo kits (Qiagen) on a BioRobot 8000 (Qiagen). For GPCA, DNA concentration was measured for all clones and were diluted to a final concentration of 25ng/µl. Three different GPCA and two different MAPPIT experiments were performed.

GPCA

On the first day of the assay, ~30,000 to 40,000 HEK293T cells were seeded in each well of a 96 well microtiter plate (Greiner Bio-One). After 24 hours of incubation at 37°C, confluent cells were transfected with 300ng of pSPICA-N1-ORF and pSPICA-N2-ORF vectors using polyethylenimine (PEI). After another 24 hours of incubation at 37°C, cells were washed with PBS supplemented with calcium and magnesium chloride. To each well, 40µl of 5x diluted Renilla lysis buffer (Promega) was added. The plate was then covered with aluminum foil and placed under agitation at 900 rpm for 30 minutes at 37°C for cell lysis. Luciferase activity was measured on a TriStar Berthold Microplate reader by adding 50µl of Renilla luciferase substrate (Renilla Luciferase Assay System, Promega) per well, with a measurement time of 4 seconds.

MAPPIT

HEK293T cells were grown in 384-well plates and co-transfected with a luciferase reporter and plasmids for both bait and prey fusion proteins. Twenty-four hours post-transfection, cells were either stimulated with ligand (erythropoietin) or left untreated, then incubated for an additional 24 hours before luciferase activity was measured in duplicate. The MAPPIT validation experiment

was deemed valid, if both bait and prey were successfully cloned into expression vectors and bait expression was detected. “Fold-induction” values (signal from stimulated cells divided by signal from unstimulated cells) were calculated for each tested pair, and two negative controls (no bait with prey and bait with no prey). Each tested pair was assigned a quantitative score: the fold-induction value of the pair divided by the maximum of the fold-induction value of the two negative controls.

Experimental benchmarking of public PPI datasets

PPIs extracted from the biophysical maps described in **Table S3** have been tested in assays Y2H v4, GPCA and MAPPIT following the same experimental procedures as described above. A summary of the number of tested pairs in each dataset is available in **Table S3**. Samples, if used, were drawn randomly.

#10 Direct or indirect contact in a complex structure

We queried Interactome3D (version 2020_01) (Mosca et al., 2013) for complexes involving three or more proteins with an experimental structure available. For all combinations of protein pairs within a complex, Interactome3D calculated the number of residue-residue contacts by accounting for hydrogen bonds, van der Waals interactions, and salt and disulfide bridges. We defined protein pairs with five or more contacts as direct, and remaining pairs as indirect. Separately for each given PPI dataset, the fraction of direct PPIs was calculated as the number of direct PPIs divided by the number of direct and indirect PPIs.

#12 K_d dataset

Yeast PPIs with measured dissociation constant (K_d) values were obtained from the PDBbind database (Liu et al., 2015) 2017 release and from (Kastritis et al., 2011). In the case where multiple values existed for a pair, the geometric mean was used.

#28 PPIs in KEGG pathways and in the four gold standard inner- and outer-complexome datasets

We collected PPIs from KEGG annotated as activation, inhibition, phosphorylation, dephosphorylation, ubiquitination, glycosylation, methylation, binding/association, complex as

defined by KEGG. Gene expression relations and enzyme-enzyme relations were excluded. The four gold standard inner- and outer-complexome PPI datasets are 1) direct co-complex PPIs using the intersection between protein complex dataset collected by Costanzo *et al.* 2016 filtered with three or more subunits and direct interactions from Interactome3D, 2) co-complex pairs annotated in 5 KEGG yeast pathways Cell Cycle, Meiosis, MAPK Signaling pathway, Autophagy and Mitophagy (KEGG co-complex), 3) PPIs regulating activation or inhibition from the same 5 KEGG yeast pathways (KEGG regulation), 4) high-quality kinase-substrate pairs from the Yeast KID database (http://www.moseslab.csb.utoronto.ca/KID/) (Sharifpoor et al. 2011) with score greater or equal to 6.4 (p-value < 0.01).

#25 List of genes of unknown function

A list of 979 *S. Cerevisiae* genes of unknown function was obtained from Table S9 of (Wood et al., 2019).

#27 Properties used to rank proteins

1. Number of publications per gene was extracted from the gene2pubmed file from NCBI, downloaded on 2018-08-01.
2. Protein abundance information was downloaded from PaxDB (https://pax-db.org) undetected pairs were given an abundance of 0.
3. Gene essentiality information was downloaded from the Saccharomyces Genome Deletion Project (https://www.yeastgenome.org).
4. Conservation score was derived by combining data from HomoloGene (ftp://ftp.ncbi.nih.gov/pub/HomoloGene/build68) and (Carvunis et al., 2012). For a gene with homologs in HomoloGene, its conservation score is the number of distinct non-*S. cerevisiae* species that it shares the same homologene group with plus 9, assuming that it is conserved in the 10 Ascomycota species. For genes without homologs in HomoloGene, we used classification proposed in Carvunis *et al* where genes were scored from 1-10 based on their conservation throughout the Ascomycota phylogeny. Genes without homologs in HomoloGene and that did not appear in the Carvunis data were given a score of 0.

5. Fraction of intrinsic disorder of a protein was calculated as the length of its disordered region as predicted by IUPred2A (https://iupred2a.elte.hu) divided by its total length.
6. Complex size was the number of different protein subunits taken from the complexome dataset. If a protein was a member of multiple complexes, the size of the largest complex was used.

QUANTIFICATION AND STATISTICAL ANALYSIS

#4 Treatment of heterodimers and homodimers (self-interactions)

Unless otherwise noted, only heterodimers were used for analyses and homodimers were excluded. The rationale is because we're trying to explore the functional relationships between interacting genes/proteins and the functional relationship of a gene with itself is conceptually different and often impossible to measure, i.e. all PCC values of profiles would be 1.0 by definition.

#6 Calculation of recovery rates in Y2H v4, MAPPIT and GPCA

In Y2H v4, MAPPIT and GPCA assays, pairs were scored positive or negative based on thresholds set by the highest scoring scRRS-v2 pair in the corresponding experiment. Pairs without valid quantitative scores were dropped, and recovery rates were calculated as the number of positive pairs over the sum of the positive and negative pairs. The error bars on the recovery rates were calculated using a Bayesian model (a binomial likelihood with a uniform prior), taking the central 68.27% interval of *Beta* ($p + 1$, $n + 1$), where p and n are the number of pairs testing positive and negative, respectively. P-values for difference in recovery between two datasets tested in the same experiment are calculated using Fisher's exact test, two-sided in all cases except when testing a dataset against the scPRS-v2 / scRRS-v2 positive or negative controls, where a one-sided test is used.

#11 Calculation of interface areas of PPIs

We retrieved experimental structures, or either complete or domain-based models (using Interactome3D version 2018_04) (Mosca et al., 2013). For each subunit in a complex structure,

we defined its interaction interface as the residues for which the Accessible Surface Area (ASA) changed more than 1 $Å^2$ between the bound and unbound state.

#13 Interaction 2D histogram heat maps

For a particular gene/protein property and a network, we ranked all proteins using that property. Tied values were sorted randomly. The proteins were split into an equal number of bins, creating 2D bins of the protein-by-protein space. Values on the diagonal were multiplied by a factor of $N^2 / (N^2 / 2 - N / 2)$, where N is the number of proteins in the bin, to correct for the smaller number of possible pairwise combinations, since edges were undirected. Self-interactions were not used.

#21 Sequence identification of interacting ORFs

We used an existing computational pipeline to process demultiplexed paired-end reads returned from Illumina sequencing and identify the matching ORF pairs corresponding to Y2H-positive colonies. Paired-end reads are in fastq format, with one read, R1, containing a part of the ORF sequence and the other paired read, R2, containing the well index. We used Bowtie 2[16] (v2.2.3) to align all R1 reads to reference sequences and extracted the well-identifying indices from the R2 reads. AD-ORFs and DB-ORFs that shared the same well indices were paired together and called FiPPS. To identify likely true AD/DB pairs, we developed a “SWIM score” (Luck et al., 2020) S that takes into account the AD and DB reads in each well, total reads returned from the sequencing run, and other factors.

$$S = \frac{2}{\frac{a+M}{x} + \frac{d+N}{y}}$$

where x and y are read counts of an AD-ORF and DB-ORF in a given well respectively, a and d are total read counts of all aligned AD-ORF and DB-ORF in that well, and M and N are pseudo-counts for AD and DB respectively, which were constant for each sequencing batch but varied for different batches. We then selected FiPPs for pairwise testing using a cutoff that balances the risk of testing too many false positives FiPPs versus not testing too many true positive FiPPs. The cutoff varied for different screens and sequencing runs to adjust for slight variations in the screening and sequencing protocol.

#22 *Calculation of enrichment for connecting proteins in the same pathway, subcellular compartment and complex*

Subcellular compartment data was obtained from CYCLoPS (Koh et al., 2015), using the WT data, annotating a protein to a compartment if it has any non-zero value in any of the three repeats. Pathways were obtained from KEGG (Kanehisa et al., 2019). Complexes were obtained from CYC2008 (Pu et al., 2009). The number of PPIs that connected two different proteins in the same compartment, pathway or complex was divided by the mean value for 1,000 degree-preserved randomized networks, generated using the Viger and Latapy algorithm implementation through python iGraph (Viger and Latapy, 2005), and CI values were taken from the innermost 68.27% of the random networks.

#23 *SAFE network visualization*

We used the SAFE network visualization tool (v1.5) (Baryshnikova, 2016). The layouts were generated with Cytoscape (v3.4.0) (Shannon et al., 2003) using the edge-weighted spring embedded layout. GO terms were downloaded from SGD database (version on Jan 17th 2019) and GO terms enriched with $P < 0.05$ were colored and labeled. SAFE analysis was run with the default option except layoutAlgorithm = none (using layout generated by Cytoscape), neighborhoodRadius = 200, and neighborhoodRadiusType = absolute.

#24 *Estimates of the complete yeast interactome size*

We used three estimates, relying on partially overlapping assumptions and data, made by independent groups, that predicted the yeast protein binary interactome contains between ~18,000 and ~38,000 direct binary interactions, corresponding to ~0.1-0.2% of all ~18,000,000 possible protein pair combinations.

- From (Yu et al., 2008) 18,000 (13,500-22,500 95% CI). Taken from Page 107: “we estimated that the yeast binary interactome consists of ~18,000 +- 4500 interactions (SOM VI)” From SOM VI the +/- refers to the 95% CI.
- From (Stumpf et al., 2008) 28,472 (26,650-30,460 95% CI). Taken from the Uetz et al. numbers from Table 1. We use the estimate made using Uetz et al. because three of the other datasets contain indirect protein-protein associations (Ho et al., Gavin et al. and

DIP) and the estimate using Ito et al. uses the full dataset, mainly made up of the 'Ito-noncore' subset that was shown to be of poor quality when retested Y2H and PCA (Yu et al., 2008).

- From (Sambourg and Thierry-Mieg, 2010) 37,600 (32,252-43,472 95% CI). Taken from Page 6: "Taken together, this allows to estimate the size of the binary yeast interactome at ~ 37,600 interactions (95% confidence inter- val: 32252-43472, constructed with the normal approximation method [31])."

One relatively minor difference between the estimates is that Stumpf et al. are considering only heterodimeric PPIs whereas Yu et al. and Sambourg et al. are also counting homodimeric PPIs and so we account for this when estimating the fraction of predicted interactome mapped by excluding homodimers for the Stumpf et al. estimate and including them for the Yu et al. and Sambourg et al. estimates.

#26 *Prediction of gene functions using guilt-by-association approach*

In the guilt-by-association approach the function of a node is inferred from the function of its neighbors. In particular, for each node we count the number of its neighbors annotated with a given function (n). This score is then compared to a random benchmark, obtained by randomizing the network 10,000 times in a degree-preserved way. Calculating the *z*-score, $z = (n - \bar{n}) / \sigma$, is the traditional way of such comparison, obtained by standardizing the original score with the expectation value ($\bar{n}$) and standard deviation (σ) of the score that would be expected by chance. Yet, the *z*-score is not free from degree biases and prefers low degree nodes with extremely small σ. We therefore apply a related measure, called the effect size. The effect size $n - (\bar{n} + \alpha\sigma)$ is obtained by comparing the original score with the reasonably expected value of the random benchmark, estimated as the mean value ($\bar{n}$) and α-times the standard deviation (σ). In practice, we use α = 2, selecting the same candidates as a traditional z-score threshold of $z \geq 2$, but ordering them based on the amount of signal beyond random expectations to avoid a bias towards low-degree nodes. Functional annotations of genes with GO Biological Process terms were obtained as described above and further restricted to annotations with the experimental evidence codes EXP, IDA,IPI, IMP, IGI, IEP, HTP, HDA, HMP, HGI, and HEP.

#29 Protein complex subnetworks

For each protein complex, direct interactions were defined by I3D-exp-20, described above, indirect association were all protein-protein combinations where both proteins appeared in the same experimental structure but not in direct contact.

#30 Degree distribution and degree-degree correlation plots

Degree distributions were plotted according to Chapter 4, Advanced Topic 3.B of Network Science (Barabási, 2016), on a log-log scale with logarithmic binning, with the unbinned data shown in grey. Degree-degree correlation plots were made similarly, plotting the degree of a node against the average degree of its neighbors. The degree assortativity coefficient r is shown (Newman, 2003).

#31 Overlap calculation between biophysical and functional networks

For each PPI network and several KEGG pathways, we measured the overlap in number of interactions with each of the genetic networks defined above, discarding homodimeric PPIs. For each pairwise comparison, the PPI network was trimmed to interactions where both proteins were present in the search space of the genetic network. The error bars were calculated using a Bayesian model (a binomial likelihood with a uniform prior), taking the central 68.27% interval of $Beta(p + 1, n + 1)$, where p and n are the number of pairs testing positive and negative, respectively.

#32 Date and party hubs

Coexpression data was obtained from COXPRESdb (Obayashi et al., 2019). To ensure robustness against the exact definition of date and party hubs, three different cutoffs were used, hubs were defined as either in the top 5% or 10% degree proteins in each network or a fixed cutoff of degree ≥ 10 was used. PCC cutoffs of 0.3 and 0.35 were used, where proteins with a mean coexpression PCC across all partners above the cutoff were party hubs and below the cutoff were date hubs.

#33 Overlap by degree plots

We perform a logistic regression of whether a biophysical interaction is also connected in the corresponding functional map against the higher of the two degrees of the interacting proteins in the biophysical network. The max degree per PPI variable is log2 transformed. Shaded error bands represent 95% CI. Binned data is also shown, with 10 evenly sized bins, with the binned data displayed on the x-axis at the mean max degree value of the bin.

References

Bader, G.D., Betel, D., and Hogue, C.W.V. (2003). BIND: the Biomolecular Interaction Network Database. Nucleic Acids Res. *31*, 248–250.

Barabási, A.-L. (2016). Network Science (Cambridge University Press).

Baryshnikova, A. (2016). Systematic Functional Annotation and Visualization of Biological Networks. Cell Syst *2*, 412–421.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res. *28*, 235–242.

Calderone, A., Castagnoli, L., and Cesareni, G. (2013). mentha: a resource for browsing integrated protein-interaction networks. Nat. Methods *10*, 690–691.

Carvunis, A.-R., Rolland, T., Wapinski, I., Calderwood, M.A., Yildirim, M.A., Simonis, N., Charloteaux, B., Hidalgo, C.A., Barbette, J., Santhanam, B., et al. (2012). Proto-genes and de novo gene birth. Nature *487*, 370–374.

Cassonnet, P., Rolloy, C., Neveu, G., Vidalain, P.-O., Chantier, T., Pellet, J., Jones, L., Muller, M., Demeret, C., Gaud, G., et al. (2011). Benchmarking a luciferase complementation assay for detecting protein complexes. Nat. Methods *8*, 990–992.

Chatr-Aryamontri, A., Oughtred, R., Boucher, L., Rust, J., Chang, C., Kolas, N.K., O'Donnell, L., Oster, S., Theesfeld, C., Sellam, A., et al. (2017). The BioGRID interaction database: 2017 update. Nucleic Acids Res. *45*, D369–D379.

Costanzo, M., VanderSluis, B., Koch, E.N., Baryshnikova, A., Pons, C., Tan, G., Wang, W., Usaj, M., Hanchard, J., Lee, S.D., et al. (2016). A global genetic interaction network maps a wiring diagram of cellular function. Science *353*, aaf1420–aaf1420.

Dreze, M., Monachello, D., Lurin, C., Cusick, M.E., Hill, D.E., Vidal, M., and Braun, P. (2010). Chapter 12 - High-Quality Binary Interactome Mapping. In Methods in Enzymology, (Academic Press), pp. 281–315.

Eisinger, D.P., Dick, F.A., Denke, E., and Trumpower, B.L. (1997). SQT1, which encodes an essential WD domain protein of Saccharomyces cerevisiae, suppresses dominant-negative

mutations of the ribosomal protein gene QSR1. Mol. Cell. Biol. *17*, 5146–5155.

Eyckerman, S., Verhee, A., der Heyden, J.V., Lemmens, I., Ostade, X.V., Vandekerckhove, J., and Tavernier, J. (2001). Design and application of a cytokine-receptor-based interaction trap. Nat. Cell Biol. *3*, 1114–1119.

Finley, D., Ulrich, H.D., Sommer, T., and Kaiser, P. (2012). The ubiquitin-proteasome system of Saccharomyces cerevisiae. Genetics *192*, 319–360.

Gavin, A.-C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.-M., Cruciat, C.-M., et al. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature *415*, 141–147.

Gavin, A.-C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dümpelfeld, B., et al. (2006). Proteome survey reveals modularity of the yeast cell machinery. Nature *440*, 631–636.

Hallez, R., Mignolet, J., Van Mullem, V., Wery, M., Vandenhaute, J., Letesson, J.-J., Jacobs-Wagner, C., and De Bolle, X. (2007). The asymmetric distribution of the essential histidine kinase PdhS indicates a differentiation event in Brucella abortus. EMBO J. *26*, 1444–1455.

Hillenmeyer, M.E., Fung, E., Wildenhain, J., Pierce, S.E., Hoon, S., Lee, W., Proctor, M., St Onge, R.P., Tyers, M., Koller, D., et al. (2008). The chemical genomic portrait of yeast: uncovering a phenotype for all genes. Science *320*, 362–365.

Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.-L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., et al. (2002). Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature *415*, 180–183.

Hu, Y., Rolfs, A., Bhullar, B., Murthy, T.V.S., Zhu, C., Berger, M.F., Camargo, A.A., Kelley, F., McCarron, S., Jepson, D., et al. (2007). Approaching a complete repository of sequence-verified protein-encoding clones for Saccharomyces cerevisiae. Genome Res. *17*, 536–543.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc. Natl. Acad. Sci. U. S. A. *98*, 4569–4574.

James, P., Halladay, J., and Craig, E.A. (1996). Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. Genetics *144*, 1425–1436.

Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K., and Tanabe, M. (2019). New approach for understanding genome variations in KEGG. Nucleic Acids Res. *47*, D590–D595.

Kastritis, P.L., Moal, I.H., Hwang, H., Weng, Z., Bates, P.A., Bonvin, A.M.J.J., and Janin, J. (2011). A structure-based benchmark for protein-protein binding affinity. Protein Sci. *20*, 482–491.

Kerrien, S., Aranda, B., Breuza, L., Bridge, A., Broackes-Carter, F., Chen, C., Duesbury, M., Dumousseau, M., Feuermann, M., Hinz, U., et al. (2012). The IntAct molecular interaction

database in 2012. Nucleic Acids Res. *40*, D841–D846.

Koh, J.L.Y., Chong, Y.T., Friesen, H., Moses, A., Boone, C., Andrews, B.J., and Moffat, J. (2015). CYCLoPs: A Comprehensive Database Constructed from Automated Analysis of Protein Abundance and Subcellular Localization Patterns in Saccharomyces cerevisiae. G3 *5*, 1223–1232.

Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P., et al. (2006). Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature *440*, 637–643.

Licata, L., Briganti, L., Peluso, D., Perfetto, L., Iannuccelli, M., Galeota, E., Sacco, F., Palma, A., Nardozza, A.P., Santonico, E., et al. (2012). MINT, the molecular interaction database: 2012 update. Nucleic Acids Res. *40*, D857–D861.

Liu, Z., Li, Y., Han, L., Li, J., Liu, J., Zhao, Z., Nie, W., Liu, Y., and Wang, R. (2015). PDB-wide collection of binding data: current status of the PDBbind database. Bioinformatics *31*, 405–412.

Luck, K., Kim, D.-K., Lambourne, L., Spirohn, K., Begg, B.E., Bian, W., Brignall, R., Cafarelli, T., Campos-Laborie, F.J., Charloteaux, B., et al. (2020). A reference map of the human binary protein interactome. Nature *580*, 402–408.

Meldal, B.H.M., Bye-A-Jee, H., Gajdoš, L., Hammerová, Z., Horácková, A., Melicher, F., Perfetto, L., Pokorný, D., Lopez, M.R., Türková, A., et al. (2019). Complex Portal 2018: extended content and enhanced visualization tools for macromolecular complexes. Nucleic Acids Res. *47*, D550–D558.

Mosca, R., Céol, A., and Aloy, P. (2013). Interactome3D: adding structural details to protein networks. Nat. Methods *10*, 47–53.

Newman, M.E.J. (2003). Mixing patterns in networks. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. *67*, 026126.

Obayashi, T., Kagaya, Y., Aoki, Y., Tadaka, S., and Kinoshita, K. (2019). COXPRESdb v7: a gene coexpression database for 11 animal species supported by 23 coexpression platforms for technical evaluation and evolutionary inference. Nucleic Acids Res. *47*, D55–D62.

Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., Campbell, N.H., Chavali, G., Chen, C., del-Toro, N., et al. (2014). The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Res. *42*, D358–D363.

Plaschka, C., Hantsche, M., Dienemann, C., Burzinski, C., Plitzko, J., and Cramer, P. (2016). Transcription initiation complex structures elucidate DNA opening. Nature *533*, 353–358.

Pu, S., Wong, J., Turner, B., Cho, E., and Wodak, S.J. (2009). Up-to-date catalogues of yeast protein complexes. Nucleic Acids Res. *37*, 825–831.

Rolland, T., Taşan, M., Charloteaux, B., Pevzner, S.J., Zhong, Q., Sahni, N., Yi, S., Lemmens, I., Fontanillo, C., Mosca, R., et al. (2014). A proteome-scale map of the human interactome network. Cell *159*, 1212–1226.

Rual, J.-F., Hirozane-Kishikawa, T., Hao, T., Bertin, N., Li, S., Dricot, A., Li, N., Rosenberg, J., Lamesch, P., Vidalain, P.-O., et al. (2004). Human ORFeome version 1.1: a platform for reverse proteomics. Genome Res. *14*, 2128–2135.

Salwinski, L., Miller, C.S., Smith, A.J., Pettit, F.K., Bowie, J.U., and Eisenberg, D. (2004). The Database of Interacting Proteins: 2004 update. Nucleic Acids Res. *32*, D449–D451.

Sambourg, L., and Thierry-Mieg, N. (2010). New insights into protein-protein interaction data lead to increased estimates of the S. cerevisiae interactome size. BMC Bioinformatics *11*, 605.

Scaiola, A., Peña, C., Weisser, M., Böhringer, D., Leibundgut, M., Klingauf-Nerurkar, P., Gerhardy, S., Panse, V.G., and Ban, N. (2018). Structure of a eukaryotic cytoplasmic pre-40S ribosomal subunit. EMBO J. *37*.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Stumpf, M.P.H., Thorne, T., de Silva, E., Stewart, R., An, H.J., Lappe, M., and Wiuf, C. (2008). Estimating the size of the human interactome. Proc. Natl. Acad. Sci. U. S. A. *105*, 6959–6964.

Tarassov, K., Messier, V., Landry, C.R., Radinovic, S., Serna Molina, M.M., Shames, I., Malitskaya, Y., Vogel, J., Bussey, H., and Michnick, S.W. (2008). An in vivo map of the yeast protein interactome. Science *320*, 1465–1470.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., et al. (2000). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature *403*, 623–627.

Venkatesan, K., Rual, J.-F., Vazquez, A., Stelzl, U., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Zenkner, M., Xin, X., Goh, K.-I., et al. (2009). An empirical framework for binary interactome mapping. Nat. Methods *6*, 83–90.

Viger, F., and Latapy, M. (2005). Efficient and Simple Generation of Random Simple Connected Graphs with Prescribed Degree Sequence. In Computing and Combinatorics, (Springer Berlin Heidelberg), pp. 440–449.

Walhout, A.J., and Vidal, M. (1999). A genetic strategy to eliminate self-activator baits prior to high-throughput yeast two-hybrid screens. Genome Res. *9*, 1128–1134.

Wood, V., Lock, A., Harris, M.A., Rutherford, K., Bähler, J., and Oliver, S.G. (2019). Hidden in plain sight: what remains to be discovered in the eukaryotic proteome? Open Biol. *9*, 180241.

Xu, J., Lahiri, I., Wang, W., Wier, A., Cianfrocco, M.A., Chong, J., Hare, A.A., Dervan, P.B., DiMaio, F., Leschziner, A.E., et al. (2017). Structural basis for the initiation of eukaryotic transcription-coupled DNA repair. Nature *551*, 653–657.

Yu, H., Braun, P., Yildirim, M.A., Lemmens, I., Venkatesan, K., Sahalie, J., Hirozane-Kishikawa, T., Gebreab, F., Li, N., Simonis, N., et al. (2008). High-quality binary protein interaction map of the yeast interactome network. Science *322*, 104–110.

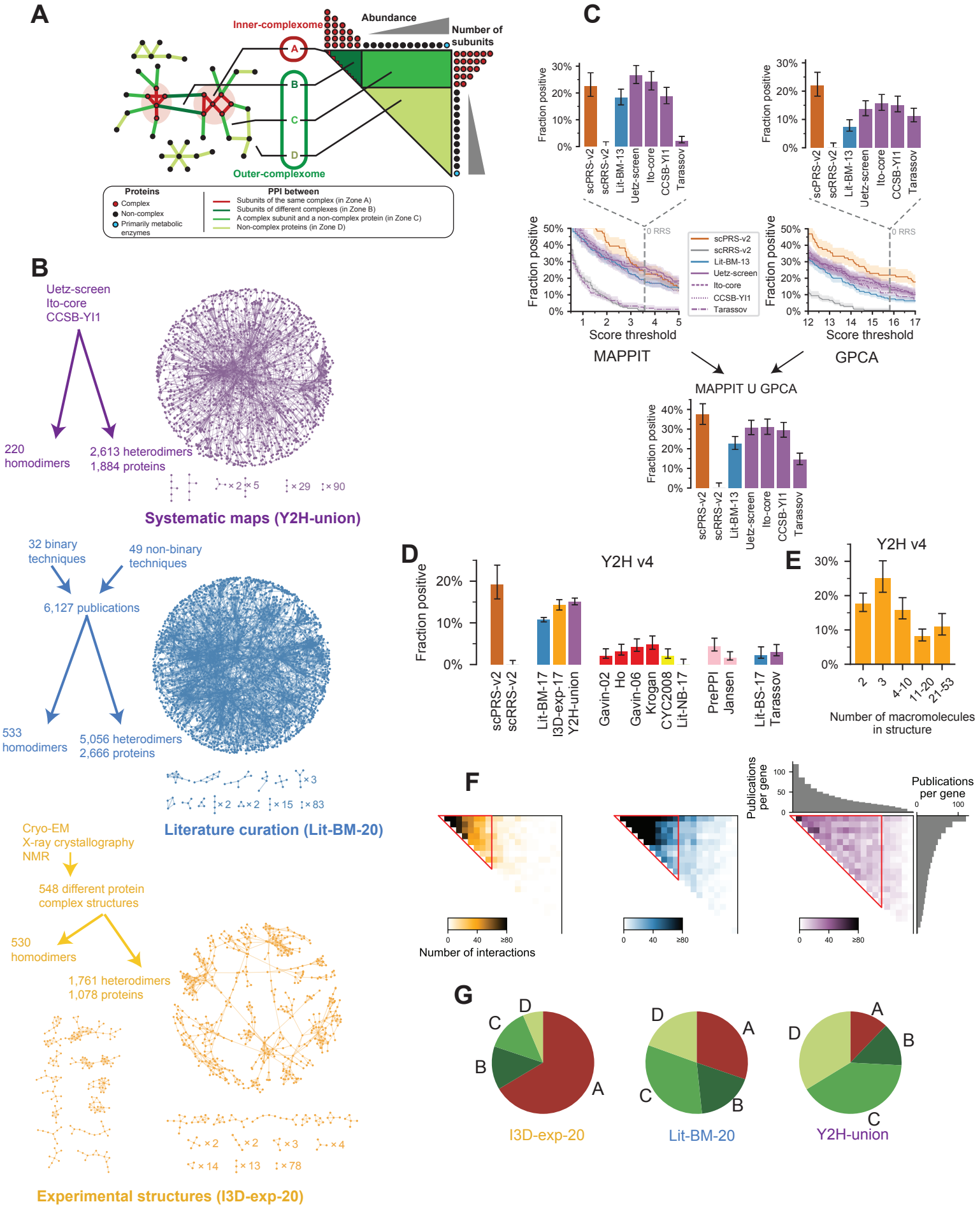

A
Inner-complexome
Abundance
Number of subunits
Outer-complexome
Proteins
Complex
Non-complex
Primarily metabolic enzymes
PPI between
Subunits of the same complex (in Zone A)
Subunits of different complexes (in Zone B)
A complex subunit and a non-complex protein (in Zone C)
Non-complex proteins (in Zone D)
B
Uetz-screen
Ito-core
CCSB-YI1
220 homodimers
2,613 heterodimers
1,884 proteins
Systematic maps (Y2H-union)
32 binary techniques
49 non-binary techniques
6,127 publications
533 homodimers
5,056 heterodimers
2,666 proteins
Literature curation (Lit-BM-20)
Cryo-EM
X-ray crystallography
NMR
548 different protein complex structures
530 homodimers
1,761 heterodimers
1,078 proteins
Experimental structures (I3D-exp-20)
C
Fraction positive
Score threshold
0 RRS
scPRS-v2
scRRS-v2
Lit-BM-13
Uetz-screen
Ito-core
CCSB-YI1
Tarassov
MAPPIT
GPCA
MAPPIT U GPCA
D
Y2H v4
Fraction positive
scPRS-v2
scRRS-v2
Lit-BM-17
I3D-exp-17
Y2H-union
Gavin-02
Ho
Gavin-06
Krogan
CYC2008
Lit-NB-17
PrePPI
Jansen
Lit-BS-17
Tarassov
E
Y2H v4
Number of macromolecules in structure
F
Publications per gene
Number of interactions
G
I3D-exp-20
Lit-BM-20
Y2H-union

Figure 1

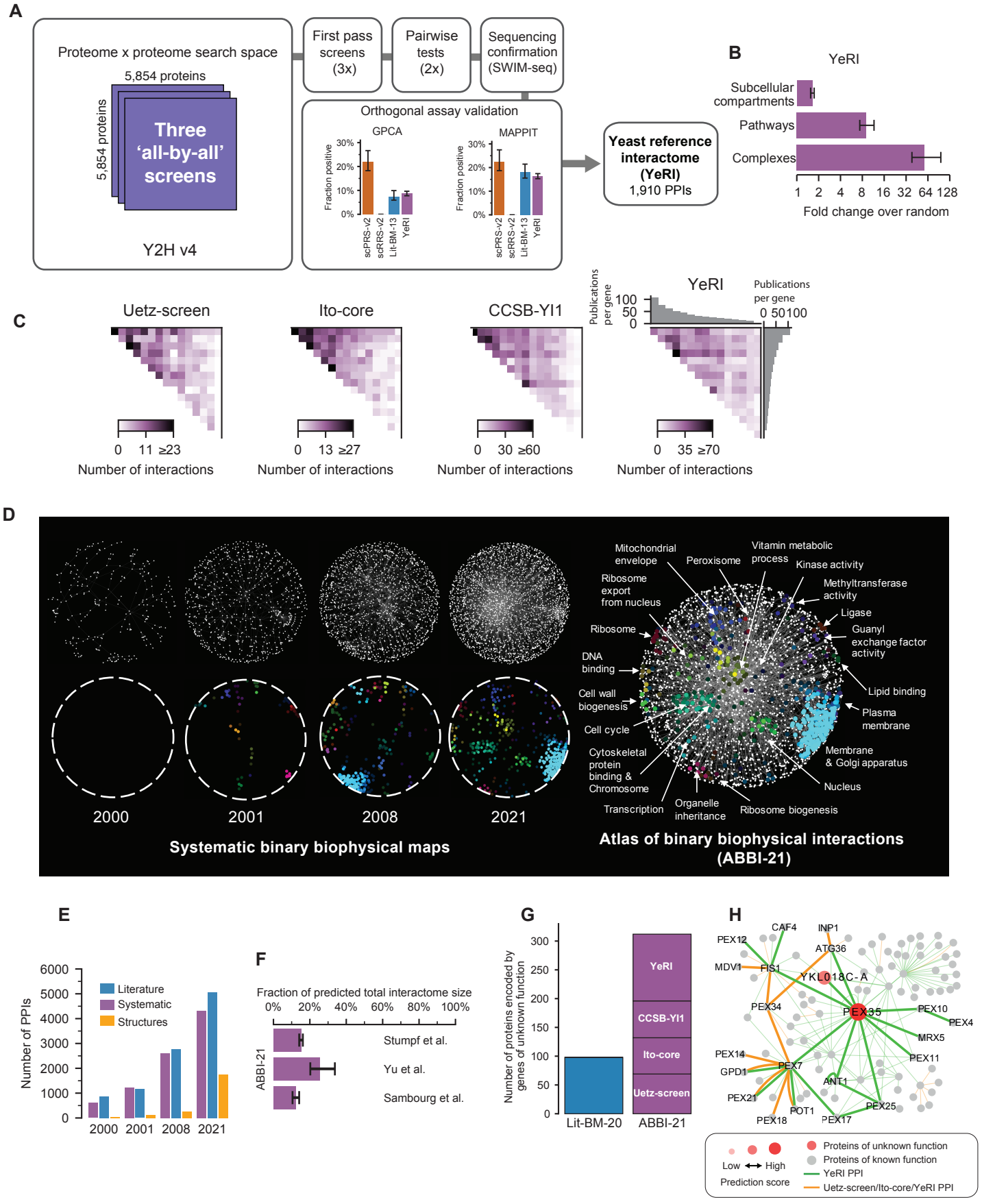

A
Proteome x proteome search space
5,854 proteins
5,854 proteins
Three 'all-by-all' screens
Y2H v4
First pass screens (3x)
Pairwise tests (2x)
Sequencing confirmation (SWIM-seq)
Orthogonal assay validation
GPCA
MAPPIT
Fraction positive
30%
20%
10%
0%
scPRS-v2
scRRS-v2
Lit-BM-13
YeRI
Yeast reference interactome (YeRI)
1,910 PPIs
B
YeRI
Subcellular compartments
Pathways
Complexes
1 2 4 8 16 32 64 128
Fold change over random
C
Uetz-screen
Ito-core
CCSB-YI1
YeRI
Publications per gene
100
50
0
0 50100
0 11 ≥23
Number of interactions
0 13 ≥27
Number of interactions
0 30 ≥60
Number of interactions
0 35 ≥70
Number of interactions
D
2000
2001
2008
2021
Systematic binary biophysical maps
Mitochondrial envelope
Peroxisome
Vitamin metabolic process
Kinase activity
Ribosome export from nucleus
Methyltransferase activity
Ligase
Ribosome
Guanyl exchange factor activity
DNA binding
Lipid binding
Cell wall biogenesis
Plasma membrane
Cell cycle
Membrane & Golgi apparatus
Cytoskeletal protein binding & Chromosome
Nucleus
Transcription
Organelle inheritance
Ribosome biogenesis
Atlas of binary biophysical interactions (ABBI-21)
E
Number of PPIs
6000
5000
4000
3000
2000
1000
0
Literature
Systematic
Structures
2000 2001 2008 2021
F
Fraction of predicted total interactome size
0% 20% 40% 60% 80% 100%
ABBI-21
Stumpf et al.
Yu et al.
Sambourg et al.
G
Number of proteins encoded by genes of unknown function
300
250
200
150
100
50
0
YeRI
CCSB-YI1
Ito-core
Uetz-screen
Lit-BM-20
ABBI-21
H
CAF4
INP1
PEX12
ATG36
MDV1
FIS1
YKL018C-A
PEX34
PEX35
PEX10
PEX4
MRX5
PEX14
PEX11
GPD1
PEX7
ANT1
PEX21
POT1
PEX25
PEX18
PEX17
Low High
Prediction score
Proteins of unknown function
Proteins of known function
YeRI PPI
Uetz-screen/Ito-core/YeRI PPI

Figure 2

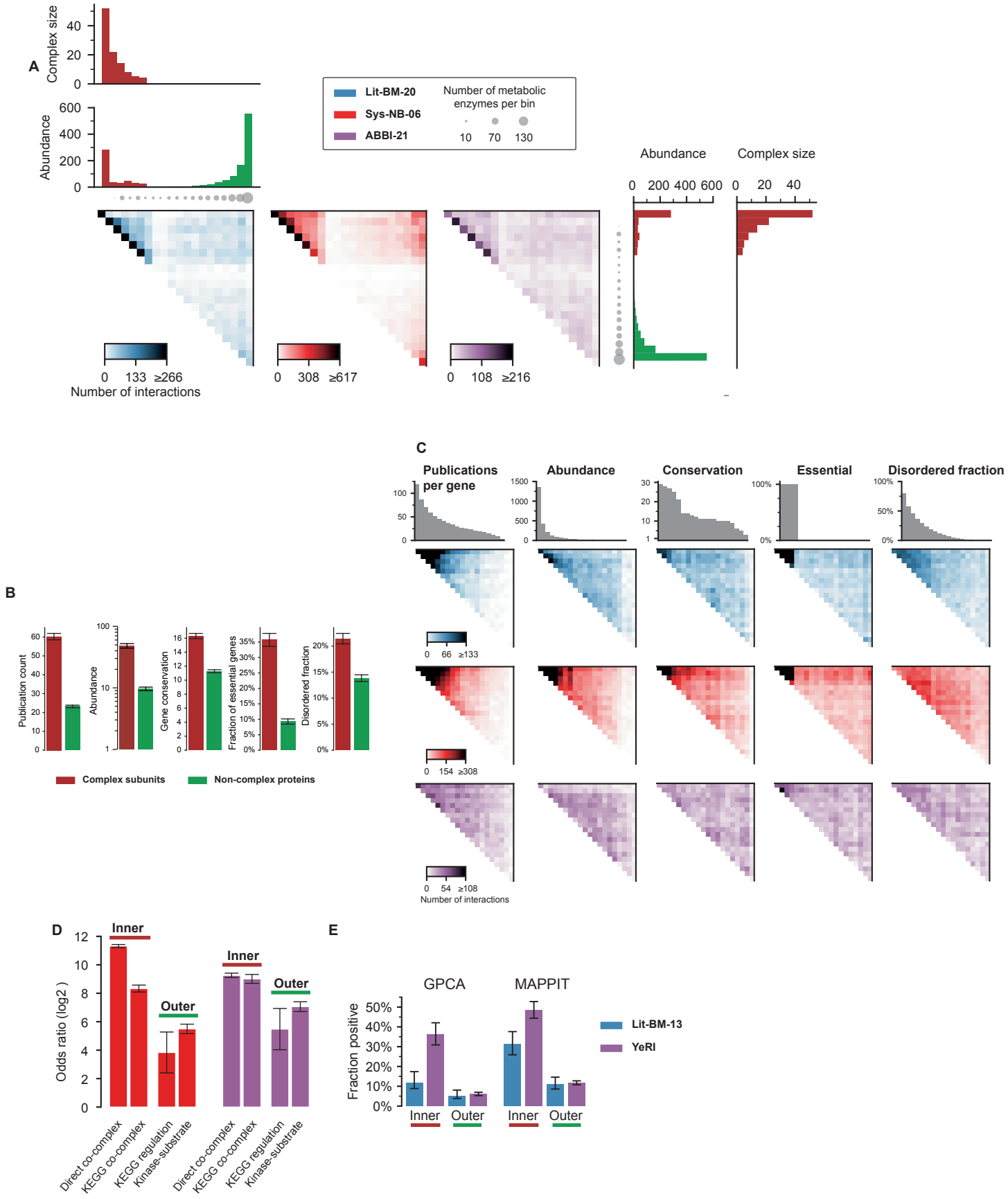

A
Complex size
Abundance
Lit-BM-20
Sys-NB-06
ABBI-21
Number of metabolic enzymes per bin
10 70 130
Abundance
Complex size
0 133 ≥266
Number of interactions
0 308 ≥617
0 108 ≥216
B
Publication count
Abundance
Gene conservation
Fraction of essential genes
Disordered fraction
Complex subunits
Non-complex proteins
C
Publications per gene
Abundance
Conservation
Essential
Disordered fraction
0 66 ≥133
0 154 ≥308
0 54 ≥108
Number of interactions
D
Odds ratio (log2)
Inner
Outer
Direct co-complex
KEGG co-complex
KEGG regulation
Kinase-substrate
E
GPCA
MAPPIT
Fraction positive
Inner
Outer
Lit-BM-13
YeRI

Figure 3

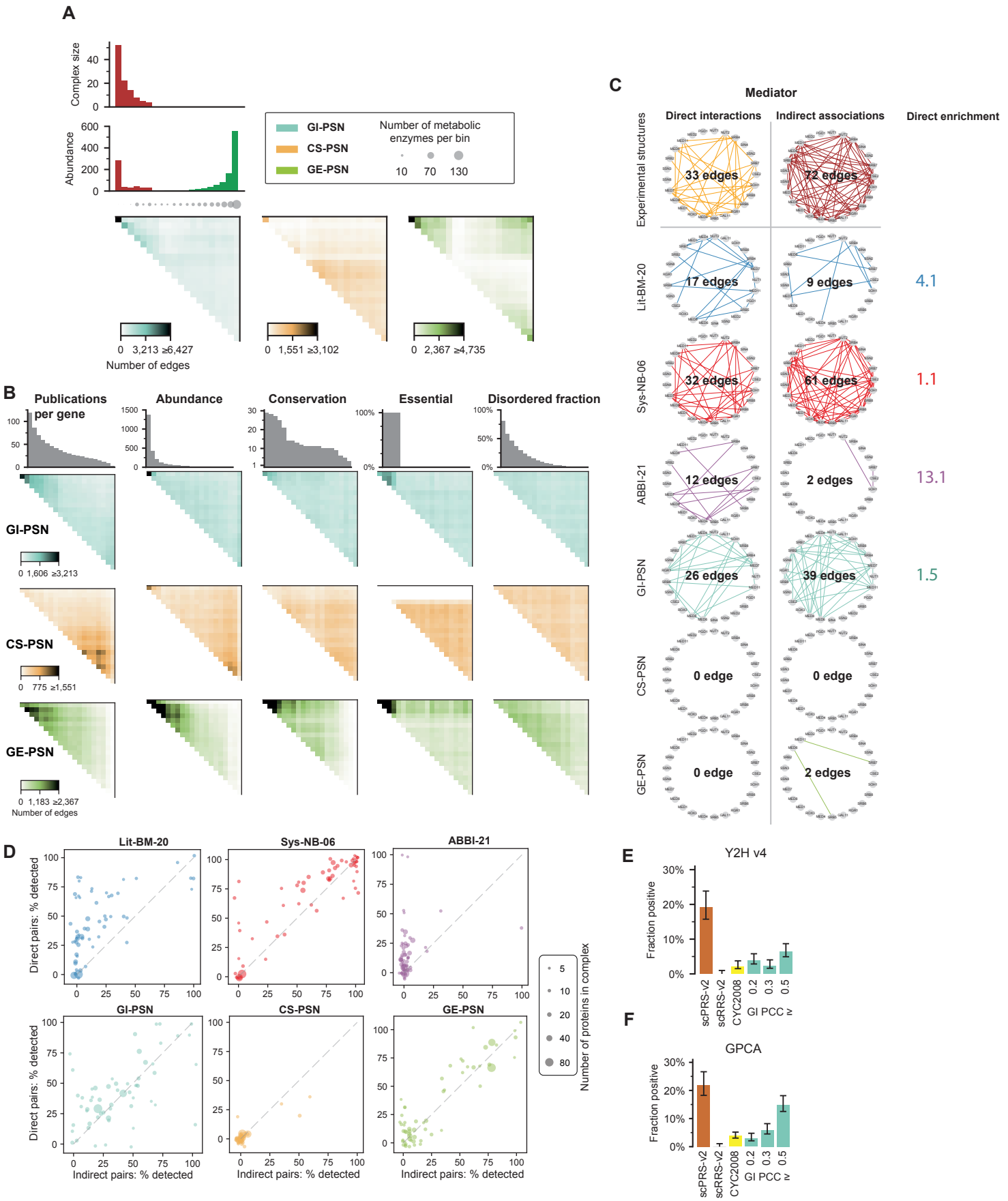

A
Complex size
Abundance
GI-PSN
CS-PSN
GE-PSN
Number of metabolic enzymes per bin
10 70 130
0 3,213 ≥6,427
Number of edges
0 1,551 ≥3,102
0 2,367 ≥4,735
B
Publications per gene
Abundance
Conservation
Essential
Disordered fraction
GI-PSN
0 1,606 ≥3,213
CS-PSN
0 775 ≥1,551
GE-PSN
0 1,183 ≥2,367
Number of edges
C
Mediator
Direct interactions
Indirect associations
Direct enrichment
Experimental structures
33 edges
72 edges
Lit-BM-20
17 edges
9 edges
4.1
Sys-NB-06
32 edges
61 edges
1.1
ABBI-21
12 edges
2 edges
13.1
GI-PSN
26 edges
39 edges
1.5
CS-PSN
0 edge
0 edge
GE-PSN
0 edge
2 edges
D
Lit-BM-20
Sys-NB-06
ABBI-21
GI-PSN
CS-PSN
GE-PSN
Direct pairs: % detected
Indirect pairs: % detected
Number of proteins in complex
E
Y2H v4
Fraction positive
scPRS-v2
scRRS-v2
CYC2008
GI PCC ≥
F
GPCA
Fraction positive
scPRS-v2
scRRS-v2
CYC2008
GI PCC ≥

Figure 4

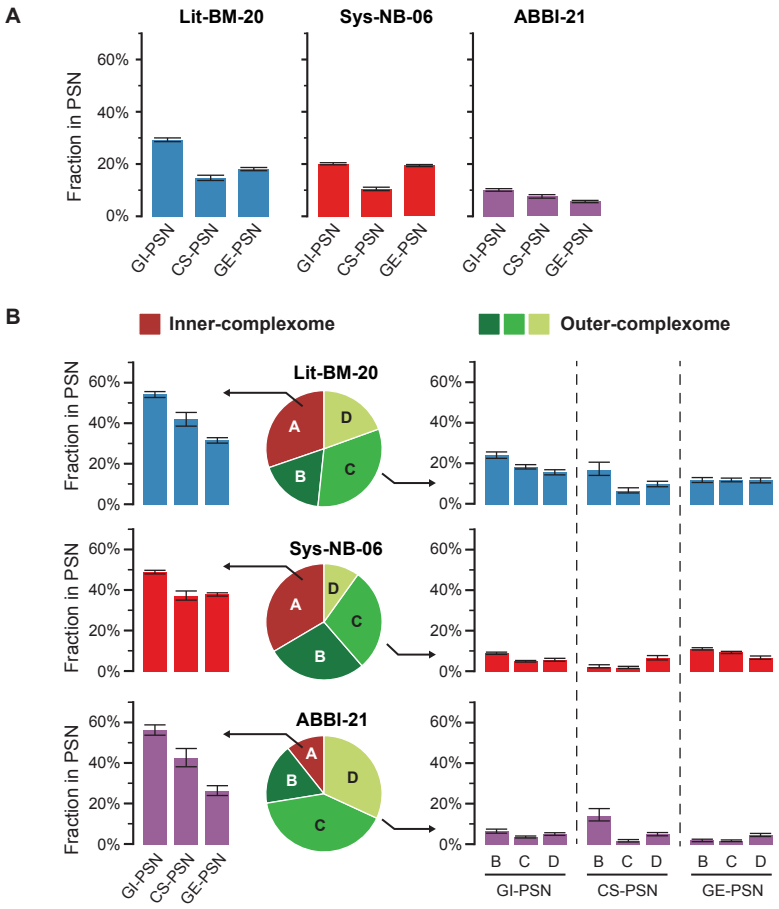

A
Lit-BM-20
Sys-NB-06
ABBI-21
Fraction in PSN
60%
40%
20%
0%
GI-PSN
CS-PSN
GE-PSN
B
Inner-complexome
Outer-complexome
Lit-BM-20
Sys-NB-06
ABBI-21
A
B
C
D
B C D
GI-PSN
CS-PSN
GE-PSN

Figure 5

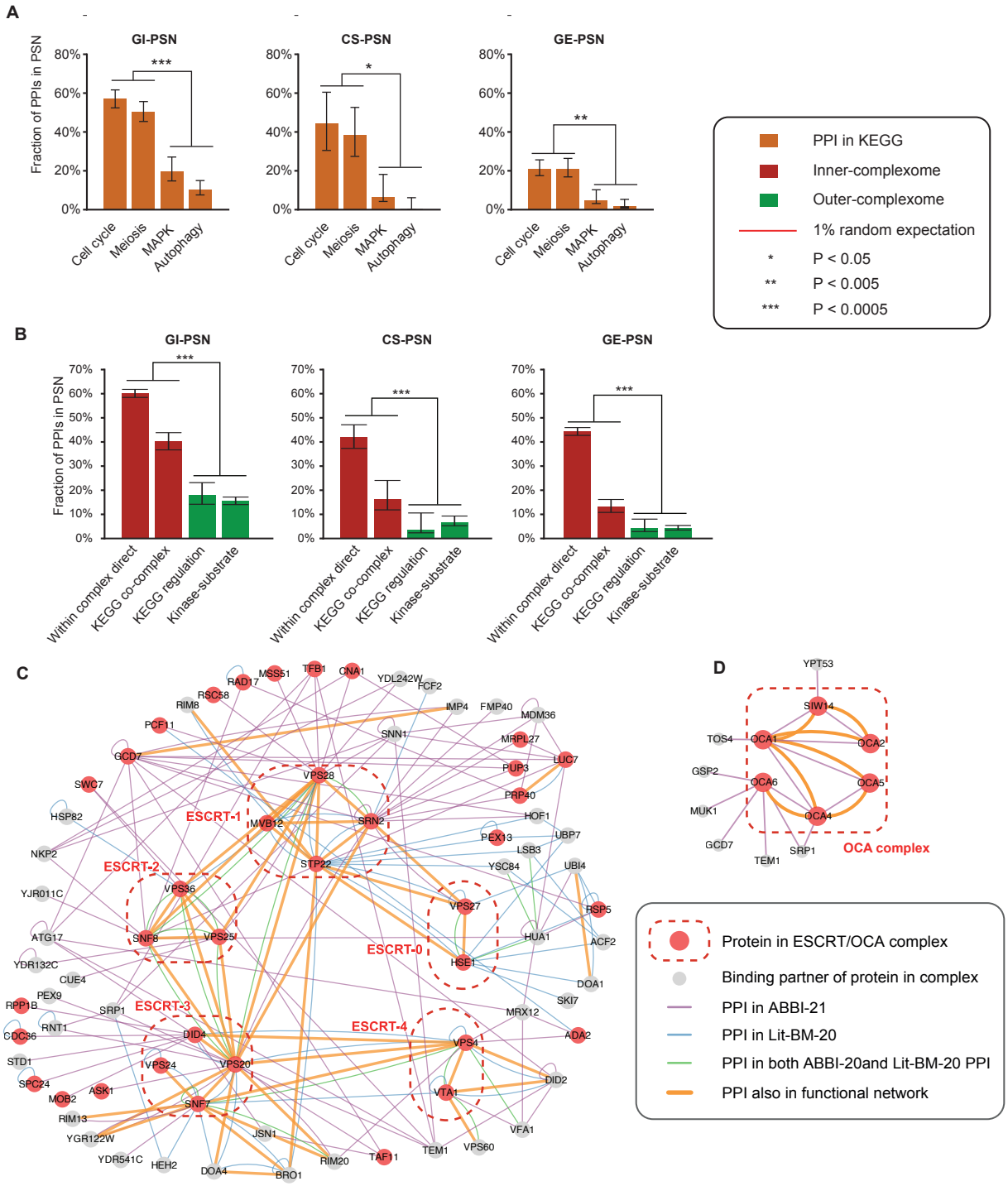

A
GI-PSN
CS-PSN
GE-PSN
Fraction of PPIs in PSN
80%
60%
40%
20%
0%
***
*
**
Cell cycle
Meiosis
MAPK
Autophagy
PPI in KEGG
Inner-complexome
Outer-complexome
1% random expectation
* P < 0.05
** P < 0.005
*** P < 0.0005
B
70%
50%
30%
10%
Within complex direct
KEGG co-complex
KEGG regulation
Kinase-substrate
C
ESCRT-0
ESCRT-1
ESCRT-2
ESCRT-3
ESCRT-4
D
OCA complex
Protein in ESCRT/OCA complex
Binding partner of protein in complex
PPI in ABBI-21
PPI in Lit-BM-20
PPI in both ABBI-20and Lit-BM-20 PPI
PPI also in functional network

Figure 6

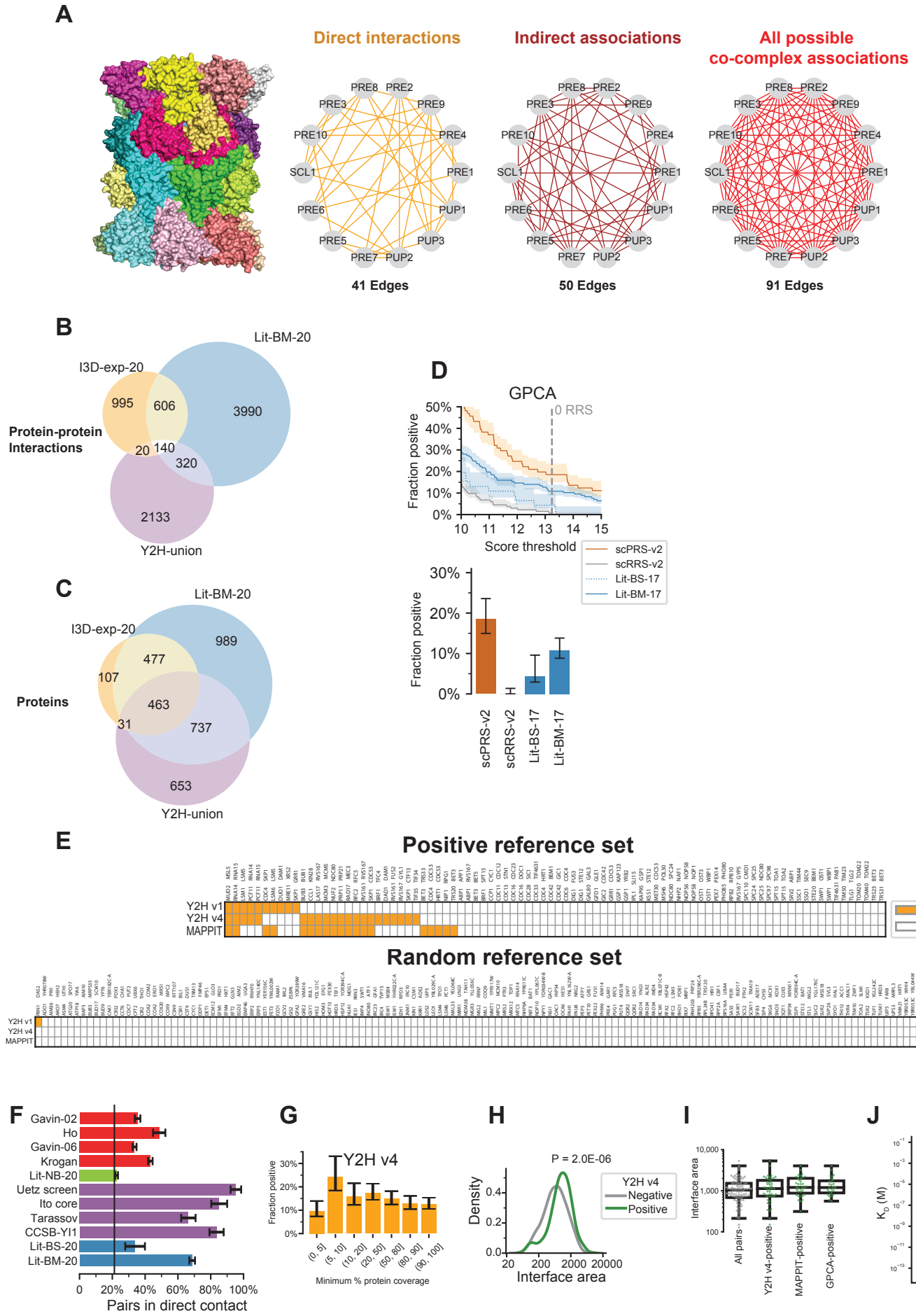
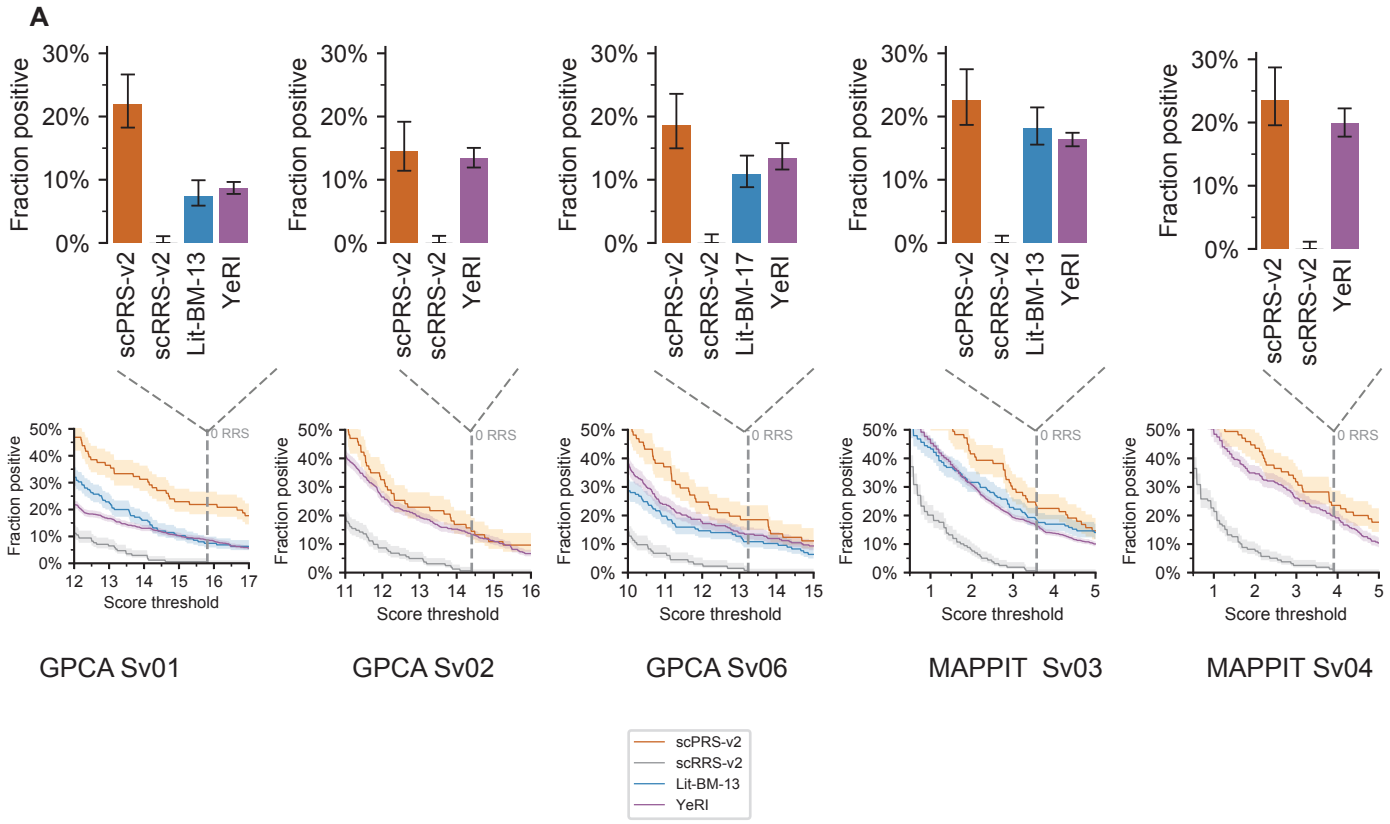
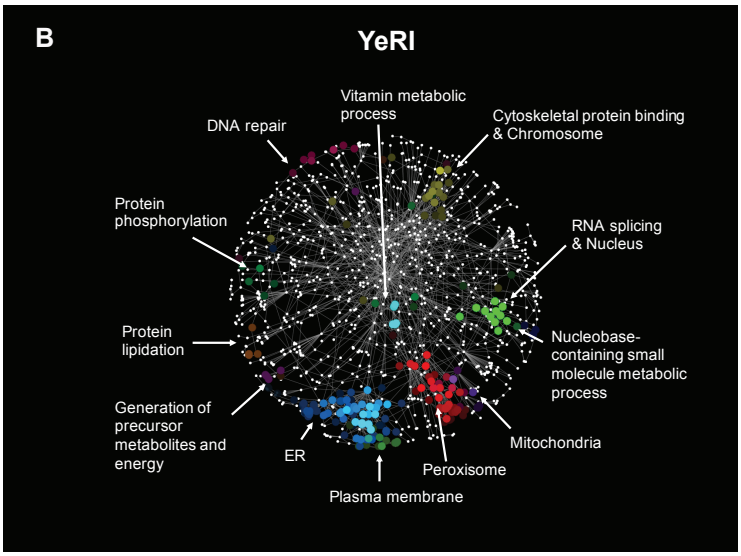

A
Direct interactions
Indirect associations
All possible co-complex associations
PRE8 PRE2 PRE3 PRE9 PRE10 PRE4 SCL1 PRE1 PRE6 PUP1 PRE5 PUP3 PRE7 PUP2
41 Edges
50 Edges
91 Edges
B
Lit-BM-20
I3D-exp-20
Y2H-union
Protein-protein Interactions
995
606
3990
20
140
320
2133
C
Lit-BM-20
I3D-exp-20
Y2H-union
Proteins
989
477
107
463
31
737
653
D
GPCA
0 RRS
Fraction positive
Score threshold
scPRS-v2
scRRS-v2
Lit-BS-17
Lit-BM-17
E
Positive reference set
Y2H v1
Y2H v4
MAPPIT
Positive
Negative
Random reference set
F
Gavin-02
Ho
Gavin-06
Krogan
Lit-NB-20
Uetz screen
Ito core
Tarassov
CCSB-YI1
Lit-BS-20
Lit-BM-20
Pairs in direct contact
G
Y2H v4
Fraction positive
Minimum % protein coverage
H
P = 2.0E-06
Y2H v4
Negative
Positive
Density
Interface area
I
Interface area
All pairs
Y2H v4-positive
MAPPIT-positive
GPCA-positive
J
Y2H v4
Positive
Negative

Figure S1


A
Fraction positive
30%
20%
10%
0%
scPRS-v2
scRRS-v2
H
Lit-BM-13
Lit-BM-17
YeRI
0 RRS
50%
40%
Score threshold
GPCA Sv01
GPCA Sv02
GPCA Sv06
MAPPIT Sv03
MAPPIT Sv04
scPRS-v2
scRRS-v2
Lit-BM-13
YeRI


B
YeRI
Vitamin metabolic process
DNA repair
Cytoskeletal protein binding & Chromosome
Protein phosphorylation
RNA splicing & Nucleus
Protein lipidation
Nucleobase-containing small molecule metabolic process
Generation of precursor metabolites and energy
ER
Mitochondria
Peroxisome
Plasma membrane

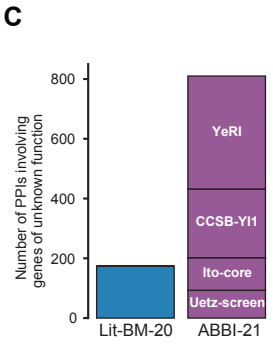

C
Number of PPIs involving genes of unknown function
800
600
400
200
0
YeRI
CCSB-YI1
Ito-core
Uetz-screen
Lit-BM-20
ABBI-21

Figure S2

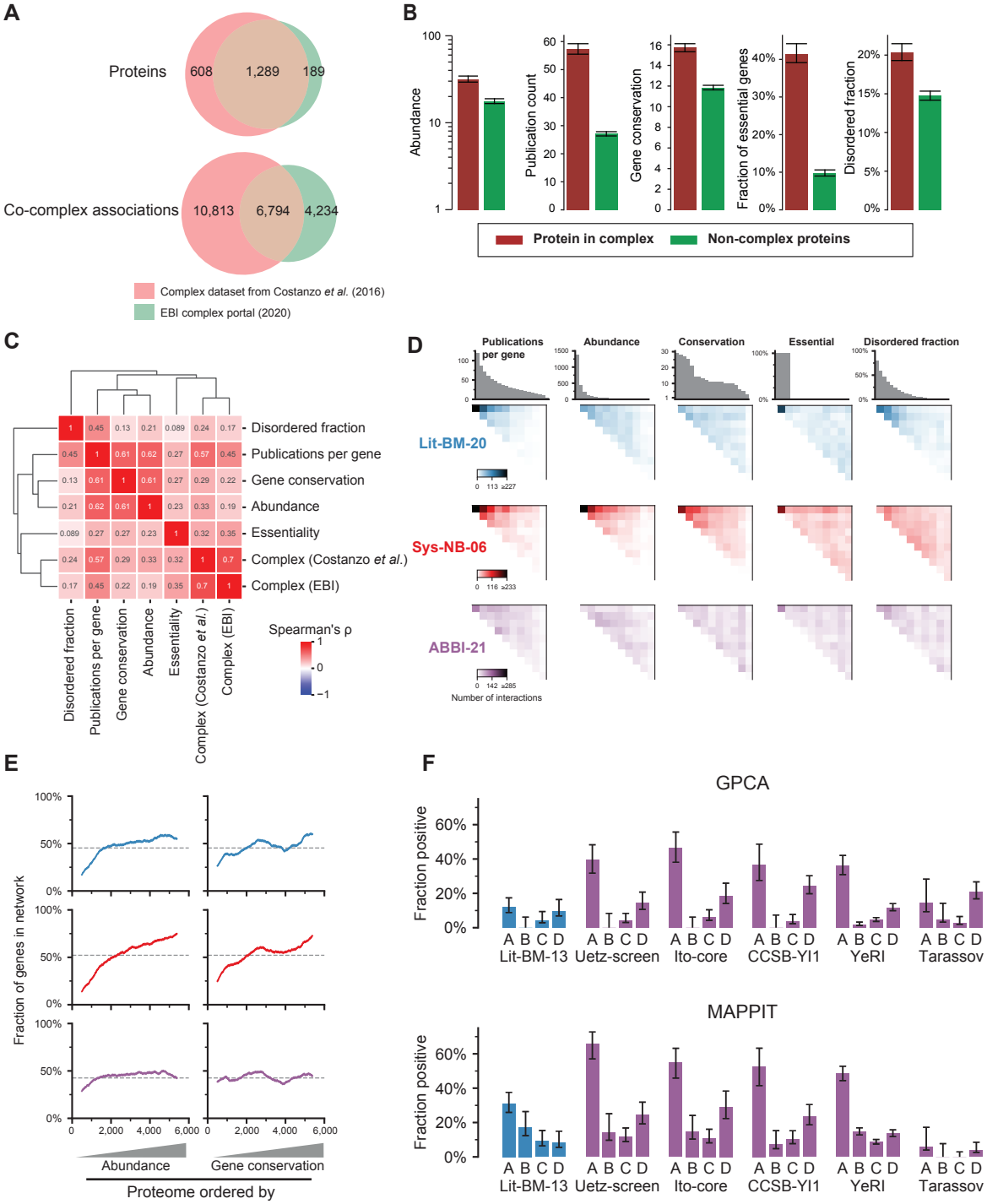

A
Proteins
608
1,289
189
Co-complex associations
10,813
6,794
4,234
Complex dataset from Costanzo et al. (2016)
EBI complex portal (2020)
B
Abundance
Publication count
Gene conservation
Fraction of essential genes
Disordered fraction
Protein in complex
Non-complex proteins
C
Disordered fraction
Publications per gene
Gene conservation
Abundance
Essentiality
Complex (Costanzo et al.)
Complex (EBI)
Spearman's ρ
D
Publications per gene
Abundance
Conservation
Essential
Disordered fraction
Lit-BM-20
Sys-NB-06
ABBI-21
Number of interactions
E
Fraction of genes in network
Abundance
Gene conservation
Proteome ordered by
F
GPCA
MAPPIT
Fraction positive
Lit-BM-13
Uetz-screen
Ito-core
CCSB-YI1
YeRI
Tarassov

Figure S3

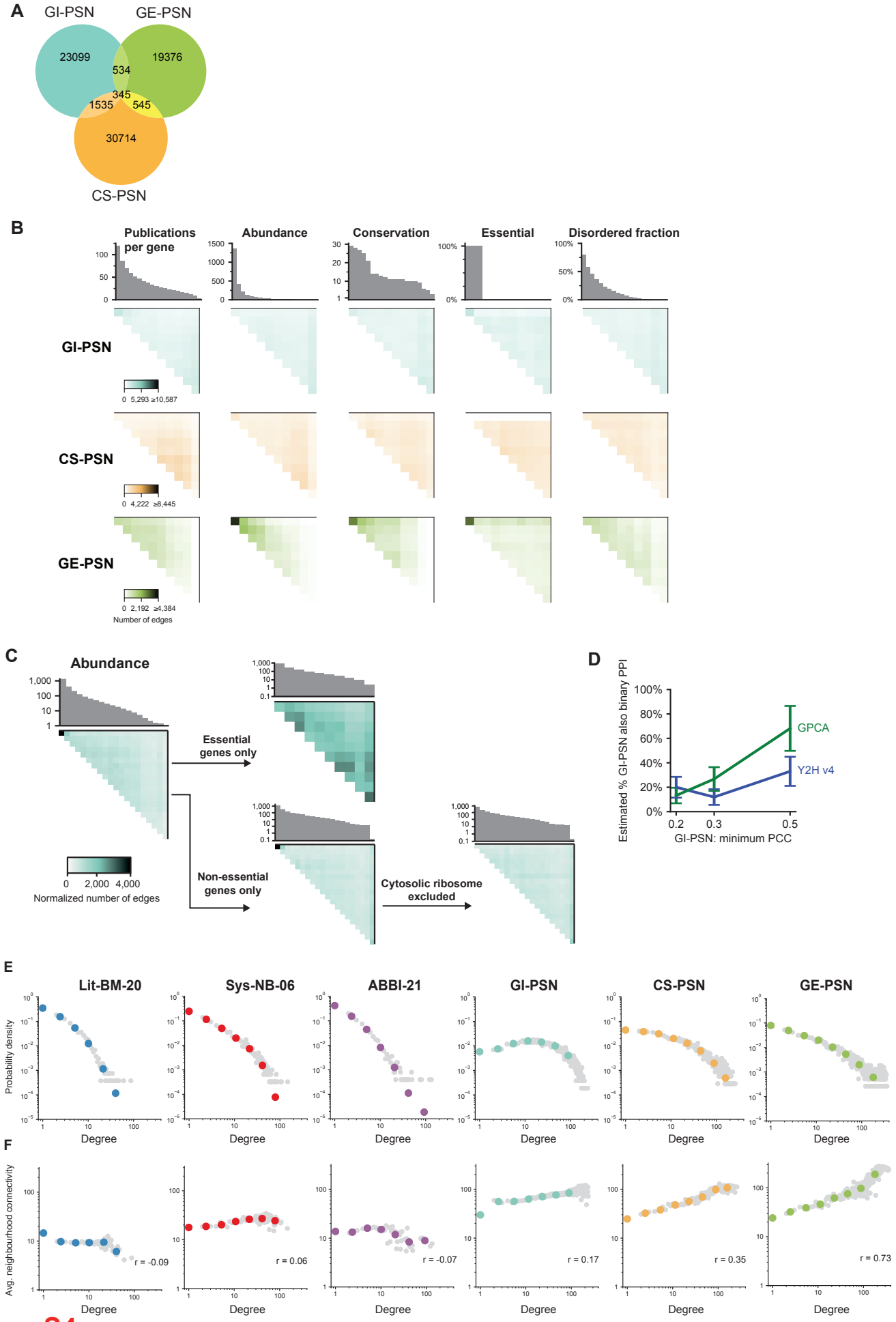

A
GI-PSN
GE-PSN
CS-PSN
23099
19376
534
345
1535
545
30714
B
Publications per gene
Abundance
Conservation
Essential
Disordered fraction
GI-PSN
CS-PSN
GE-PSN
0 5,293 ≥10,587
0 4,222 ≥8,445
0 2,192 ≥4,384
Number of edges
C
Abundance
Essential genes only
Non-essential genes only
Cytosolic ribosome excluded
0 2,000 4,000
Normalized number of edges
D
Estimated % GI-PSN also binary PPI
GPCA
Y2H v4
GI-PSN: minimum PCC
E
Lit-BM-20
Sys-NB-06
ABBI-21
GI-PSN
CS-PSN
GE-PSN
Probability density
Degree
F
Avg. neighbourhood connectivity
r = -0.09
r = 0.06
r = -0.07
r = 0.17
r = 0.35
r = 0.73

Figure S4

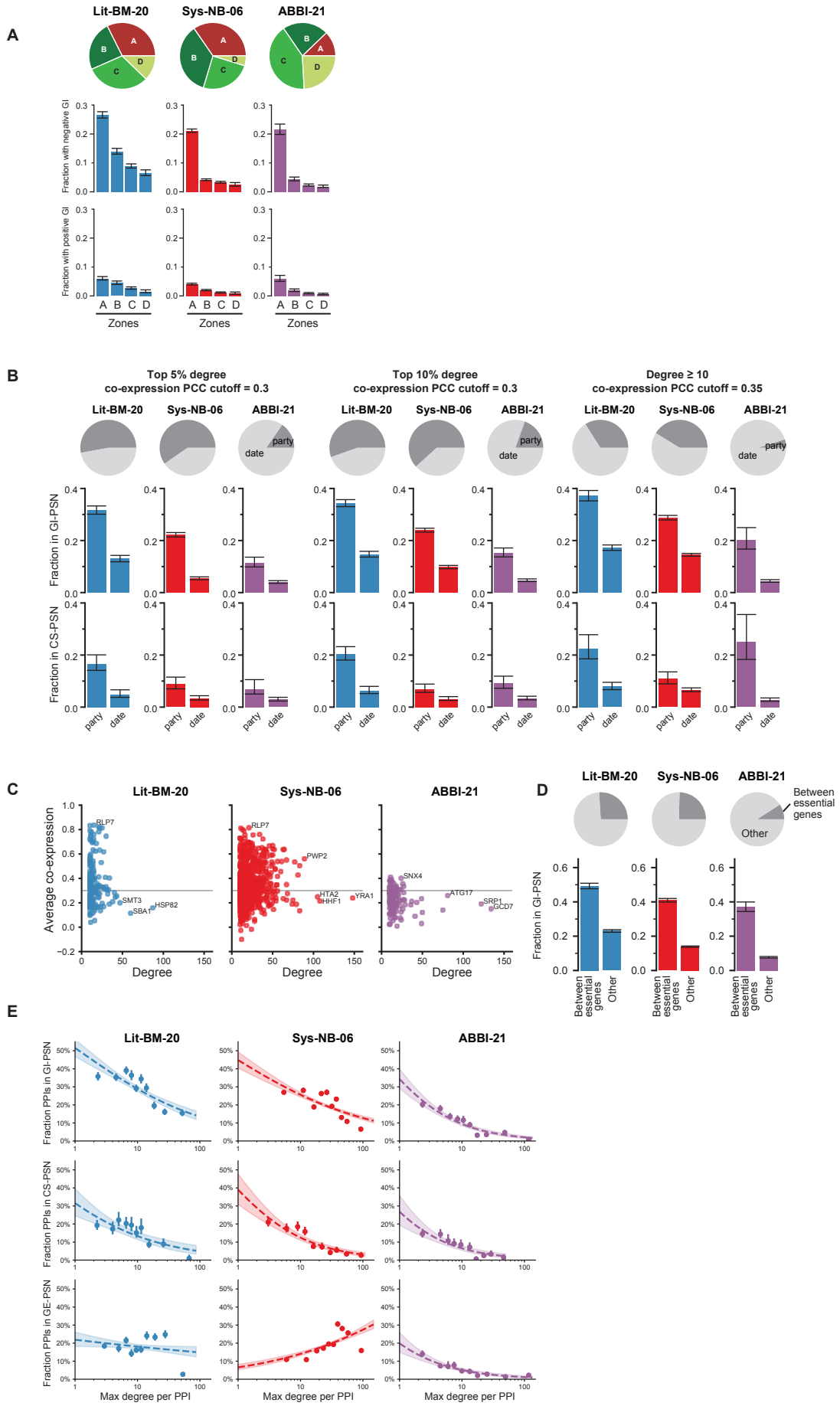

A
Lit-BM-20
Sys-NB-06
ABBI-21
Fraction with negative GI
Fraction with positive GI
Zones
B
Top 5% degree
co-expression PCC cutoff = 0.3
Top 10% degree
co-expression PCC cutoff = 0.3
Degree ≥ 10
co-expression PCC cutoff = 0.35
Fraction in GI-PSN
Fraction in CS-PSN
party
date
C
Average co-expression
Degree
RLP7
SMT3
SBA1
HSP82
PWP2
HTA2
HHF1
YRA1
SNX4
ATG17
SRP1
GCD7
D
Between essential genes
Other
E
Fraction PPIs in GI-PSN
Fraction PPIs in CS-PSN
Fraction PPIs in GE-PSN
Max degree per PPI

Figure S5