Protein-protein interaction is a major source of epistasis in genetic interaction networks

Xavier Castellanos-Girouard[1,2], Adrian WR Serohijos[1,2], Stephen W Michnick[1,2*]

[1]Department of Biochemistry, Université de Montréal; Montreal, H3T 1J4, Canada.

[2]Robert-Cedergren Center for Bioinformatics and Genomics, Université de Montréal; Montreal, H3T 1J4, Canada.

* Email: stephen.michnick@umontreal.ca

Abstract

Genetic interaction (GI) and protein-protein interaction (PPI) networks have proved among the most important tools for inference of gene function and genotype-phenotype relationship[1–6]. It remains unclear, however, how these two networks are related to each other. Here we demonstrate that the strengths of epistatic genetic interactions between two genes strongly correlate with the binding free energies of interactions between their cognate proteins, for both yeast and human genomes. Consequently, we show that GI and PPI networks can reciprocally predict each other. Further, we observe that functional divergence and redundancy in duplicated genes (paralogs) are reflected in the binding affinity of their interactions with other proteins and in their genetic interaction strengths. Finally, we demonstrate that the overall topologies of GI and PPI networks significantly overlap in two topological features: Modules, in which genes/proteins are organized by common function, and connectors, which link modules to each other. This opens avenues for new integrated network approaches in understanding cellular processes and flow of genomic information.

Introduction

Determining how genes cooperate to shape phenotype is a central goal in molecular biology. At the cellular level, our understanding of functional relationships between genes is best reflected in genetic interaction (GI) and protein-protein interaction (PPI) networks[1,3]. These two networks are highly complementary; GI networks link broad functional relationships between genes via their contribution to a phenotype, whereas PPI networks are maps of physical contacts between proteins. However, how these two networks are related to each other remains an open question[1,7].

GI networks are based on systematic or pooled pairwise deletion or disruption of genes and a genetic interaction is scored as effects on some trait, such as cell proliferation rates, that are less than or greater than the product of the effects of deleting the individual genes, a phenomenon called epistasis[8,9]. Epistasis is a major source of complexity in predicting genotype-phenotype relationships, making the elucidation of its molecular mechanism a significant endeavor[2,10–13]. GI networks also serve important applications such as identifying unique genetic vulnerabilities in cancer cells and inferring gene function by organizing genes into functional groups[1,14–16]. PPIs are also useful for inferring gene functions, linking proteins of unknown, to those of known function[3,17]. Finally, about two-thirds of mutations that cause Mendelian human diseases also disrupt PPIs, linking the integrity of PPI networks to disease[18]. Considering their central role in understanding normal and pathological cellular processes, modelling the relationship between GI and PPI networks has been a longstanding goal in systems biology[7,19].

Recent quantitative proteomic studies provide clues of a deeper physical relationship between PPI and GI networks[20–22]. The number of complexes formed between two proteins is determined by their abundances and their binding affinity. Hein et al. used mass spectrometry to measure the ratios of the abundances of the two binding partners (abundance stoichiometry) and of total numbers of protein pairs that form complexes (interaction stoichiometry), a number proportional to binding dissociation constants (K_d)[20]. The ranges of these binding and abundance stoichiometries were remarkably broad, between ~10^{-7}- 10^{1} (Fig. 1d, regions 2 and 4)[20]. Additionally, well annotated and known stable protein complexes were shown to be enriched around 1:1 abundance and interaction stoichiometries, suggesting a balance of the relative concentration of proteins that form such stable PPIs. The strong negative epistasis that can occur among genes that code for the subunits of essential and stable protein complexes raises the question

of whether the broad range of binding stoichiometries could be proportional to epistatic strengths[1,7].

In addition, Cho et al. demonstrated that a quantitative measure of cellular spatial organization similarity, measured by spinning disk confocal microscopy, of interacting proteins is proportional to interaction stoichiometry[21]. These results imply that higher order physical organization of proteins in the cell is also determined by binding strength. This relationship also defines topological features of cellular PPI networks in which high interaction stoichiometry PPIs formed modules of functionally related proteins while the low interaction stoichiometry pairs formed transient interactions that connect modules[20,21]. GI networks also show a highly modular organization and the correspondence of modular structures between PPI and GI networks has been established[1,7,23]. Is it possible that there is a similar correspondence of PPI protein and GI pairs in connector regions of PPI and GI networks?

Finally, it has been demonstrated that common patterns of epistasis between pairs of mutations in subunits of protein complexes and a set of gene knockout strains correlate with the distance between the mutated residues in protein complexes[24–26]. Consistent with the notion that mutations of residues within functional sites, including hydrophobic core, allosteric, catalytic or binding sites, should produce similar phenotypes and equally, be more strongly energetically coupled to each other[27–29].

Following from these three observations, here, we test the hypotheses that the strengths of PPI and GI correlate and that PPI networks and GI networks overlap in their topological features (modules and connectors).

Results

Epistasis varies with protein abundance and PPI stoichiometries

Since a complete systematic GI network has been constructed for the budding yeast *Saccharomyces cerevisiae* (henceforth "yeast), we chose to make our initial comparison of PPI binding and epistatic strengths for this organism[1]. We first constructed a protein abundance versus PPI binding protein stoichiometry map based on data from a recent affinity purification-mass spectrometry and a meta-analysis of protein abundance measurements[20–22,30]. The distribution of yeast PPIs in the stoichiometry plot (Fig. 1b, Supplementary data 1) is qualitatively similar to that of human proteins (Fig. 1c, Supplementary data 2), the major difference being a higher frequency of low stoichiometry interactions for human proteins. This higher frequency of weak interactions could reflect relaxation of selection pressure that occurred in the transition from unicellular to multicellular organisms as a consequence of reduced population, allowing for the increased rate of emergence of new genomic features, such as gene duplicates and accrual of new gene functions[31–33]. It could also reflect the increase in requirements of connector PPIs to coordinate activities of the expanded number of functional modules in metazoans. Furthermore, we also found that region 1 (Fig. 1b, d) in the yeast stoichiometry plot is enriched for stable complexes (Extended Data Fig. 1), consistent with findings for human proteins[20]. Notwithstanding the evolutionary distance between the two organisms, the similarity in their stoichiometry maps may reflect the conservation of core cellular processes, particularly the molecular mechanisms that mediate them, across yeast and humans[34].

We next sought to explore how GIs distribute across the PPI stoichiometry map. To this end, we merged the GI and PPI networks using the criterion that if the genes comprising a GI also code for proteins that interact, the GI score was assigned to that PPI (Supplementary Data 2, 3). Yeast GIs were taken from a near-complete digenic knockout screen[1] and human GIs were taken from the largest double knockdown study in human cell lines (Jurkat and K562) to date[35]. Of note, the measured GI network in human cell lines is only comprised of 472 genes and consequently only about 0.05% of all possible double knockouts were measured, a much smaller sample than the data for yeast. By visualizing average GI strengths throughout the stoichiometry plots, we observed that negative GIs are strongest for PPIs near 1:1 interaction and abundance stoichiometries (Fig. 1e, left panel). This observation is consistent with previous analyses in yeast showing that negative GIs are enriched between genes whose proteins assemble into complexes[1]. Some examples are

highlighted in Extended Data Table 1. This pattern also holds in human cells (Fig. 1f, left panel, Extended Data Fig. 2a), but to a lesser extent, likely due to the smaller sample size. In contrast, the strength of positive GIs seems to be distributed more uniformly across the stoichiometry map. This suggests that positive GIs are not predominantly driven by PPI and may be better explained in combination with other mechanisms. Indeed, suppression mechanisms involving mRNA decay and protein degradation, as well as specific patterns of disruption in metabolic pathways have been proposed as important sources of positive GIs[1,36–38].

To further investigate the distribution of GI strength throughout the stoichiometric landscape, we divided interactions into 4 regions previously outlined by Hein et al. (Fig. 1d): region 1 consists of interacting protein pairs of equal abundances and 1:1 binding ratios (enriched for stable complexes); region 2 consists of protein pairs of equal abundances but low binding ratios (enriched for transient interactions); region 3 consists of protein pairs of unequal abundances and 1:1 binding ratios, and region 4 consists of protein pairs of unequal abundances and low binding ratios[20]. As suggested by the stoichiometry plots, the distribution of negative GIs within region 1 is significantly higher than that of other regions ($P < 0.01$, Wilcoxon rank-sum test) (Fig. 1g, h, left panels, Extended Data Fig. 2c). The distribution of positive GI values is also highest in region 1 in yeast and human cell lines (Fig. 1g, h, right panels), in agreement with previous observations that complexes, composed mostly of non-essential genes, are enriched for positive GIs (Extended Data Table 2)[1]. Interestingly, region 3 that is characterized by high interaction stoichiometry and an excess of affinity purified versus bait protein show the lowest number of GIs among the four regions (Fig. 1g and h, and Extended Data Fig. 2c), reflecting that the number of protein complexes formed is not the sole determinant of GI strength. Taken together, these results show that there is robust overlap between GI and PPI networks.

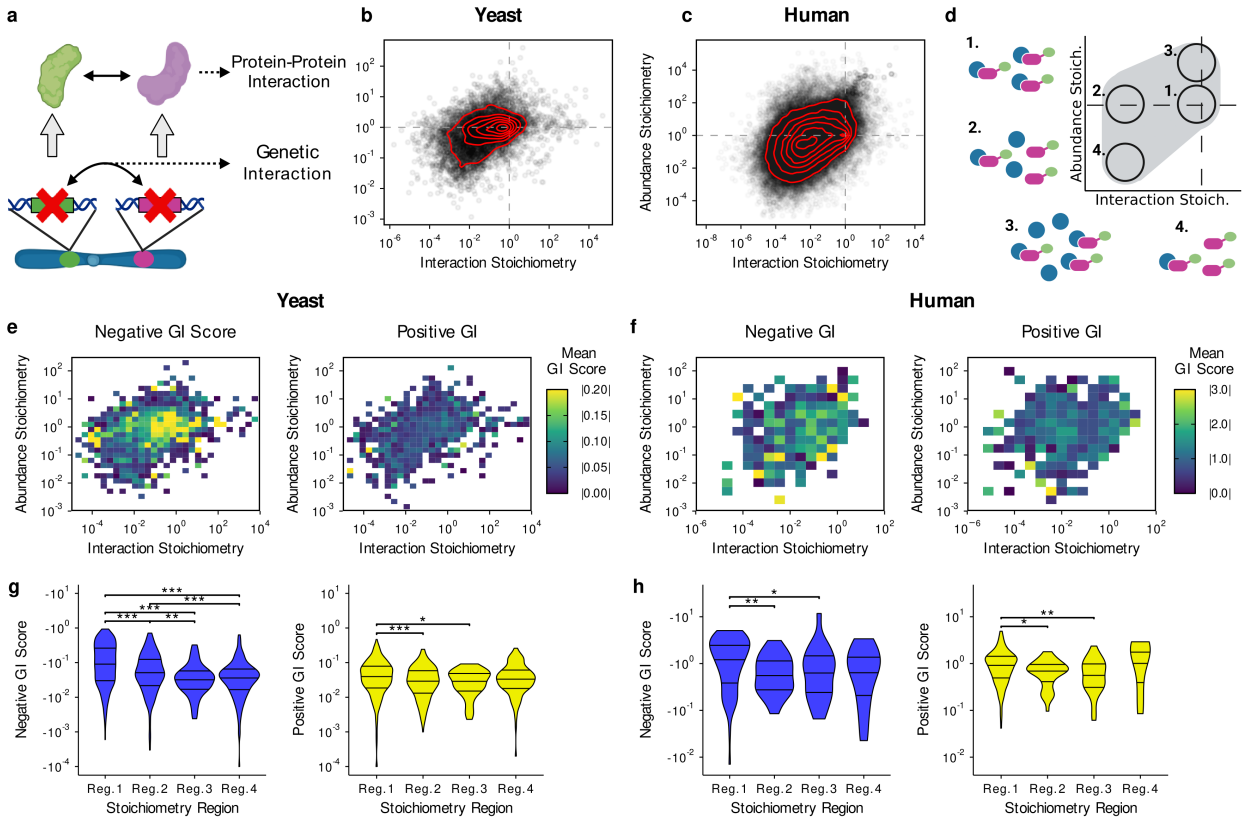


Figure 1. Strong negative GIs map to PPIs of equal protein stoichiometry. a, Genetic interactions are generated by comparing fitness defects in genotypes with single or double gene knockout/knockdowns, and protein-protein interactions are defined as physical contacts between proteins. Illustration created with BioRender. **b, c**, Protein-protein interactions (grey points) are plotted according to two stoichiometries: Interaction Stoichiometry and Abundance Stoichiometry for yeast (b) and human cell lines (c). **d**, The stoichiometry plot can be separated into 4 regions representing PPIs with different ranges of protein stoichiometry values. Based on figure from reference 20. Illustration created with BioRender. **e (yeast), f (human),** Mean negative and positive genetic interaction scores for PPIs throughout the stoichiometry plot. GI scores in human were taken from a screen in K562 cell lines. **g, (yeast), h, (human),** Distribution of GI scores within the 4 defined regions of the stoichiometry plot. Significance levels are annotated as: * for $p \leq 0.05$, ** for $p \leq 0.01$, and *** for $p \leq 0.001$.

145 **Negative epistasis correlates with PPI binding strength**

146 Substantial efforts have been made to model how the physicochemical properties of proteins such
147 as folding stability and PPI binding energies drive epistasis[29,39,40]. Whilst these approaches explain
148 how epistasis arises from point mutations for single proteins or pairs of proteins in PPIs, the extent
149 to which the strengths of GIs for whole gene knockout (or knock down) could be explained by
150 binding free energies is unknown.

151 Interaction stoichiometry can provide an indirect measure of PPI strength if protein pairs are of
152 equal abundance, however, interaction stoichiometry is inherently limited by the abundance of the
153 bait in an affinity purification-mass spectrometry (AP-MS) experiment[20,21]. This is reflected in the
154 triangular shape of stoichiometry maps (Fig. 1b to d), where protein pairs of very low abundance
155 stoichiometry have a strong tendency to not form high interaction stoichiometry interactions (Fig.
156 1b and c). Abundance is therefore a confounding factor when using interaction stoichiometry as a
157 proxy for binding strength.

158 To obtain direct measures of PPI binding affinities, we devised a simple approach to estimate
159 equilibrium dissociation constants (K_d) from protein stoichiometries and protein abundances
160 measured by mass spectrometry. Briefly, the number of complexes formed by a pair of proteins
161 was estimated using the measured interaction stoichiometry and cellular abundances. The number
162 of complexes were then used to estimate the number of unbound copies of each interactor. Finally,
163 copy numbers were converted to concentrations, from which second order equilibrium dissociation
164 constants (K_d) were calculated (Supplementary Data 5, 6). We validated our results by comparing
165 the estimated values to *in vitro* measurements of dissociation constants from literature and found
166 a moderate but significant correlation of 0.40 (Pearson's r, *p*-value = 5.2×10^{-4}, Extended Data
167 Figure 3a, Supplementary Data 7), which was comparatively better than that of interaction
168 stoichiometry (Pearson's r = -0.29, *p*-value = 0.016, Extended Data Fig. 3b). The theoretically
169 expected relationship between K_d and interaction stoichiometry is inversely proportional. It has
170 been noted that important criteria to assess reliability (*i.e.*, controlling for titration and
171 demonstrating equilibrium) in equilibrium binding experiments often go unreported throughout
172 literature, and different experimental approaches can sometimes yield non-overlapping values of
173 K_d[41,42]. Moreover, experiments are generally performed *in vitro* and under conditions that do not
174 reflect native cellular environments. Considering these important sources of variability, we

175 reasoned that our values fall within an acceptable range of precision. To our knowledge, this is the
176 first report of a proteome-wide estimate of *in vivo* binding affinities.

177 A density heat map relating GI scores and K_d shows a visible bias towards negative GIs for high
178 affinity interactions in yeast and human (Extended Data Fig. 4a). To quantitatively evaluate the
179 relationship between K_d and GI strength, we binned the PPIs into quantiles according to K_d and
180 calculated the mean GI strength within each bin (Extended Data Fig. 4b, c). The mean strengths
181 of negative GIs correlate remarkably and significantly with K_d for both yeast (Fig. 2a, Spearman's
182 $\rho = -0.98$, p-value = 2.2×10^{-5}) and human PPIs (Fig. 2b, Spearman's $\rho = 0.95$, p-value = 0.014).
183 A weak but significant correlation is also observable for positive GIs in yeast (Fig. 2c, Spearman's
184 $\rho = 0.76$, p-value = 0.008), which is likely driven by an enrichment of significant positive GIs
185 within complexes composed of non-essential genes (Extended Data Fig. 5c, Extended Data Table
186 2)[1,7]. At the molecular level, this trend could be explained by a mechanism where the absence of a
187 single protein (single gene knockout) is enough to completely disrupt the assembly of a complex
188 in which it is embedded[43]. In this case, the growth defect for the double knockout would be
189 expected to be less than that of the product of the two single knockout. Specifically, in protein
190 biochemistry, the presence of a binding partner can increase the folding stability of a protein by
191 providing nucleation sites for its unfolded species, an effect that is heightened for stronger PPIs[44].
192 The single gene knockout could destabilize the other gene, consequently, the double knockout is
193 not substantially more drastic than the single knockout.

194 The shapes of the curves are important in that they take the same forms as single site binding
195 curves, with a flat lower end, dynamic range covering about two orders of magnitude, and flat
196 saturation range. We speculate that these curves may represent an effective average binding curve
197 with, for negative GIs, centered around average K_ds of 1 μM and 10 μM for mouse and human,
198 respectively and mean GI scores for both species at about -1.4. The curve for positive GIs in yeast
199 is notable in that the half-maximal Mean GI Score is shifted to higher affinity (0.1 μM), consistent
200 with the postulate that positive GIs are stronger for higher affinity PPIs[44].

201 The strong correlation of mean PPI K_d and negative GI suggests that PPI binding affinity is a major
202 driver of negative epistasis between the genes that code for the interacting proteins. These results
203 imply that one network could be predicted from the other. To test this idea, we devised an approach
204 to reconstruct GI network matrices from binding data alone (see Methods, Extended Data Fig. 6).
205 We observed that the clusters along the diagonal of a matrix reconstructed from PPI binding data

in yeast qualitatively matches that of experimental results in genetic screens (Fig. 2e, f)[1]. Overall, these results indicate that there exists a strong relationship between the quantitative features of GI and PPI networks, and that the modular structure of quantitative PPI networks seem to match that observed in GI networks.

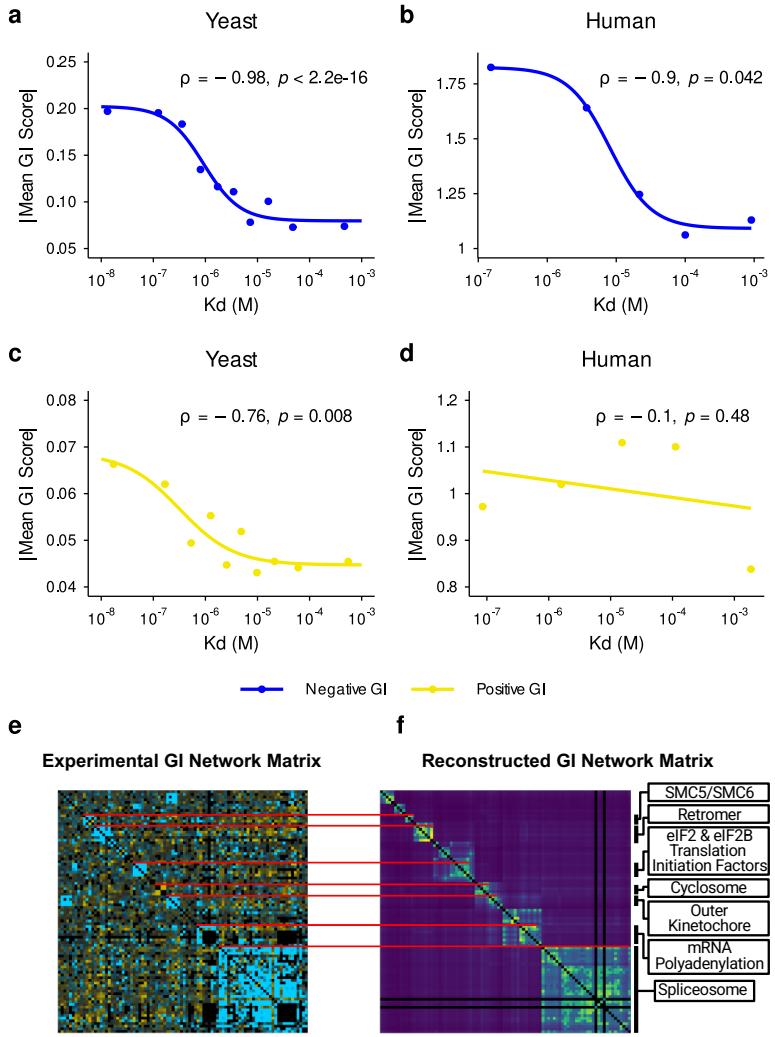


Figure 2. Genetic interactions correlate with PPI binding affinity. a, Binning PPIs according to K_ds reveals a significant monotonic relationship between K_d and negative GIs. PPIs in yeast were binned into deciles (10 quantiles). **b**, The binning analysis described in (a) was repeated for negative genetic interaction data from K562 human cell lines. Data were binned into quintiles (5 quantiles) to adjust for data sparsity. **c**, **d**, The binning analyses described in (a) and (b) were performed on positive genetic interactions for yeast (c) and the human K562 cell line (d). The GI versus K_d trend was fit to a sigmoidal equation except for panel d where no sigmoidal fit was possible, a linear trend was used instead. *P*-values for Spearman rank correlations coefficients

were computed with the following alternative hypothesis: Mean GI increases as K_ds decrease. Correlations for unbinned data are reported in Extended Data Table 3. **e,** A section of the yeast GI network matrix, generated experimentally by synthetic genetic array (SGA). **f**, A reconstructed GI network matrix, generated computationally from a PPI network weighted by K_ds. Hierarchical clustering was performed on the reconstructed matrix, and the resulting order of genes was applied to the experimental matrix. Full matrices are available on Figshare[45,46].

Redundancy and divergence of paralogs are reflected in genetic epistasis and PPI affinities

In our results above, we hypothesized that PPI, specifically binding affinity (K_d), is an accurate proxy for cellular function and thus, it quantitatively reflects epistasis across the genome. Since the number of protein complexes is a function of abundance and K_d, this poses a specific prediction that genomic events resulting in a maintained decrease or increase in abundance over macro-evolutionary time scales will be accompanied by a change in K_d. In particular, gene duplication events, which could effectively double the number of cellular copies of a protein, could relax selection on K_d and still maintain the same number of complexes. Correspondingly, due to the relaxed selection on K_d, we also expect a reduction in genetic interaction.

To test this hypothesis, we identified paralogs in our dataset and grouped them into four distinct types: 1) PPIs in which neither protein has a paralog, 2) PPIs in which one of the two (but not both) proteins has a paralog, 3) PPIs in which both proteins have a paralog, and 4) PPIs in which a protein interacts with its own paralog (paralogous heteromers); (Fig. 3, legend). Consistent with our expectation and previous analyses of GIs in paralogs, duplicated genes show weaker GIs with other genes than non-duplicates genes (Fig. 3a)[47]. This effect seems to be additive overall, as gene pairs where both genes have paralogs show weaker GIs (Fig. 3a, green). Remarkably, a similar trend is observed for binding affinity, where proteins encoded by duplicated genes tend to form weaker (high K_d) PPIs (Fig. 3b).

Intriguingly, two distinct modes are observable in the distribution of K_ds for interactions in which both proteins have a paralog (Fig. 3b). This result is reminiscent of results from analyses of tri-genic GIs involving duplicated genes, in which functionally redundant paralogs form more trigenic interactions (as opposed to digenic interactions) than functionally divergent paralogs, thereby resulting in a distinct bimodal distribution[48]. To test whether the separation of paralogs into

249 redundant and divergent groups drives the bimodality observed in the K_ds, we computed the
250 sequence divergence (defined here as the number of non-synonymous mutations per non-
251 synonymous site) between genes and their paralogs and separated them accordingly into "high"
252 and "low" divergence paralogs (Extended Data Fig. 8, Supplementary Data 8). For interactions in
253 which only one protein has a paralog, highly divergent paralogs participate in stronger interactions
254 (lower K_d) than paralogs with similar amino acid sequences (Fig. 3c); this is also true for
255 interactions in which both proteins have a paralog (Fig. 3d).

256 It is known that some duplicated genes have conserved sequences and functions over evolutionary
257 time[49,50]. However, protein abundance is expected to increase in the short term after gene
258 duplication events, resulting in an imbalance of concentrations of protein complex components,
259 which in turn negatively impacts fitness[51,52]. To reconcile functional conservation and
260 concentration imbalance, it has been proposed that for short evolutionary time scales, post-
261 translational expression attenuation is an important mechanism for robustness against an abrupt
262 abundance increase following gene duplication and, over larger time scales, reduced expression
263 allows for maintenance of redundant duplicated genes[53,54].

264 However, reduced expression is not observed for all gene duplication events, suggesting other
265 mechanisms may be at play[54]. In light of our results, we propose a model to explain how, on longer
266 evolutionary time scales, protein complexes adapt to increased expression of their subcomponents
267 after gene duplication (Fig. 3e). We hypothesize that following a duplication event, paralogous
268 genes that conserve their sequence will conserve their interaction partners but participate in weaker
269 interactions overall to balance the binding equilibrium such that the number of complexes stays
270 the same (Fig. 3e, Fate 1). This can happen, for instance, when mutations occur in a few important
271 amino acids at the binding interface or allosteric sites of paralogs, or if their interacting partners
272 diverge such that their binding affinity for the duplicated genes is decreased. Furthermore, the
273 subset of highly divergent paralogs participating in high affinity interactions (Fig. 3c and d) can
274 be explained by a second scenario, where one of the duplicated genes diverges and loses (or
275 drastically reduces the affinity for) interacting partners while the other maintains high affinity
276 interactions (Fig. 3e, Fate 2). This model is both compatible with previous works showing that
277 abundance is decreased following gene duplication[53,54] and explains the tendency of highly
278 divergent paralogs to make high affinity PPIs.

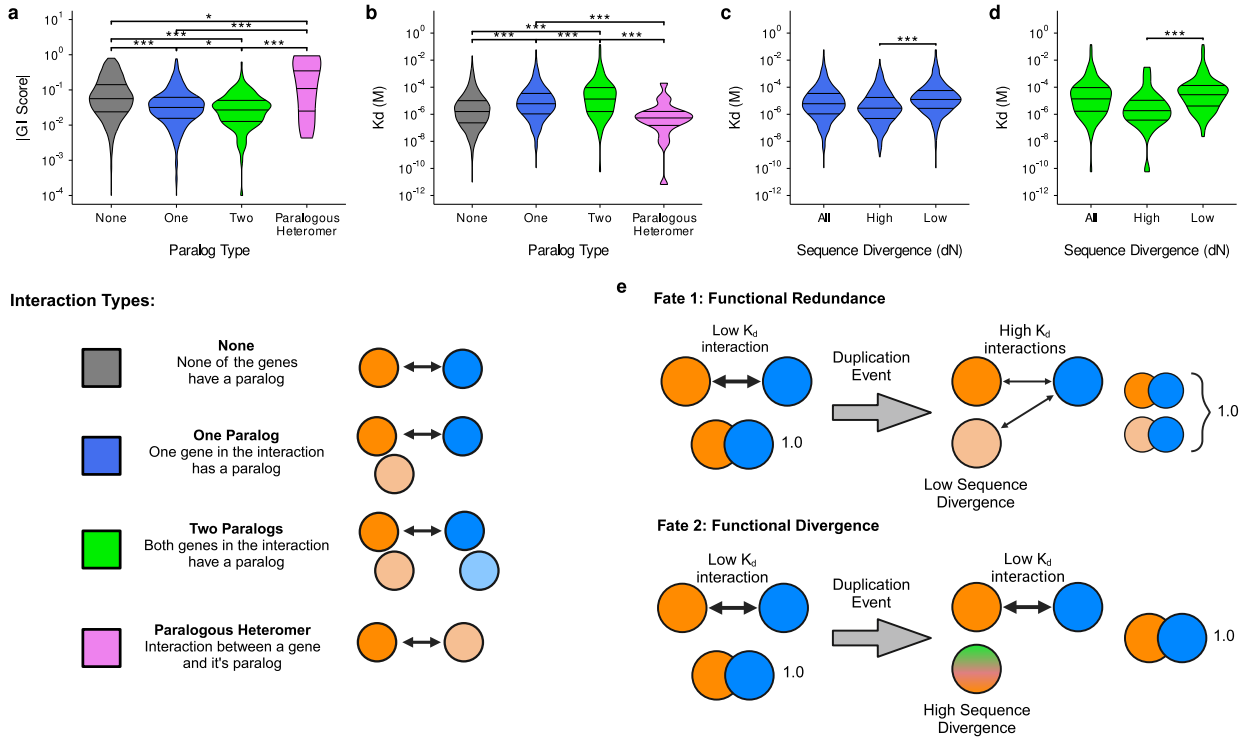


Figure 3. Functionally redundant paralogs participate in low binding affinity PPIs. a, GI scores for gene pairs which have one or two paralogs are lower than those without paralogs. Distributions for individual GI type (positive or negative) are included in Extended Data Fig. 7. **b,** Dissociation constants (K_d) between proteins involved in a PPI are higher (lower affinity) if one or both proteins have a paralog. **c, d,** PPIs in which one (c) or both (d) proteins have a paralog separated according to sequence divergence. Significance levels are annotated as: * for $p \leq 0.05$, ** for $p \leq 0.01$, and *** for $p \leq 0.001$. **e,** Hypothesized evolutionary model where duplicated genes with low sequence diverge (Fate 1) participate in weaker interactions over time. For duplicated genes with high sequence divergence (Fate 2), one paralog loses (or greatly reduces) their affinity for common interactors. Illustrations in legend and (e) were made with BioRender.

Compositions of module and connector regions of the PPI network overlap with corresponding GI network regions

The classical interpretation of GI and PPI overlap mainly focuses on GIs within and between modules[4,7,19]; commonly referred to as the "within pathway" and "between pathway" models, respectively[17]. Indeed, protein-protein interaction modules ("complexes") are reflected in the functional modules in GI networks[1,7,23]. However, not all proteins that participate in strong

interactions form stable complexes of specific function; some participate in transient interactions that "connect" protein complexes together[55]. The degree to which these "connectors" in PPI networks are reflected in GI networks has not been explored. We hypothesize that the ability of certain proteins to connects complexes in PPI networks also holds for GI networks. In other words, if a protein connects two complexes (modules) in the PPI network, it will likely connect the same two functional modules in GI networks.

At the cellular level, functional modules involving proteins are typically metabolic pathways, signaling pathways or protein complexes[56]. However, we specifically limit our analysis of protein modules to solely include stable protein complexes in yeast, as defined by the ComplexPortal database[57]. Furthermore, GI modules were taken from an analysis of the double knockout GI network used in the previous sections[1]. For both GIs and PPIs, connectors were defined as the set of genes or proteins that did not fall into a module. As a preliminary step, we sought to determine the extent to which proteins/genes in modules overlapped between the two networks using this classification. We observe a significant overlap between the modules and for connectors for both networks (*p*-value = 1×10^{-49}, hypergeometric test) (Fig. 4b, left panels).

GI modules map directly to PPI modules

We next sought to determine whether GI modules could be mapped one-to-one with protein complexes. To this end, we evaluated the similarity in gene composition of all possible GI module and PPI module pairings and identified module pairs with the highest degree of overlap. Strikingly, of the 112 GI modules, 88 could be mapped to a protein complex (Fig. 4b, upper left panel, Supplementary Data 9), which is significantly more than would be expected by chance (hypergeometric test *p*-value = 7×10^{-160}, Fig. 4c). This analysis is complementary to that of Costanzo et al.[1], in that we use a set of manually curated complexes, as opposed to GO enrichment, to demonstrate correspondence between GI and PPI modules. The remaining GI modules are likely attributable to pathways not involving stable complexes. For instance, a GI module composed of ERG13 and MVD1 was determined to be enriched for steroid and isoprenoid synthesis but was not mapped to any known complex. It is also important to note that the large fraction of non-overlapping protein complexes reflect the redundancy inherent to the ComplexPortal database, in which several variants or subunits of a complex are included as separate entries.

327

328 ***Connector proteins in PPI networks are analogous to connector genes in GI networks***

329 The ability of connector proteins to bridge protein complexes has been highlighted in recent mass
330 spectrometry screens of protein-protein interactions in human cells[20,21,55]. We reasoned that if most
331 GI modules overlap strongly with protein complexes, some GI connectors may have topological
332 properties similar to that of PPI connectors. The ability to bridge protein complexes being the most
333 salient topological property of protein connectors, we focused on systematically identifying this
334 motif network-wide for both GI and PPIs. Towards this aim, we devised an algorithm consisting
335 of four main steps as illustrated in Fig. 4a for a single pair of modules from each network: 1)
336 Mapping GI and PPI modules 1:1 as described above, 2) Finding the shortest paths between all
337 pairs of modules; this was done independently in GI and PPI networks, 3) Identifying connector
338 nodes (genes or proteins) in the shortest paths that are common between both networks, and 4)
339 combining the common components into an integrated network. We identified 555 shortest path
340 connectors common to both networks (Fig. 4b, lower right panel, Supplementary Data 10).
341 Comparing this degree of overlap to that of randomized networks, we find that our results are
342 extremely unlikely to arise by chance (hypergeometric test p-value = 4×10^{-53}, Fig. 4d). Overall,
343 our results demonstrate that connectors bridging modules are a topological feature shared by both
344 GI networks and PPI networks.

345

346 ***Integrating PPI and GI networks highlights the functionality of connector genes and proteins***

347 Previous analyses combining GIs and protein complexes have focused on GIs within and between
348 complexes and the functional interpretations thereof[1,7,19]. Our approach to integrating PPI and GI
349 networks into a "module-connector-module" structure could reveal novel functional relationships
350 between the two. In fact, our integrated network reveals a core set of modules shared between both
351 networks and highlights known functionality of well-studied connectors. For instance, a
352 subnetwork of our integrated network correctly identifies the karyopherin kap123p as an important
353 bridge between the nuclear pore and various complexes that are transported and/or localized in the
354 nucleus (Fig. 4e). The integrated network also places RAS2 (Ras2p) adjacent to the adenylyl
355 cyclase complex and the polarisome, which aligns with their respective roles in cell growth[58–61].
356 Interestingly, the integrated network also indicates that RAS2 connects the adenylyl cyclase

complex to mitochondrial complexes, supporting a recently proposed mechanism in which Ras2p in the mitochondria promotes apoptosis via the cAMP/PKA pathway[62]. These examples highlight the potential of analyzing motifs in the connector region for generating new hypotheses on the functions of genes.

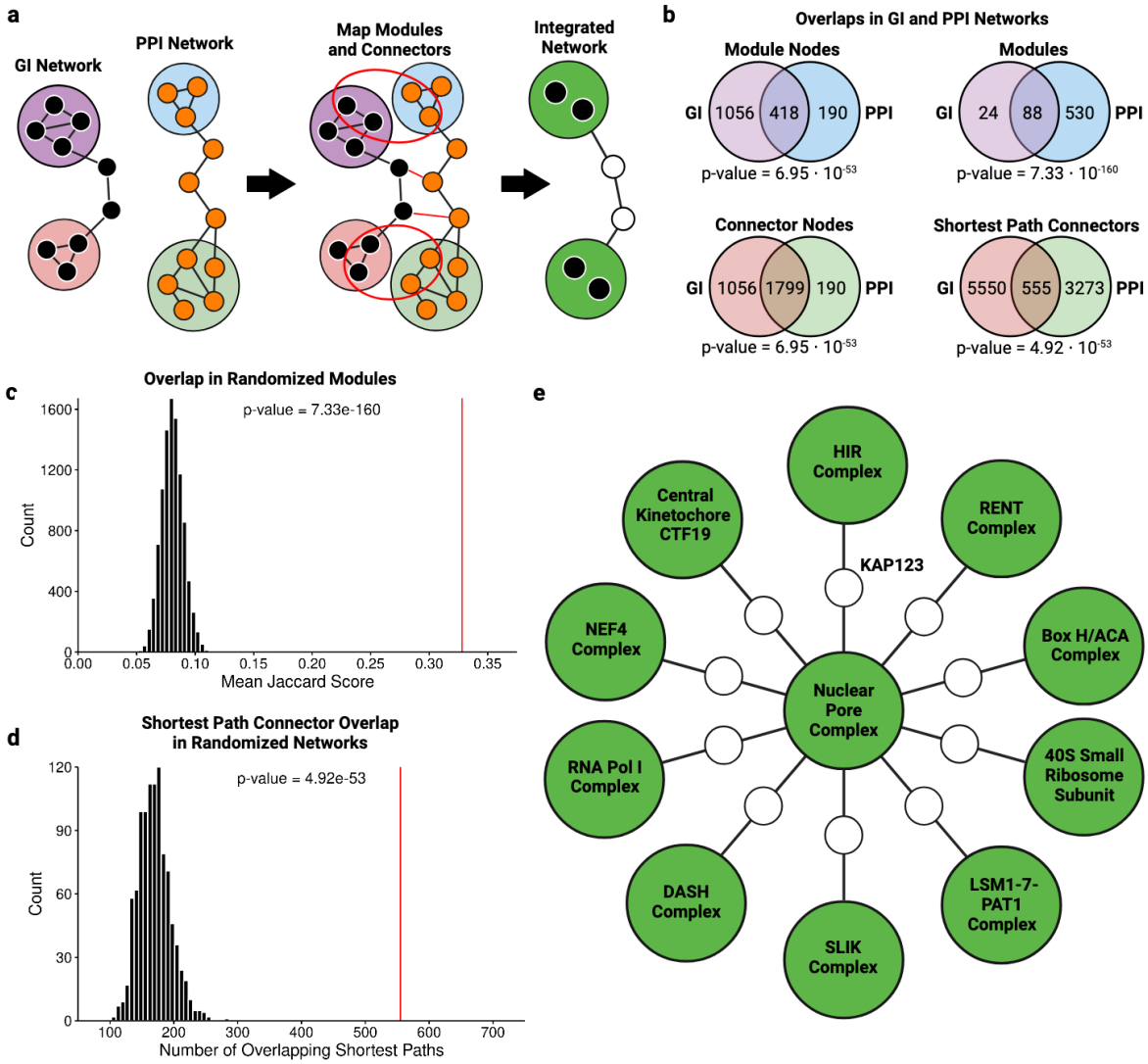


Figure 4. Integrating genetic interactions networks and protein-protein interaction networks using common topological structures. a, Brief description of the algorithm devised to identify connectors of interest. First, shortest paths between modules in GI and PPI networks are determined. Modules of GI and PPI networks are mapped 1:1 and their common nodes are identified; common shortest path nodes are also identified. The final integrated network is comprised of common components in the two networks. **b**, Summary statistics of different overlapping components in GI and PPI networks. **c**, A statistical validation of overlaps between

modules in GI and PPI networks by comparing the observed overlap with that of randomized modules. **d**, A statistical validation of overlaps between shortest path in GI and PPI networks by comparing the number of observed shortest path overlaps to that of randomized networks. **e**, A subset of the integrated network centered around the nuclear pore complex and the karyopherin KAP123. Illustrations in a, b and d were made using BioRender.

375 ## Discussion

376 Determining how GI and PPI networks are related is an outstanding problem in systems and cell
377 biology and its resolution has profound implications for our understanding of the genotype-
378 phenotype relationship. While some relationship between binding affinity and GI strength could
379 fit intuitive expectation, a quantitative relationship between binding affinity and the continuum of
380 GI values has never been demonstrated. In this study, we examined the two properties that define
381 biological networks (weights and topology) to establish a rigorous framework through which PPI
382 and GI networks can be analyzed, predicted, and integrated. Although we focus on networks
383 generated with data from the model organism *Saccharomyces cerevisiae*, we find that the
384 principles established in yeast translate well to human networks.

385 The relationship between binding energies and GIs was further examined by estimating PPI
386 dissociation constants proteome-wide, which we found to be strongly correlated with values from
387 literature. These K_d estimations were used as a baseline to construct new models relating protein
388 binding to genetic interactions. Indeed, binning PPIs according to binding information enabled
389 both a reconstruction of a yeast GI network which qualitatively matched experimental results and
390 the identification of an equation relating the general quantitative relationship between GIs and K_d:

391 $$\varepsilon = \frac{a}{1 + e^{b \cdot \ln(K_d) + d}} + c \quad (1)$$

392 Where K_d is the dissociation constant of a PPI, ε is epistasis (multiplicative model) between the
393 interacting proteins' genes, and *a, b, c* and *d* are free parameters. Importantly, these parameters
394 have tangible biological and biophysical meanings. For instance, the parameter *a* sets the range of
395 epistasis values. Further, the parameter *b* (replaced by kT below), the parameter *d*, and K_d form
396 the Gibbs free energy of binding equation:

397 $$\Delta G = -kT \ln(K_d) + d \quad (2)$$

398 We note that equation 1, relating K_d and epistasis is a sigmoidal function, having the same form
399 as the equation relating K_d and fraction bound. The implication of this correspondence between
400 GI and PPI networks is profound: Tracing non-additive effects (epistasis) to their molecular
401 thermodynamic underpinnings implies the possibility of modelling effects of individual genetic
402 variations on a given trait (genotype-to-phenotype relationship). A caveat of our analysis is that
403 epistasis is defined specifically in the context of pairwise gene deletions (or temperature sensitive

404 mutant gene knock downs) as opposed to other type of genetic changes such as point mutations,
405 recombination, or structural variants. Moreover, our model for the reconstruction of a complete GI
406 network (Fig. 2f) explains approximately 2% of the entire variance in the negative GI network,
407 including GIs between genes whose proteins do not interact with each other. Although a better
408 understanding of how genetic perturbations percolate through the PPI network would likely
409 increase explainability, it is likely that GIs are also strongly driven by other molecular mechanisms
410 such as substrate and product dependencies between enzymes in metabolic networks, and
411 transcription factor-promoter relations in gene regulatory networks.

412 Certain limitations exist as well regarding our estimates of dissociation constants. First, each PPI
413 is treated as completely independent of any other, leading to deviations from the actual values. For
414 instance, competitive binding will systematically lead to underestimations of binding strength,
415 especially for very strong interactions. Second, under our current model, interaction
416 stoichiometries over 1:1, which may occur if more than one prey protein binds to a single bait
417 protein, lead to negative K_d values which are physically impossible. The number of interactions
418 which are discarded as a result represents a non-negligible fraction of the interactome, which may
419 bias the distribution of our estimated K_d. Third, our ability to assess the precision of our estimates
420 is limited to the quality of reference binding affinity data in the literature. Fourth, we assume all
421 reported interactions in our datasets to be direct. In reality, AP-MS approaches also detect indirect
422 interactions which are not readily distinguishable from direct ones.

423 Previous approaches aiming to identify common topological features between GI networks and
424 PPI networks have mostly focused on GIs within or between protein modules[4,7,19]. However, not
425 all genes and proteins cluster into highly connected groups. Although it has long been known that
426 certain proteins serve as "connectors" between protein modules[63], the extent to which this
427 "module-connector-module" motif occurs in GI networks had not been explored. We demonstrate
428 that the common topological organization in GI and PPI networks extends further than overlap in
429 modules, but also the connectors that bridge these modules. By merging modules and connectors
430 from both GI and PPI networks an integrated network was constructed, where the functional role
431 of connector genes/proteins are highlighted and can be used a novel tool for functional inference
432 that will be valuable for predicting GIs in other organisms.

433 Finally, the quantitative and topological relationships we have established in this study could
434 provide means of interpreting and predicting fundamental relationships between genes underlying

435 normal and disease phenotypes, as well as identify specific vulnerabilities or changes in PPIs that
436 could predict potential therapeutic targets for diseases.
437

438 # Methods

439 **Collecting yeast and human genetic and protein-protein interaction data**

440 For yeast, raw synthetic genetic array (SGA) genetic interaction data was downloaded from
441 TheCellMap[1,64] in matrix format. Data from essential and non-essential strain crosses (Essential
442 vs. Essential, Essential vs. Nonessential, Nonessential vs. Nonessential) were combined. In
443 instances where multiple GIs were measured for a pair of genes, only the entry with the most
444 significant *P*-value was retained. For human cell lines, GI data was retrieved from a digenic
445 knockdown screen of 472x472 genes in Jurkat and K562 cells[35] (table S5 of reference 35). For
446 both yeast and human data, unless otherwise indicated, analyses were performed using all GIs
447 regardless of significance scores.

448

449 Yeast protein-protein interactions were taken from a near exhaustive whole-proteome affinity
450 enrichment coupled to an affinity purification-mass spectrometry (AP-MS) screen[22]
451 (Yeast_Interactome_Edges.csv from supplementary file 2 of reference 22). Human protein-protein
452 interactions and stoichiometries were taken from two quantitative AP-MS screens: one in which
453 genes were tagged with GFP and expressed on bacterial artificial chromosomes in HeLa cells[20]
454 (table S2 in reference 20), and a second in which genes are endogenously tagged in HEK 293T
455 cells with the small fragment of the split $mNeonGreen_2$ tag[21] (table S4 in reference 21). Both
456 human cell line interactomes were concatenated but common interactions between the two screens
457 were kept as separate data points for all analyses.

458

459 **Computing interaction stoichiometries and abundance stoichiometries in yeast**

460 Interaction stoichiometries were calculated from data generated in an AP-MS screen[22]
461 using a previously described strategy[20]. The starting preprocessed table containing normalized and
462 imputed label-free quantification (LFQ) intensities from the AP-MS screen were provided by the
463 authors. Briefly, computing interaction stoichiometries was done in three mains steps: 1)
464 Subtracting the median intensity of a protein across all sample, 2) Dividing LFQ intensities by the
465 number of theoretically observable peptides of lengths between 7 and 30 amino acids, 3) dividing
466 the values observed for the prey protein by that of the bait protein. Importantly, the number of
467 theoretically observed peptides were computed using sequences from the Uniprot database[65]
468 (https://www.uniprot.org/proteomes/UP000002311) using LysC only digestion (as opposed to
469 LysC and Tryptic digestion). Negative interaction stoichiometries (which arise because of

normalization in step 1) were dropped from the dataset. The final dataset includes interaction stoichiometries for 7348 unique PPIs.

Protein abundances were taken from a dataset unifying measurements from various sources[30] (table S4 in reference 30). Abundance stoichiometries were determined by calculating the ratio between the median copy numbers of the prey over that of the bait, as described previously[20]. Abundance measurements for both interactors were available for 7336 out of the 7348 PPIs for which interaction stoichiometry values were calculated. The 22 interactions which had interaction stoichiometry values, but not abundances stoichiometry values, were conserved in the dataset for analyses in which abundance stoichiometry was not relevant.

Analysis of genetic interactions and protein-protein interaction stoichiometries

The relationship between GIs and stoichiometries was explored in yeast and human cell lines (see first section for data collection). To merge GI and PPI stoichiometry datasets, tables were merged such that if genes comprising a GI also code for proteins that interact, the GI score is assigned to that PPI. Four stoichiometric regions were graphically defined within stoichiometry plots according to previous criteria[20]. Hein et al. defined region 1 as the space enclosed by a circle centered at (-0.5, 0) where the abscissa is interaction stoichiometry, the ordinate is abundance stoichiometry, and coordinates are in base 10 logarithmic scale. Using these definitions as a reference, region 3 was defined using circles of identical radii (0.75) centered at (-0.5, 1.5), region 2 was centered at (-2.5, 0) for yeast and (-3.5, 0) for human, and region 4 was centered at (-2.5, -1.5) for yeast and (-3.5, -1.5) for human.

Estimating protein-protein interaction dissociation constants

To estimate dissociation constants from quantitative mass spectrometry data, we first assumed all individual PPIs to be independent of all others (no competitive binding) and their binding dynamics to be describable by a two-state system:

$$A + B \rightleftharpoons AB \tag{1}$$

$$K_d = \frac{[A][B]}{[AB]} \equiv \frac{[Bait][Prey]}{[Bait::Prey]} \tag{2}$$

501

502 Where K_d is the dissociation constant, [A] the molar concentration of unbound protein A, [B] the
503 concentration of unbound protein B, and [AB] the concentration of the complex formed by proteins
504 A and B. In a given mass spectrometry experiment, one of the two proteins will be the bait (tagged
505 protein retained on column) and the other a prey.

506

507 Next, we sought to estimate the number of bait-prey complexes using interaction stoichiometries.
508 By definition, interaction stoichiometry is a proxy measurement of the ratio of prey protein bound
509 to bait protein over the total abundance of bait protein. Under the previously mentioned
510 assumptions, the amount of prey bound to bait is equal to the number of bait-prey complexes.

511

512 $$Interaction\ Stoichiometry = \frac{Prey_{LFQ_{Intensity}}}{Bait_{LFQ_{Intensity}}} \approx \frac{Prey_{Bound\ to\ Bait}}{Bait_{Total}} = \frac{Bait::Prey}{Bait_{Total}} \quad (3)$$

513

514 Therefore, we used the product of interaction stoichiometry and total abundance of bait (in copy
515 numbers) to estimate the number of bait-prey complexes in a given cell.

516

517 $$Bait::Prey = \frac{Bait::Prey}{Bait_{Total}} \cdot Bait_{Total} = (Interaction\ Stoichiometry) \cdot Bait_{Total} \quad (4)$$

518

519 For yeast PPIs, the cellular copy numbers of bait proteins were taken from the same source as that
520 described to calculate abundance stoichiometries[30]. For PPIs in human cell lines, abundances were
521 taken from whole proteome quantification experiments which were performed alongside the AP-
522 MS measurements of interaction stoichiometries[20,21]
523 (opencell.czbiohub.org/data/datasets/opencell-protein-abundance.csv [downloaded April 4th
524 2022]; Hein et al., table S3).

525

526 Next, the number of unbound bait and prey proteins were estimated by subtracting the number of
527 bait-prey complexes[40,66]. In a few instances, the estimated number of bait-prey complexes was
528 higher than that of the total abundance of bait or prey, resulting in a negative number of unbound
529 proteins. This almost only arose when interaction stoichiometry is above 1. These negative values
530 were removed.

531 $$Bait_{Unbound} = Bait_{Total} - Bait::Prey \quad (5)$$

532

533 $$Prey_{Unbound} = Prey_{Total} - Bait::Prey \quad (6)$$

534

535 Copy numbers of bait proteins, prey proteins and bait-prey complexes were next converted to
536 molar concentrations (mol/L) using Avogadro's constant and cell volumes. Below is an example
537 formula for bait concentrations; concentrations prey and bait-prey complexes were calculated in
538 the same way. Dissociation constants were calculated using the resulting concentrations.

539

540 $$[Bait] = \frac{Bait_{unbound}/(6.022 \cdot 10^{23}\ mol^{-1})}{Cell\ Volume} \quad (7)$$

541

542 $$K_d = \frac{[Bait][Prey]}{[Bait::Prey]} \quad (8)$$

543

544 A cell volume of 4.2 $\times 10^{-14}$ Liters was used for yeast PPIs (which were measured in haploid
545 cells)[67]. A cell volume of 2.6 $\times 10^{-14}$ Liters was used for PPIs measured in HeLa cells by Hein et
546 al.[68] Cho et al. measured PPIs and stoichiometries in HEK293 cells, for which no experimental
547 measurement of volume was found. However, the diameter of HEK293 cells was measured to be
548 approximately 13 micrometers (Bionumbers ID: 108893). Therefore, an estimated cell volume of
549 1.15 $\times 10^{-14}$ Liters was calculated from this diameter by approximating the shape of a cell to that
550 of a sphere. The above-mentioned values were all found via the BioNumbers database[69].

551

552 **Validating estimated dissociation constants using values from literature**

553 Experimental dissociation constants from literature were taken from the protein-protein
554 complex subset of the PDBbind database (2020 version)[70,71]. Taxonomic identification numbers
555 were extracted from the PDB file of each database entry. Only interactions involving proteins from
556 budding yeast and human were retained. PDB files which listed more than two proteins were
557 removed, as it would not be possible to assign dissociation constants to the correct protein pair.

558

559 Systematic names or IDs for proteins and genes are absent from the PDBbind database. For this
560 reason, the PDB ID for each complex was used to access their structure's RCSB Protein Data Bank

561 web page and retrieve the Uniprot IDs of the proteins within the structure. Web scraping for the
562 data published in this article was performed on April 23rd 2024, using the *BeautifulSoup4 Python3*
563 package. For yeast PPIs, Uniprot IDs were then converted to systematic ORF IDs using the
564 *org.Sc.sgd.db R* package. For human PPIs, a Uniprot to HGNC symbol conversion table was
565 downloaded (on February 28th 2024) from the "Custom downloads" page of the HGNC website[72].
566 Information on the original source, which is often different from that of the structure, and details
567 on the experimental setup of each dissociation constant measurement is also absent from the
568 database. Although this limits quality control, it is known from curated sets for structure-based
569 modelling of PPI affinities that common methods to measure PPI binding affinities are best for
570 interactions in the micromolar to nanomolar range[42]. Therefore, measurements were limited to
571 dissociation constants no smaller than 1 nM and no greater than 1 mM (1 nM $\leq K_d <$ 1 mM). The
572 final reference dataset contains 505 K_d measurements for human PPIs and 59 for yeast PPIs.
573
574 Estimated K_d values for PPIs were merged with experimental values for the reference dataset. For
575 PPIs in human cell lines, 70 estimated K_d values were mapped to values from literature. Only 4
576 estimated K_d values were mapped to values from literature for yeast PPI.
577
578 **Comparison of mass spectrometry K_d estimates in yeast using PCA K_d estimates**
579 Colony intensities and z-scores were collected from a genome-wide screen for protein-
580 protein interaction using a survival-selection protein-fragment complementation assay (PCA)[73]
581 (table S4 of the referenced article) and filtered to only contain the final 2770 protein-protein
582 interactions (table S3 of the referenced article) reported in the article. Next colony intensities were
583 collected from a proteome-wide abundance screen based on the GFP-PCA method[74] (Data Set S2
584 of the referenced article). GFP-PCA intensities were merged to cellular copy numbers from a
585 compilation of abundances in yeast[30] to generate a linear regression model relating PCA colony
586 intensities to copy numbers. Protein-protein interaction PCA intensities were then rescaled to a
587 range similar to that of the abundance GFP-PCA intensities. The linear regression model was then
588 used to convert protein-protein interaction intensities to numbers of complexes. With the number
589 of copies of complexes and individual protein copy numbers at hand, dissociation constants were
590 estimated as described in the "Estimating protein-protein interaction dissociation constants"
591 section. In total, 1480 K_d estimates were calculated from PCA data, of which 48 also had K_d
592 estimated from mass spectrometry data.

593

594 **Reconstructing a GI network matrix from a PPI binding network**

595 Yeast GI and PPI dissociation constant (K_d) tables were merged such that if genes
596 comprising a GI also code for proteins that interact, the GI score is assigned to that PPI. The final
597 table contains 3265 interactions, each with a GI score and a K_d value. A graph was then generated,
598 using individual proteins as nodes/vertices (V_{PPI}), PPIs as edges (E_{PPI}), and the reciprocal negative
599 logarithm of K_d values as weights (W_{PPI}):

600

601 $$G_{PPI} = \{V_{PPI}, E_{PPI}, W_{PPI}\} \quad (9)$$

602

603 $$w_{PPI}(\{u, v\} \in E_{PPI}) = \frac{1}{-log10(K_{d,u,v})} \quad (10)$$

604

605 Where u, $v \in V_{PPI}$, $K_{d,u,v}$ is the dissociation constant between the interacting proteins u and v, and
606 $w_{PPI}(e)$ is the weight assigned to the edge e (where $e = \{u, v\}$ connects u and v).

607

608 To establish an equation relating weights and GI scores, PPIs were first binned into quantiles
609 according to their weights. The mean weight and mean GI score within each quantile were then
610 calculated and an exponential equation was used to fit the data (See Extended Data Fig. 6a for
611 example with 25 quantiles). The root mean square error (RMSE) of the fit for different numbers
612 of quantiles were tested to find the most optimal (Extended Data Fig. 6b). A dip of RMSE values
613 and narrowing of RMSE spread was observable at around 25 quantiles, which justified its choice
614 as the ideal number of quantiles for the final fitting (Extended Data Fig. 6a, c).

615

616 To mimic the completeness of GI networks, the PPI graph was converted to an "association" matrix
617 similar to that described by Rives and Galitsky[75]. Briefly, the *shortest_path_length* function of the
618 *Networkx* python library was used on G_{PPI}, using W_{PPI} as the weight parameter, to generate a $|V| \times$
619 $|V|$ matrix (d) in which d[i,j] denotes the length of the shortest path between the nodes (proteins) i
620 and j. This all-pairs-shortest-path (APSP) matrix was used as a graph (G_{APSP}), such that:

621

622 $$G_{APSP} = \{V_{APSP}, E_{APSP}, W_{APSP}\} \quad (11)$$

623

$$w_{APSP}(\{x,y\}) = shortest_path_length(u,v) \quad (12)$$

Where $\{u,v\} \in E_{PPI}$, $\{x,y\} \in E_{APSP}$, and both edges represent the same protein interaction. The exponential equation relating means of quantized weights and GI scores was then used on W_{APSP} to generate a new graph (G_{GI_pred}) with weights W_{GI} containing estimated GI scores. The adjacency matrix of G_{GI_pred} was clustered with average linkage hierarchical clustering using Pearson correlation in Cluster 3.0 (*74*).

Analysis of dissociation constants in yeast paralogs

Whole genome duplication (WGD) paralogs were collected from Kellis et al.[76] (Supplementary file 8 of the cited article) and small-scale duplication (SSD) paralogs were collected form VanderSluis et al.[47] Paralogs were separated into 3 categories as described in the main text. The rate of non-synonymous mutations per non-synonymous site (dN) was used as a proxy for sequence divergence. To calculate dN values, paralog pairs were aligned pairwise at the protein level with ClustalW (version 2.1) using the PAM protein weight matrix[77]. Using the *reverse.align* function of the *seqinr* R package, protein-coding nucleic acid sequences were next aligned using protein alignments as a guide[78]. The reverse alignments of each paralog pair were then used to compute dN using the *kaks* function of *seqinr*. Amino acid sequences and coding nucleic acid sequences were downloaded from the SGD sequence archives (http://sgd-archive.yeastgenome.org/sequence/S288C_reference/, March 2024) for the S288C reference strain. A sequence divergence threshold of 0.2 was selected to categorize paralog pairs into "high" and "low" sequence divergence (dN) (Extended Data Fig. 8a). For protein-protein interactions where both interactors have a paralog, the interaction was categorized into "high" when both interactors had a dN value above 0.2.

649 **Mapping GI Modules to PPI Modules**

650 GI clusters were downloaded from an analysis of a near-complete double knockout study
651 in yeast[1] (Data File S6 of the cited article). Yeast protein complexes were downloaded from a
652 species-specific dataset of the ComplexPortal database[57]; protein Uniprot IDs were converted to
653 systematic yeast ORFs using the *org.Sc.sgd.db* R library[79]. All GI clusters were compared to all
654 PPI complexes by their common ORFs, the degree of overlap was scored using the Jaccard index.
655 For each GI cluster, the protein complex which had the highest overlap (Jaccard Index) was
656 retained.

657

658 If a GI cluster overlapped better with connector nodes (proteins outside complexes), it was dropped
659 from the dataset. To validate the significance of the overlap, GI clusters were randomized and
660 mapped to protein complexes. GI clusters were randomized by retaining the size of the clusters
661 and randomly redistributing ORFs. The randomization was performed for 10 000 iterations on a
662 general-purpose computer cluster. A p-value was derived from a z-score in which the mean Jaccard
663 score from the experimental dataset was compared to the distribution of mean Jaccard scores of
664 randomized clusters (Fig. 4c).

665

666 **Identifying Nodes Connecting GI and PPI Modules**

667 The GI and PPI networks used for the analysis of "connector" nodes were taken from the
668 near-complete double knockout screen in yeast and whole-proteome affinity enrichment coupled
669 to mass spectrometry screen described in the *Collecting yeast and human genetic and protein-*
670 *protein interaction data* section. Only negative genetic interactions were used for the analysis in
671 Fig. 4. Because shortest paths are sensitive to noise, we used the moderately stringent cutoff of GI
672 score < -0.08 and p-value < 0.05 recommended by the authors[1]. Identifying nodes (gene/proteins)
673 connecting modules in GI and PPI networks was performed in 4 main steps. First, genetic modules
674 and protein modules are mapped 1-to-1 as described in the *Mapping GI Modules to PPI Modules*
675 section above.

676

677 The 1-to-1 mapping can be expressed mathematically, where M_{GI} and M_{PPI} are each a set
678 containing all modules in the GI and PPI network, respectively, and a module is itself a set of
679 genes/proteins. Connectors can be defined as the sets C_{GI} and C_{PPI}, which contain all the connector

genes/proteins in the GI and PPI networks respectively. Importantly, a gene/protein cannot be both a connector and within a module. In other words:

$$C_{GI} \cap \bigcup M_{GI} = \emptyset \tag{13}$$

$$C_{PPI} \cap \bigcup M_{PPI} = \emptyset \tag{14}$$

We note that here the large union symbol denotes the union of all genes/proteins that are within modules into a single large set. Next, a module $X \in M_{GI}$ is related (mapped) to a module $A \in M_{PPI}$ if and only if $J(X, A) > J(X, C_{PPI})$, and $J(X, A) > J(X, B)$ for all $B \in M_{PPI}$. This relation (mapping) is denoted as R, where $(X, A) \in R$ for all modules X in M_{GI} and for all modules A in M_{PPI}. Note that J denotes the Jaccard index:

$$J(X, A) = \frac{X \cap A}{X \cup A} \tag{15}$$

Second, for all combinations of modules within a network, the shortest path between all pairs of nodes were computed. This was done independently for GI and PPI networks. In mathematical terms, for a graph $G = \{V, E, W\}$ (representing the GI or PPI network), let $\{a, b, c\}$ be a set of genes within module A such that $A = \{a, b, c\}$ and $\{a, b, c\} \subset$ of V, and let $\{x, y, z\}$ be a set of genes within module B such that $B = \{x, y, z\}$ and $\{x, y, z\} \subset V$. Importantly, $B \cap A = \emptyset$. For any pair $(u, v) \in A \times B$, the shortest paths between nodes u and v were computed using Dijkstra's algorithm (*all_short_paths* function of the *networkx* python package).

Third, connector nodes that bridge the same modules in both GI and PPI networks were systematically identified. If the sequence of nodes (i, j, k) is the shortest path between module A and module B in the PPI network (G_{PPI}) and the sequence (k, m, n) is the shortest path between module X and module Y in the GI network (G_{GI}), then $(i, j, k) \cap (k, m, n)$ is the shortest path common to both networks. Note that here modules A and X, as well as B and Y are mapped to each other; $(A, X) \in R$ and $(B, Y) \in R$. Importantly, to avoid including nodes from the source and target modules in the shortest path, nodes belonging to a module were removed from the shortest path.

710

711 Fourth, the overlapping features in both networks *i.e.,* modules and shortest paths connectors, are
712 combined to generate an integrated network. To determine if the extent of the overlap in connectors
713 is due to network topology as opposed to random chance, the approach described above was
714 performed on randomized networks. The networks were randomized by first conserving edges
715 (interactions) within modules and then performing degree preserved randomization for all other
716 edges. Overlapping connectors were identified for 1 000 random GI and PPI networks on a general-
717 purpose computer cluster. A p-value was derived from a z-score in which the number of
718 overlapping connectors from the experimental dataset was compared to the distribution of
719 overlapping connectors from randomized networks (Fig. 4d).

720

Acknowledgments

The authors thank Anastasia Baryshnikova for advice on preprocessing of synthetic genetic array genetic interaction data. We also thank Christian R. Landry, Ben Lehner, Elena Kuzmin, and Brenda Andrews for comments on the manuscript.

Funding

Natural Sciences and Engineering Research Council of Canada – Masters Canada Graduate Scholarship (XCG). Fonds de Recherche du Québec en Santé – Masters Training Scholarship (XCG). Canadian Institute of Health Research grants MOP-G-408523 (AWRS). Natural Sciences and Engineering Research Council (NSERC) of Canada grant RGPIN-2016-06566 (AWRS). Canada Research Chairs (SWM, AWRS). CIHR grant PJT-185976 and NSERC grant RGPIN-03216-2021 (SWM).

Author contributions

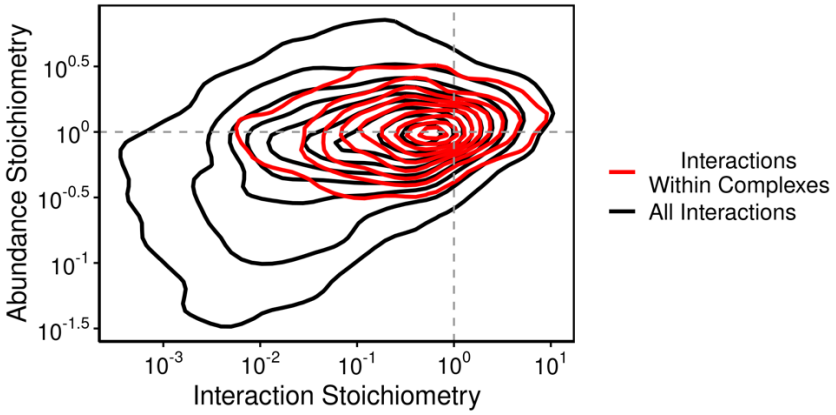
SWM, AWRS and XCG conceptualized this study. XCG generated the computational pipeline. XCG, AWRS and SWM analyzed and interpreted data. XCG, AWRS and SWM wrote the manuscript.

Competing interests

Authors declare that they have no competing interests.

Data and materials availability

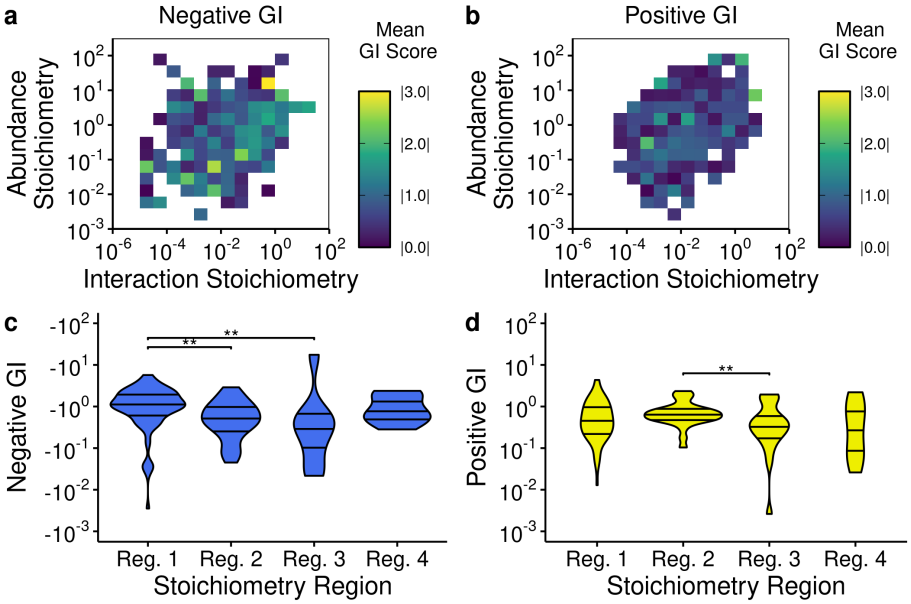
Full clustered Genetic Interaction Matrices are available on Figshare (*42-43*). All final and important intermediate data files are provided online and on GitHub (see link below). Original data sources are cited in Methods. Python, R and bash scripts are available in the following GitHub repository: https://github.com/Xavier-Castellanos-Girouard/PPI_Binding_Epistasis.



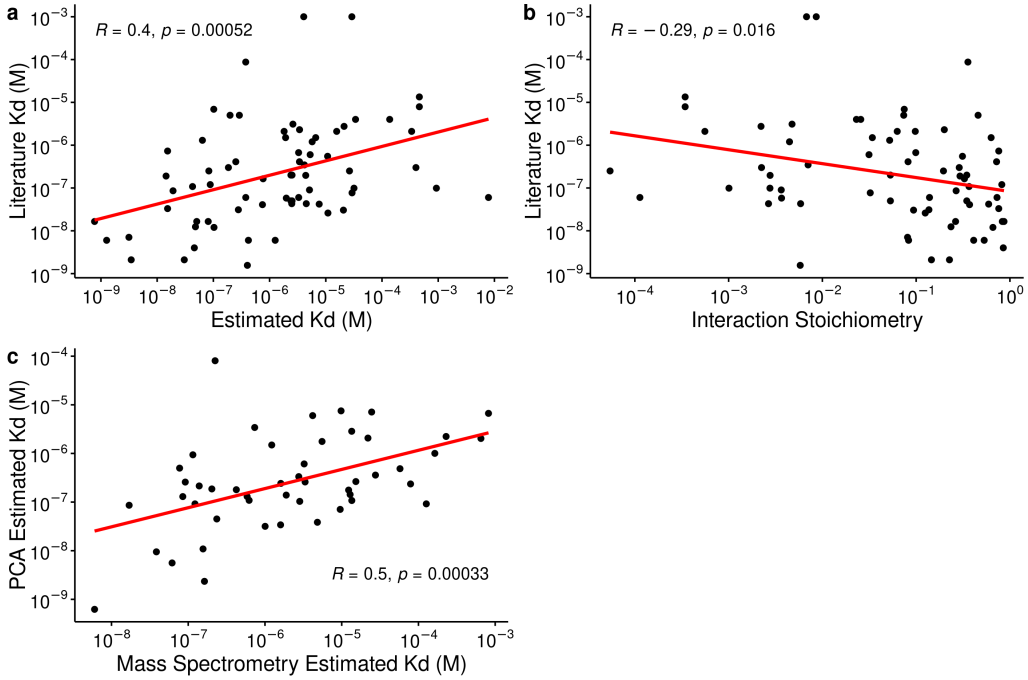
744

745 **Extended Data Figure 1. Stable protein complexes in *Saccharomyces cerevisiae* are enriched**
746 **for 1:1 stoichiometries.** Yeast PPIs occurring between proteins of stable complexes as annotated
747 by the ComplexPortal database (*55*).

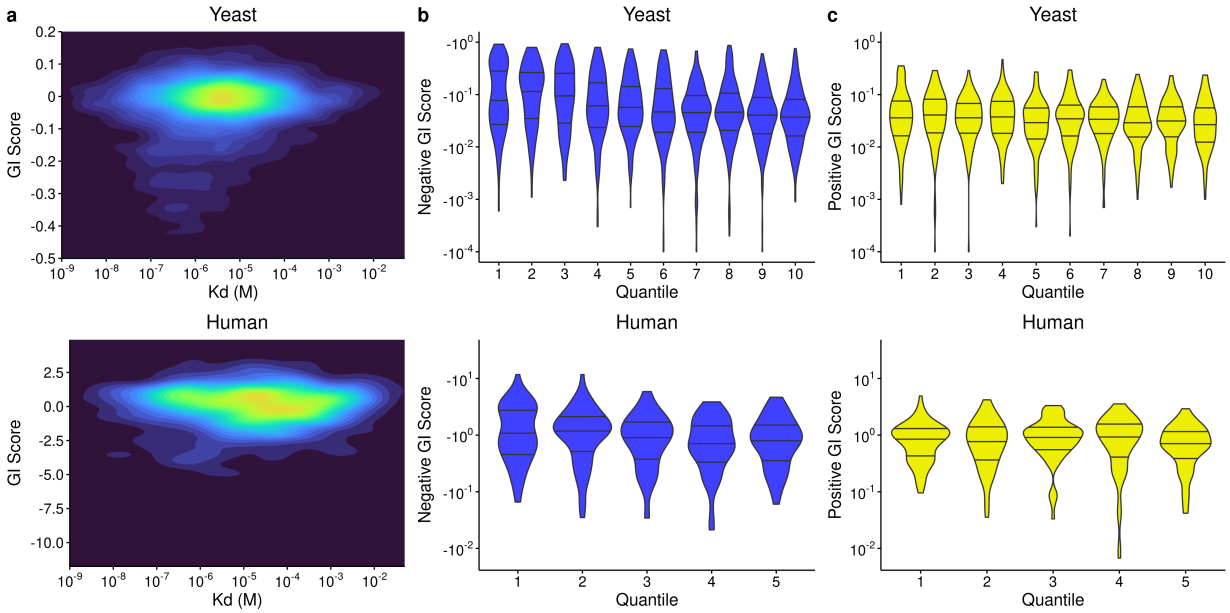
748



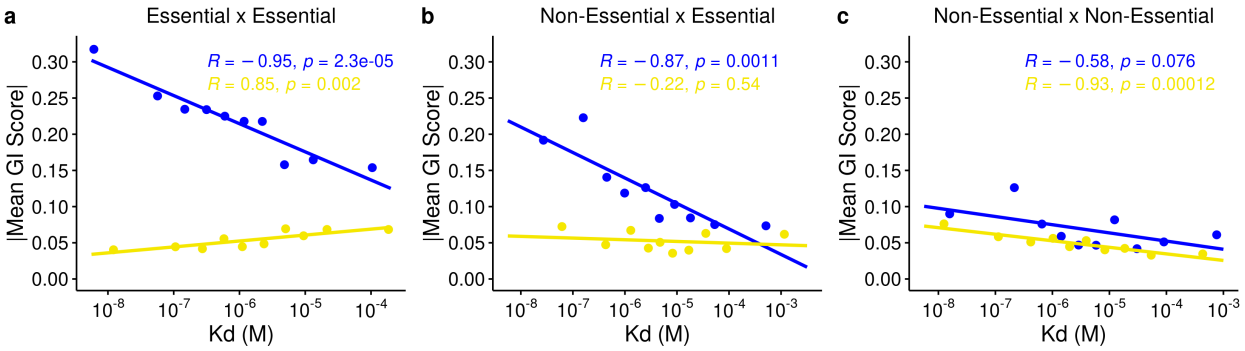
Extended Data Figure 2. The PPI stoichiometry landscape of GIs in the Jurkat cell line. a, b, Mean negative and positive genetic interaction scores for PPIs throughout the stoichiometry plot. GI scores in human were taken from a screen in Jurkat cell lines. **c, d,** Distribution of negative (c) and positive (d) GI scores within the 4 defined regions of the stoichiometry plot. Significance levels are annotated as: ** for $p \leq 0.01$.



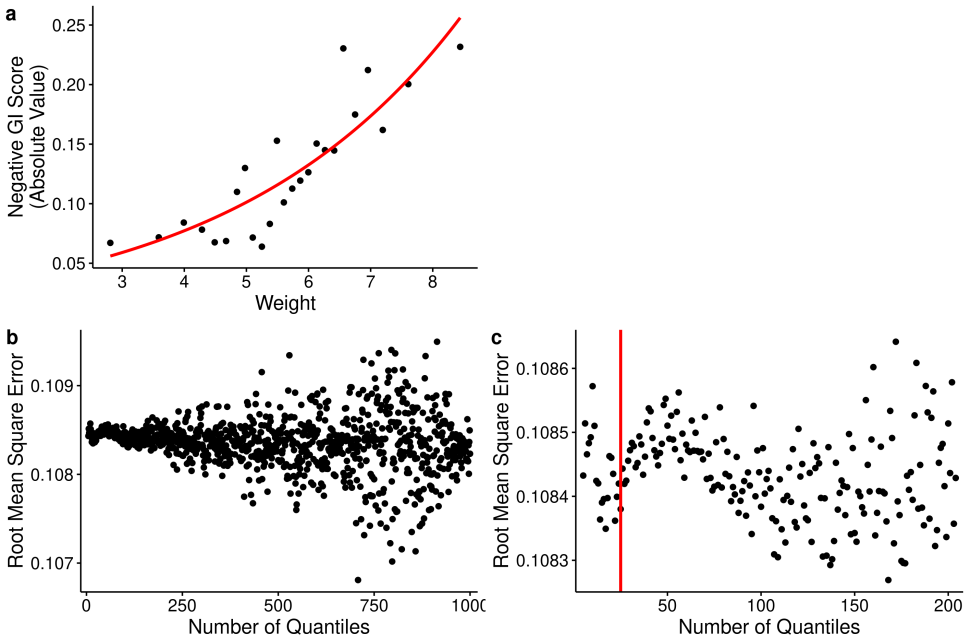
Extended Data Figure 3. Validation of dissociation constants estimated from quantitative mass spectrometry measurements. Pearson correlation was used in all panels. **a,** Estimated PPI dissociation constants estimated from quantitative mass spectrometry data in human cell lines correlate moderately but very significantly with values from literature. **b,** Interaction stoichiometries from quantitative mass spectrometry data in human cell lines correlate weakly with values from literature. **c,** Dissociation constants estimated from quantitative mass spectrometry data in yeast correlate moderately but very significantly with those estimated from protein-fragment complementation (PCA) results.



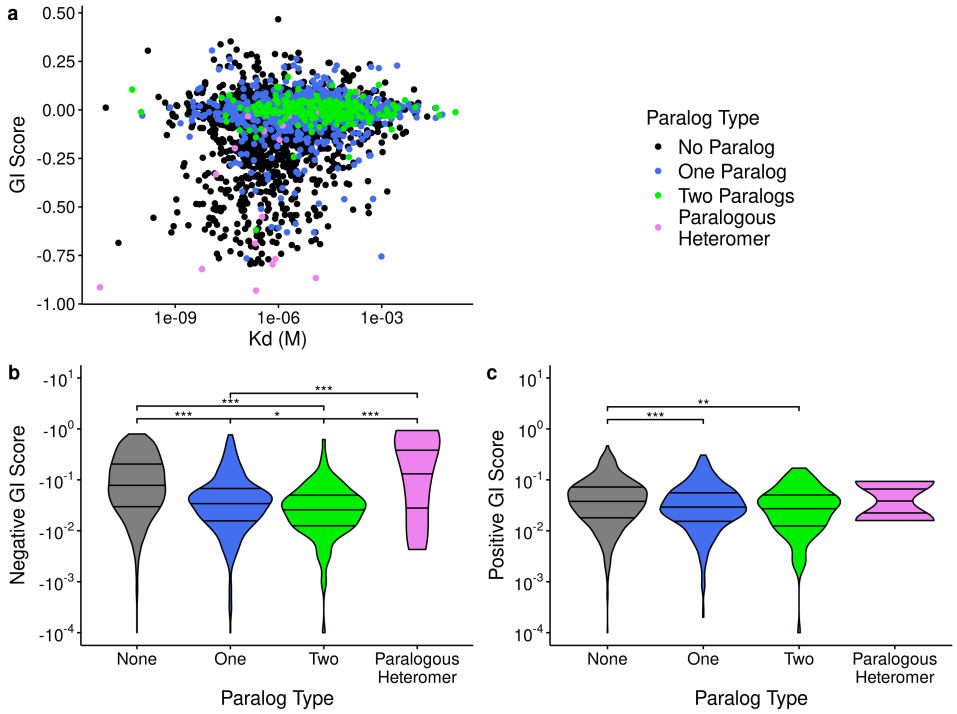
Extended Data Figure 4. High affinity protein-protein interactions are biased towards strong GIs. a, Density plot of genetic interaction scores as a function of their protein-protein interaction affinity (K_d) for yeast and human (K562 cell line). **b,** Distribution of negative GI scores across K_d quantiles (quantile 1 is the lowest K_d; strongest affinity) for yeast and human. **c,** Distribution of positive GI scores across K_d quantiles (quantile 1 is the lowest K_d) for yeast and human.



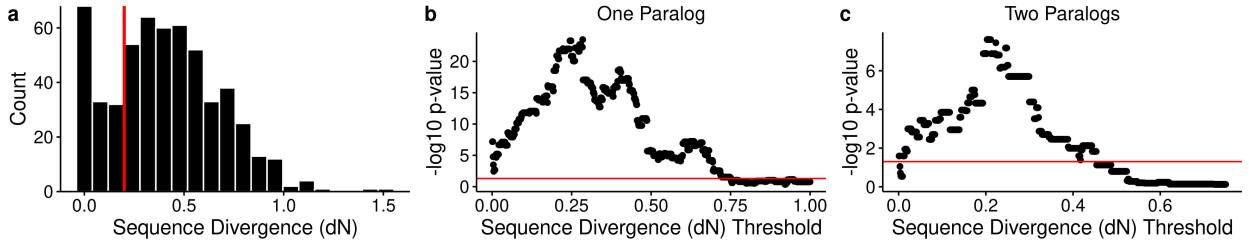
Extended Data Figure 5. Correlations between binned genetic interactions and K_d for genes pairs of different essentialities. a-c, Pearson correlation coefficients between dissociation constants (K_d) and GIs are computed for genes pairs where both genes are essential (a), only one gene is essential (b), or both genes are non-essential (c). *p*-values are computed from two tailed test. All trendlines are fit using a linear regression.



Extended Data Figure 6. Fitting an equation relating binned PPI weights (derived from K_d values) to GI scores. a, Yeast PPIs are binned into 25 quantiles according to their network weights. The average weight and GI score is calculated for every quantile and an exponential equation is used to fit the data. **b,** The fitting procedure described in (a) was tested with different numbers of quantiles to visualize their root mean square errors (RMSE). **c,** A narrowing of RMSE spread and a dip in RMSE values can be observed at around 25 quantiles.



Extended Data Figure 7. Distribution of positive and negative genetic interaction scores according to PPIs separated by their paralog composition. a, Distribution of estimated K_d and GIs (yeast dataset) according to the PPI's paralog composition/type. **b,** Distribution of negative GI scores for different interaction type. **c,** Distribution of positive GI scores for different interaction type.



Extended Data Figure 8. Dividing protein-protein interactions according to paralog sequence divergence. a, Sequence divergence (dN) between paralogous protein/gene pairs retrieved from literature. A dN threshold of 0.2 separates the two visible modes in the distribution. **b, c,** The statistically significant differences in dissociation constants between high divergence and low divergence are robust to the chosen dN threshold. This holds for PPIs in which one protein has a paralog (b) and PPIs in which both proteins have a paralog (c). *p*-values are calculated using Wilcoxon rank sum test. Red line indicates a *p*-value of 0.05.

805 **Extended Data Table 1[a]: Essential Complexes Enriched for Negative GIs**

ComplexPortal ID	Common Name
CPX-1890	Exocyst
CPX-1364	SMC5/6 complex
CPX-1041	DASH complex
CPX-1656	Transcription factor TFIIIC complex
CPX-607	Arp2/3 protein complex
CPX-1186	MIND complex
CPX-1869	Nuclear condensin complex
CPX-1831	Eukaryotic translation initiation factor 3 complex
CPX-2262	20S proteasome core particle
CPX-1895	mRNA cleavage factor complex

806 **[a]Examples were taken from Costanzo et al. (2016)**[1]

807

808 **Extended Data Table 2[a]: Non-Essential Complexes Enriched for Positive GIs**

ComplexPortal ID	Common Name
CPX-779	Elongator holoenzyme complex
CPX-1840	Golgi transport complex
CPX-568	CAF-1 complex
CPX-1197	FAR complex
CPX-1342	Set3C
CPX-1884	HDA complex
CPX-1193	Vacuolar proton-transporting V-type ATPase
CPX-1412	Ubp3-Bre5 ubiquitin hydrolase complex
CPX-1655	i-AAA complex
CPX-1178	Cytoplasmic dynein complex

809 **[a]Examples were taken from Costanzo et al. (2016)**[1]

810

811 **Extended Data Table 3: Correlation coefficients between unbinned K_d and GI values**

Organism	GI Type	Correlation Measure	Coefficient	p-value
Yeast	Negative	Pearson	0.22	1.3×10^{-23}
		Spearman	0.24	1.1×10^{-26}
	Positive	Pearson	0.083	2.8×10^{-3}
		Spearman	0.091	1.0×10^{-3}
Human	Negative	Pearson	0.098	0.11
		Spearman	0.13	0.027
	Positive	Pearson	0.038	0.48
		Spearman	0.019	0.73

812

813

References

1. Costanzo, M. *et al.* A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**, aaf1420–aaf1420 (2016).
2. Costanzo, M. *et al.* Global Genetic Networks and the Genotype-to-Phenotype Relationship. *Cell* **177**, 85–100 (2019).
3. Sharan, R., Ulitsky, I. & Shamir, R. Network-based prediction of protein function. *Molecular Systems Biology* **3**, 88 (2007).
4. Beyer, A., Bandyopadhyay, S. & Ideker, T. Integrating physical and genetic maps: from genomes to interaction networks. *Nat Rev Genet* **8**, 699–710 (2007).
5. Fang, G. *et al.* Discovering genetic interactions bridging pathways in genome-wide association studies. *Nat Commun* **10**, 4274 (2019).
6. Barrio-Hernandez, I. *et al.* Network expansion of genetic associations defines a pleiotropy map of human cell biology. *Nat Genet* **55**, 389–398 (2023).
7. VanderSluis, B. *et al.* Integrating genetic and protein–protein interaction networks maps a functional wiring diagram of a cell. *Current Opinion in Microbiology* **45**, 170–179 (2018).
8. Mani, R., St.Onge, R. P., Hartman, J. L., Giaever, G. & Roth, F. P. Defining genetic interaction. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 3461–3466 (2008).
9. Phillips, P. C. Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* **9**, 855–867 (2008).
10. Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 1193–1198 (2012).
11. Sackton, T. B. & Hartl, D. L. Genotypic Context and Epistasis in Individuals and Populations. *Cell* **166**, 279–287 (2016).

838 12. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–
839 753 (2009).
840 13. Aguirre, L., Hendelman, A., Hutton, S. F., McCandlish, D. M. & Lippman, Z. B.
841 Idiosyncratic and dose-dependent epistasis drives variation in tomato fruit size. *Science* **382**,
842 315–320 (2023).
843 14. Mair, B., Moffat, J., Boone, C. & Andrews, B. J. Genetic interaction networks in cancer
844 cells. *Current Opinion in Genetics & Development* **54**, 64–72 (2019).
845 15. Shen, J. P. & Ideker, T. Synthetic Lethal Networks for Precision Oncology: Promises and
846 Pitfalls. *Journal of Molecular Biology* **430**, 2900–2912 (2018).
847 16. Costanzo, M. *et al.* The Genetic Landscape of a Cell. *Science* **327**, 425–431 (2010).
848 17. Oliver, S. Guilt-by-association goes global. *Nature* **403**, 601–602 (2000).
849 18. Sahni, N. *et al.* Widespread Macromolecular Interaction Perturbations in Human Genetic
850 Disorders. *Cell* **161**, 647–660 (2015).
851 19. Kelley, R. & Ideker, T. Systematic interpretation of genetic interactions using protein
852 networks. *Nat Biotechnol* **23**, 561–566 (2005).
853 20. Hein, M. Y. *et al.* A Human Interactome in Three Quantitative Dimensions Organized by
854 Stoichiometries and Abundances. *Cell* **163**, 712–723 (2015).
855 21. Cho, N. H. *et al.* OpenCell: Endogenous tagging for the cartography of human cellular
856 organization. *Science* **375**, eabi6983 (2022).
857 22. Michaelis, A. C. *et al.* The social and structural architecture of the yeast protein interactome.
858 *Nature* **624**, 192–200 (2023).
859 23. Bandyopadhyay, S., Kelley, R., Krogan, N. J. & Ideker, T. Functional Maps of Protein
860 Complexes from Quantitative Genetic Interaction Data. *PLoS Comput Biol* **4**, e1000065
861 (2008).

862 24. Collins, S. R. *et al.* Functional dissection of protein complexes involved in yeast
863 chromosome biology using a genetic interaction map. *Nature* **446**, 806–810 (2007).
864 25. Braberg, H. *et al.* From Structure to Systems: High-Resolution, Quantitative Genetic
865 Analysis of RNA Polymerase II. *Cell* **154**, 775–788 (2013).
866 26. Braberg, H. *et al.* Genetic interaction mapping informs integrative structure determination of
867 protein complexes. *Science* **370**, eaaz4910 (2020).
868 27. Halabi, N., Rivoire, O., Leibler, S. & Ranganathan, R. Protein Sectors: Evolutionary Units of
869 Three-Dimensional Structure. *Cell* **138**, 774–786 (2009).
870 28. Marks, D. S. *et al.* Protein 3D Structure Computed from Evolutionary Sequence Variation.
871 *PLoS ONE* **6**, e28766 (2011).
872 29. Faure, A. J. *et al.* Mapping the energetic and allosteric landscapes of protein binding
873 domains. *Nature* **604**, 175–183 (2022).
874 30. Ho, B., Baryshnikova, A. & Brown, G. W. Unification of Protein Abundance Datasets Yields
875 a Quantitative Saccharomyces cerevisiae Proteome. *Cell Systems* **6**, 192-205.e3 (2018).
876 31. Lynch, M. & Conery, J. S. The Origins of Genome Complexity. *Science* **302**, 1401–1404
877 (2003).
878 32. Lynch, M. & Force, A. The Probability of Duplicate Gene Preservation by
879 Subfunctionalization. *Genetics* **154**, 459–473 (2000).
880 33. Crow, K. D. What Is the Role of Genome Duplication in the Evolution of Complexity and
881 Diversity? *Molecular Biology and Evolution* **23**, 887–892 (2006).
882 34. Dolinski, K. & Botstein, D. Orthology and Functional Conservation in Eukaryotes. *Annu.*
883 *Rev. Genet.* **41**, 465–507 (2007).
884 35. Horlbeck, M. A. *et al.* Mapping the Genetic Landscape of Human Cells. *Cell* **174**, 953-
885 967.e22 (2018).

886 36. Van Leeuwen, J. *et al.* Exploring genetic suppression interactions on a global scale. *Science*
887 **354**, aag0839 (2016).
888 37. He, X., Qian, W., Wang, Z., Li, Y. & Zhang, J. Prevalent positive epistasis in Escherichia
889 coli and Saccharomyces cerevisiae metabolic networks. *Nat Genet* **42**, 272–276 (2010).
890 38. Segrè, D., DeLuna, A., Church, G. M. & Kishony, R. Modular epistasis in yeast metabolism.
891 *Nat Genet* **37**, 77–83 (2005).
892 39. Xie, X., Sun, X., Wang, Y., Lehner, B. & Li, X. Dominance vs epistasis: the biophysical
893 origins and plasticity of genetic interactions within and between alleles. *Nat Commun* **14**,
894 5551 (2023).
895 40. Diss, G. & Lehner, B. The genetic landscape of a physical interaction. *eLife* **7**, e32472
896 (2018).
897 41. Jarmoskaite, I., AlSadhan, I., Vaidyanathan, P. P. & Herschlag, D. How to measure and
898 evaluate binding affinities. *eLife* **9**, e57264 (2020).
899 42. Kastritis, P. L. *et al.* A structure-based benchmark for protein–protein binding affinity.
900 *Protein Science* **20**, 482–491 (2011).
901 43. Diss, G., Dubé, A. K., Boutin, J., Gagnon-Arsenault, I. & Landry, C. R. A Systematic
902 Approach for the Genetic Dissection of Protein Complexes in Living Cells. *Cell Reports* **3**,
903 2155–2167 (2013).
904 44. Wright, P. E. & Dyson, H. J. Linking folding and binding. *Current Opinion in Structural*
905 *Biology* **19**, 31–38 (2009).
906 45. Castellanos-Girouard, X. *GI_experimental_full.Png*. 11784216 Bytes (figshare, 2024).
907 doi:10.6084/M9.FIGSHARE.27020701.
908 46. Castellanos-Girouard, X. *GI_reconstructed_full.Png*. 6541289 Bytes (figshare, 2024).
909 doi:10.6084/M9.FIGSHARE.27020704.

910 47. VanderSluis, B. *et al.* Genetic interactions reveal the evolutionary trajectories of duplicate
911 genes. *Molecular Systems Biology* **6**, 429 (2010).
912 48. Kuzmin, E. *et al.* Exploring whole-genome duplicate gene retention with complex genetic
913 interaction analysis. *Science* **368**, eaaz5667 (2020).
914 49. Nowak, M. A., Boerlijst, M. C., Cooke, J. & Smith, J. M. Evolution of genetic redundancy.
915 *Nature* **388**, 167–171 (1997).
916 50. Vavouri, T., Semple, J. I., Garcia-Verdugo, R. & Lehner, B. Intrinsic Protein Disorder and
917 Interaction Promiscuity Are Widely Associated with Dosage Sensitivity. *Cell* **138**, 198–208
918 (2009).
919 51. Rice, A. M. & McLysaght, A. Dosage-sensitive genes in evolution and disease. *BMC Biol*
920 **15**, 78 (2017).
921 52. Papp, B., Pál, C. & Hurst, L. D. Dosage sensitivity and the evolution of gene families in
922 yeast. *Nature* **424**, 194–197 (2003).
923 53. Ascencio, D. *et al.* Expression attenuation as a mechanism of robustness against gene
924 duplication. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2014345118 (2021).
925 54. Qian, W., Liao, B.-Y., Chang, A. Y.-F. & Zhang, J. Maintenance of duplicate genes and their
926 functional redundancy by reduced expression. *Trends in Genetics* **26**, 425–430 (2010).
927 55. Michnick, S. W. & Levy, E. D. The modular cell gets connected. *Science* **375**, 1093–1094
928 (2022).
929 56. Pereira-Leal, J. B., Levy, E. D. & Teichmann, S. A. The origins and evolution of functional
930 modules: lessons from protein complexes. *Phil. Trans. R. Soc. B* **361**, 507–517 (2006).
931 57. Meldal, B. H. M. *et al.* Complex Portal 2018: extended content and enhanced visualization
932 tools for macromolecular complexes. *Nucleic Acids Res* **47**, D550–D558 (2019).
933 58. Tamanoi, F. Ras Signaling in Yeast. *Genes & Cancer* **2**, 210–215 (2011).

934 59. Cazzanelli, G. *et al.* The Yeast Saccharomyces cerevisiae as a Model for Understanding RAS
935 Proteins and their Role in Human Tumorigenesis. *Cells* **7**, 14 (2018).
936 60. Ho, J. & Bretscher, A. Ras Regulates the Polarity of the Yeast Actin Cytoskeleton through
937 the Stress Response Pathway. *MBoC* **12**, 1541–1555 (2001).
938 61. Xie, Y. & Miao, Y. Polarisome assembly mediates actin remodeling during polarized yeast
939 and fungal growth. *Journal of Cell Science* **134**, jcs247916 (2021).
940 62. Bonomelli, B., Busti, S., Martegani, E. & Colombo, S. Active Ras2 in mitochondria
941 promotes regulated cell death in a cAMP / PKA pathway-dependent manner in budding
942 yeast. *FEBS Letters* **597**, 298–308 (2023).
943 63. Han, J.-D. J. *et al.* Evidence for dynamically organized modularity in the yeast protein–
944 protein interaction network. *Nature* **430**, 88–93 (2004).
945 64. Usaj, M. *et al.* TheCellMap.org: A Web-Accessible Database for Visualizing and Mining the
946 Global Yeast Genetic Interaction Network. *G3 Genes|Genomes|Genetics* **7**, 1539–1549
947 (2017).
948 65. The UniProt Consortium *et al.* UniProt: the Universal Protein Knowledgebase in 2023.
949 *Nucleic Acids Research* **51**, D523–D531 (2023).
950 66. Celaj, A. *et al.* Quantitative analysis of protein interaction network dynamics in yeast.
951 *Molecular Systems Biology* **13**, 934 (2017).
952 67. Jorgensen, P., Nishikawa, J. L., Breitkreutz, B.-J. & Tyers, M. Systematic Identification of
953 Pathways That Couple Cell Growth and Division in Yeast. *Science* **297**, 395–400 (2002).
954 68. Zhao, L. *et al.* Intracellular water-specific MR of microbead-adherent cells: the HeLa cell
955 intracellular water exchange lifetime. *NMR in Biomedicine* **21**, 159–164 (2008).
956 69. Milo, R., Jorgensen, P., Moran, U., Weber, G. & Springer, M. BioNumbers—the database of
957 key numbers in molecular and cell biology. *Nucleic Acids Research* **38**, D750–D753 (2010).

958 70. Liu, Z. *et al.* PDB-wide collection of binding data: current status of the PDBbind database.
959 *Bioinformatics* **31**, 405–412 (2015).
960 71. Wang, R., Fang, X., Lu, Y. & Wang, S. The PDBbind Database: Collection of Binding
961 Affinities for Protein−Ligand Complexes with Known Three-Dimensional Structures. *J.*
962 *Med. Chem.* **47**, 2977–2980 (2004).
963 72. Seal, R. L. *et al.* Genenames.org: the HGNC resources in 2023. *Nucleic Acids Research* **51**,
964 D1003–D1009 (2023).
965 73. Tarassov, K. *et al.* An in Vivo Map of the Yeast Protein Interactome. *Science* **320**, 1465–
966 1470 (2008).
967 74. Levy, E. D., Kowarzyk, J. & Michnick, S. W. High-Resolution Mapping of Protein
968 Concentration Reveals Principles of Proteome Architecture and Adaptation. *Cell Reports* **7**,
969 1333–1340 (2014).
970 75. Rives, A. W. & Galitski, T. Modular organization of cellular networks. *Proc. Natl. Acad. Sci.*
971 *U.S.A.* **100**, 1128–1133 (2003).
972 76. Kellis, M., Birren, B. W. & Lander, E. S. Proof and evolutionary analysis of ancient genome
973 duplication in the yeast Saccharomyces cerevisiae. *Nature* **428**, 617–624 (2004).
974 77. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence
975 alignments using Clustal Omega. *Molecular Systems Biology* **7**, 539 (2011).
976 78. Charif, D. & Lobry, J. R. SeqinR 1.0-2: A Contributed Package to the R Project for
977 Statistical Computing Devoted to Biological Sequences Retrieval and Analysis. in *Structural*
978 *Approaches to Sequence Evolution* (eds. Bastolla, U., Porto, M., Roman, H. E. &
979 Vendruscolo, M.) 207–232 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2007).
980 doi:10.1007/978-3-540-35306-5_10.

981 79. Carlson, M. org.Sc.sgd.db. Bioconductor
982 https://doi.org/10.18129/B9.BIOC.ORG.SC.SGD.DB (2017).
983 80. Maslov, S. & Sneppen, K. Specificity and Stability in Topology of Protein Networks.
984 *Science* **296**, 910–913 (2002).

985