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

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

1. Department of Biochemistry, Université de Montréal; Montreal, Quebec, Canada
2. Robert-Cedergren Center for Bioinformatics and Genomics, Université de Montréal; Montreal, Quebec, Canada

*Correspondence: 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[1–5]. 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[1,6].

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[7,8]. Epistasis is a major source of complexity in predicting genotype-phenotype relationships, making the elucidation of its molecular mechanism a significant endeavor[2,9–12]. 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[1,13–15]. PPIs are also useful for inferring gene functions, linking proteins of unknown, to those of known function[3,16]. Finally, about two-thirds of mutations that cause Mendelian human diseases also disrupt PPIs, linking the integrity of PPI networks to disease[17]. 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[6,18].

Recent quantitative proteomic studies provide clues of a deeper physical relationship between PPI and GI networks[19–21]. 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)[19]. The ranges of these binding and abundance stoichiometries were remarkably broad, between ~10^{-7}- 10^{1} (Figure 1d, regions 2 and 4)[19]. 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[1,6].

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[20]. 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[19,20]. GI networks also show a highly modular organization and the correspondence of modular structures between PPI and GI networks has been established[1,6,22]. 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[23–25]. 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[26–28].

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).

83 # Results

84

85 **Epistasis varies with protein abundance and PPI stoichiometries**

86

87 Since a complete systematic GI network has been constructed for the budding yeast
88 *Saccharomyces cerevisiae* (henceforth "yeast), we chose to make our initial comparison of PPI
89 binding and epistatic strengths for this organism[1]. We first constructed a protein abundance versus
90 PPI binding protein stoichiometry map based on data from a recent affinity purification-mass
91 spectrometry and a meta-analysis of protein abundance measurements[19–21,29]. The distribution of
92 yeast PPIs in the stoichiometry plot is qualitatively similar to that of human proteins (Figure 1b,
93 c), the major difference being a higher frequency of low stoichiometry interactions for human
94 proteins (Figure 1a, b). This higher frequency of weak interactions could reflect relaxation of
95 selection pressure that occurred in the transition from unicellular to multicellular organisms as a
96 consequence of reduced population, allowing for the increased rate of emergence of new genomic
97 features, such as gene duplicates and accrual of new gene functions[30–32]. It could also reflect the
98 increase in requirements of connector PPIs to coordinate activities of the expanded number of
99 functional modules in metazoans. Furthermore, we also found that region 1 (Figure 1d) in the yeast
100 stoichiometry plot is enriched for stable complexes (Supplementary Figure 1), consistent with
101 findings for human proteins[19]. Notwithstanding the evolutionary distance between the two
102 organisms, the similarity in their stoichiometry maps may reflect the conservation of core cellular
103 processes, particularly the molecular mechanisms that mediate them, across yeast and humans[33].

104

105 We next sought to explore how GIs distribute across the PPI stoichiometry map. To this end, we
106 merged the GI and PPI networks using the criterion that if the genes comprising a GI also code for
107 proteins that interact, the GI score was assigned to that PPI. Yeast GIs were taken from a near-
108 complete digenic knockout screen[1] and human GIs were taken from the largest double knockdown
109 study in human cell lines (Jurkat and K562) to date[34]. Of note, the measured GI network in human
110 cell lines is only comprised of 472 genes and consequently only about 0.05% of all possible double
111 knockouts were measured, a much smaller sample than the data for yeast. By visualizing average
112 GI strengths throughout the stoichiometry plots, we observed that negative GIs are strongest for
113 PPIs near 1:1 interaction and abundance stoichiometries (Figure 1e, left panel). This observation

114 is consistent with previous analyses in yeast showing that negative GIs are enriched between genes
115 whose proteins assemble into complexes[1]. Some examples are highlighted in Table S1. This
116 pattern also holds in human cells (Figure 2f, left panel; Supplementary Figure 2a), but to a lesser
117 extent, likely due to the smaller sample size. In contrast, the strength of positive GIs seems to be
118 distributed more uniformly across the stoichiometry map. This suggests that positive GIs are not
119 predominantly driven by PPI and may be better explained in combination with other mechanisms.
120 Indeed, suppression mechanisms involving mRNA decay and protein degradation, as well as
121 specific patterns of disruption in metabolic pathways have been proposed as important sources of
122 positive GIs[1,35–37].
123
124 To further investigate the distribution of GI strength throughout the stoichiometric landscape, we
125 divided interactions into 4 regions previously outlined by Hein *et al.* (Figure 2c): region 1 consists
126 of interacting protein pairs of equal abundances and 1:1 binding ratios (enriched for stable
127 complexes); region 2 consists of protein pairs of equal abundances but low binding ratios (enriched
128 for transient interactions); region 3 consists of protein pairs of unequal abundances and 1:1 binding
129 ratios, and region 4 consists of protein pairs of unequal abundances and low binding ratios[19]. As
130 suggested by the stoichiometry plots, the distribution of negative GIs within region 1 is
131 significantly higher than that of other regions ($P < 0.01$, Wilcoxon rank-sum test) (Figure 1f, g,
132 left panels; Supplementary Figure 2c). The distribution of positive GI values is also highest in
133 region 1 in yeast and human cell lines (Figure 1f, g, right panels), in agreement with previous
134 observations that complexes, composed mostly of non-essential genes, are enriched for positive
135 GIs (Table S2)[1]. Interestingly, region 3 that is characterized by high interaction stoichiometry and
136 an excess of affinity purified *versus* bait protein show the lowest number of GIs among the four
137 regions (Figure 1g and h; Supplementary figure 2c), reflecting that the number of protein
138 complexes formed is not the sole determinant of GI strength. Taken together, these results show
139 that there is robust overlap between GI and PPI networks.

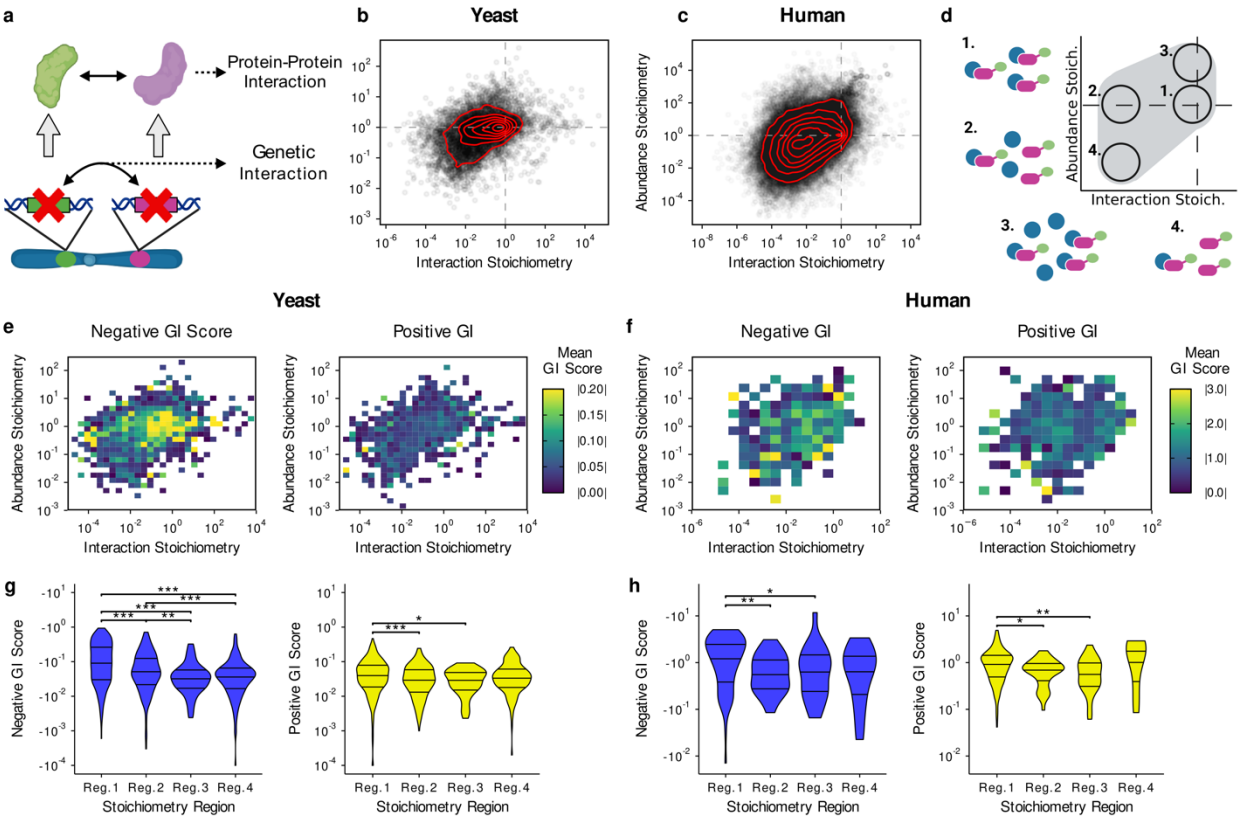


Figure 1. Strong negative GIs map to PPIs of equal protein stoichiometry. a, Genetic interactions are generated by comparing fitness defects in genotypes with single or double gene knockout/knockdowns, and protein-protein interactions are defined as physical contacts between proteins. **b, c,** Protein-protein 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. **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 $p \leq 0.001$.

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

153

154 Substantial efforts have been made to model how the physicochemical properties of proteins such
155 as folding stability and PPI binding energies drive epistasis[28,38,39]. Whilst these approaches explain
156 how epistasis arises from point mutations for single proteins or pairs of proteins in PPIs, the extent
157 to which the strengths of GIs for whole gene knockout (or knock down) could be explained by
158 binding free energies is unknown.

159

160 Interaction stoichiometry can provide an indirect measure of PPI strength if protein pairs are of
161 equal abundance, however, interaction stoichiometry is inherently limited by the abundance of the
162 bait in an affinity purification-mass spectrometry (AP-MS) experiment[19,20]. This is reflected in the
163 triangular shape of stoichiometry maps (Figure 1b-d), where protein pairs of very low abundance
164 stoichiometry have a strong tendency to not form high interaction stoichiometry interactions
165 (Figures 1a, b). Abundance is therefore a confounding factor when using interaction stoichiometry
166 as a proxy for binding strength.

167

168 To obtain direct measures of PPI binding affinities, we devised a simple approach to estimate
169 equilibrium dissociation constants (K_d) from protein stoichiometries and protein abundances
170 measured by mass spectrometry (*Methods*). Briefly, the number of complexes formed by a pair of
171 proteins was estimated using the measured interaction stoichiometry and cellular abundances. The
172 number of complexes were then used to estimate the number of unbound copies of each interactor.
173 Finally, copy numbers were converted to concentrations, from which second order equilibrium
174 dissociation constants were calculated (K_d; *Methods*, equations 1 and 2). We validated our results
175 by comparing the estimated values to *in vitro* measurements of dissociation constants from
176 literature and found a moderate but significant correlation of 0.40 (Pearson's r, *p*-value =
177 5.2×10^{-4}; Supplementary Figure 3a), which was comparatively better than that of interaction
178 stoichiometry (Pearson's r = -0.29, *p*-value = 0.016; supplementary figure 3b). The theoretically
179 expected relationship between K_d and interaction stoichiometry is inversely proportional. It has
180 been noted that important criteria to assess reliability (*i*.e., controlling for titration and
181 demonstrating equilibrium) in equilibrium binding experiments often go unreported throughout
182 literature, and different experimental approaches can sometimes yield non-overlapping values of

183 K_d[40,41]. Moreover, experiments are generally performed *in vitro* and under conditions that do not
184 reflect native cellular environments. Considering these important sources of variability, we
185 reasoned that our values fall within an acceptable range of precision. To our knowledge, this is the
186 first report of a proteome-wide estimate of *in vivo* binding affinities.
187
188 The density heat map relating GI scores and K_d shows a visible bias towards negative GIs for high
189 affinity interactions in yeast (Figure 2a). This bias is also apparent in the data from human proteins,
190 however, the sparsity of GI data for human genes makes this relationship less clear (Figure 2b).
191 To quantitatively evaluate the relationship between K_d and GI strength, we binned the PPIs into
192 quantiles according to K_d and calculated the mean GI strength within each bin. The mean strengths
193 of negative GIs correlate remarkably and significantly with K_d for both yeast (Figure 2c;
194 Spearman's $\rho = -0.98$, *p*-value = 2.2×10^{-5}) and human PPIs (Figure 2d; Spearman's $\rho = 0.95$,
195 *p*-value = 0.014). A weak but significant correlation is also observable for positive GIs in yeast,
196 which is likely driven by an enrichment of significant positive GIs within complexes composed of
197 non-essential genes (Supplementary Figure 4c; Table S2)[1,6]. At the molecular level, this trend
198 could be explained by a mechanism where the absence of a single protein (single gene knockout)
199 is enough to completely disrupt the assembly of a complex in which it is embedded[42]. Specifically,
200 in protein biochemistry, the presence of a binding partner can increase the folding stability of a
201 protein by providing nucleation sites for its unfolded species, an effect that is heightened for
202 stronger PPI[43]. The single gene knockout could destabilize the other gene, consequently, the
203 double knockout is not substantially more drastic than the single knockout.
204
205 The strong correlation of PPI and negative GI suggests that PPI binding affinity is a major driver
206 of negative epistasis between the genes that code for the interacting proteins. These results imply
207 that one network could be predicted from the other. To test this idea, we devised an approach to
208 reconstruct GI network matrices from binding data alone (*Methods*). We observed that the clusters
209 along the diagonal of a matrix reconstructed from PPI binding data in yeast qualitatively matches
210 that of experimental results in genetic screens (Figure 2ef) [1].Overall, these results indicate that
211 there exists a strong relationship between the quantitative features of GI and PPI networks, and
212 that the modular structure of quantitative PPI networks seem to match that observed in GI
213 networks.

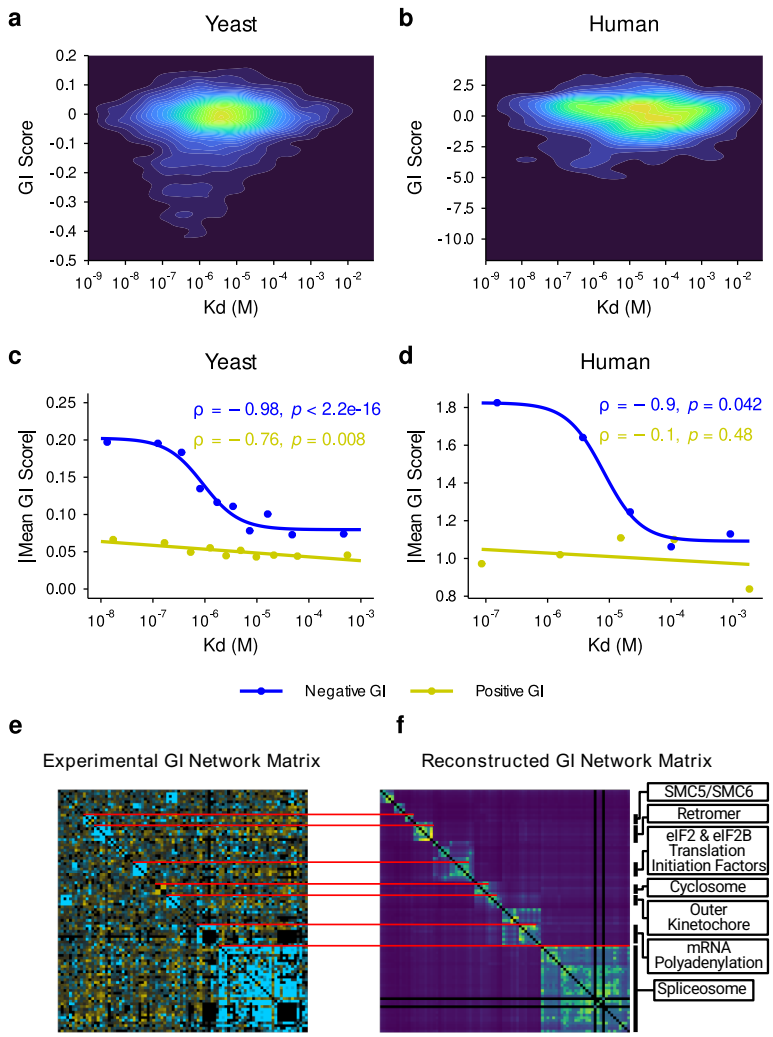


Figure 2. Genetic interactions correlate with PPI binding affinity. a, b, Density heat map of scores as a function of dissociation constants (K_d) for yeast PPIs (a), and the human K562 cell line (b) data. **c, d,** Binning PPIs according to K_ds reveals a significant monotonic relationship between K_d and negative GIs. Negative GIs versus K_d fit to a sigmoidal equation, and for positive GIs versus K_d relationship, a linear equation. Blue dots and trend line correspond to negative GI scores, and yellow to positive GI scores. *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 Table S3. **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[44,45].

227 **Redundancy and divergence of paralogs are reflected in genetic epistasis and PPI affinities**
228
229 In our results above, we hypothesized that PPI, specifically binding affinity (K_d), is an accurate
230 proxy for cellular function and thus, it quantitatively reflects epistasis across the genome. Since
231 the number of protein complexes is a function of abundance and K_d, this poses a specific prediction
232 that genomic events resulting in a maintained decrease or increase in abundance over macro-
233 evolutionary time scales will be accompanied by a change in K_d. In particular, gene duplication
234 events, which could effectively double the number of cellular copies of a protein, could relax
235 selection on K_d and still maintain the same number of complexes. Correspondingly, due to the
236 relaxed selection on K_d, we also expect a reduction in genetic interaction.
237
238 To test this hypothesis, we identified paralogs in our dataset and grouped them into four distinct
239 types: 1) PPIs in which neither protein has a paralog, 2) PPIs in which one of the two (but not both)
240 proteins has a paralog, 3) PPIs in which both proteins have a paralog, and 4) PPIs in which a
241 protein interacts with its own paralog (paralogous heteromers); (Figure 3). Consistent with our
242 expectation and previous analyses of GIs in paralogs, duplicated genes show weaker GIs with other
243 genes than non-duplicates genes (Figure 3b)[46]. This effect seems to be additive overall, as gene
244 pairs where both genes have paralogs show weaker GIs (Figure 3b; green). Remarkably, a similar
245 trend is observed for binding affinity, where proteins encoded by duplicated genes tend to form
246 weaker (high K_d) PPIs (Figure 3c).
247
248 Intriguingly, two distinct modes are observable in the distribution of K_ds for interactions in which
249 both proteins have a paralog (Figure 3c). This result is reminiscent of results from analyses of tri-
250 genic GIs involving duplicated genes, in which functionally redundant paralogs form more trigenic
251 interactions (as opposed to digenic interactions) than functionally divergent paralogs, thereby
252 resulting in a distinct bimodal distribution[47]. To test whether the separation of paralogs into
253 redundant and divergent groups drives the bimodality observed in the K_ds, we computed the
254 sequence divergence (defined here as the number of non-synonymous mutations per non-
255 synonymous site) between genes and their paralogs and separated them accordingly into "high"
256 and "low" divergence paralogs (supplementary figure 7). For interactions in which both proteins
257 have a paralog, highly divergent paralogs participate in stronger interactions (lower K_d) than

258 paralogs with similar amino acid sequences (Figure 4d); this is also true for interactions in which
259 only one protein has a paralog (Figure 4c).
260
261 It is known that some duplicated genes have conserved sequences and functions over evolutionary
262 time[48,49]. However, protein abundance is expected to increase in the short term after gene
263 duplication events, resulting in an imbalance of concentrations of protein complex components,
264 which in turn negatively impacts fitness[50,51]. To reconcile functional conservation and
265 concentration imbalance, it has been proposed that for short evolutionary time scales, post-
266 translational expression attenuation is an important mechanism for robustness against an abrupt
267 abundance increase following gene duplication and, over larger time scales, reduced expression
268 allows for maintenance of redundant duplicated genes[52,53].
269
270 However, reduced expression is not observed for all gene duplication events, suggesting other
271 mechanisms may be at play[53]. In light of our results, we propose a model to explain how, on longer
272 evolutionary time scales, protein complexes adapt to increased expression of their subcomponents
273 after gene duplication (Figure 4e). We hypothesize that following a duplication event, paralogous
274 genes that conserve their sequence will conserve their interaction partners but participate in weaker
275 interactions overall to balance the binding equilibrium such that the number of complexes stays
276 the same (Figure 4e, Fate 1). This can happen, for instance, when mutations occur in a few
277 important amino acids at the binding interface or allosteric sites of paralogs, or if their interacting
278 partners diverge such that their binding affinity for the duplicated genes is decreased. Furthermore,
279 the subset of highly divergent paralogs participating in high affinity interactions (Figure 4d) can
280 be explained by a second scenario, where one of the duplicated genes diverges and loses (or
281 drastically reduces the affinity for) interacting partners while the other maintains high affinity
282 interactions (Figure 4e; Fate 2). This model is both compatible with previous works showing that
283 abundance is decreased following gene duplication[52,53] and explains the tendency of highly
284 divergent paralogs to make high affinity PPIs.

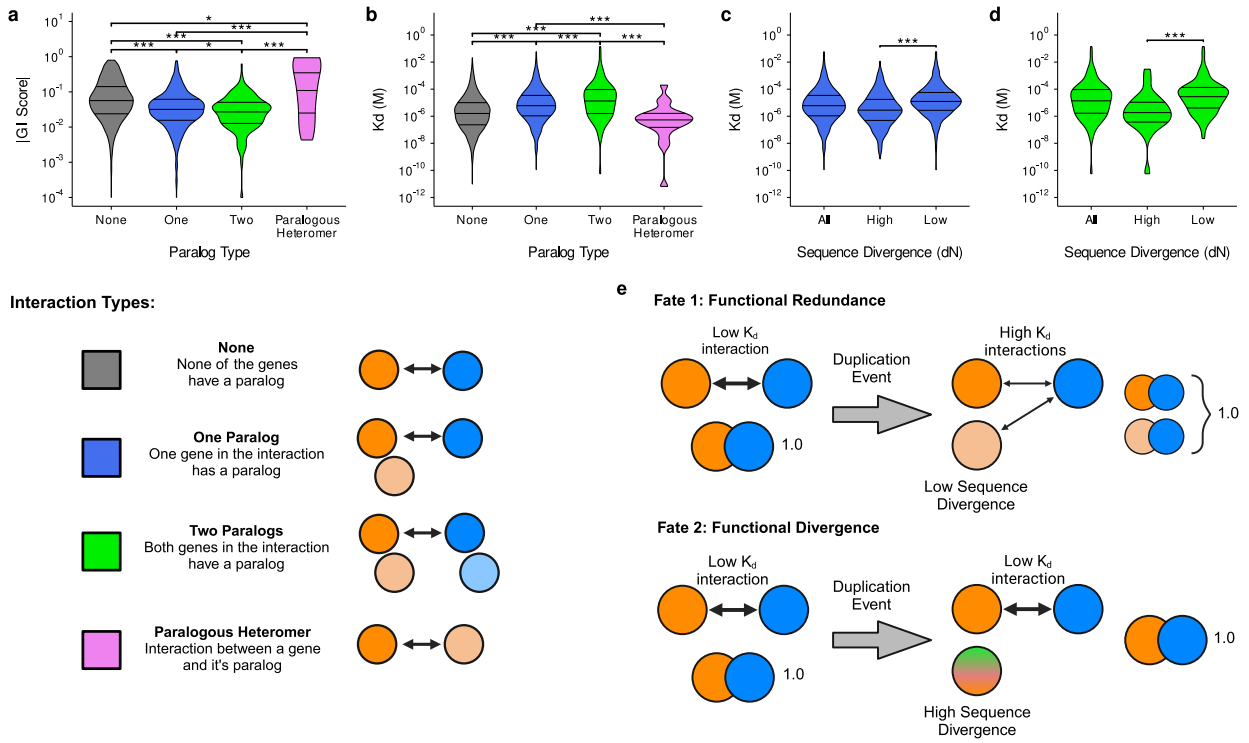


Figure 3. Functionally redundant paralogs participate in low binding affinity PPIs. a, GI scores for gene pairs which have one or two paralogs are lower than those without paralogs. Distributions for individual GI type (positive or negative) are included in supplementary figure 6. **b,** Dissociation 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.

296 **Compositions of module and connector regions of the PPI network overlap with**
297 **corresponding GI network regions**
298
299 The classical interpretation of GI and PPI overlap mainly focuses on GIs within and between
300 modules[4,6,18]; commonly referred to as the "within pathway" and "between pathway" models,
301 respectively[18]. Indeed, protein-protein interaction modules ("complexes") are reflected in the
302 functional modules in GI networks[1,6,22]. However, not all proteins that participate in strong
303 interactions form stable complexes of specific function; some participate in transient interactions
304 that "connect" protein complexes together[54]. The degree to which these "connectors" in PPI
305 networks are reflected in GI networks has not been explored. We hypothesize that the ability of
306 certain proteins to connects complexes in PPI networks also holds for GI networks. In other words,
307 if a protein connects two complexes (modules) in the PPI network, it will likely connect the same
308 two functional modules in GI networks.
309
310 At the cellular level, functional modules involving proteins are typically metabolic pathways,
311 signaling pathways or protein complexes[55]. However, we specifically limit our analysis of protein
312 modules to solely include stable protein complexes in yeast, as defined by the ComplexPortal
313 database[56]. Furthermore, GI modules were taken from an analysis of the double knockout GI
314 network used in the previous sections[1]. For both GIs and PPIs, connectors were defined as the set
315 of genes or proteins that did not fall into a module. As a preliminary step, we sought to determine
316 the extent to which proteins/genes in modules overlapped between the two networks using this
317 classification. We observe a significant overlap between the modules and for connectors for both
318 networks (P-value = 1×10^{-49}, hypergeometric test) (Figure 4b, left panels).
319
320 *GI modules map directly to PPI modules*
321
322 We next sought to determine whether GI modules could be mapped one-to-one with protein
323 complexes. To this end, we evaluated the similarity in gene composition of all possible GI module
324 and PPI module pairings and identified module pairs with the highest degree of overlap (*methods*).
325 Strikingly, of the 112 GI modules, 88 could be mapped to a protein complex (Figure 4B, upper left
326 panel), which is significantly more than would be expected by chance (hypergeometric test P-

327 value = 7×10^{-160}, Figure 4C). This analysis is complementary to that of Costanzo et al.[1], in that we
328 use a set of manually curated complexes, as opposed to GO enrichment, to demonstrate
329 correspondence between GI and PPI modules. The remaining GI modules are likely attributable to
330 pathways not involving stable complexes. For instance, a GI module composed of ERG13 and
331 MVD1 was determined to be enriched for steroid and isoprenoid synthesis but was not mapped to
332 any known complex. It is also important to note that the large fraction of non-overlapping protein
333 complexes reflect the redundancy inherent to the ComplexPortal database, in which several
334 variants or subunits of a complex are included as separate entries.
335
336 *Connector proteins in PPI networks are analogous to connector genes in GI networks*
337
338 The ability of connector proteins to bridge protein complexes has been highlighted in recent mass
339 spectrometry screens of protein-protein interactions in human cells[19,20,54]. We reasoned that if most
340 GI modules overlap strongly with protein complexes, some GI connectors may have topological
341 properties similar to that of PPI connectors. The ability to bridge protein complexes being the most
342 salient topological property of protein connectors, we focused on systematically identifying this
343 motif network-wide for both GI and PPIs. Towards this aim, we devised an algorithm consisting
344 of four main steps as illustrated in Fig. 4a for a single pair of modules from each network: 1)
345 Mapping GI and PPI modules 1:1 as described above, 2) Finding the shortest paths between all
346 pairs of modules; this was done independently in GI and PPI networks, 3) Identifying connector
347 nodes (genes or proteins) in the shortest paths that are common between both networks, and 4)
348 combining the common components into an integrated network. We identified 555 shortest path
349 connectors common to both networks (Figure 4b, lower right panel). Comparing this degree of
350 overlap to that of randomized networks, we find that our results are extremely unlikely to arise by
351 chance (hypergeometric test *P*-value = 4×10^{-53}) (Figure 1D). Overall, our results demonstrate that
352 connectors bridging modules are a topological feature shared by both GI networks and PPI
353 networks.

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

355

356 Previous analyses combining GIs and protein complexes have focussed on GIs within and between
357 complexes and the functional interpretations thereof[1,6,18]. Our approach to integrating PPI and GI
358 networks into a "module-connector-module" structure could reveal novel functional relationships
359 between the two. In fact, our integrated network reveals a core set of modules shared between both
360 networks and highlights known functionality of well-studied connectors. For instance, a
361 subnetwork of our integrated network correctly identifies the karyopherin kap123p as an important
362 bridge between the nuclear pore and various complexes that are transported and/or localized in the
363 nucleus (figure 4e). The integrated network also places *RAS2* (Ras2p) adjacent to the adenylyl
364 cyclase complex and the polarisome, which aligns with their respective roles in cell growth[57–60].
365 Interestingly, the integrated network also indicates that *RAS*2 connects the adenylyl cyclase
366 complex to mitochondrial complexes, supporting a recently proposed mechanism in which Ras2p
367 in the mitochondria promotes apoptosis *via* the cAMP/PKA pathway[61]. These examples highlight
368 the potential of analysing motifs in the connector region for generating new hypotheses on the
369 functions of genes.

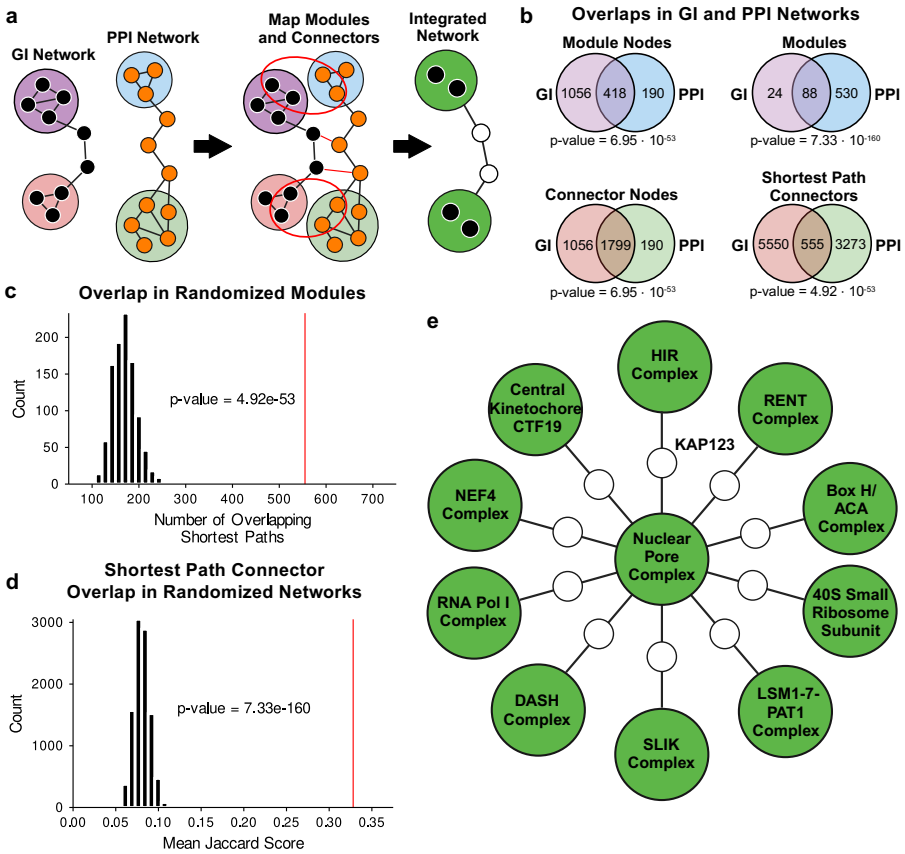


Figure 4. Integrating genetic interactions networks and protein-protein interaction networks using common topological structures. a, Brief description of the algorithm devised to identify connectors of interest. First, shortest paths between modules in GI and PPI networks are determined. Modules of GI 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.

383 **Discussion**

384

385 Determining how GI and PPI networks are related is an outstanding problem in systems biology
386 and its resolution has profound implications for our understanding of the genotype-phenotype
387 relationship. While some relationship between binding affinity and GI strength could fit intuitive
388 expectation, a quantitative relationship between binding affinity and the continuum of GI values
389 has never been demonstrated. In this study, we examined the two properties that define biological
390 networks (weights and topology) to establish a rigorous framework through which PPI and GI
391 networks can be analysed, predicted, and integrated. Although we focus on networks generated
392 using data from the model organism *Saccharomyces cerevisiae*, we find that the principles
393 established in yeast translate well to human networks.

394

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

401 $$\varepsilon^{-} = \frac{a}{1 + e^{b \cdot \ln(K_d) + d}} + c \quad (1)$$

402 Where K_d is the dissociation constant of a PPI, ε^{-} is negative epistasis (multiplicative model)
403 between the interacting proteins' genes, and *a*, *b*, *c* and *d* are free parameters. Importantly, these
404 parameters have tangible biological and biophysical meanings. For instance, the parameter *a* sets
405 the range of epistasis values. Further, the parameter *b* (replaced by kT below), the parameter *d,*
406 and K_d form the Gibbs free energy of binding equation:

407 $$\Delta G = -kT \ln(K_d) + d \quad (2)$$

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

414 mutant gene knock downs) as opposed to other type of genetic changes such as point mutations,
415 recombination, or structural variants. Moreover, our model for the reconstruction of a complete GI
416 network (Figure 2f) explains approximately 2% of the entire variance in the negative GI network,
417 including GIs between genes whose proteins do not interact with each other. Although a better
418 understanding of how genetic perturbations percolate through the PPI network would likely
419 increase explainability, it is likely that GIs are also strongly driven by other molecular mechanisms
420 such as substrate and product dependencies between enzymes in metabolic networks, and
421 transcription factor-promoter relations in gene regulatory networks.
422
423 Certain limitations exist as well regarding our estimates of dissociation constants. First, each PPI
424 is treated as completely independent of any other, leading to deviations from the actual values. For
425 instance, competitive binding will systematically lead to underestimations of binding strength,
426 especially for very strong interactions. Second, under our current model, interaction
427 stoichiometries over 1:1, which may occur if more than one prey protein binds to a single bait
428 protein, lead to negative K_d values which are physically impossible. The number of interactions
429 which are discarded as a result represents a non-negligible fraction of the interactome, which may
430 bias the distribution of our estimated K_d. Third, our ability to assess the precision of our estimates
431 is limited to the quality of reference binding affinity data in the literature. Fourth, we assume all
432 reported interactions in our datasets to be direct. In reality, AP-MS approaches also detect indirect
433 interactions which are not readily distinguishable from direct ones.
434
435 Previous approaches aiming to identify common topological features between GI networks and
436 PPI networks have mostly focused on GIs within or between protein modules[4,6,18]. However, not
437 all genes and proteins cluster into highly connected groups. Although it has long been known that
438 certain proteins serve as "connectors" between protein modules[62], the extent to which this
439 "module-connector-module" motif occurs in GI networks had not been explored. We demonstrate
440 that the common topological organization in GI and PPI networks extends further than overlap in
441 modules, but also the connectors that bridge these modules. By merging modules and connectors
442 from both GI and PPI networks an integrated network was constructed, where the functional role
443 of connector genes/proteins are highlighted and can be used a novel tool for functional inference
444 that will be valuable for predicting GIs in other organisms.

Finally, the quantitative and topological relationships we have established in this study could provide means of interpreting and predicting fundamental relationships between genes underlying normal and disease phenotypes, as well as identify specific vulnerabilities or changes in PPIs that could predict potential therapeutic targets for diseases.

Acknowledgments

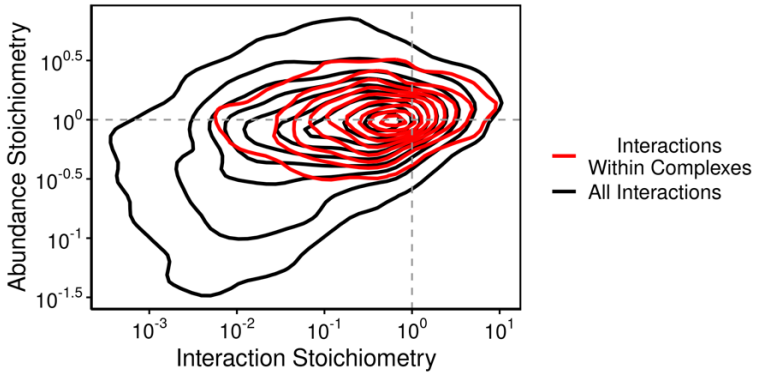
The authors thank Anastasia Baryshnikova for advice on preprocessing of synthetic genetic array genetic interaction data. We also thank Christian R. Landry, Ben Lehner, Elena Kuzmin, and Brenda Andrews for comments on the manuscript. We acknowledge funding from the following agencies: Natural Sciences 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 (NSERC) grant RGPIN-2016-06566 (AWRS); Canada Research Chairs (SWM, AWRS); CIHR grant PJT-185976 and NSERC grant RGPIN-03216-2021 (SWM).

Author Contributions

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

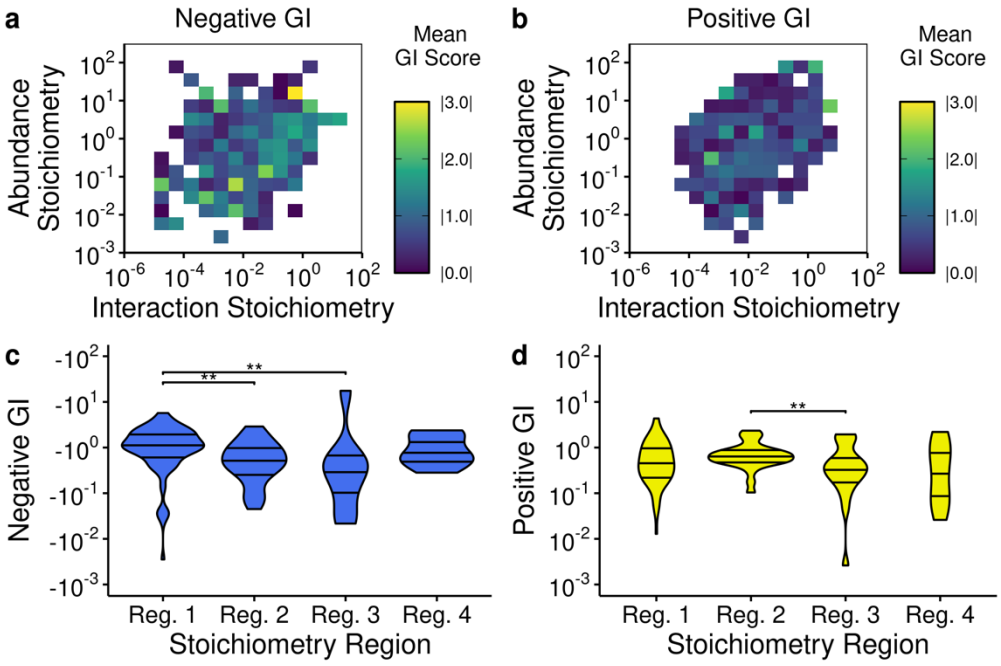
466 **Supplementary Figures**

467

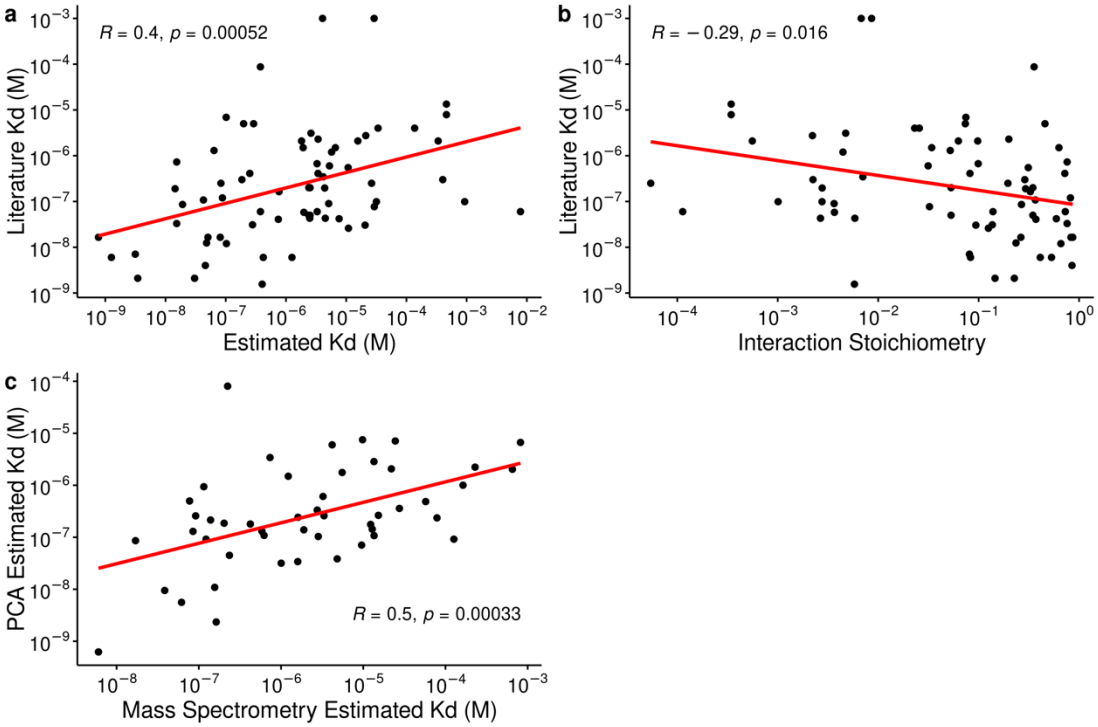


468

469 **Supplementary Figure 1. Stable protein complexes in *Saccharomyces cerevisiae* are enriched**
470 **for 1:1 stoichiometries.** Yeast PPIs occurring between proteins of stable complexes as annotated
471 by the ComplexPortal database[56].



Supplementary 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 and positive GI scores within the 4 defined regions of the stoichiometry plot. Significance levels are annotated as: ** for $p \leq 0.01$.

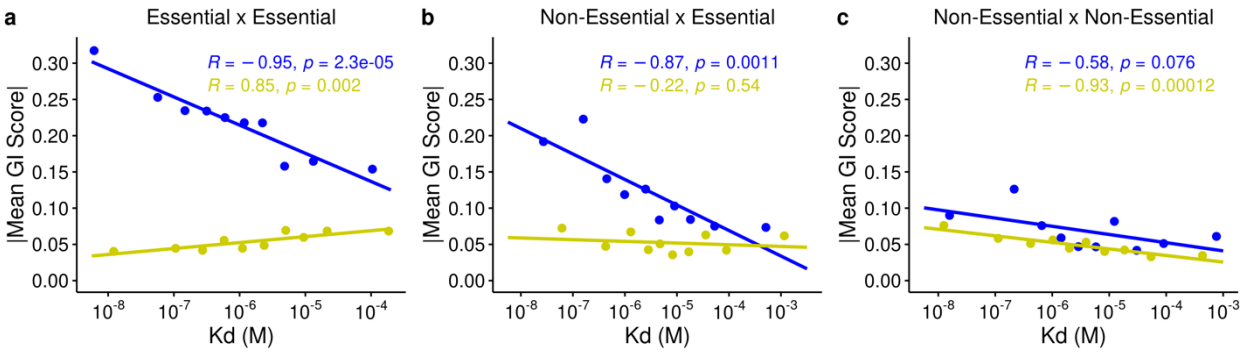


481

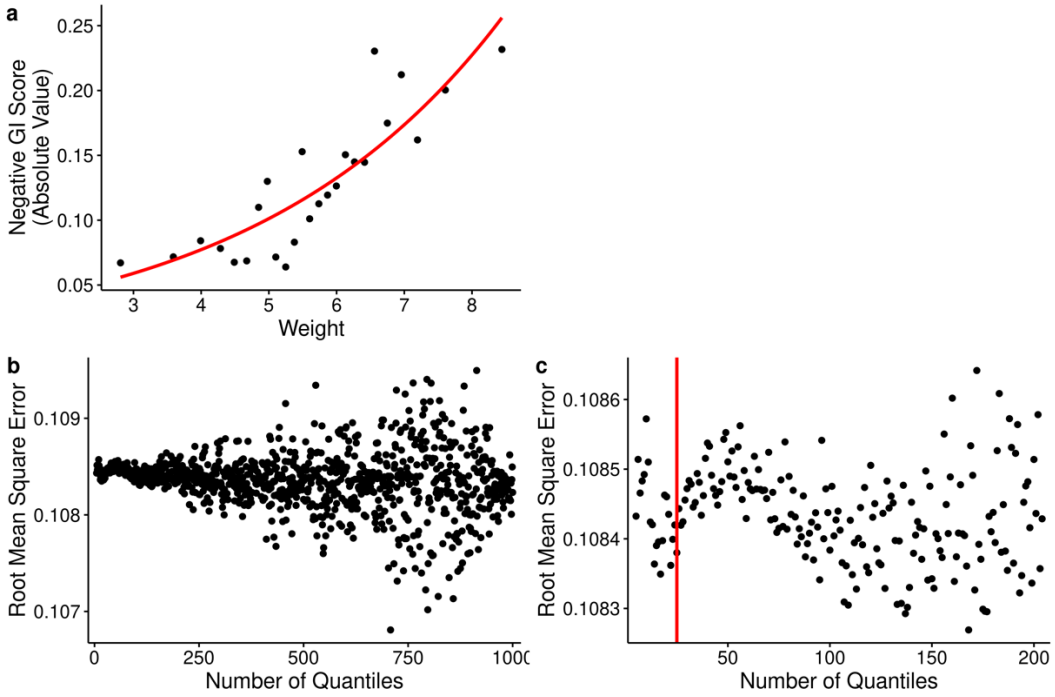
482

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

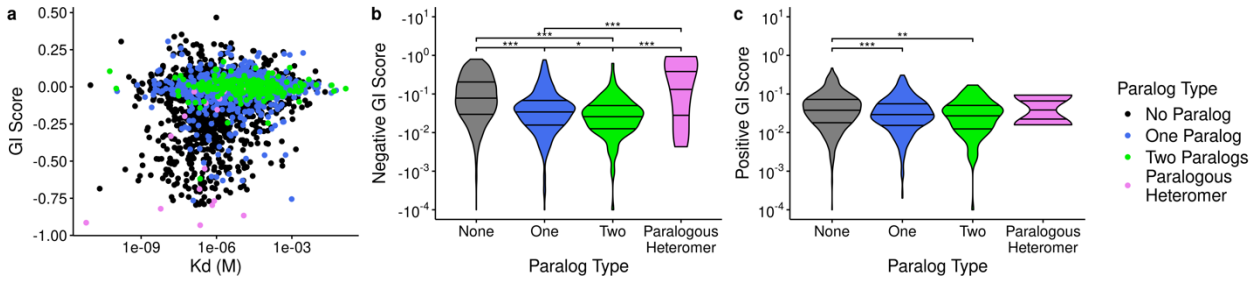
491



Supplementary Figure 4. Correlations between binned genetic interactions and K_d for genes pairs of different essentialities. P-values are computed from two tailed test. All trendlines are fit using a linear regression. 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**).

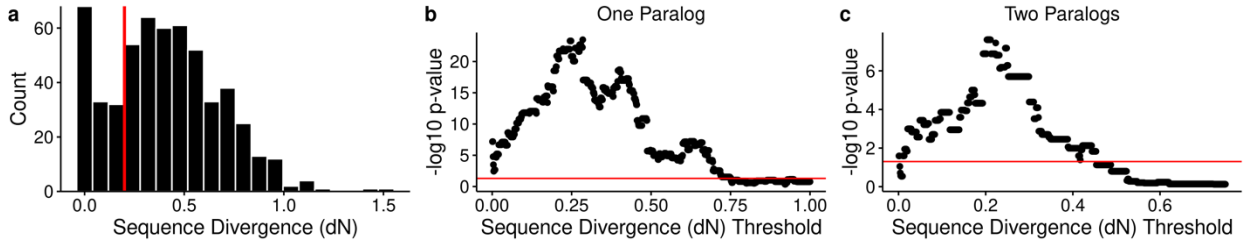


Supplementary Figure 5. Fitting an equation relating binned PPI weights (derived from K_d values) to GI scores. a, Yeast PPIs are binned into 25 quantiles according to their network weights. The average weight and GI score is calculated for every quantile and an exponential equation is used to fit the 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.



Supplementary Figure 6. 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.

512



513

514 **Supplementary Figure 7. Dividing protein-protein interactions according to paralog**
515 **sequence divergence. a,** Sequence divergence (dN) between paralogous protein/gene pairs
516 retrieved from literature. A dN threshold of 0.2 separates the two visible modes in the distribution.
517 **b, c,** The statistically significant separations of dissociation constants between high divergence
518 and low divergence observed in Figure 3 (panels c and d) are robust to the chosen dN threshold.
519 P-values are calculated using Wilcoxon rank sum test. Red line indicates a P-value of 0.05.

520 **Supplementary 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

521 **[a]Examples were taken from Costanzo et al. (2016)**[1]
522
523 **Supplementary 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

524 **[a]Examples were taken from Costanzo et al. (2016)**[1]
525

526 **Supplementary 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

527
528

529 # Methods

530

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

532

533 For yeast, raw synthetic genetic array (SGA) genetic interaction data was downloaded from
534 TheCellMap[1,63] in matrix format. Data from essential and non-essential strain crosses (Essential
535 vs. Essential, Essential vs. Nonessential, Nonessential vs. Nonessential) were combined. In
536 instances where multiple GIs were measured for a pair of genes, only the entry with the most
537 significant *P*-value was retained. For human cell lines, GI data was retrieved from a digenic
538 knockdown screen of 472x472 genes in Jurkat and K562 cells[34] (table S5 of reference 34). For
539 both yeast and human data, unless otherwise indicated, analyses were performed using all GIs
540 regardless of significance scores.

541

542 Yeast protein-protein interactions were taken from a near exhaustive whole-proteome affinity
543 enrichment coupled to an affinity purification-mass spectrometry (AP-MS) screen[21]
544 (Yeast_Interactome_Edges.csv from supplementary file 2 of the reference 28). Human protein-
545 protein interactions and stoichiometries were taken from two quantitative AP-MS screens: one in
546 which genes were tagged with GFP and expressed on bacterial artificial chromosomes in HeLa
547 cells[19] (table S2 of the reference 19), and a second in which genes are endogenously tagged in
548 HEK 293T cells with the small fragment of the split $mNeonGreen_2$ tag[20] (table S4 of the reference
549 20). Both human cell line interactomes were concatenated but common interactions between the
550 two screens were kept as separate data points for all analyses.

551

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

553

554 Interaction stoichiometries were calculated from data generated in an AP-MS screen[21] using a
555 previously described strategy[19]. The starting preprocessed table containing normalized and
556 imputed label-free quantification (LFQ) intensities from the AP-MS screen were provided by the
557 authors. Briefly, computing interaction stoichiometries was done in three mains steps: 1)
558 Subtracting the median intensity of a protein across all sample, 2) Dividing LFQ intensities by the
559 number of theoretically observable peptides of lengths between 7 and 30 amino acids, 3) dividing

560 the values observed for the prey protein by that of the bait protein. Importantly, the number of
561 theoretically observed peptides were computed using sequences from the Uniprot database[64]
562 (https://www.uniprot.org/proteomes/UP000002311) using LysC only digestion (as opposed to
563 LysC and Tryptic digestion). Negative interaction stoichiometries (which arise because of
564 normalization in step 1) were dropped from the dataset. The final dataset includes interaction
565 stoichiometries for 7348 unique PPIs.
566
567 Protein abundances were taken from a dataset unifying measurements from various sources [65]
568 (table S4 of the referenced article). Abundance stoichiometries were determined by calculating the
569 ratio between the median copy numbers of the prey over that of the bait, as described previously[19].
570 Abundance measurements for both interactors were available for 7336 out of the 7348 PPIs for
571 which interaction stoichiometry values were calculated. The 22 interactions which had interaction
572 stoichiometry values, but not abundances stoichiometry values, were conserved in the dataset for
573 analyses in which abundance stoichiometry was not relevant.
574
575 **Analysis of genetic interactions and protein-protein interaction stoichiometries**
576
577 The relationship between GIs and stoichiometries was explored in yeast and human cell lines (see
578 first section for data collection). To merge GI and PPI stoichiometry datasets, tables were merged
579 such that if genes comprising a GI also code for proteins that interact, the GI score is assigned to
580 that PPI. Four stoichiometric regions were graphically defined within stoichiometry plots
581 according to previous criteria[19]. Hein et al. defined region 1 as the space enclosed by a circle
582 centered at (-0.5, 0) where the abscissa is interaction stoichiometry, the ordinate is abundance
583 stoichiometry, and coordinates are in base 10 logarithmic scale. Using these definitions as a
584 reference, region 3 was defined using circles of identical radii (0.75) centered at (-0.5, 1.5), region
585 2 was centered at (-2.5, 0) for yeast and (-3.5, 0) for human, and region 4 was centered at (-2.5, -
586 1.5) for yeast and (-3.5, -1.5) for human.
587

588

589 **Estimating protein-protein interaction dissociation constants**

590

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

594

595 $$A + B \rightleftharpoons AB \tag{1}$$

596

597 $$K_d = \frac{[A][B]}{[AB]} \equiv \frac{[Bait][Prey]}{[Bait::Prey]} \tag{2}$$

598

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

603

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

608

609 $$Interaction\ Stoichiometry = \frac{Prey_{LFQ_{Intensity}}}{Bait_{LFQ_{Intensity}}} \approx \frac{Prey_{Bound\ to\ Bait}}{Bait_{Total}} = \frac{Bait::Prey}{Bait_{Total}} \tag{3}$$

610

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

613

614 $$Bait::Prey = \frac{Bait::Prey}{Bait_{Total}} \cdot Bait_{Total} = (Interaction\ Stoichiometry) \cdot Bait_{Total} \tag{4}$$

615

616 For yeast PPIs, the cellular copy numbers of bait proteins were taken from the same source as that
617 described to calculate abundance stoichiometries[65]. For PPIs in human cell lines, abundances were
618 taken from whole proteome quantification experiments which were performed alongside the AP-
619 MS measurements of interaction stoichiometries[19,20]
620 (opencell.czbiohub.org/data/datasets/opencell-protein-abundance.csv [downloaded April 4th
621 2022]; Hein et al., table S3)[19].
622
623 Next, the number of unbound bait and prey proteins were estimated by subtracting the number of
624 bait-prey complexes[39,66]. In a few instances, the estimated number of bait-prey complexes was
625 higher than that of the total abundance of bait or prey, resulting in a negative number of unbound
626 proteins. This almost only arose when interaction stoichiometry is above 1. These negative values
627 were removed.
628
629 $$Bait_{Unbound} = Bait_{Total} - Bait::Prey \quad (5)$$
630
631 $$Prey_{Unbound} = Prey_{Total} - Bait::Prey \quad (6)$$
632
633 Copy numbers of bait proteins, prey proteins and bait-prey complexes were next converted to
634 molar concentrations (mol/L) using Avogadro's constant and cell volumes. Below is an example
635 formula for bait concentrations; concentrations prey and bait-prey complexes were calculated in
636 the same way. Dissociation constants were calculated using the resulting concentrations.
637
638 $$[Bait] = \frac{Bait_{unbound}/(6.022 \cdot 10^{23}\ mol^{-1})}{Cell\ Volume} \quad (7)$$
639
640 $$K_d = \frac{[Bait][Prey]}{[Bait::Prey]} \quad (8)$$
641
642 A cell volume of 4.2×10^{-14} Liters was used for yeast PPIs (which were measured in haploid
643 cells)[67]. A cell volume of 2.6×10^{-14} Liters was used for PPIs measured in HeLa cells by Hein et
644 al.[68] Cho et al. measured PPIs and stoichiometries in HEK293 cells, for which no experimental

645 measurement of volume was found. However, the diameter of HEK293 cells was measured to be
646 approximately 13 micrometers (Bionumbers ID: 108893). Therefore, an estimated cell volume of
647 1.15×10^{-14} Liters was calculated from this diameter by approximating the shape of a cell to that
648 of a sphere. The above-mentioned values were all found via the BioNumbers database[69].
649
650 **Validating estimated dissociation constants using values from literature**
651
652 Experimental dissociation constants from literature were taken from the protein-protein complex
653 subset of the PDBbind database (2020 version)[70,71]. Taxonomic identification numbers were
654 extracted from the PDB file of each database entry. Only interactions involving proteins from
655 budding yeast and human were retained. PDB files which listed more than two proteins were
656 removed, as it would not be possible to assign dissociation constants to the correct protein pair.
657
658 Systematic names or IDs for proteins and genes are absent from the PDBbind database. For this
659 reason, the PDB ID for each complex was used to access their structure's RCSB Protein Data Bank
660 web page and retrieve the Uniprot IDs of the proteins within the structure. Web scraping for the
661 data published in this article was performed on April 23rd 2024, using the *BeautifulSoup4 Python3*
662 package. For yeast PPIs, Uniprot IDs were then converted to systematic ORF IDs using the
663 *org.Sc.sgd.db R* package. For human PPIs, a Uniprot to HGNC symbol conversion table was
664 downloaded (on February 28th 2024) from the "Custom downloads" page of the HGNC website[72].
665 Information on the original source, which is often different from that of the structure, and details
666 on the experimental setup of each dissociation constant measurement is also absent from the
667 database. Although this limits quality control, it is known from curated sets for structure-based
668 modelling of PPI affinities that common methods to measure PPI binding affinities are best for
669 interactions in the micromolar to nanomolar range[41]. Therefore, measurements were limited to
670 dissociation constants no smaller than 1 nM and no greater than 1 mM ($1\ nM \leq K_d < 1\ mM$). The
671 final reference dataset contains 505 K_d measurements for human PPIs and 59 for yeast PPIs.
672
673 Estimated K_d values for PPIs were merged with experimental values for the reference dataset. For
674 PPIs in human cell lines, 70 estimated K_d values were mapped to values from literature. Only 4
675 estimated K_d values were mapped to values from literature for yeast PPI.
676

677 **Comparison of mass spectrometry K_d estimates in yeast using PCA K_d estimates**
678
679 Colony intensities and z-scores were collected from a genome-wide screen for protein-protein
680 interaction using a survival-selection protein-fragment complementation assay (PCA) [73](table S4
681 of the referenced article) and filtered to only contain the final 2770 protein-protein interactions
682 (table S3 of the referenced article) reported in the article. Next colony intensities were collected
683 from a proteome-wide abundance screen based on the GFP-PCA method[74] (Data Set S2 of the
684 reference article). GFP-PCA intensities were merged to cellular copy numbers from a compilation
685 of abundances in yeast[65] to generate a linear regression model relating PCA colony intensities to
686 copy numbers. Protein-protein interaction PCA intensities were then rescaled to a range similar to
687 that of the abundance GFP-PCA intensities. The linear regression model was then used to convert
688 protein-protein interaction intensities to numbers of complexes. With the number of copies of
689 complexes and individual protein copy numbers at hand, dissociation constants were estimated as
690 described in the "Estimating protein-protein interaction dissociation constants" section. In total,
691 1480 K_d estimates were calculated from PCA data, of which 48 also had K_d estimated from mass
692 spectrometry data.
693
694 **Reconstructing a GI network matrix from a PPI binding network**
695
696 Yeast GI and PPI dissociation constant (K_d) tables were merged such that if genes comprising a
697 GI also code for proteins that interact, the GI score is assigned to that PPI. The final table contains
698 3265 interactions, each with a GI score and a K_d value. A graph was then generated, using
699 individual proteins as nodes/vertices (V_{PPI}), PPIs as edges (E_{PPI}), and the reciprocal negative
700 logarithm of K_d values as weights (W_{PPI}):
701
702 $$G_{PPI} = \{V_{PPI}, E_{PPI}, W_{PPI}\} \quad (9)$$
703
704 $$w_{PPI}(\{u, v\} \in E_{PPI}) = \frac{1}{-log10(K_{d,u,v})} \quad (10)$$
705 Where $u, v \in V_{PPI}$, $K_{d,u,v}$ is the dissociation constant between the interacting proteins u and v, and
706 $w_{PPI}(e)$ is the weight assigned to the edge e (where $e = \{u, v\}$ connects u and v).
707

708 To establish an equation relating weights and GI scores, PPIs were first binned into quantiles
709 according to their weights. The mean weight and mean GI score within each quantile were then
710 calculated and an exponential equation was used to fit the data (See supplementary Figure 5a for
711 example with 25 quantiles). The root mean square error (RMSE) of the fit for different numbers
712 of quantiles were tested to find the most optimal (Supplementary Figure 5b). A dip of RMSE
713 values and narrowing of RMSE spread was observable at around 25 quantiles, which justified its
714 choice as the ideal number of quantiles for the final fitting (Supplementary Figure 5c and 5a).
715
716 To mimic the completeness of GI networks, the PPI graph was converted to an "association" matrix
717 similar to that described by Rives and Galitsky[75]. Briefly, the *shortest_path_length* function of the
718 *Networkx* python library was used on G_{PPI}, using W_{PPI} as the weight parameter, to generate a $|V| \times$
719 $|V|$ matrix (d) in which d[i,j] denotes the length of the shortest path between the nodes (proteins) i
720 and j. This all-pairs-shortest-path (APSP) matrix was used as a graph (G_{APSP}), such that:
721
722 $$G_{APSP} = \{V_{APSP}, E_{APSP}, W_{APSP}\} \quad (11)$$
723
724 $$w_{APSP}(\{x, y\}) = shortest_path_length(u, v) \quad (12)$$
725
726 Where $\{u, v\} \in E_{PPI}$, $\{x, y\} \in E_{APSP}$, and both edges represent the same protein interaction. The
727 exponential equation relating means of quantized weights and GI scores was then used on W_{APSP}
728 to generate a new graph (G_{GI_pred}) with weights W_{GI} containing estimated GI scores. The adjacency
729 matrix of G_{GI_pred} was clustered with average linkage hierarchical clustering using Pearson
730 correlation in Cluster 3.0[76].
731
732 **Analysis of dissociation constants in yeast paralogs**
733
734 Whole genome duplication (WGD) paralogs were collected from Kellis et al.[77] (Supplementary
735 file 8) and small-scale duplication (SSD) paralogs were collected form VanderSluis et al.[46]
736 Paralogs were separated into 3 categories as described in the main text. The rate of non-
737 synonymous mutations per non-synonymous site (dN) was used as a proxy for sequence
738 divergence. To calculate dN values, paralog pairs were aligned pairwise at the protein level with

739 ClustalW (version 2.1) using the PAM protein weight matrix. Using the *reverse.align* function of
740 the *seqinr* R package, protein-coding nucleic acid sequences were next aligned using protein
741 alignments as a guide. The reverse alignments of each paralog pair were then used to compute dN
742 using the *kaks* function of *seqinr*. Amino acid sequences and coding nucleic acid sequences were
743 downloaded from the SGD sequence archives (http://sgd-
744 archive.yeastgenome.org/sequence/S288C_reference/, March 2024) for the S288C reference
745 strain. A sequence divergence threshold of 0.2 was selected to categorize paralog pairs into "high"
746 and "low" sequence divergence (dN) (Supplementary Figure 7a). For protein-protein interactions
747 where both interactors have a paralog, the interaction was categorized into "high" when both
748 interactors had a dN value above 0.2.
749
750 **Mapping GI Modules to PPI Modules**
751
752 GI clusters were downloaded from an analysis of a near-complete double knockout study in yeast[1]
753 (Data File S6). Yeast protein complexes were downloaded from a species-specific dataset of the
754 ComplexPortal database[56]; protein Uniprot IDs were converted to systematic yeast ORFs using the
755 *org.Sc.sgd.db* R library. All GI clusters were compared to all PPI complexes by their common
756 ORFs, the degree of overlap was scored using the Jaccard index. For each GI cluster, the protein
757 complex which had the highest overlap (Jaccard Index) was retained.
758
759 If a GI cluster overlapped better with connector nodes (proteins outside complexes), it was dropped
760 from the dataset. To validate the significance of the overlap, GI clusters were randomized and
761 mapped to protein complexes. GI clusters were randomized by retaining the size of the clusters
762 and randomly redistributing ORFs. The randomization was performed for 10 000 iterations on a
763 general-purpose computer cluster. A p-value was derived from a z-score in which the mean Jaccard
764 score from the experimental dataset was compared to the distribution of mean Jaccard scores of
765 randomized clusters (Figure 4c).
766

767

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

769

770 The GI and PPI networks used for the analysis of "connector" nodes were taken from the near-
771 complete double knockout screen in yeast and whole-proteome affinity enrichment coupled to
772 mass spectrometry screen described in the *Collecting yeast and human genetic and protein-protein*
773 *interaction data* section. Only negative genetic interactions were used for the analysis in figure 4.
774 Because shortest paths are sensitive to noise, we used the moderately stringent cutoff of GI score
775 < -0.08 and p-value < 0.05 recommended by the authors[1]. Identifying nodes (gene/proteins)
776 connecting modules in GI and PPI networks was performed in 4 main steps. First, genetic modules
777 and protein modules are mapped 1-to-1 as described in the *Mapping GI Modules to PPI Modules*
778 section above.

779

780 The 1-to-1 mapping can be expressed mathematically, where M_{GI} and M_{PPI} are each a set
781 containing all modules in the GI and PPI network, respectively, and a module is itself a set of
782 genes/proteins. Connectors can be defined as the sets C_{GI} and C_{PPI}, which contain all the connector
783 genes/proteins in the GI and PPI networks respectively. Importantly, a gene/protein cannot be both
784 a connector and within a module. In other words:

785 $$C_{GI} \cap \bigcup M_{GI} = \emptyset \tag{13}$$

786 $$C_{PPI} \cap \bigcup M_{PPI} = \emptyset \tag{14}$$

787 We note that here the large union symbol denotes the union of all genes/proteins that are within
788 modules into a single large set. Next, a module $X \in M_{GI}$ is related (mapped) to a module $A \in M_{PPI}$
789 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)
790 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
791 J denotes the Jaccard index:

792 $$J(X, A) = \frac{X \cap A}{X \cup A} \tag{15}$$

793 Second, for all combinations of modules within a network, the shortest path between all pairs of
794 nodes were computed. This was done independently for GI and PPI networks. In mathematical
795 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.

Fourth, the overlapping features in both networks *i.e.,* modules and shortest paths connectors, are combined to generate an integrated network. To determine if the extent of the overlap in connectors is due to network topology as opposed to random chance, the approach described above was performed on randomized networks. The networks were randomized by first conserving edges (interactions) within modules and then performing degree preserved randomization[78] for all other edges. Overlapping connectors were identified for 1 000 random GI and PPI networks on a general-purpose computer cluster. A p-value was derived from a z-score in which the number of overlapping connectors from the experimental dataset was compared to the distribution of overlapping connectors from randomized networks (Figure 4d).

Code and Data Availability

All original scripts as well as information on how to properly access data and run the code is provided in a GitHub repository dedicated to this article (https://github.com/Xavier-Castellanos-Girouard/PPI_Binding_Epistasis).

823 1. Costanzo, M. *et al.* A global genetic interaction network maps a wiring diagram of cellular
824 function. *Science* **353**, aaf1420–aaf1420 (2016).
825 2. Costanzo, M. *et al.* Global Genetic Networks and the Genotype-to-Phenotype Relationship.
826 *Cell* **177**, 85–100 (2019).
827 3. Sharan, R., Ulitsky, I. & Shamir, R. Network-based prediction of protein function. *Molecular*
828 *Systems Biology* **3**, 88 (2007).
829 4. Beyer, A., Bandyopadhyay, S. & Ideker, T. Integrating physical and genetic maps: from
830 genomes to interaction networks. *Nat Rev Genet* **8**, 699–710 (2007).
831 5. Fang, G. *et al.* Discovering genetic interactions bridging pathways in genome-wide
832 association studies. *Nat Commun* **10**, 4274 (2019).
833 6. VanderSluis, B. *et al.* Integrating genetic and protein–protein interaction networks maps a
834 functional wiring diagram of a cell. *Current Opinion in Microbiology* **45**, 170–179 (2018).
835 7. Mani, R., St.Onge, R. P., Hartman, J. L., Giaever, G. & Roth, F. P. Defining genetic
836 interaction. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 3461–3466 (2008).
837 8. Phillips, P. C. Epistasis — the essential role of gene interactions in the structure and
838 evolution of genetic systems. *Nat Rev Genet* **9**, 855–867 (2008).
839 9. Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. The mystery of missing heritability:
840 Genetic interactions create phantom heritability. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 1193–
841 1198 (2012).
842 10. Sackton, T. B. & Hartl, D. L. Genotypic Context and Epistasis in Individuals and
843 Populations. *Cell* **166**, 279–287 (2016).
844 11. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–
845 753 (2009).

846 12. Aguirre, L., Hendelman, A., Hutton, S. F., McCandlish, D. M. & Lippman, Z. B.
847 Idiosyncratic and dose-dependent epistasis drives variation in tomato fruit size. *Science* **382**,
848 315–320 (2023).
849 13. Mair, B., Moffat, J., Boone, C. & Andrews, B. J. Genetic interaction networks in cancer
850 cells. *Current Opinion in Genetics & Development* **54**, 64–72 (2019).
851 14. Shen, J. P. & Ideker, T. Synthetic Lethal Networks for Precision Oncology: Promises and
852 Pitfalls. *Journal of Molecular Biology* **430**, 2900–2912 (2018).
853 15. Costanzo, M. *et al.* The Genetic Landscape of a Cell. *Science* **327**, 425–431 (2010).
854 16. Oliver, S. Guilt-by-association goes global. *Nature* **403**, 601–602 (2000).
855 17. Sahni, N. *et al.* Widespread Macromolecular Interaction Perturbations in Human Genetic
856 Disorders. *Cell* **161**, 647–660 (2015).
857 18. Kelley, R. & Ideker, T. Systematic interpretation of genetic interactions using protein
858 networks. *Nat Biotechnol* **23**, 561–566 (2005).
859 19. Hein, M. Y. *et al.* A Human Interactome in Three Quantitative Dimensions Organized by
860 Stoichiometries and Abundances. *Cell* **163**, 712–723 (2015).
861 20. Cho, N. H. *et al.* OpenCell: Endogenous tagging for the cartography of human cellular
862 organization. *Science* **375**, eabi6983 (2022).
863 21. Michaelis, A. C. *et al.* The social and structural architecture of the yeast protein interactome.
864 *Nature* **624**, 192–200 (2023).
865 22. Bandyopadhyay, S., Kelley, R., Krogan, N. J. & Ideker, T. Functional Maps of Protein
866 Complexes from Quantitative Genetic Interaction Data. *PLoS Comput Biol* **4**, e1000065
867 (2008).

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

890 34. Horlbeck, M. A. *et al.* Mapping the Genetic Landscape of Human Cells. *Cell* **174**, 953-
891 967.e22 (2018).
892 35. Van Leeuwen, J. *et al.* Exploring genetic suppression interactions on a global scale. *Science*
893 **354**, aag0839 (2016).
894 36. He, X., Qian, W., Wang, Z., Li, Y. & Zhang, J. Prevalent positive epistasis in Escherichia
895 coli and Saccharomyces cerevisiae metabolic networks. *Nat Genet* **42**, 272–276 (2010).
896 37. Segrè, D., DeLuna, A., Church, G. M. & Kishony, R. Modular epistasis in yeast metabolism.
897 *Nat Genet* **37**, 77–83 (2005).
898 38. Xie, X., Sun, X., Wang, Y., Lehner, B. & Li, X. Dominance vs epistasis: the biophysical
899 origins and plasticity of genetic interactions within and between alleles. *Nat Commun* **14**,
900 5551 (2023).
901 39. Diss, G. & Lehner, B. The genetic landscape of a physical interaction. *eLife* **7**, e32472
902 (2018).
903 40. Jarmoskaite, I., AlSadhan, I., Vaidyanathan, P. P. & Herschlag, D. How to measure and
904 evaluate binding affinities. *eLife* **9**, e57264 (2020).
905 41. Kastritis, P. L. *et al.* A structure-based benchmark for protein–protein binding affinity.
906 *Protein Science* **20**, 482–491 (2011).
907 42. Diss, G., Dubé, A. K., Boutin, J., Gagnon-Arsenault, I. & Landry, C. R. A Systematic
908 Approach for the Genetic Dissection of Protein Complexes in Living Cells. *Cell Reports* **3**,
909 2155–2167 (2013).
910 43. Wright, P. E. & Dyson, H. J. Linking folding and binding. *Current Opinion in Structural*
911 *Biology* **19**, 31–38 (2009).

912 44. Castellanos-Girouard, X. *GI_experimental_full.Png*. 11784216 Bytes (figshare, 2024).
913 doi:10.6084/M9.FIGSHARE.27020701.
914 45. Castellanos-Girouard, X. *GI_reconstructed_full.Png*. 6541289 Bytes (figshare, 2024).
915 doi:10.6084/M9.FIGSHARE.27020704.
916 46. VanderSluis, B. *et al.* Genetic interactions reveal the evolutionary trajectories of duplicate
917 genes. *Molecular Systems Biology* **6**, 429 (2010).
918 47. Kuzmin, E. *et al.* Exploring whole-genome duplicate gene retention with complex genetic
919 interaction analysis. *Science* **368**, eaaz5667 (2020).
920 48. Nowak, M. A., Boerlijst, M. C., Cooke, J. & Smith, J. M. Evolution of genetic redundancy.
921 *Nature* **388**, 167–171 (1997).
922 49. Vavouri, T., Semple, J. I., Garcia-Verdugo, R. & Lehner, B. Intrinsic Protein Disorder and
923 Interaction Promiscuity Are Widely Associated with Dosage Sensitivity. *Cell* **138**, 198–208
924 (2009).
925 50. Rice, A. M. & McLysaght, A. Dosage-sensitive genes in evolution and disease. *BMC Biol*
926 **15**, 78 (2017).
927 51. Papp, B., Pál, C. & Hurst, L. D. Dosage sensitivity and the evolution of gene families in
928 yeast. *Nature* **424**, 194–197 (2003).
929 52. Ascencio, D. *et al.* Expression attenuation as a mechanism of robustness against gene
930 duplication. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2014345118 (2021).
931 53. Qian, W., Liao, B.-Y., Chang, A. Y.-F. & Zhang, J. Maintenance of duplicate genes and their
932 functional redundancy by reduced expression. *Trends in Genetics* **26**, 425–430 (2010).
933 54. Michnick, S. W. & Levy, E. D. The modular cell gets connected. *Science* **375**, 1093–1094
934 (2022).

935 55. Pereira-Leal, J. B., Levy, E. D. & Teichmann, S. A. The origins and evolution of functional
936 modules: lessons from protein complexes. *Phil. Trans. R. Soc. B* **361**, 507–517 (2006).
937 56. Meldal, B. H. M. *et al.* Complex Portal 2018: extended content and enhanced visualization
938 tools for macromolecular complexes. *Nucleic Acids Res* **47**, D550–D558 (2019).
939 57. Tamanoi, F. Ras Signaling in Yeast. *Genes & Cancer* **2**, 210–215 (2011).
940 58. Cazzanelli, G. *et al.* The Yeast Saccharomyces cerevisiae as a Model for Understanding RAS
941 Proteins and their Role in Human Tumorigenesis. *Cells* **7**, 14 (2018).
942 59. Ho, J. & Bretscher, A. Ras Regulates the Polarity of the Yeast Actin Cytoskeleton through
943 the Stress Response Pathway. *MBoC* **12**, 1541–1555 (2001).
944 60. Xie, Y. & Miao, Y. Polarisome assembly mediates actin remodeling during polarized yeast
945 and fungal growth. *Journal of Cell Science* **134**, jcs247916 (2021).
946 61. Bonomelli, B., Busti, S., Martegani, E. & Colombo, S. Active Ras2 in mitochondria
947 promotes regulated cell death in a cAMP / PKA pathway-dependent manner in budding
948 yeast. *FEBS Letters* **597**, 298–308 (2023).
949 62. Han, J.-D. J. *et al.* Evidence for dynamically organized modularity in the yeast protein–
950 protein interaction network. *Nature* **430**, 88–93 (2004).
951 63. Usaj, M. *et al.* TheCellMap.org: A Web-Accessible Database for Visualizing and Mining the
952 Global Yeast Genetic Interaction Network. *G3 Genes|Genomes|Genetics* **7**, 1539–1549
953 (2017).
954 64. The UniProt Consortium *et al.* UniProt: the Universal Protein Knowledgebase in 2023.
955 *Nucleic Acids Research* **51**, D523–D531 (2023).
956 65. Ho, B., Baryshnikova, A. & Brown, G. W. Unification of Protein Abundance Datasets Yields
957 a Quantitative Saccharomyces cerevisiae Proteome. *Cell Systems* **6**, 192-205.e3 (2018).

958 66. Celaj, A. *et al.* Quantitative analysis of protein interaction network dynamics in yeast.
959 *Molecular Systems Biology* **13**, 934 (2017).
960 67. Jorgensen, P., Nishikawa, J. L., Breitkreutz, B.-J. & Tyers, M. Systematic Identification of
961 Pathways That Couple Cell Growth and Division in Yeast. *Science* **297**, 395–400 (2002).
962 68. Zhao, L. *et al.* Intracellular water-specific MR of microbead-adherent cells: the HeLa cell
963 intracellular water exchange lifetime. *NMR in Biomedicine* **21**, 159–164 (2008).
964 69. Milo, R., Jorgensen, P., Moran, U., Weber, G. & Springer, M. BioNumbers—the database of
965 key numbers in molecular and cell biology. *Nucleic Acids Research* **38**, D750–D753 (2010).
966 70. Liu, Z. *et al.* PDB-wide collection of binding data: current status of the PDBbind database.
967 *Bioinformatics* **31**, 405–412 (2015).
968 71. Wang, R., Fang, X., Lu, Y. & Wang, S. The PDBbind Database: Collection of Binding
969 Affinities for Protein−Ligand Complexes with Known Three-Dimensional Structures. *J.*
970 *Med. Chem.* **47**, 2977–2980 (2004).
971 72. Seal, R. L. *et al.* Genenames.org: the HGNC resources in 2023. *Nucleic Acids Research* **51**,
972 D1003–D1009 (2023).
973 73. Tarassov, K. *et al.* An in Vivo Map of the Yeast Protein Interactome. *Science* **320**, 1465–
974 1470 (2008).
975 74. Levy, E. D., Kowarzyk, J. & Michnick, S. W. High-Resolution Mapping of Protein
976 Concentration Reveals Principles of Proteome Architecture and Adaptation. *Cell Reports* **7**,
977 1333–1340 (2014).
978 75. Rives, A. W. & Galitski, T. Modular organization of cellular networks. *Proc. Natl. Acad. Sci.*
979 *U.S.A.* **100**, 1128–1133 (2003).

980 76. De Hoon, M. J. L., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software.
981 *Bioinformatics* **20**, 1453–1454 (2004).
982 77. Kellis, M., Birren, B. W. & Lander, E. S. Proof and evolutionary analysis of ancient genome
983 duplication in the yeast Saccharomyces cerevisiae. *Nature* **428**, 617–624 (2004).
984 78. Maslov, S. & Sneppen, K. Specificity and Stability in Topology of Protein Networks.
985 *Science* **296**, 910–913 (2002).
986