Binary Interactome Models of Inner- Versus Outer-Complexome Organization

Luke Lambourne,[1-3,23] Anupama Yadav,[1-3,23] Yang Wang,[1-3,23] Alice Desbuleux,[1-4,23] 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 W. Mee,[5,11] Mohamed Helmy,[5] Yves Jacob,[1,12,13] Irma Lemmens,[14,15] Thomas Rolland,[1-3] Atina G. Cote,[1,5-7] Marinella Gebbia,[1,5-7] Nishka Kishore,[1,5-7] Jennifer J. Knapp,[1,5-7] Joseph C. Mellor,[1,5-7,16] Jüri Reimand,[6,11,17] Jan Tavernier,[14,15] Michael E. Cusick,[1-3] Pascal Falter-Braun,[1-3,18,19] Kerstin Spirohn,[1-3] Quan Zhong,[1-3,20] Patrick Aloy,[8,21] Tong Hao,[1-3] Benoit Charloteaux,[1-3] Frederick P. Roth,[1,5-7,22] 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]Department of Physics and Astronomy, Northwestern University, Evanston, IL, USA
[11]Computational Biology Program, Ontario Institute for Cancer Research, 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]seqWell, Beverly, MA, USA
[17]Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada
[18]Institute of Network Biology (INET), Helmholtz Center Munich, German Research Center for Environmental Health, Munich-Neuherberg, Germany
[19]Microbe-Host Interactions, Faculty of Biology, Ludwig-Maximilians-University (LMU) Munich, Planegg-Martinsried, Germany
[20]Department of Biological Sciences, Wright State University, Dayton, OH, USA
[21]Institució Catalana de Recerca I Estudis Avançats (ICREA), Barcelona, Catalonia, Spain
[22]Department of Computer Science, University of Toronto, Toronto, ON, Canada
[23]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.)

SUMMARY

The interactome is often conceived of primarily as a collection of hundreds of multimeric machines, collectively referred to as the “complexome”. However, a large proportion of the interactome exists outside of the complexome, or in the “outer-complexome”, and may account for most of the functional plasticity exhibited by cellular systems. To compare features of inner- versus outer-complexome organization, we systematically generated a yeast all-by-all binary interactome map, integrated it with previous binary maps, and compared the resulting interactome “atlas” with systematic co-complex association and functional similarity network datasets. Direct protein-protein interactions in the inner-complexome tend to be readily detected in multiple 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, exhibit higher levels of functional heterogeneity. Thus, a small proportion of the interactome corresponds to a stable, functionally homogeneous, inner-complexome, forming quaternary structure, while a much greater proportion consists of transient interactions between pairs of functionally heterogeneous proteins in the outer-complexome, forming quinary structure.

INTRODUCTION

Intracellular organization relies on large numbers of macromolecular interactions forming extremely complex interactome networks that underlie highly diverse functional relationships between proteins, RNA molecules and DNA. Several efforts have been described to map and model, at proteome-scale, global and local properties of biophysical interactome networks such as protein-protein interaction (PPI) networks (Cafarelli et al., 2017; Havugimana et al., 2017; Luck et al., 2017), including the recent release of a first-of-its-kind binary reference map of the human binary protein interactome (Luck et al., 2020). Although extremely useful in our quest to understand cellular organizational principles (Bludau and Aebersold, 2020; Vidal et al., 2011), current models describing global properties of PPI networks, for the most part, omit to consider the range of biophysical properties exhibited by different cellular PPIs.

On one side of that range are relatively stable, direct PPIs taking place in highly constrained "quaternary" structures forming protein complexes, which are often assembled through contacting pairs of proteins via discrete interaction-mediating domains. The protein interactome is often conceived of, primarily, as a collection of hundreds of such multimeric machines, collectively referred to as the "complexome" (Deshaies et al., 2002). However, these stable quaternary PPIs are only a subset of the binary protein-protein interactome (Luck et al., 2020; Vidal, 2001). On the other side of the range of biophysical properties, are weaker transient interactions that are much more dependent on aspects of the overall cellular milieu, such as molecular crowding, and that are more likely to underlie the overall organization and compartmentalization in cells and help sustain biochemical pathways and signaling cascades (Cohen and Pielak, 2017; Davey, 2019; Guin and Gruebele, 2019). Four decades ago, the term "*quinary structure*" was suggested to describe this "fifth level of organization", referring to macromolecular interactions that, although potentially highly functionally relevant,

might be “transient *in vivo”* (Edelstein, 1980; McConkey, 1982; Vaĭnshteĭn, 1973). As predicted by 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”.

How these two fundamental aspects of PPIs interconnect at the scale of the whole interactome remains largely unresolved. Are these two parts of the interactome distinguishable in terms of fundamental principles underlying their global organization? The part of the proteome involved in processing or manufacturing tends to be highly abundant (Liebermeister et al., 2014), yet only represents a relatively small fraction of all encoded proteins. In contrast, the rest of the proteome, including regulatory and control systems, encompasses a relatively large number of proteins, the majority of which are expressed at relatively low abundance (Liebermeister et al., 2014). Might the strikingly distinct roles of these interconnected systems correspond to distinct types of sub-proteome organization? To what extent might the quaternary interactome, either constitutively expressed or condition-dependent, operate in a relatively constant, robust, and persistent manner, while the regulatory quinary interactome would exhibit greater flexibility, plasticity, environmental responsiveness, and, perhaps, evolvability?

A eukaryotic proteome can be operationally organized into three major classes of proteins: i) subunits of protein complexes; ii) highly abundant non-complex proteins; and iii) non-complex proteins expressed at low abundance (Figure 1A). Protein complexes include molecular machines involved in genetic information processing such as DNA replication, RNA transcription, and protein translation and degradation. Examples range in size from the gigantic ribosome and the more moderately sized mediator to smaller complexes such as telomerase. Approximately 60-80% of proteins, depending on the species, are not constituent parts of any molecular machine (Meldal et al., 2019; Ruepp et al., 2010) and can thus be considered “non-complex” proteins. Non-complex proteins can be divided into two groups, those above the mean abundance and those below.

Among the most abundant non-complex proteins are metabolic enzymes, which, in yeast cells, make up 30% of proteins by molarity, but are only 10% of all encoded proteins (Cherry et al., 2012; Wang et al., 2015). The third class of proteins, scarce under normal conditions, make up 60% of encoded proteins but less than 10% of the proteins, by molarity, in the yeast cell.

Here, we investigate organizational principles 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 have different properties or how such interactions relate to functional relationships is not well understood. Likewise, whether alternative organizational properties might be expected between highly abundant or scarce proteins remains mostly unresolved.

To address these questions, we selected *S. cerevisiae*, which, as the most extensively studied eukaryotic model system, has available the most comprehensive and diverse systematic datasets of functional relationships between genes as well as comprehensive, yet incomplete, maps of biophysical interactions between proteins. In this study we: i) generated for the first time an “all-versus-all” systematic binary interactome map; ii) integrated it with three previously available binary maps; and iii) compared the resulting systematic binary interactome “atlas” to: i) a systematic co-complex association network map; and ii) three different global functional profile similarity network (PSN) maps reporting genetic interactions, condition-specific phenotypes and gene expression. We find strong support for an emerging model in which a relatively small proportion of the interactome corresponds to the stable, quaternary, functionally coherent, inner-complexome, while the vast majority of the interactome consists of quinary interactions between functionally heterogeneous proteins forming the outer-complexome.

RESULTS

Visualizing the Complexome

To visualize the distribution of biological networks between inner- and outer-complexome, we organized the global proteome-by-proteome space according to two key protein properties: co-complex membership and abundance (Figure 1A). First, we ordered all proteins that are subunits of a complex in decreasing order of the size of the corresponding complex. Then the remaining non-complex proteins were ordered by increasing abundance. In addition to an integrated dataset of abundance under normal conditions (Wang et al., 2015), we used a yeast complexome dataset (Costanzo et al., 2016), filtering for complexes containing three or more distinct protein subunits, resulting in 339 complexes, containing 1,897 different proteins out of the total 5,883 yeast proteins.

Protein complexes span a range of sizes. Examples include: the ribosomal large subunit, nuclear pore, mediator, spliceosome, TFIIH, prefoldin, calcineurin, and telomerase with 81, 52, 27, 16, 10, 6, 4, and 3 unique protein subunits, respectively. The abundance distribution of proteins is such that a small number of encoded proteins make up a large fraction of expressed proteins. Of non-complex proteins, 89% are below the mean abundance, by molarity. Just two proteins, pyruvate kinase encoded by *CDC19* and the plasma membrane P2-type H+-ATPase transporter encoded by *PMA1*, account for more than two percent of the total proteins, by molarity, in the cell. At the other end of the spectrum, lowly expressed proteins such as Cln3, an important cell-cycle regulator, and Ime1, a master regulator of meiosis, are four orders of magnitude lower in abundance than Cdc19 and Pma1.

We then extended this organization of the one-dimensional proteome space to the two-dimensional proteome-by-proteome space of ~18 million protein-protein pairwise combinations (Figure 1B). We refer to this as a “complexogram”, within which we define four different “zones” (Figure 1B; STAR Methods #3): 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 protein belongs to a different complex; iii) Zone C represents all pairwise combinations between complex subunits and non-complex proteins; and iv) Zone D corresponds to all pairwise combinations between non-complex proteins.

The yeast inner-complexome, Zone A, with its 339 complexes corresponds to a maximum of ~17,600 possible protein pairings, representing 0.1% of the approximately 18 million possible protein pairs (see STAR Methods #2). In contrast, the outer-complexome encompasses 99.9% of the whole yeast interactome space of possible pairings with Zone B, C, and D representing 10.3%, 43.7%, and 45.9%, respectively (Figure 1B). Finally, in both Zone C and D, subzones can be defined relative to abundance levels with: i) greater than 30% of the space representing pairwise combinations of proteins of below the mean abundance; and ii) only 1.5% of the space corresponding to pairs where both proteins are above the mean abundance (Figure 1B).

Current Status of the Yeast Binary Interactome Landscape

Next, we assessed available biophysical datasets, to find the most suitable datasets with which to study the interactome in the inner- and outer-complexome. Maps of biophysical relationships between pairs of proteins report either co-complex associations, where proteins are in the same complex but not necessarily in direct contact, or “binary” PPIs, for which two interaction partners are likely to be in direct contact (Figure S1A). It is important to distinguish between these two types of protein-protein relationship when investigating the inner- and outer-complexome, as the two will show different patterns, with one example being that large protein complexes contain many more indirect associations than direct contacts. In this study we focus on binary PPIs, as they are the more fundamental of the two. 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 2A; STAR Methods #4, #20): i) systematic yeast two-hybrid (Y2H) proteome-scale maps, collectively reporting 2,613 heterodimeric "Y2H-union" PPIs (Ito et al., 2001; Uetz et al., 2000; Yu et al., 2008); ii) literature-curated binary pairs supported by multiple pieces of 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 providing definitive evidence of physical, direct interaction between 1,761 experimentally derived heterodimeric Interactome3D or "I3D-exp-20" pairs (Mosca et al., 2013) (Figures 2A, S1B-C). Systematic Y2H datasets were released in three publications each using a different single version of the Y2H method. In contrast, Lit-BM-20 was derived from 6,127 curated publications that collectively used 81 different experimental interaction detection methods (Table S1). Y2H is the most widely used binary PPI assay in the literature, with 76% (3,986/5,056) of heterodimeric Lit-BM-20 interactions being supported by Y2H evidence (Figure 2A). Lit-BM-20 contains only 4% of all literature-curated pairs because the vast majority of curated pairs correspond to pairs lacking binary evidence or pairs supported by only a single experiment. Other datasets considered were: "Tarassov", a proteome-scale dataset generated using a dihydrofolate reductase protein complementation assay (DHFR PCA) (Tarassov et al., 2008); four systematic co-complex maps (Gavin et al., 2002, 2006; Ho et al., 2002; Krogan et al., 2006); a complexome dataset (Pu et al., 2009); and two sets of predicted PPIs (Jansen et al., 2003; Zhang et al., 2013) (Table S2).

To better understand the quality of literature-curated versus systematic datasets before using them to investigate principles of interactome organization, we used two orthogonal assays: 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) (Figure 2B, STAR Methods #14, #21, Table S3). We assessed genuine binary interaction content by comparing the recovery of pairs selected from each dataset to that of 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 #5, Table S4). These two assays were not used to discover any of the PPIs in the tested datasets: literature-curated, Y2H-union and Tarassov. Across a wide range of stringency for both MAPPIT and GPCA, all three systematic Y2H datasets constituting Y2H-union, "Uetz-screen" (Uetz et al., 2000), "Ito-core" (Ito et al., 2001), and "CCSB-YI1" (Yu et al., 2008), as well as "Lit-BM-13", a random sample from literature binary PPIs available in 2013, validated close to scPRS-v2 (Figure 2B), indicating that the Y2H-union and Lit-BM maps are of high quality. Using the score set by the highest scoring of the 198 scRRS-v2 pairs as a threshold, MAPPIT recovered scPRS-v2, Lit-BM-13, and systematic Y2H pairs at rates statistically indistinguishable from one another, between 18% and 27%. At a similarly stringent threshold, GPCA recovered significantly fewer Lit-BM-13 pairs than Y2H-union pairs (Figure 2B). Interestingly, Tarassov, the DHFR PCA dataset, 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 2B). As shown previously for human PPIs (Rolland et al., 2014), a sample of the putative yeast binary PPIs supported by only a single piece of evidence in the literature in (Lit-BS) were recovered at a low rate, not statistically different from RRS (P = 0.06, Fisher's exact test), of only 4% in GPCA (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 2B bottom panel).

We wished to further investigate binary PPI information using Interactome3D, literature curation, and systematic binary maps. We therefore experimentally retested all pairs available in 2017 for these three datasets (I3D-exp-17, Lit-BM-17, Y2H-union), along with random samples of other datasets, thus testing a total of 8,999 pairs (Figure 2C, Table S5). We used a different Y2H version, Y2H v4 (see STAR Methods #6-7), that had not been used for systematic screening. Y2H v4 detects 19% of scPRS-v2 pairs

under conditions in which 0/198 scRRS-v2 are recovered (Figure 2C, Table S6). 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 #7), further supporting the hypothesis that different versions of the same assay differ in their ability to detect different interactions (Choi et al., 2019). Y2H v4 recovered 11%, 14%, and 15% of Interactome3D, literature curation, and systematic binary pairs, respectively. Strikingly, Y2H-union was indistinguishable from both scPRS-v2 (P = 0.2, Fisher's exact test) and I3D-exp-17 (P = 0.6, Fisher's exact test), our benchmark of true direct, heterodimeric binary interactions, and all three performed slightly better than Lit-BM-17 (P = 0.006, Fisher's exact test) (Figure 2C), again demonstrating that the biophysical quality of systematic binary interaction maps is at least as good, if not superior to that of literature-curated binary interactions, suggesting that they are usable for the purpose of investigating interactome organization.

In striking contrast, pairs from co-complex association datasets were detected at much lower rates than those from for binary literature-curated with multiple evidence or systematic binary datasets, although significantly higher than the negative control scRRS-v2 (median P = 0.018, Fisher's exact test) (Figure 2C). This result suggests that a large proportion of protein pairs in co-complex association datasets are indirect, distal associations (Figure S1A), both in literature-curated information as in CYC2008 (Pu et al., 2009), a carefully curated dataset, and in affinity-purification followed by mass spectrometry (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 pairs (P = 0.27, Fisher's exact test); and iii) the four systematic, non-binary, AP-MS co-complex association maps currently available (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 4% vs 17%, P = 0.0002, Fisher's exact test) (Figure 2C). 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 pairs, from either literature-curation or systematic experiments (Figure S1F, STAR Methods #15, Table S7A). Finally, two datasets of predicted PPIs, PrePPI and Jansen et al (Jansen et al., 2003; Zhang et al., 2013), tested positive at low levels of 4% and 2%, respectively (Figure 2C).

Testing a comprehensive set of Interactome3D pairs with Y2H v4 also demonstrated that, although the number of subunits involved in forming large protein complexes appears to have some impact on the rate of interaction recovery by Y2H v4 (Figure 2D), binary assays can readily detect pairs of interacting proteins even in large complexes. Second, although our data was 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 fragments (P = 0.14, Kolmogorov-Smirnov test) (Figure S1G). Third, although we observed a trend towards larger surface areas among Y2H v4 positives compared to Y2H v4 negatives, Y2H v4 could detect interactions with interface areas ranging widely from 100 to 10,000 $Å^2$ (Figure S1H, STAR Methods #22, Table S7B). Y2H v4 was better able to detect PPIs with small interaction interfaces than MAPPIT and GPCA (Figure S1I). Finally, Y2H v4 detected interactions with K_d values up into the micromolar range (Figure S1J, STAR Methods #16), consistent with previous findings that Y2H can identify even weak interactions (Estojak et al., 1995).

To investigate the overall homogeneity of coverage, we displayed binary PPIs in a representation of the proteome-by-proteome space organized by ranking proteins in both dimensions according to their number of publications (Rolland et al., 2014) (Figure 2E, STAR Methods #23). A "dense zone" matrix of 15% of yeast protein pairs contains 80% of Interactome3D pairs and a matrix of 25% of yeast protein pairs contains 80% of

literature-curated PPIs. In striking contrast, Y2H-union interacting pairs are distributed more homogeneously across the space, such that a matrix of 50% of pairs is needed to contain 80% of the PPIs (Figure 2E). 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 2F).

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

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

Although we have shown that systematic mapping strategies yield high-quality PPI data that could inform us about inner- versus outer-complexome organization and other properties of the interactome, the three maps currently available do not fully cover the whole yeast interactome (Figure 2E). This is due to the “search space incompleteness” of the three screens performed so far (Yu et al., 2008). CCSB-YI1 and Ito-core were obtained using incomplete sets of open reading frames (ORFs), or “ORFeome” collections (Walhout et al., 2000), so that only ~70-75% of the search space was screened in each study and only ~60% screened by both.

To improve coverage of the yeast binary interactome, we compiled a nearly complete ORFeome collection of high-quality ORFs by verifying an already existing, yet incomplete collection of 4,933 ORFs, the FLEXGene collection (Hu et al., 2007), and cloning an additional 921 ORFs, corresponding to a 19% increase. The resulting nearly complete ORFeome collection covers 99% of the annotated yeast protein-coding genes

(5,854/5,883) (See STAR Methods #1, #8, Table S8A). To maximize the potential for novel discovery relative to the three previous binary systematic maps, we used the Y2H v4 version described above. To maximize coverage, we systematically screened a space of 27 million bait-prey combinations three times independently (Figure 3A, STAR Methods #9-11). 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 #13). The quality of putatively interacting pairs that were positive in both attempts, and for which the identity of both ORFs could be sequence-confirmed (see STAR Methods #12, #24), was assessed using both MAPPIT and GPCA validation (see STAR Methods #14). To maximize the 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, and that exhibited validation at a rate similar to Lit-BM-20 in both MAPPIT and GPCA (Figures 3A, S2A, Table S8B). YeRI displays significant enrichment for interactions between proteins that share cellular compartment, pathway, or protein complex annotations (Figure 3B, STAR Methods #25), 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 the number of publications as described above (Rolland et al., 2014). Compared to Uetz-screen, Ito-core, and CCSB-YI1, YeRI covers the search space more homogeneously (Figure 3C). Since we showed that all four maps, Uetz-screen, Ito-core, CCSB-YI1, and YeRI validate to the same extent using 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

dataset (Ito et al., 2001; Uetz et al., 2000; Yu et al., 2008). ABBI-21, with the addition of YeRI, is of similar size to Lit-BM-20 (Figure 3D, Table S8C).

To test whether functionally related proteins tend to be closely connected in ABBI-21, we used the Spatial Analysis of Functional Enrichment (SAFE) method (Baryshnikova, 2016) (see STAR Methods #26). By progressively combining systematic maps based on their chronological order of release, we show that the numbers of functionally enriched clusters increase substantially with the addition of each map (Figure 3E), and that running SAFE on YeRI alone finds 12 functional modules (Figure S2B). In total, ABBI-21 covers 12-25% of the estimated yeast binary interactome (Figure 3F, STAR Methods #27) (Sambourg and Thierry-Mieg, 2010; Stumpf et al., 2008; Yu et al., 2008).

Despite being one of the most well-studied organisms, the function of almost one-sixth of yeast genes remains unknown (Wood et al., 2019) (see STAR Methods #18). We investigated the number of PPIs involving the products of these uncharacterized 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 3G). Given the lack of progress in characterizing the functions of these genes, systematically mapped PPIs provide information to infer their cellular roles. We predicted functions for genes of unknown function with a guilt-by-association approach using GO term annotations of their interaction partners (Luck et al., 2020) (see STAR Methods #28, Table S9) in ABBI-21. The lag between the release of publications describing gene functions and the curation of that information into GO terms sometimes results in a small number of genes which appear in the GO-term based list of genes as “genes of unknown”, while in fact, they have been assigned functions already. These cases present an opportunity to test the accuracy of our predictions. For example, a gene of unknown function *YGR168C* is also known as *PEX35* owing to its recently

demonstrated role as a regulator of peroxisomal abundance (Yofe et al., 2017). Using ABBI-21, we predicted *YGR168C* to be involved in peroxisomal protein import machinery, showcasing the ability of ABBI-21 to accurately predict gene function. Pex35 has 23 PPIs in ABBI-21, all from YeRI, out of which eight are proteins involved in peroxisomal biology (Figure 3H). Pex35 also interacts with another protein encoded by a gene of unknown function, *YKL018C-A*/*MCO12*, which we predict to be involved in peroxisomal abundance as well (Figure 3H). Another example of the efficacy of using a guilt by association approach with our systematic PPI network to predict gene function is *YJR015W*, the product of which was recently demonstrated to be localized to the ER (Koh et al., 2015). Indeed, we predicted *YJR015W* as a putative facilitator of endoplasmic reticulum (ER) transport activity, based on its protein interactions with ER secretory pathway components such as Sec11, Spc1, and Sar1.

In summary, these results illustrate how systematically generated binary PPI maps or atlases, in particular YeRI and ABBI-21, with their (nearly) all-by-all screening coverage, provide valuable information on even the least well understood genes, and thus are a good basis to understand global organizational principles of the interactome.

High Discrepancy Among Alternative Views of Interactome Organization

To understand potential differences of functional relationships between large numbers of pairs of physically interacting proteins within the inner- and outer-complexome, we turned to three different systematic genome-wide functional genomic profiling approaches which determined profiles for most yeast genes based on: i) positive and negative genetic interactions observed in double mutants bearing knock-out (KO) and/or thermosensitive alleles (Costanzo et al., 2016); ii) growth of KOs of non-essential genes across over 1,000 chemical and environmental stress conditions (Hillenmeyer et al., 2008); and iii) transcriptome-wide measurements of gene expression over thousands of samples (Obayashi et al., 2019) (Figure 4A). Taking the 1% highest Pearson's correlation coefficient (PCC) values, we derived three types of PSNs: i) ~135,000 genetic interaction profile similarities forming a "GI-PSN" (Costanzo et al., 2016); ii)

~65,000 condition sensitivity-based similarities constituting a “CS-PSN” (Hillenmeyer et al., 2008); and iii) ~99,000 gene expression similarities leading to a “GE-PSN” (Obayashi et al., 2019). While these three PSNs exhibit statistically significant overlaps ($P < 0.001$, one-sided empirical test, restricted to genes tested in all PSNs), these overlaps are small, with only 4% of edges connected in more than one PSN, suggesting that different PSNs identify complementary functional similarities (Figure S3A).

To understand what functional networks can tell us about the relationships of physically interacting proteins in the inner- and outer-complexome, we examined how their connections are distributed throughout the proteome-by-proteome space. First, we extended our analysis of the coverage of the space of pairwise protein combinations according to the number of publications per gene from Lit-BM-20 literature-curated binary information and our systematic binary atlas ABBI-21 (Figure 2E and 3C) to a systematic non-binary co-complex association dataset derived from the three systematic co-complex association maps available as of 2006 “Sys-NB-06” (Gavin et al., 2002, 2006; Krogan et al., 2006) (Figure 4B). Unlike ABBI-21, and similar to Lit-BM-20, Sys-NB-06 exhibits a strong dense zone, with 50% of the space ranked by publication count containing 80% of the information (Figure 4B). Although Sys-NB-06 was obtained using sociologically unbiased, systematic approaches, we’ve seen previously that systematic co-complex association maps show a bias towards highly studied proteins, which is mediated through a bias towards highly expressed proteins, which are also more highly studied (Luck et al., 2020). To further explore the relationship between experimental coverage of the interactome and protein complexes, we examined four additional intrinsic protein properties potentially associated with complex membership: i) abundance; ii) evolutionary conservation; iii) essentiality; and iv) disorder, first by comparing the average values of complex subunits and non-complex proteins using two different complex datasets (Figure S3B, C, STAR Methods #19, Table S10). 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 S3C, D), with strong correlations between these protein

properties (Figure S3E). We then ranked the proteome-by-proteome space according to these variables and compared the coverage of the different binary PPI and co-complex association maps (Figure 4B). In all cases, ABBI-21 covers the interactome more homogeneously, across multiple biological properties, compared to Lit-BM-20 and Sys-NB-06. ABBI-21 does show some depletion for the highest abundance proteins as well as the lowest abundance, least studied and least conserved proteins (Figure 4B). In the lowest abundance or conservation zones, the number of proteins with at least one interactor is significantly higher for ABBI-21 than Lit-BM-20 or Sys-NB-06 (Figure S3F).

Having identified such fundamental differences between systematic binary and co-complex approaches, we then turned to the three systematic PSN functional networks. One of the most striking observations is the dense zone exhibited by GE-PSN within the spaces corresponding to highly abundant or conserved proteins as well as essential genes (Figure 4B). This similarity between Sys-NB-06 and GE-PSN is likely because both experimental strategies are highly dependent on endogenous gene expression levels. The second observation is that GI-PSN also shows a higher density of functional relationships amongst extremely well-studied and highly abundant proteins (Figure 4B), albeit restricted to a smaller zone than Sys-NB-06. This was unexpected since this network was generated systematically, independently of any sociological bias. Upon investigation, we found this was due to a combination of higher connection density for both essential genes (Costanzo et al., 2016) and highly abundant non-essential ribosomal subunits in the GI-PSN network (Figure S3G). Many yeast ribosomal proteins retained paralogs after the 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 for genetic interactions to capture biological processes with high functional redundancy. Lastly, CS-PSN was obtained from a homozygous gene deletion collection that does not include essential genes, which could explain some of the observed patterns.

Next we investigated to what extent co-complex and binary biophysical networks identify direct interactions versus indirect associations using available 3D structures of complexes, and thus experimental information of direct contacts between proteins, and subsequently compared this within the three functional networks. As an example, the experimentally solved 3D structure of mediator shows 33 direct contacts between its 25 subunits, which is considerably less than its 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 complex's size, 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 reports roughly equally direct binary interactions and indirect co-complex association between proteins (Figure 4C). Likewise, the genetic interaction profile network GI-PSN finds both direct interactions and indirect associations (Figure 4C), with a preference towards direct PPIs (P = 0.02, two-sided Fisher's exact test), as has been observed previously (Meldal et al., 2021). GI-PSN connecting indirect associations is understandable, as all the proteins in the complex collectively contribute to a common function, irrespective of whether they are in direct contact or not. We then investigated this trend across all different protein complexes for which a 3D structure is available (Figure 4D) and observed that binary PPI datasets primarily find direct-contact pairs, whereas Sys-NB-06, GI-PSN, and GE-PSN connect both direct-contact and indirect co-complex association pairs, with a tendency towards direct PPIs. Among all six datasets analyzed here, ABBI-21 is the most enriched for direct PPIs vs indirect associations (P = 0.0002, two-sided Wicoxon signed-rank test) (Figure 4D).

After observing that within protein complexes, functional networks connect both direct PPIs and indirect associations, we then tested GI-PSN edges using two orthogonal binary PPI assays, Y2H v4 and GPCA, thus providing direct experimental estimates of the fraction of genetic interaction profile similarity relationships that correspond to binary PPIs. In random samples of GI-PSN pairs selected across a range

of PCC cutoffs (Figures 4E, 4F, S3H), the positive test rate increases proportionally to the PCC threshold in GPCA, and an increase is consistent with the Y2H v4 data. On average the positive rate 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 4E, 4F), and consistent with the reported higher average PCC values for direct than indirect (Meldal et al., 2021). Importantly, at the point where the direct binary PPI content substantially exceeds that of protein complexes, at PCC ≥ 0.5, the GI-PSN contains only 841 edges (Figure 4G).

In summary, we identified two alternative views of interactome organization. The first view provided by biophysical co-complex association and functional gene expression similarity data suggests a highly heterogeneous distribution of the interactome in the outer-complexome, while the other view provided by biophysical binary interaction and functional gene interaction similarity data suggests a more homogenous distribution in the outer-complexome.

Inner- Versus Outer-Complexome Interactome Organization

Having identified intrinsic differences and commonalities between biophysical binary and co-complex association networks and functional genetic interaction and gene expression networks, we then wanted to assess their distribution between in the inner- and outer-complexome, and so we compared their “edge density deviation” (Figure 5A). Edge density deviation for a given subspace is calculated as the fold difference between observed interactions or associations in that subspace, relative to what would be expected if the whole proteome-by-proteome space had been covered uniformly. For example, in the inner-complexome, 0.1% of the proteome-by-proteome space, Lit-BM-20’s expected value is 5 PPIs (0.1% of 5,056 total PPIs) and its real value is 1,530 PPIs, so its edge density deviation is 300-fold (Figure 5A). Likewise, Sys-NB-06’s edge density deviation is 330-fold (Figure 5A). Obviously, this high density in the inner-complexome is expected, where co-complex association maps are detecting

complexes, defined by independently curated experiments. Noticeably, ABBI-21's edge density deviation in the inner-complexome is 105-fold, which illustrates yet again that binary approaches can readily identify direct contacts inside protein complexes. In fact, ABBI-21 reports binary interactions in Zone A that represent ~15% of all direct contacts available from structural data, a proportion that is on par with its estimated coverage of the full binary interactome (Figure 3F). Finally, GI-PSN, CS-PSN, and GE-PSN are also enriched in the inner-complexome, although to a lesser extent than the three biophysical networks (Figure 5A), which further illustrates the high level of functional coherence in the inner-complexome.

We then used the same approach to characterize Zone B, C and D in the outer-complexome and observed that (Figure 5B): i) all interactome mapping approaches except for CS-PSN identify a positive edge density deviation in Zone B; ii) Zone D is particularly depleted of interactions or associations in Lit-BM-20, Sys-NB-06, and GE-PSN; iii) Zone D is much closer to uniformly distributed in ABBI-21 and GI-PSN; and finally iv) the approach that uses the largest number of exceptional conditions, CS-PSN is actually slightly enriched in Zone D, after accounting for the untested essential genes (Figure S3I).

To further explore the distribution of these networks relative to both co-complex membership and protein abundance levels, we turned to our complexogram strategy described above (Figure 1A-B). Having ranked all 5,883 proteins of the yeast proteome into 20 bins (Figure 1A), the whole yeast proteome-by-proteome space can now be subdivided into 210 "tiles". For example, the tile "C1/R1" (column 1/row 1), corresponds to pairwise combinations between subunits of the largest complexes – ribosome related and nuclear pore – and C20/R8 are between the most abundant and some of the least abundant, non-complex proteins (Figure 1A).

The first observation that stems from the resulting complexograms is that all three biophysical datasets show a strong, statistically significant ($P < 0.05$, permutation test), enrichment for interactions or associations in tiles C1/R1, C2/R2, C3/R3, C4/R4,

C5/R5, and C6/R6 (Figure 5C, D). We observe the same for GI-PSN and similarly, statistically significant enrichment in C1/R1, C2/R2, C3/R3, C4/R4 for GE-PSN. So again, we see the expected high density of biophysical and functional relationships in the inner-complexome.

The situation is different in the three zones of the outer-complexome. Although Sys-NB-06 associations are significantly enriched in Zone B, which for the most part mirrors what is observed in GE-PSN, in contrast ABBI-21, Lit-BM-20 and GI-PSN are more uniformly distributed in this zone (Figure 5C, D). This result suggests that functionally relevant inter-complex interactions might be widespread, yet relatively less detectable by binary biophysical approaches. In Zone C, Sys-NB-06 is depleted within the subspace corresponding to low abundance proteins (C8-14/R1-7), but enriched within the subspace corresponding to high abundance (C19-20/R1-5). In striking contrast, systematic binary interactions from ABBI-21 are more uniform throughout Zone C, similar to what is observed in Lit-BM-20 (Figure 5C). All three functional networks in the low abundance subzone point to a possible depletion of functional relationships, while GE-PSN reports significantly enriched associations in the high abundance subzone of Zone C (Figure 5D). Thus, the extent to which complex subunits mediate functionally relevant interactions with non-complex proteins remains overall unclear.

The most striking observation can be made in Zone D, particularly in the relatively large subzone corresponding to low abundance (see tiles R8-18/C8-18). This subzone, which in total represents ~30% of the proteome-by-proteome space, is strongly depleted in Sys-NB-06 while almost entirely uniformly distributed in ABBI-21, again on par with what is observed in Lit-BM-20, except for subzones involving extremely low abundance non-complex proteins (C8-14/R8-9) (Figure 5C). Finally, the high abundance subzone corresponding to tiles C19-20/R19-20 is also worth mentioning since it appears significantly enriched in Sys-NB-06 and to a lesser extent in Lit-BM-20, while being more uniformly populated in ABBI-21 (Figure 5C). In the low

abundance subzone of Zone D, GI-PSN associations appear to confirm the uniform distribution observed for ABBI-21 (Figure 5D), and thus should trigger a renewed interest in further investigating this large subspace of the yeast interactome. As shown above, GE-PSN and Sys-NB-06 appear to agree on the possibility that this zone is highly depleted. However, their dependency on endogenous expression might be the main reason for this discrepancy. CS-PSN's depletion in Zone C and enrichment in Zone D is dampened after correcting for the untested essential genes (Figures S3J, K).

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 excluded Lit-BM-20 from this analysis as the reference PPIs sets are also literature derived. We extracted four sets of gold-standard functionally characterized PPIs from different datasets (see STAR Methods #17), 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 shows higher uniformity between inner- and outer-complexome pairs than Sys-NB-06, which is more biased towards capturing inner-complexome pairs (Figure 5E).

To better understand the observed differences between inner- and outer-complexome, we investigated whether ABBI-21 PPIs from either inner- or outer-complexome are of similar biophysical quality by comparing their recovery rates in MAPPIT and GPCA using Lit-BM-13 as a benchmark (Figures 5F, S4A, S4B). While

ABBI-21 validates at a higher rate than Lit-BM-13 in the inner-complexome, PPI pairs from both ABBI-21 and Lit-BM-13 datasets show lower recovery rates in the outer-complexome than in the inner-complexome ($P = 3 \times 10^{-12}$, Fisher's exact test). The fact that both our literature benchmark and ABBI-21 behave similarly in the outer-complexome demonstrates that ABBI-21 pairs in the outer-complexome are of good biophysical quality, suggesting that the difference in recovery rates between inner- and -outer-complexome stems from differing biophysical factors, e.g. interaction affinity or post-translational modification dependency. These observations are consistent with our previous observations that within-complex PPIs are detected more frequently in Y2H screens and PPIs detected in more screens test positive in validation assays at higher rates, independent of data quality (Luck et al., 2020). The striking observation of inner-complexome PPIs being more readily detected by different PPI assays suggests that inner-complexome PPIs tend to be overrepresented in interactome maps relative to their proportion in the real interactome.

In summary, we observe a high discrepancy among alternative views of inner- versus outer-complexome organization obtained from co-complex association and coexpression as opposed to binary interaction maps and genetic interaction profiles (Figure 5G). These results are most likely explained by a technical bias of AP-MS and co-expression towards highly abundant proteins and relatively stable associations. Compared to Sys-NB-06 or Lit-BM-20, ABBI-21 is the most uniform biophysical map in the outer-complexome, for proteins of both high and low abundance.

A Large Proportion of the Interactome Consists of Interactions Between Functionally Heterogeneous Proteins in the Outer-Complexome

To further investigate global relationships between biophysical and functional interactome networks, we computed the fraction of protein pairs from Lit-BM-20, Sys-NB-06, and ABBI-21, where the corresponding gene pairs are also connected in the functional networks (Figures 6A, STAR Methods #30). In the global interactome, pairs from Lit-BM-20 and Sys-NB-06 are more likely to connect proteins that are also

connected in PSNs, compared to ABBI-21 (Figure 6B). We observed similar patterns for all three biophysical datasets when directly using the positive and negative genetic interactions instead of the profile similarities (Figure S5A). In general, pairs of 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 (Figures 2B,C and 3A), and functional networks are enriched in highly abundant, conserved inner-complexome direct and indirect pairs (Figure 4). Therefore, the differential likelihood of pairs from different biophysical maps to be connected in the three PSNs suggests varying functional relationships between interacting proteins in the three biophysical maps rather than differences in the fraction of true positive biophysical interactions.

Investigating the difference in the fraction of PPIs connected in the functional networks, we note that ABBI-21 and Lit-BM-20 are not directly comparable as the four individual maps in ABBI-21 were produced by systematically testing proteome-scale-by-proteome-scale search spaces. In contrast, 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, contributes the largest number of PPIs that intersect with the functional networks. It reports 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 which protein pairs to test rather than the specific interactions detected. In contrast,

ABBI-21 is free of this bias, since it was generated by standardized testing of each protein pair, and so the rate of overlap of ABBI-21 with functional networks should more accurately reflect the rate for the real binary protein interactome.

We investigated the functional relationships between proteins interacting in the inner- and outer-complexome by comparing the ability of PSNs to connect pairs from the inner- and outer-complexome within the three biophysical maps (Figure 6C). 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 6C). By contrast, interactions from all three zones of the outer-complexome, B, C, and D, in all three biophysical maps are found to a much lower extent in the functional networks (Figure 6D). This demonstrates a higher tendency for functional networks to preferentially connect constitutive, co-complex interactions rather than the interactions of the outer-complexome. These three biophysical datasets differ substantially in their proportion of inner-complexome pairs. More than one-quarter of Lit-BM-20 interactions and Sys-NB-06 associations are between proteins that are subunits of the same complex (Zone A) compared to around one-tenth for ABBI-21 interactions (Figure 6C). This difference contributes to the lower aggregate fraction of ABBI-21 PPIs connected in functional networks, relative to Lit-BM-20 and Sys-NB-06 (Figure 6B).

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). In relation to the distinction between inner- and outer-complexome, party hubs tend to be in complexes, whereas date hubs tend to be outside complexes (Kim et al., 2006). We observed a similar trend for party and date hubs as for the inner- and outer-complexome. Systematic binary maps have mostly date-hubs, whereas literature-curated maps have more party-hubs (Figure S5B, STAR Methods #31), 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). To check if our results were robust, we repeated our analysis using a range of different cutoffs, to define date and party hubs, and observed consistent results across all the different cutoffs (Figures S5B, S5C). Our results suggest that while the literature-curated PPIs and co-complex associations are biased towards high-affinity constitutive interactions, a systematically generated map of the binary interactome captures more 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 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. The fraction of PPIs connected in the functional networks decreases as the interacting proteins' degree increases across the different biophysical networks (Figure S5E, STAR Methods #32). With the notable exception of Sys-NB-06 in GE-PSN due to the tightly correlated expression of the largest protein complexes' subunits. We would expect high-degree proteins to be less functionally similar to their binding partners, as they are more pleiotropic (Yu et al., 2008). Together these results indicate that systematic binary maps have a higher tendency to identify transient, context-specific interactions.

Because functional networks show a substantially higher tendency to connect protein pairs in the inner-complexome than the outer-complexome (Figure 6C), they likely 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 (Figure 6E). 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 MAPK and Autophagy (Figure 6E).

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 5E. 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 6F). 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 6C) that the inner-complexome tends to consist of functionally similar interacting proteins. In contrast, the outer-complexome (in which the majority of systematic binary maps are located) tends to consist of interactions between functionally heterogeneous proteins necessary for intracellular crosstalk. This conclusion is supported by the high level of pathway cross-talk and pleiotropy revealed by a binary map of the plant signaling network and interactome-informed phenotypic assays (Altmann et al., 2020). We should be clear that when we refer to PPIs between "functionally heterogeneous" proteins, that it's still likely that these proteins carry out a

common function through their interaction but just that the overall function of the two proteins is not so similar. As an example, in a PPI of the nuclear pore component importin-α and one of its cargo proteins, being transported to the nucleus is likely to be crucial for the function of the cargo but at the aggregate level the overall function of two genes is different.

Network diagrams of the biophysical maps, split into the four Zones (Figure 7A) show modules formed by the larger protein complexes in Sys-NB-06, due to the high recovery of both direct and indirectly contacting co-complex associations. Whereas the binary PPI maps don't display such clear modules, related to them finding mainly direct and not indirectly contacting interactions. In Zone D, we see a denser network in Sys-NB-06 and ABBI-21 than in Lit-BM-20.

Integrating Biophysical and Genetic Network Maps Aids Understanding Cellular Organization

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 five ESCRT complexes' core constituents interact biophysically in both ABBI-21 and Lit-BM-20 and are highly interconnected in functional networks (Figure 7B). 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 binary maps can help us understand how proteins within and outside complexes function together to mediate various biological processes. One such example is 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 and other non-complex endosomal proteins like Nkp2 (Figure 7B). 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). 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). This complex's six components are connected biophysically and in the functional networks, exhibiting similar genetic interaction and condition sensitivity profiles (Figure 7C). Although inner-complexome interactions with OCA may well be critical for its function, they do not explain the complex's stress-specificity. Outer-complexome interactions of OCA proteins do not 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 7C). 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.

Altogether, 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 the inner- and outer-complexome.

DISCUSSION

The observations presented in this paper suggest fundamental differences of organization between the inner-complexome, containing mostly quaternary structures that are highly detectable by affinity purification approaches, and the outer-complexome, which has a greater tendency towards quinary structures, "largely destroyed by cell fractionation" (McConkey, 1982), but detectable by binary assays in living cells that reconstitute functional proteins such as 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.

A concept related to the quaternary vs quinary divide comes from Jacob and Monod, who noted that: as important as lactose permease and β-galactosidase (β-gal) 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). From this example emerged 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*, encoding regulators that control and coordinate these biochemical activities.

Our results establish that the biophysical and functional dichotomy between "structural" and "regulatory" genes is 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 constitutive housekeeping functions or pleiotropic regulatory processes is a property of its interactions. For example, Rpl8A, a component of the large ribosomal subunit, 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.

Thus, the inner-complexome 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 the 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 that 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 remain 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). A.Y. was supported by a Deborah F. Allinger Fellowship awarded to CCSB.

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. Sequencing to identify interacting proteins was carried out by A.G.C., M.G., N.K., J.J.K., and J.C.M. 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., 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 co-authors. 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.

COMPETING INTERESTS

J.C.M. is a founder and CEO of seqWell, Inc; F.P.R. and M.V. are shareholders and scientific advisors of seqWell, Inc.

FIGURE LEGENDS

Figure 1. The Complexogram Strategy

(A) The yeast proteome ordered by the number of subunits, decreasing, of the corresponding complex, for proteins in a complex, and by abundance, increasing, for non-complex proteins. After dividing into 20 bins, the number of proteins in each of three general functional categories (Liebermeister et al., 2014), is shown by the area of the circles. Not Mapped corresponds to unannotated proteins.

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

Figure 2. Current Status of the Yeast Binary Interactome Landscape

(A) Curation of the three binary PPI datasets.

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

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

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

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

(E) 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 space containing 80% of PPIs in the dataset.

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

In all panels homodimeric interactions are excluded.

Figure 3. 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, ordered by the number of publications per protein, from all four systematic binary interactome maps.

(D) Increase in the number of high-quality binary PPIs from available experiments in Interactome-3D (yellow), Lit-BM (blue), and systematic studies (purple). Note: our literature and structure datasets only go as far as 2020.

(E) 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).

(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 4. High Discrepancy Among Alternative Views of Interactome Organization

(A) Schematic of the different functional profile similarity networks.

(B) Heatmaps of protein interactions and associations in biophysical and functional networks, ordered by the 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.

(C) Using experimental structures of mediator, the direct interactions and indirect associations that are contained within the different biophysical and functional networks. Direct enrichment is defined as the ratio of the fraction of detected direct interactions over the fraction of detected indirect associations.

(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).

(G) The number of edges in GI-PSN at the tested PCC cutoffs.

Figure 5. Inner- Versus Outer-Complexome Organization

(A) Edge density fold change in the inner- and outer-complexome in the biophysical and functional maps.

(B) Edge density fold change in the Zones B, C and D of the outer-complexome in the biophysical and functional maps.

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

(D) Statistically significant enrichment and depletion of edges per bin in (C), calculated by random permutation of the order of the proteins.

(E) 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.

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

(G) Schematic summarizing the two patterns of distribution from (D).

Figure 6. A Large Proportion of the Interactome Consists of Interactions Between Functionally Heterogeneous Proteins in the Outer-Complexome

(A) Schematic showing pairs from biophysical maps also connected in functional networks.

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

(C) The fraction of pairs in the inner- and outer-complexome in the three biophysical maps connected in functional networks. The pie charts show the proportion of pairs in the inner- and outer-complexome within the three biophysical maps.

(D) The fraction of pairs from the outer-complexome, Zones B, C, and D, also connected in functional networks.

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

(F) The fraction of PPIs from the four gold-standard biophysical interaction datasets (see Figure 5E) connected in functional networks.

Figure 7. Integrating Biophysical and Genetic Network Maps Aids Understanding Cellular Organization

(A) Three different biophysical networks, separately showing only PPIs/co-complex associations in each zone.

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

(C) 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 2

(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 contacting 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.

(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 3

(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 Interactome Organization, Related to Figures 4 and 5

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

(B) 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.

(C-D) 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 Costanzo et al 2016 (D) the and EBI complex portal (D). Error bars denote 95% CI.

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

(F) 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.

(G) 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.

(H) 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.

(I-K) Distribution of CS-PSN after restricting to genes tested in the experiment: (I) edge density fold change in the outer-complexome (Zones B, C and D), (J) heatmap of the number of connected gene pairs with the proteome first ordered by size of protein complexes in which the proteins are involved then ordered by protein abundance, and, (K) statistically significant enrichment and depletion of pairs in (J).

Figure S4. , Related to Figure 5

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

Figure S5. A Large Proportion of the Interactome Mainly Consists of Interactions Between Functionally Heterogeneous Proteins in the Outer-Complexome, Related to Figure 6

(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, and the fraction of pairs from biophysical maps of each category connected in functional networks.

(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

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.

Altmann, M., Altmann, S., Rodriguez, P.A., Weller, B., Elorduy Vergara, L., Palme, J., Marín-de la Rosa, N., Sauer, M., Wenig, M., Villaécija-Aguilar, J.A., et al. (2020). Extensive signal integration by the phytohormone protein network. Nature *583*, 271–276.

Baryshnikova, A. (2016). Systematic functional annotation and visualization of biological networks. Cell Syst *2*, 412–421.

Bludau, I., and Aebersold, R. (2020). Proteomic and interactomic insights into the molecular basis of cell functional diversity. Nat. Rev. Mol. Cell Biol. *21*, 327–340.

Cafarelli, T.M., Desbuleux, A., Wang, Y., Choi, S.G., De Ridder, D., and Vidal, M. (2017). Mapping, modeling, and characterization of protein-protein interactions on a proteomic scale. Curr. Opin. Struct. Biol. *44*, 201–210.

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, 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.

Cohen, R.D., and Pielak, G.J. (2017). A cell is more than the sum of its (dilute) parts: A brief history of quinary structure. Protein Sci. *26*, 403–413.

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.

Davey, N.E. (2019). The functional importance of structure in unstructured protein regions. Curr. Opin. Struct. Biol. *56*, 155–163.

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.

Edelstein, S.J. (1980). Patterns in the quinary structures of proteins. Plasticity and inequivalence of individual molecules in helical arrays of sickle cell hemoglobin and tubulin. Biophys. J. *32*, 347–360.

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.

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.

Guin, D., and Gruebele, M. (2019). Weak Chemical Interactions That Drive Protein Evolution: Crowding, Sticking, and Quinary Structure in Folding and Function. Chem. Rev. *119*, 10691–10717.

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.

Havugimana, P.C., Hu, P., and Emili, A. (2017). Protein complexes, big data, machine learning

and integrative proteomics: lessons learned over a decade of systematic analysis of protein interaction networks. Expert Rev. Proteomics *14*, 845–855.

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.

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.

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.

Jansen, R., Yu, H., Greenbaum, D., Kluger, Y., Krogan, N.J., Chung, S., Emili, A., Snyder, M., Greenblatt, J.F., and Gerstein, M. (2003). A Bayesian networks approach for predicting protein-protein interactions from genomic data. Science *302*, 449–453.

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.

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.

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., Sheynkman, G.M., Zhang, I., and Vidal, M. (2017). Proteome-Scale Human Interactomics. Trends Biochem. Sci. *42*, 342–354.

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.

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. (2021). 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.

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.

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.

Ruepp, A., Waegele, B., Lechner, M., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C., and Mewes, H.-W. (2010). CORUM: the comprehensive resource of mammalian protein complexes--2009. Nucleic Acids Res. *38*, D497–D501.

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.

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.

Vaĭnshteĭn, B.K. (1973). Three-dimensional electron microscopy of biological macromolecules. Usp. Fiziol. Nauk *109*, 455–497.

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. (2001). A biological atlas of functional maps. Cell *104*, 333–339.

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.

Wang, M., Herrmann, C.J., Simonovic, M., Szklarczyk, D., and von Mering, C. (2015). Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell-lines. Proteomics *15*, 3163–3168.

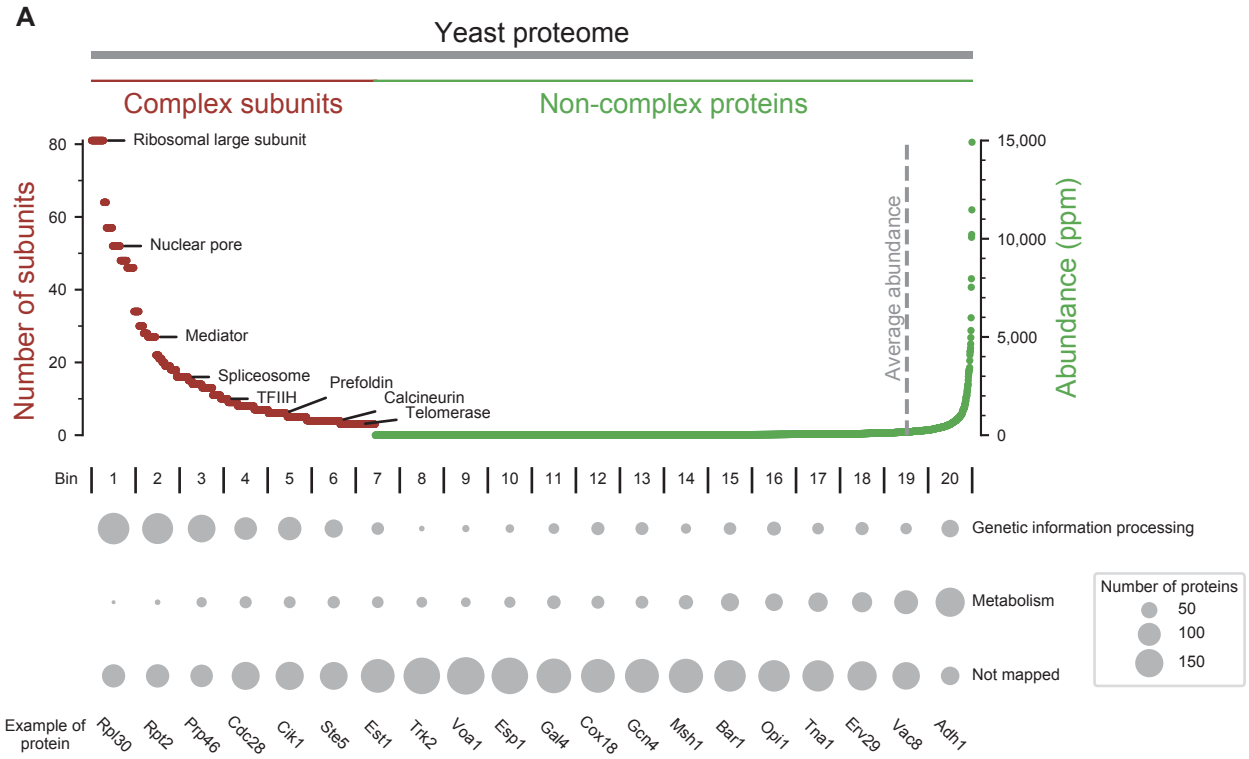
Wolfe, K.H., and Shields, D.C. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. Nature *387*, 708–713.

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.

Zhang, Q.C., Petrey, D., Garzón, J.I., Deng, L., and Honig, B. (2013). PrePPI: a structure-informed database of protein-protein interactions. Nucleic Acids Res. *41*, D828–D833.


A
Yeast proteome
Complex subunits
Non-complex proteins
Number of subunits
Ribosomal large subunit
Nuclear pore
Mediator
Spliceosome
TFIIH
Prefoldin
Calcineurin
Telomerase
Average abundance
Abundance (ppm)
Bin
Genetic information processing
Metabolism
Not mapped
Number of proteins
50
100
150
Example of protein
Rpl30
Rpt2
Prp46
Cdc28
Cik1
Ste5
Est1
Trk2
Voa1
Esp1
Gal4
Cox18
Gcn4
Msh1
Bar1
Opi1
Tna1
Erv29
Vac8
Adh1

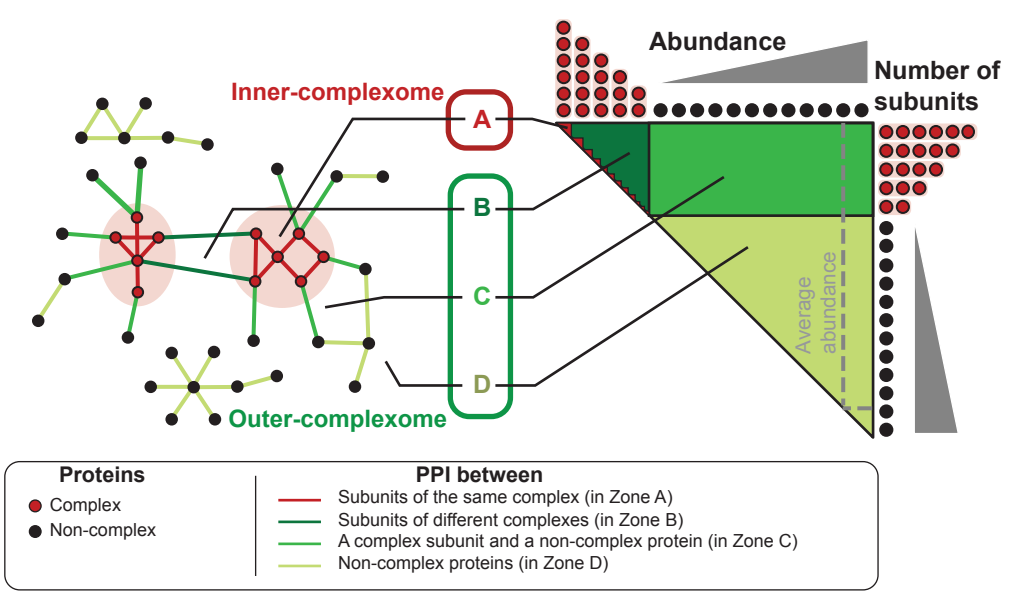

B
Abundance
Number of subunits
Inner-complexome
Outer-complexome
A
B
C
D
Average abundance
Proteins
Complex
Non-complex
PPI between
Subunits of the same complex (in Zone A)
Subunits of different complexes (in Zone B)
A complex subunit and a non-complex protein (in Zone C)
Non-complex proteins (in Zone D)

Figure 1

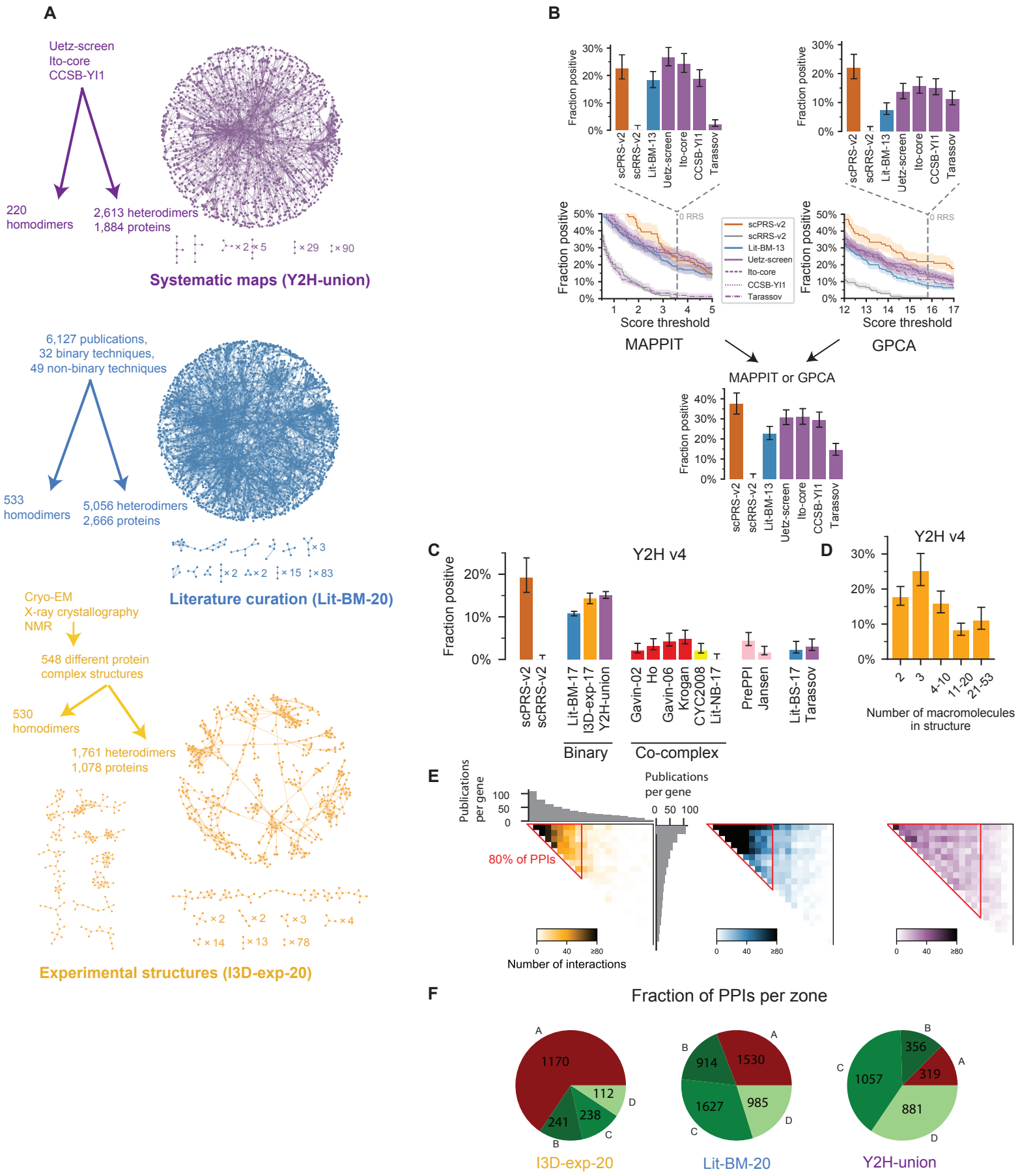

A
Uetz-screen
Ito-core
CCSB-YI1
220 homodimers
2,613 heterodimers
1,884 proteins
Systematic maps (Y2H-union)
6,127 publications,
32 binary techniques,
49 non-binary techniques
533 homodimers
5,056 heterodimers
2,666 proteins
Literature curation (Lit-BM-20)
Cryo-EM
X-ray crystallography
NMR
548 different protein complex structures
530 homodimers
1,761 heterodimers
1,078 proteins
Experimental structures (I3D-exp-20)
B
Fraction positive
scPRS-v2
scRRS-v2
Lit-BM-13
Uetz-screen
Ito-core
CCSB-YI1
Tarassov
Score threshold
0 RRS
MAPPIT
GPCA
MAPPIT or GPCA
C
Y2H v4
Binary
Co-complex
Lit-BM-17
I3D-exp-17
Y2H-union
Gavin-02
Ho
Gavin-06
Krogan
CYC2008
Lit-NB-17
PrePPI
Jansen
Lit-BS-17
D
Number of macromolecules in structure
E
Publications per gene
80% of PPIs
Number of interactions
F
Fraction of PPIs per zone
I3D-exp-20
Lit-BM-20
Y2H-union

Figure 2

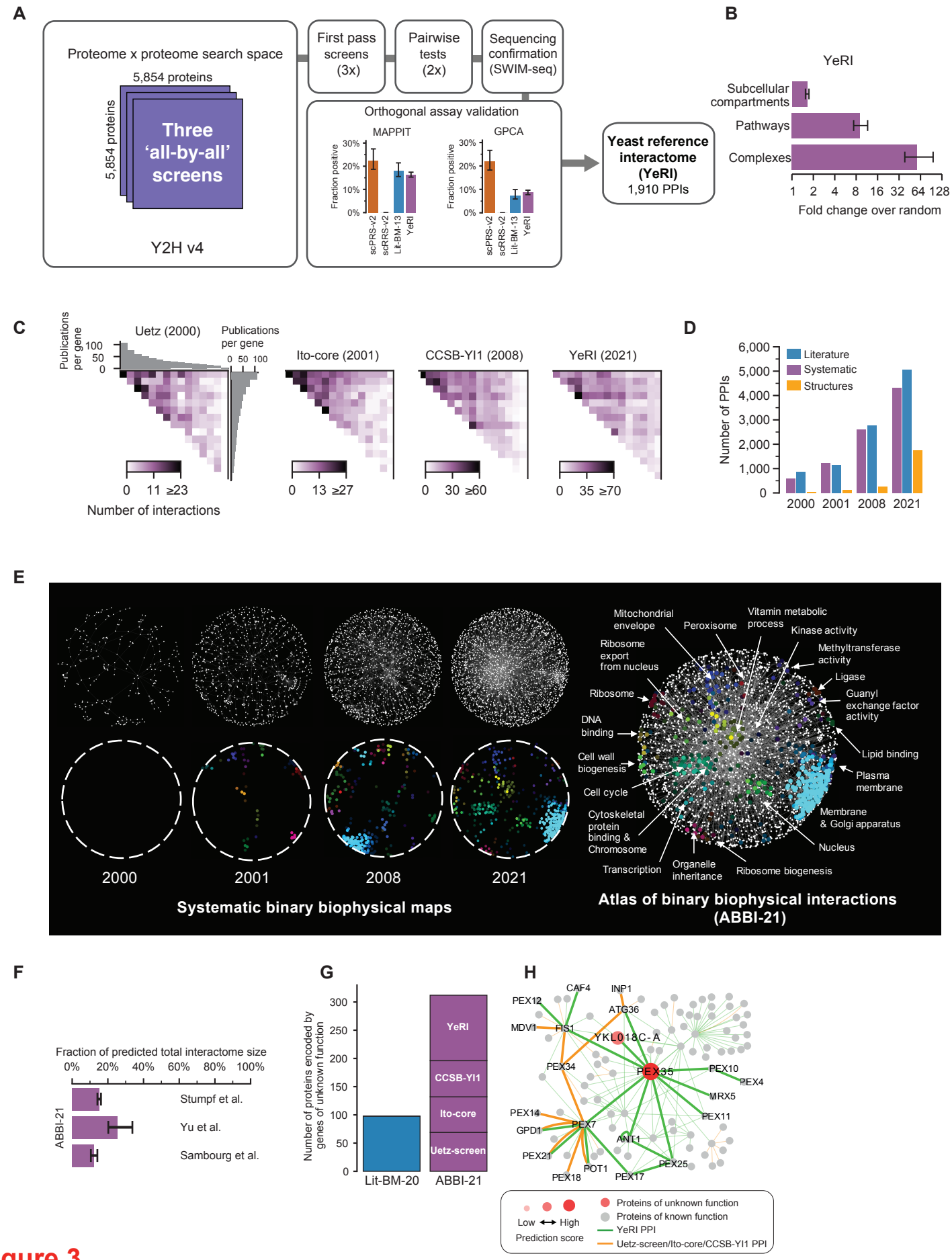

A
Proteome x proteome search space
5,854 proteins
5,854 proteins
Three 'all-by-all' screens
Y2H v4
First pass screens (3x)
Pairwise tests (2x)
Sequencing confirmation (SWIM-seq)
Orthogonal assay validation
MAPPIT
GPCA
Fraction positive
30%
20%
10%
0%
scPRS-v2
scRRS-v2
Lit-BM-13
YeRI
Yeast reference interactome (YeRI)
1,910 PPIs
B
YeRI
Subcellular compartments
Pathways
Complexes
1 2 4 8 16 32 64 128
Fold change over random
C
Publications per gene
Uetz (2000)
Publications per gene
100
50
0
0 50 100
Ito-core (2001)
CCSB-YI1 (2008)
YeRI (2021)
0 11 ≥23
0 13 ≥27
0 30 ≥60
0 35 ≥70
Number of interactions
D
Number of PPIs
6,000
5,000
4,000
3,000
2,000
1,000
0
Literature
Systematic
Structures
2000 2001 2008 2021
E
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)
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
PEX25
POT1
PEX18
PEX17
Low High
Prediction score
Proteins of unknown function
Proteins of known function
YeRI PPI
Uetz-screen/Ito-core/CCSB-YI1 PPI

Figure 3

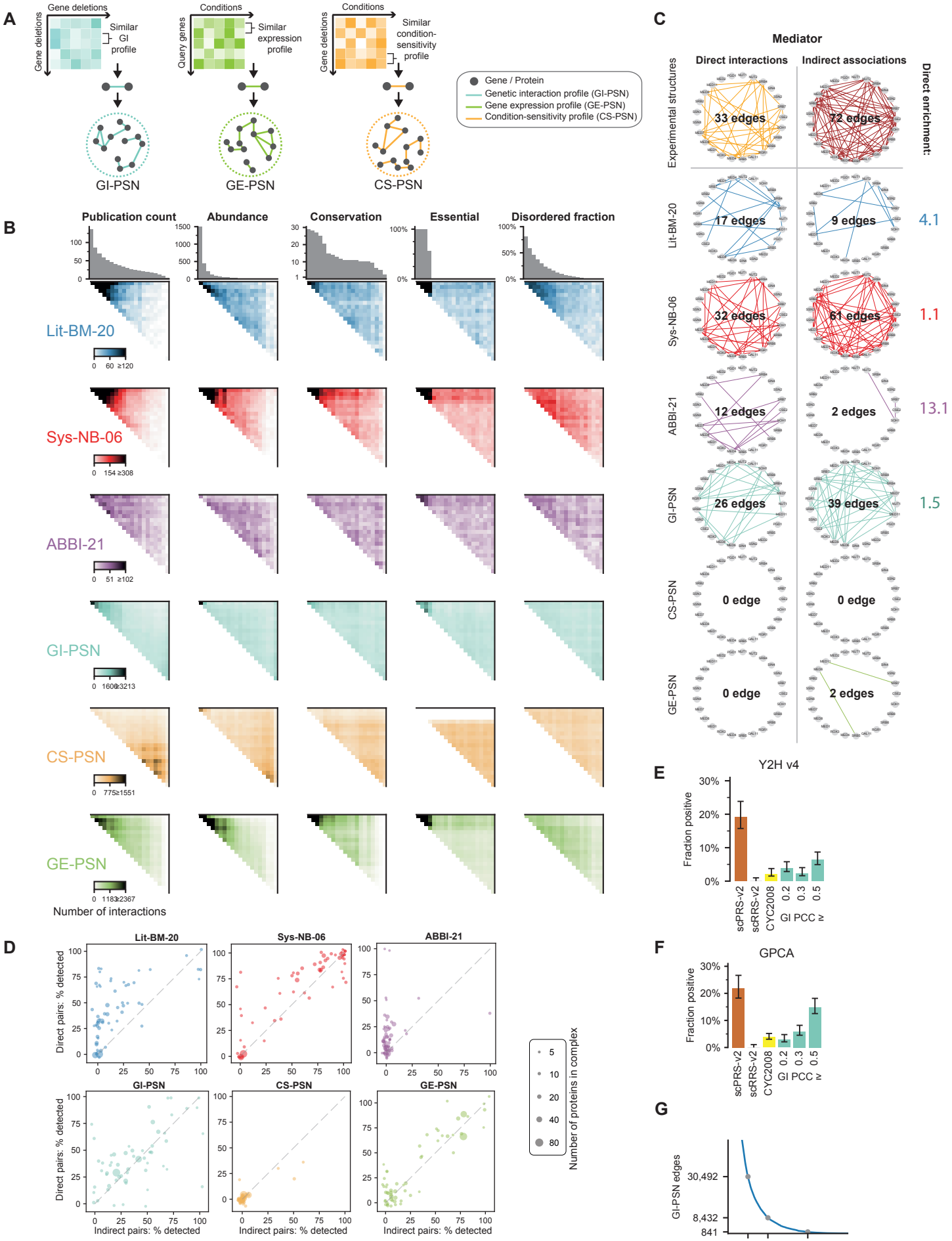

A
Gene deletions
Gene deletions
Similar GI profile
Conditions
Query genes
Similar expression profile
Conditions
Gene deletions
Similar condition-sensitivity profile
GI-PSN
GE-PSN
CS-PSN
Gene / Protein
Genetic interaction profile (GI-PSN)
Gene expression profile (GE-PSN)
Condition-sensitivity profile (CS-PSN)
B
Publication count
Abundance
Conservation
Essential
Disordered fraction
Lit-BM-20
Sys-NB-06
ABBI-21
GI-PSN
CS-PSN
GE-PSN
Number of interactions
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
G
GI-PSN edges
30,492
8,432
841
GI PCC threshold

Figure 4

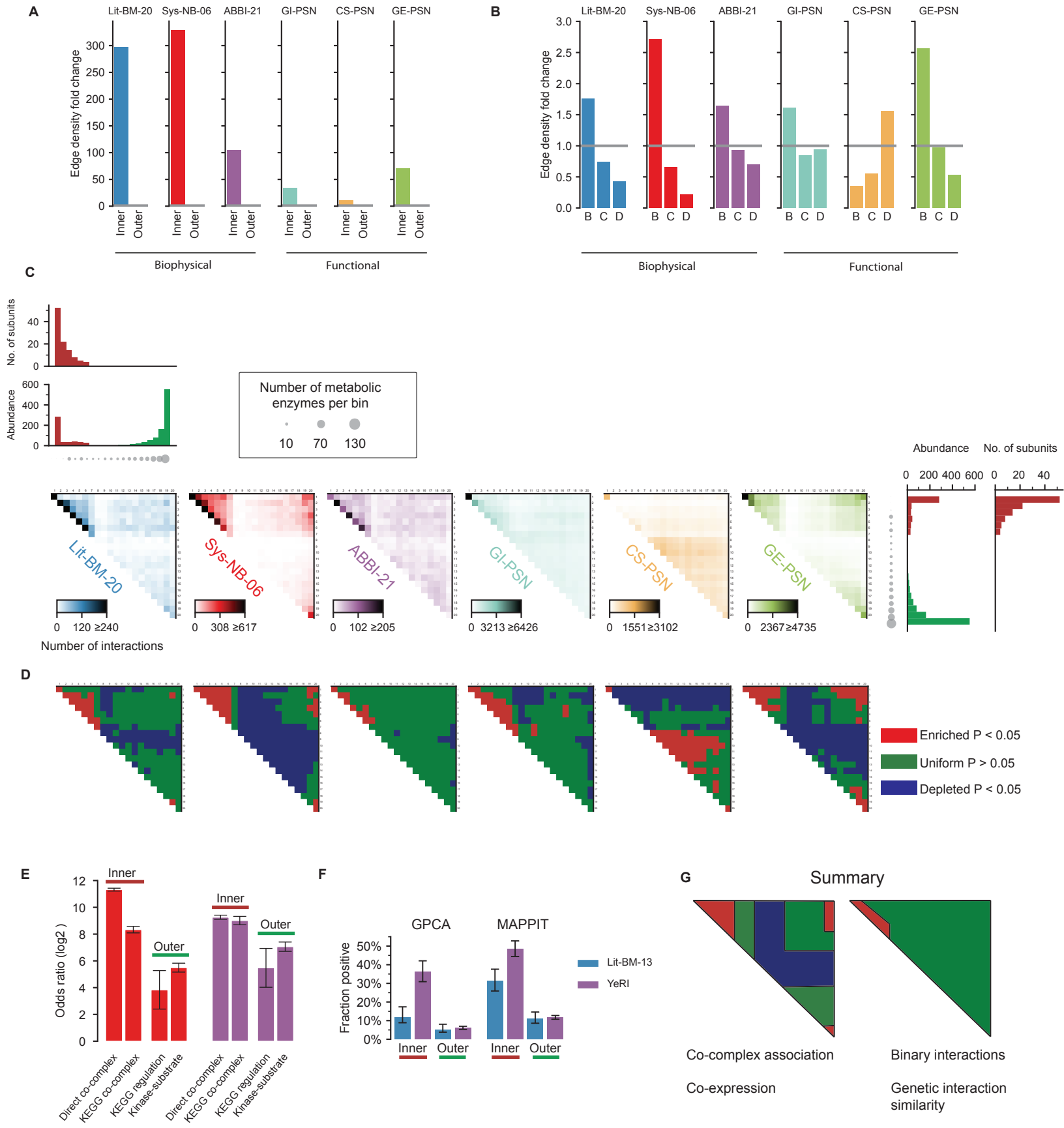

A
Lit-BM-20
Sys-NB-06
ABBI-21
GI-PSN
CS-PSN
GE-PSN
Edge density fold change
Inner
Outer
Biophysical
Functional
B
B C D
C
No. of subunits
Abundance
Number of metabolic enzymes per bin
10 70 130
Number of interactions
0 120 ≥240
0 308 ≥617
0 102 ≥205
0 3213 ≥6426
0 1551 ≥3102
0 2367 ≥4735
D
Enriched P < 0.05
Uniform P > 0.05
Depleted P < 0.05
E
Odds ratio (log2)
Direct co-complex
KEGG co-complex
KEGG regulation
Kinase-substrate
F
GPCA
MAPPIT
Fraction positive
Lit-BM-13
YeRI
G
Summary
Co-complex association
Co-expression
Binary interactions
Genetic interaction similarity

Figure 5

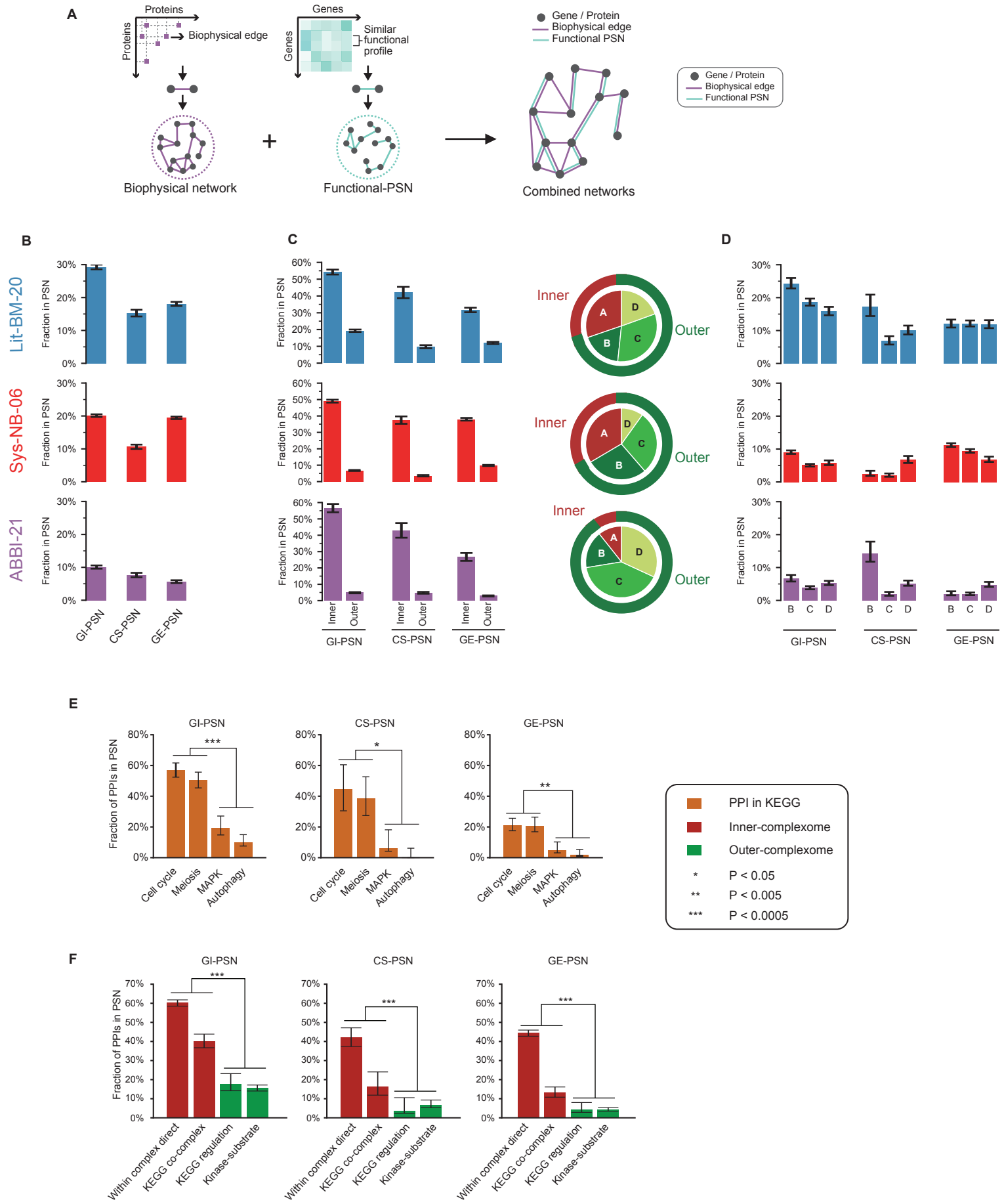

A
Proteins
Proteins
Biophysical edge
Genes
Genes
Similar functional profile
Gene / Protein
Biophysical edge
Functional PSN
Biophysical network
Functional-PSN
Combined networks
B
C
D
Lit-BM-20
Sys-NB-06
ABBI-21
Fraction in PSN
Inner
Outer
GI-PSN
CS-PSN
GE-PSN
E
Fraction of PPIs in PSN
Cell cycle
Meiosis
MAPK
Autophagy
PPI in KEGG
Inner-complexome
Outer-complexome
* P < 0.05
** P < 0.005
*** P < 0.0005
F
Within complex direct
KEGG co-complex
KEGG regulation
Kinase-substrate

Figure 6

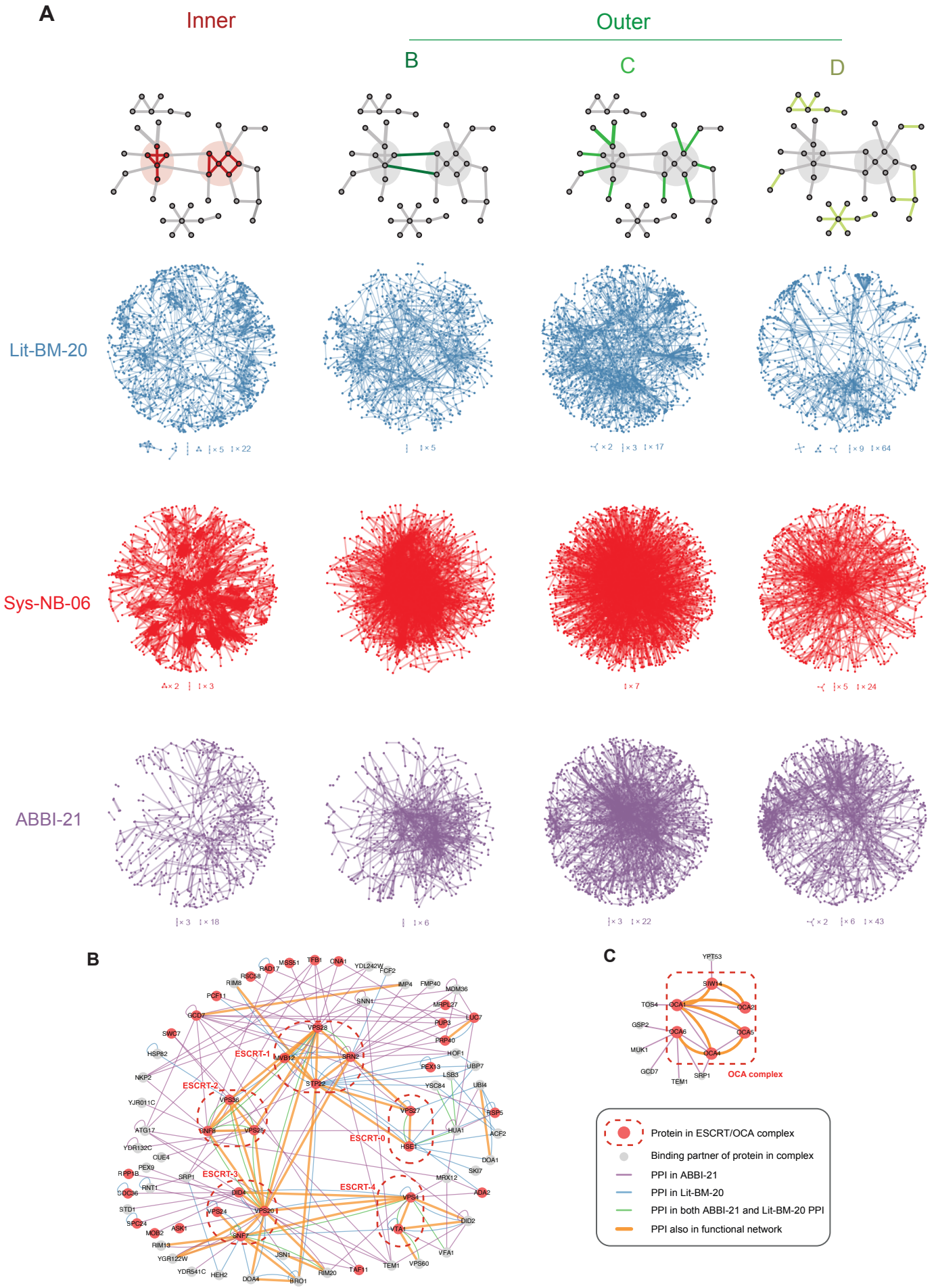

A
Inner
Outer
B
C
D
Lit-BM-20
Sys-NB-06
ABBI-21
B
ESCRT-0
ESCRT-1
ESCRT-2
ESCRT-3
ESCRT-4
C
OCA complex
Protein in ESCRT/OCA complex
Binding partner of protein in complex
PPI in ABBI-21
PPI in Lit-BM-20
PPI in both ABBI-21 and Lit-BM-20 PPI
PPI also in functional network

Figure 7

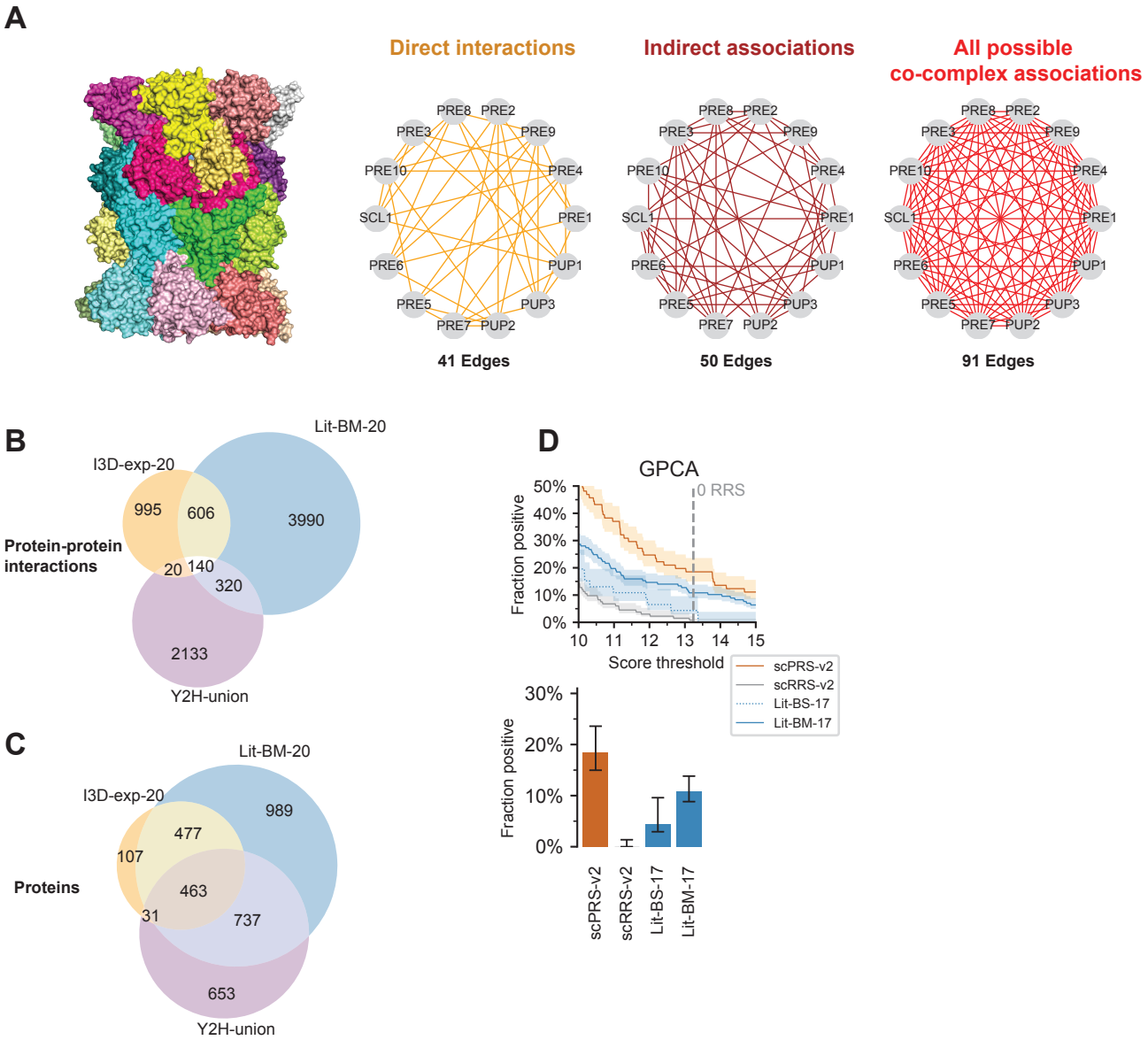
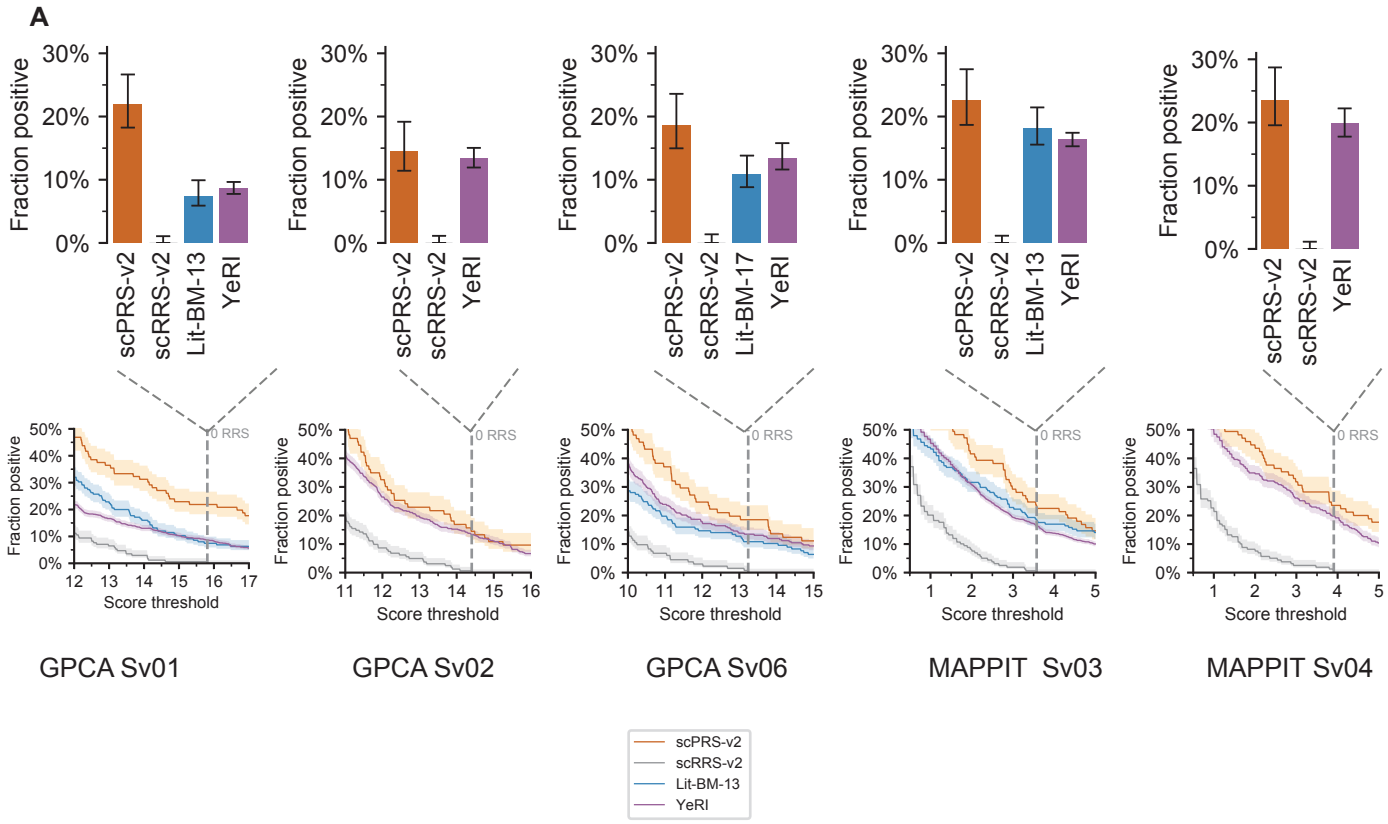
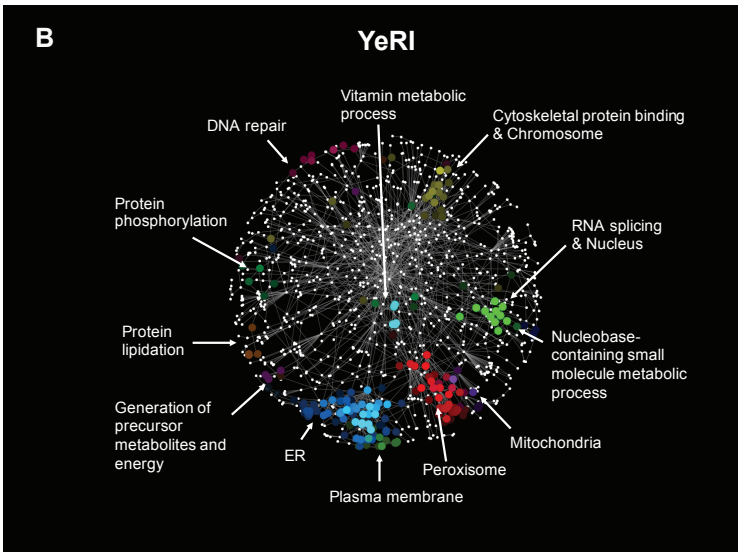

A
Direct interactions
Indirect associations
All possible co-complex associations
PRE8 PRE2 PRE3 PRE9 PRE10 PRE4 SCL1 PRE1 PRE6 PUP1 PRE5 PUP3 PRE7 PUP2
41 Edges
50 Edges
91 Edges
B
Lit-BM-20
I3D-exp-20
995
606
3990
Protein-protein interactions
20
140
320
2133
Y2H-union
C
Lit-BM-20
I3D-exp-20
989
477
107
Proteins
463
31
737
653
Y2H-union
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

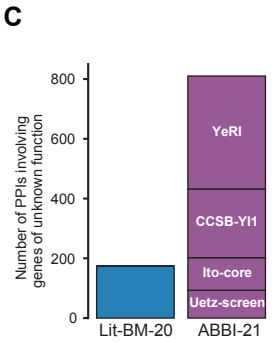

C
Number of PPIs involving genes of unknown function
800
600
400
200
0
YeRI
CCSB-YI1
Ito-core
Uetz-screen
Lit-BM-20
ABBI-21

Figure S2

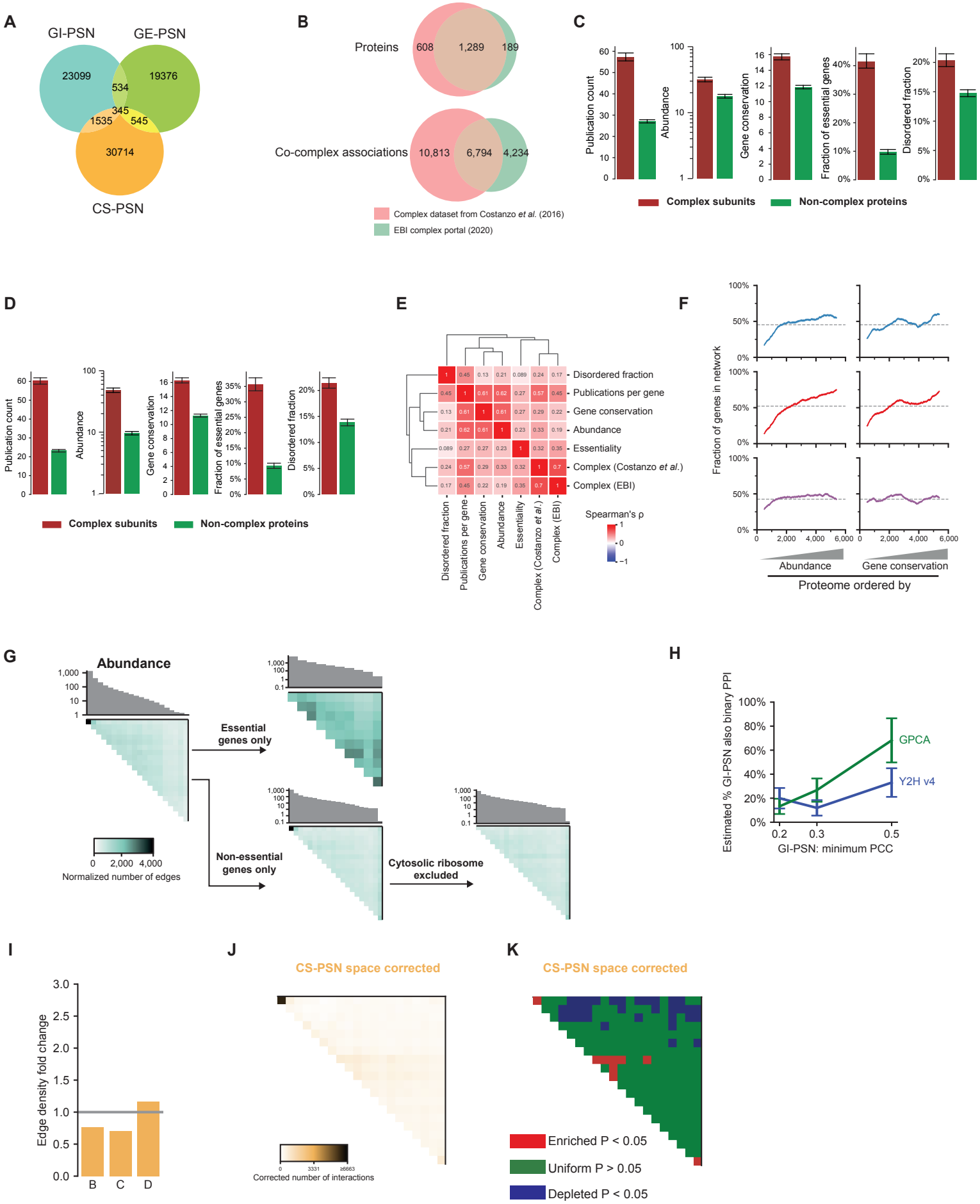

A
GI-PSN
GE-PSN
CS-PSN
23099
19376
30714
534
345
1535
545
B
Proteins
608
1,289
189
Co-complex associations
10,813
6,794
4,234
Complex dataset from Costanzo et al. (2016)
EBI complex portal (2020)
C
Publication count
Abundance
Gene conservation
Fraction of essential genes
Disordered fraction
Complex subunits
Non-complex proteins
D
Publication count
Abundance
Gene conservation
Fraction of essential genes
Disordered fraction
Complex subunits
Non-complex proteins
E
Disordered fraction
Publications per gene
Gene conservation
Abundance
Essentiality
Complex (Costanzo et al.)
Complex (EBI)
Spearman's ρ
F
Fraction of genes in network
Abundance
Gene conservation
Proteome ordered by
G
Abundance
Normalized number of edges
Essential genes only
Non-essential genes only
Cytosolic ribosome excluded
H
Estimated % GI-PSN also binary PPI
GPCA
Y2H v4
GI-PSN: minimum PCC
I
Edge density fold change
B
C
D
J
CS-PSN space corrected
Corrected number of interactions
K
CS-PSN space corrected
Enriched P < 0.05
Uniform P > 0.05
Depleted P < 0.05

Figure S3

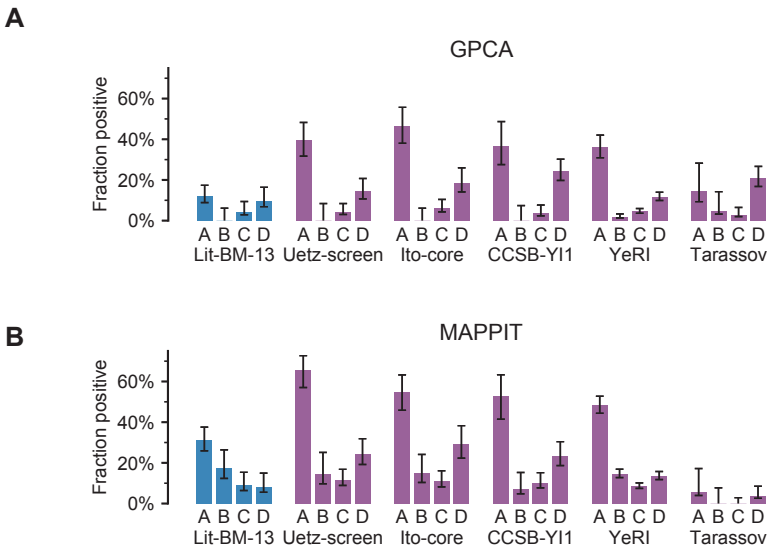

A
GPCA
Fraction positive
60%
40%
20%
0%
A B C D
Lit-BM-13
Uetz-screen
Ito-core
CCSB-YI1
YeRI
Tarassov
B
MAPPIT
Fraction positive
60%
40%
20%
0%
A B C D
Lit-BM-13
Uetz-screen
Ito-core
CCSB-YI1
YeRI
Tarassov

Figure S4

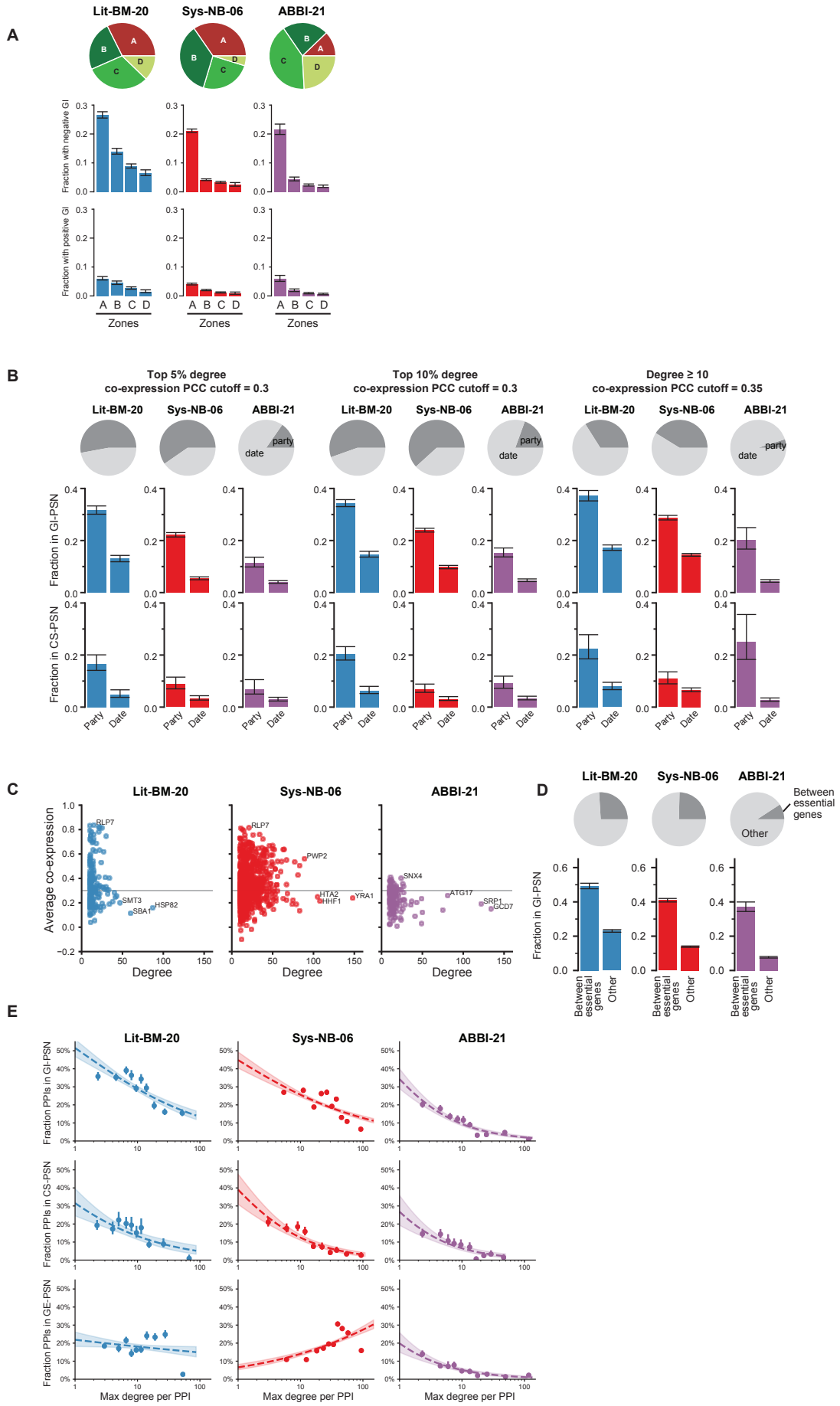

A
Lit-BM-20
Sys-NB-06
ABBI-21
Fraction with negative GI
Fraction with positive GI
Zones
B
Top 5% degree
co-expression PCC cutoff = 0.3
Top 10% degree
co-expression PCC cutoff = 0.3
Degree ≥ 10
co-expression PCC cutoff = 0.35
party
date
Fraction in GI-PSN
Fraction in CS-PSN
Party
Date
C
Average co-expression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

Figure S5

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 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 (Dreze et al., 2010; Yu et al., 2008). Both strains 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 YEPD or synthetic drop out media, supplemented as needed and incubated at 30°C.

Bacterial strains

Chemically competent DH5α or DB3.1 *E. coli* cells were used for all bacterial transformations in this study. Transformed cells were cultured in Luria Broth or Terrific Broth, supplemented with antibiotics (50 µg/ml of ampicillin, spectinomycin or kanamycin) as needed and incubated at 37°C.

Human cell lines

Human embryonic kidney HEK293T cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) 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

#1 Yeast open reading frames

The list of yeast ORFs 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. All datasets analyzed have been restricted to these 5,883 ORFs and previous ORF names that appear as aliases for one of these ORFs have been mapped to their corresponding 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 three or more different protein subunits, resulting in 339 complexes containing 1,897 different proteins. In some analyses, results were compared with an alternative yeast protein complex dataset was downloaded from the EBI Complex Portal (Meldal et al., 2019) on May 5th 2020, and also filtered for those containing three or more different protein subunits.

#3 Assigning protein pairwise combinations to individual 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, which we refer to as the inner-complexome, as all pairwise combinations of proteins within protein complexes. 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 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 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 yeast ribosome, which accounts for nearly 20% of the proteomic mass (Liebermeister et al., 2014), is encoded by only 2% of all genes and all combinations between ribosomal proteins correspond to ~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.

#4 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 PPIs from Uetz et al ((Ito et al., 2001; Uetz et al., 2000) that was obtained from a proteome-scale systematic Y2H screen, excluding a smaller-scale, relatively biased, targeted experiment with a smaller number of well-studied bait proteins. 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 or twice. 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 as follows: Uetz-screen: 645 PPIs; Ito-core: 816 PPIs; CCSB-YI: 1,772 PPIs. The union of these maps (Y2H-union) contains 1.933 nodes and 2,833 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 (containing data from IntAct, MINT and DIP) and BioGRID release 3.5.187 (from 2020-06-25). We excluded evidence corresponding to the eight systematic, proteome-scale co-complex association 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). Data was filtered to ensure valid IDs for UniProt accession numbers, Pubmed IDs and PSI-MI terms. Each piece of evidence for a protein pair had to 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 can arise 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

identical 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 protein pairs corresponding to binary interactions (as opposed to co-complex associations), we developed a manual classification of the PSI-MI interaction detection method terms (Rolland et al., 2014). Our classification has since been updated 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 'non-binary'. Where 'invalid' corresponds to PSI-MI terms that are not considered valid experimental protein-protein interaction detection methods, 'binary' corresponds to terms that detect binary protein-protein interactions and 'non-binary' corresponds to terms that detect potentially indirect associations. An example term in the "invalid" category is "colocalization". All protein pairs annotated with "invalid" terms were excluded. 'Binary' versus 'non-binary' evidence was used to categorize protein pairs in the literature-curated dataset as follows. Pairs with no binary experimental evidence were classified as "Lit-NB", corresponding to 100,940 pairs. Pairs with a single piece of binary evidence and no other evidence were classified as "Lit-BS", corresponding to 14,477 pairs. Finally, pairs with two or more pieces of evidence including at least one binary evidence were classified as "Lit-BM", corresponding to 5,589 pairs.

Previous literature-curated datasets generated in 2017 and 2013 were used as a source dataset for pairs experimentally tested in GPCA, MAPPIT and Y2H-v4 (see Engineering of new Y2H destination vectors) experiments. These were generated and processed as above with small differences. Lit-BM-17 and Lit-BS-17 were obtained via the mentha resource data file dated August 28th 2017 (Calderone et al., 2013). Lit-BM-13/Lit-BS-13/Lit-NB-13 were generated 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.
(Kerrien et al., 2012)

Direct PPIs with experimental structures

The most definitive proof that a pair of interacting proteins are in physical direct contact is the availability of a three-dimensional (3D) structure of their interface. We used the subset of Interactome3D (Mosca et al., 2013) restricted to experimental structures, excluding homology models. The dataset from the January 2020 release of Interactome3D, referred to as “I3D-exp-20”, was used for most computational analyses. The dataset from the June 2017 release, “I3D-exp-17”, was experimentally tested in its entirety using Y2H v4 (see Engineering of new Y2H destination vectors). The date assigned to PPIs was obtained from the PDB database taking their earliest release date for all PDB structures from the “complete” Interactome3D dataset.

Note on the overlap between I3D-exp-20 and Lit-BM-20 PPIs

There were a surprisingly large number of pairs in I3D-exp-20 and not in Lit-BM-20 (1,015 pairs in the difference of I3D-exp-20 from Lit-BM-20 and 746 pairs in the union, see Figure S1B). These pairs are mostly cryo-EM structures (77% Electron Microscopy in the difference vs 36% in the intersection) of larger complexes (median number of entities per structure of 18 in the difference vs 4 in the intersection). The reason for this is that in the generation of the literature-curated datasets (see section Literature-curated biophysical datasets), we don’t use the structural data for direct contacts, we base the binary vs non-binary distinction on the experimental method used and we classify Cryo-EM as non-binary since we don’t know if the reported pairs are in direct contact or not.

STAR Methods Table 1. Tested biophysical datasets

						Number of pairs attempted in:		
Category			**Dataset**	**Co-complex associations**	**Binary PPIs**	**Y2H v4**	**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		
		Literature	Lit-NB-17	71,260		151		
	Binary	Proteome x proteome	Ito-core		816	738	199	199
			Uetz-screen		645	470	193	193
			CCSB-YI1		1,772	1,536	200	200
			Tarassov		2,761	199	199	199
		With structure	I3D-exp-17		1,787	1,231		
		Literature	Lit-BS-17		13,981	146	149	
			Lit-BM-13		4,115		584	168
			Lit-BM-17		4,623	4,128		
Predicted			Jansen	9,870		200		
			PrePPI[b]		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] high confidence subset

*Negative control of random pairs of proteins

Functional profile similarity networks (PSNs)

Genetic interaction similarity profile data (GI-PSN) were extracted from Costanzo *et al*. 2016 (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 (Hillenmeyer et al., 2008). The log of growth ratios from the homozygous deletion data were used to calculate PCC for each pair of genes. 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.

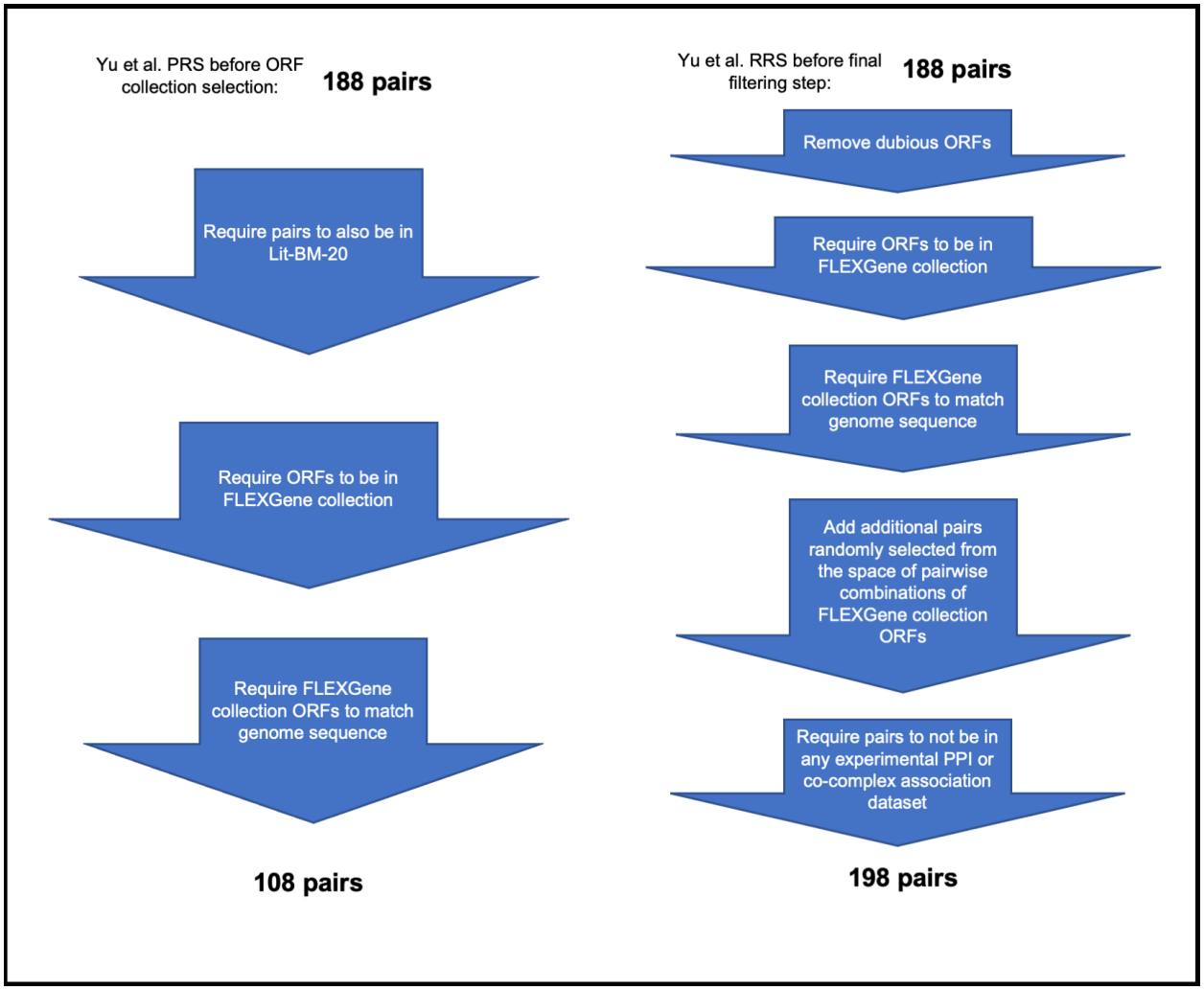
STAR Methods Table 2. Numbers of genes and interactions in the top 1% percentile of the genetic maps

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

#5 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 positive and random* reference sets (scPRS-v2 and scRRS-v2 respectively). 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 the 188 PPIs, we filtered those pairs to also be in Lit-BM-20, then to have both ORFs in the FLEXGene collection (Hu et al., 2007) resulting in a final scPRS-v2 of 108 PPIs. Of 188 RRS pairs in Yu *et al*., we removed all ORFs annotated as dubious, then required they have both ORFs in the FLEXGene 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 FLEXGene collection. Since the RRS is used as a negative control, we then filtered out any pairs that appeared in any of the experimental

PPI or co-complex association datasets, which resulted in removing one pair that appeared in Lit-NB-20 resulting in a final scRRS-v2 of 198 pairs.



#6 Engineering of new Y2H destination vectors

Gateway compatible 2µ high-copy destination vectors pVV212 and pVV213 (Hallez et al., 2007) with a Gal4 DNA binding domain and a Gal4 activation domain, respectively, were modified to be compatible with our standard Y2H vectors pDEST-DB and pDEST-AD-CYH2 (Dreze et al., 2010) with respect to the *LEU2* and *TRP1* as selectable markers. The resulting destination vectors pDEST-DB-QZ212 and pDEST-AD-QZ213 also carry *CAN1* or *CYH2* genes as counterselectable markers, respectively. The *CYH2* and *CAN1* counterselectable markers facilitate plasmid shuffling for the identification of auto-activators (Vidalain et al., 2004). Gateway LR reactions between yeast ORFs

flanked by attL1 and attL2 sites with the attR1 and attR2 sites of pDEST-DB-QZ212 and pDEST-AD-QZ213 result in attB1 and attB2 sites flanking yeast ORFs now fused downstream of either the Gal4 DB or Gal4 AD sequences of the respective destination vector. See **STAR Methods Table 3** for detailed information.

STAR Methods Table 3. Yeast destination vectors

Name	**pDEST-DB**	**pDEST-DB -QZ212**	**pDEST-AD -*CYH2***	**pDEST-AD -QZ213**
Fusion partner (aa)	Gal4-DB (1-147)	Gal4-DB (1-147)	Gal4-AD (768-881)	Gal4-AD (768-881)
Fusion location	N-terminus	N-terminus	N-terminus	N-terminus
Yeast Promoter (nt)	Truncated *ADH1* (-701 to +1)	Truncated ADH1 (-410 to +1)	Truncated *ADH1* (-701 to +1)	Truncated *ADH1* (-410 to +1)
Yeast replication of origin	CEN	2μ	CEN	2μ
Linker sequence between 3' of Gal4 element and Gateway cloning site (aa)	SRSNQ	PEFPS	GGSNQ	ICMAYPYDVPDY ASLGGHMAMEA PS
Yeast terminator	*ADH1* Term	*ADH1* Term	*ADH1* Term	*ADH1* Term
***E. coli* selection marker**	Ampicillin	Ampicillin	Ampicillin	Ampicillin
Yeast auxotrophic selection marker	Leucine	Leucine	Tryptophan	Tryptophan

#7 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-QZ212 and pDEST-AD-QZ213 was named Y2H version 4 (Y2H v4). Y2H v1 - v3 can be found in Luck et al, Nature, 2020 (Luck et al., 2020a). The performance of Y2H v4 was compared to Y2H v1, which consists of destination clones in pDEST-AD-*CYH2* and pDEST-DB, and was used to generate CCSB-YI1 (Yu et al., 2008). The Y2H assay was performed as described previously (Dreze et al., 2010; Rolland et al., 2014). Briefly, Y8930:pDEST-DB-QZ212-ORF and Y8800:pDEST-AD-QZ213-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 days incubation at 30°C, yeast strains growing on SC-Leu-Trp-His+3AT solid media and not on SC-Leu-His+3AT+CHX media were scored as positives. The interacting pairs were identified based on plate position.

#8 Generation of an expanded yeast ORFeome collection

Yeast FLEXGene clone collection (Hu et al., 2007) of full length ORFs cloned in either pDONR201 or pDONR221, both Kan^R, contains 4,933 ORFs, after removal of redundant ORFs and ORFs that no longer match SGD-annotated ORFs (version 2014) (https://www.yeastgenome.org/). For the remaining 950 SGD-annotated ORFs not in Yeast FLEXGene, entry clones were generated in-house and are referred to as supplemental ORFeome collection. ORF sequences were amplified without their 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 18-20 nucleotide ORF-specific forward and reverse PCR primers tailed with Gateway attB1 and attB2 sequences

attB1 Forward primer tail 5' GGGGACAAGTTTGTACAAAAAAGCAGGCTCCACC

attB2 Reverse primer tail 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.

#9 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-QZ212 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-QZ212 and pDEST-AD-QZ213 expression clones were transformed into yeast strains *MAT*α Y8930 and *MAT***a** Y8800, respectively (Dreze et al., 2010).

#10 Auto-activator detection for filtering before Y2H screening

We tested for auto-activation of the *GAL1::HIS3* reporter gene by AD-ORF or DB-ORF fusion proteins 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 (SC-Leu) or Tryptophan (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. Filtering for pairs that passed autoactivator screening and successful cloning resulted in a final collection which was then used for systematic screening included 4,778 DB-ORF clones and 5,700 AD-ORF clones, covering a total of 5,854 yeast ORFs.

#11 Primary yeast two-hybrid (Y2H) screening

Three replicate Y2H screens were performed. Individual *MAT*α Y8930:DB-ORFs were mated in YEPD against a pool of ~700 (FLEXGene collection) or ~200 (supplemental collection) *MAT***a** Y8800:AD ORFs. AD-ORF pool size was decreased for the supplemental collection to facilitate screening. 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). After incubation at 30°C for 3 days, strains 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 up to 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 (Dreze et al., 2010).

#12 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., 2020a). 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 products from an entire 96 well plate were pooled together and purified (Qiagen, PCR Purification Kit). These pooled amplicons from each plate were subjected to Nextera “tagmentation” using Tn5 transposase to generate a library of amplicons with random breaks to which the adapters have been ligated (Weile et al., 2017). We then re-amplified those fragments to generate a library of amplicons such that one end of each amplicon bears the well-specific tag and the other “ladder” end bears the Nextera adapter. 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) (see Sequence identification of interacting ORFs). Due to the small number of pairs to be identified, 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 (**STAR Methods Table 4**) were synthesized by Thermo Fisher Scientific, whereas the universal AD, DB and term primers were synthesized by Eurofins Genomics.

STAR Methods Table 4. 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'

#13 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. Preliminary investigations using four technical replicates demonstrated that in 97% of the cases, the quadruplicates behaved identically (data not shown). Therefore, given the high reproducibility of technical replicates, the culture was spotted only once per selective media. To increase the robustness of our approach we implemented an additional test to identify *de novo* auto-activators in which 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 respective mating plates and tested at the same time. For a batch of pairwise testing to be considered successful we required no more than 1% of RRS and between 10-25% of PRS to be scored positive.

#14 Validation in orthogonal assays

To assess the precision of various 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., 2011a). 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., 2011a), each expressing a different fragment of humanized *Gaussia princeps* luciferase (GL1 and GL2) (Tannous et al., 2005). MAPPIT expression clones were generated by LR transfer of ORFs into pMBU-I-2994 and pMBU-I-4199 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). Three different GPCA and two different MAPPIT experiments were performed.

GPCA

GPCA experiments were performed as described previously (Cassonnet et al., 2011b). Briefly, 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). DNA concentration was measured for all clones and samples were diluted to a final concentration of 25ng/µl. After a 24-hour incubation at 37°C, confluent cells were transfected with 300ng of pSPICA-N1-ORF and pSPICA-N2-ORF vectors using polyethylenimine (PEI). After a second 24-hour incubation at 37°C, cells were washed with PBS supplemented with calcium and magnesium chloride. To lyse the cells 40µl of 5x diluted Renilla lysis buffer

(Promega) were added to each well. The plate was then covered with aluminum foil and agitated 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 per well of Renilla luciferase substrate (Renilla Luciferase Assay System, Promega), with a measurement time of 4 seconds. The measurement score, RLU (relative light unit), was assigned to the tested pair.

MAPPIT

As an orthogonal validation assay, MAPPIT experiments were performed as described elsewhere (Luck et al., 2020b; Rolland et al., 2014). In short, 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 using a chemiluminescence meter. “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 **STAR Methods Table 1** 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 **STAR Methods Table 1**. Samples, if used, were drawn randomly.

#15 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. Using this annotation for each dataset, the fraction of direct PPIs was calculated as the number of direct PPIs reported in the dataset divided by the number of direct and indirect pairs reported in the dataset.

#16 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.

#17 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: i) 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 (Direct co-complex); ii) co-complex pairs annotated in 5 KEGG yeast pathways Cell Cycle, Meiosis, MAPK Signaling pathway, Autophagy and Mitophagy (KEGG co-complex); iii) PPIs regulating activation or inhibition from the same 5 KEGG yeast pathways (KEGG regulation); and iv) 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) (Kinase-substrate).

#18 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 (Wood et al., 2019), of which 950 were within the list of yeast ORFs considered for this study (see section Yeast protein-coding ORFs).

#19 Protein properties

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

#20 Treatment of heterodimers and homodimers

Unless otherwise noted, homodimers were excluded from most analyses since comparisons between physical interactions and functional relationships are obviously not applicable to single genes (all PCC values of functional profiles would be 1.0 by definition).

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

In 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. For all three assays, 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.

#22 Calculation of interface areas of PPIs

We retrieved experimental structures 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.

#23 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. Number of edges in the diagonal bins 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. Homodimeric interactions were excluded. In the case where we corrected the CS-PSN heatmaps for the untested essential genes, we divided the count in each bin by the fraction of pairs where both genes were tested in generating the CS-PSN data.

To calculate the p-values for each 2D bin, we randomly shuffled the order of the proteins 1,000 times. In each permutation of the proteome we calculated the 2D histogram counts, recorded the maximum and minimum bin count (to account for the multiple testing effect of having many bins) and calculated the p-value, for each bin, as the fraction of the random maximum/minimum counts that the observed count is above/below, multiplied by two to account for the two-tailed nature of the test. This was done separately for diagonal and off-diagonal bins because there are a different number of possible combinations of undirected edges between them.

#24 Sequence identification of interacting ORFs

We used an existing computational pipeline (Luck et al., 2020a) to process demultiplexed paired-end reads returned from Illumina sequencing and identify the interacting ORF pairs from the Y2H screen. 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., 2020a) *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.

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

Subcellular compartment data was (Koh et al., 2015) 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.

#26 *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 (Baryshnikova, 2016) 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.

#27 *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 (Yu et al., 2008).

- From Yu *et al.* 2008 (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)."

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.

#28 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, $z = (n - \overline{n}) \div \sigma$, is the traditional way of such comparison, obtained by standardizing the original score with the expectation value ($\overline{n}$) 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 $n - (\bar{n} + \alpha\sigma)\underline{n}$ 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 associations were all protein-protein combinations where both proteins appeared in the same experimental structure but not in direct contact.

#30 Overlap calculation between biophysical and functional networks

For each biophysical network and several KEGG pathways, we measured the fraction of interactions that are also connected in each of the functional networks defined above, discarding homodimeric PPIs. We calculated the overlap by dividing the number of interactions in the PPI network also found in the functional network by the total number of interactions in the PPI network where both proteins were present in the search space of the functional 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.

#31 Date and party hubs

Co-expression 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 proteins with a degree in the top 5% or 10% in each network, or those with degree ≥ 10. 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.

#32 Overlap by degree plots

For each combination of a biophysical and functional network, we conducted a logistic regression, on the dataset of biophysical interactions, where the binary dependent variable represents whether or not the two proteins of the biophysical interaction are also connected by an edge in the functional network, and the single independent variable is the higher of the two degrees, in the biophysical network, of the interacting proteins. The max degree per PPI variable is log2 transformed. Only PPIs where the pair of proteins were tested in generating the functional network were used. 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.

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. (2011a). Benchmarking a luciferase complementation assay for detecting protein complexes. Nat. Methods *8*, 990–992.

Cassonnet, P., Rolloy, C., Neveu, G., Vidalain, P.-O., Chantier, T., Pellet, J., Jones, L., Muller, M., Demeret, C., Gaud, G., et al. (2011b). 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.

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., 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. (2020a). A reference map of the human binary protein interactome. Nature *580*, 402–408.

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. (2020b). 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.

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.

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.

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.

Tannous, B.A., Kim, D.-E., Fernandez, J.L., Weissleder, R., and Breakefield, X.O. (2005). Codon-optimized Gaussia luciferase cDNA for mammalian gene expression in culture and in vivo. Mol. Ther. *11*, 435–443.

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.

Vidalain, P.-O., Boxem, M., Ge, H., Li, S., and Vidal, M. (2004). Increasing specificity in high-throughput yeast two-hybrid experiments. Methods *32*, 363–370.

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.

Weile, J., Sun, S., Cote, A.G., Knapp, J., Verby, M., Mellor, J.C., Wu, Y., Pons, C., Wong, C., van Lieshout, N., et al. (2017). A framework for exhaustively mapping functional missense variants. Mol. Syst. Biol. *13*, 957.

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.