



Published in final edited form as:

Nat Biotechnol. 2005 May ; 23(5): 561. doi:10.1038/nbt1096.

## Systematic interpretation of genetic interactions using protein networks

Ryan Kelley<sup>1,2</sup> and Trey Ideker<sup>1,2</sup>

<sup>1</sup> Program in Bioinformatics, University of California, San Diego, 9500 Gilman Dr., San Diego, California 92093-0412, USA

<sup>2</sup> Department of Bioengineering, University of California, San Diego, 9500 Gilman Dr., San Diego, California 92093-0412, USA

### Abstract

Genetic interaction analysis, in which two mutations have a combined effect not exhibited by either mutation alone, is a powerful and widespread tool for establishing functional linkages between genes. In the yeast *Saccharomyces cerevisiae*, ongoing screens have generated >4,800 such genetic interaction data. We demonstrate that by combining these data with information on protein-protein, protein-DNA or metabolic networks, it is possible to uncover physical mechanisms behind many of the observed genetic effects. Using a probabilistic model, we found that 1,922 genetic interactions are significantly associated with either between- or within-pathway explanations encoded in the physical networks, covering ~40% of known genetic interactions. These models predict new functions for 343 proteins and suggest that between-pathway explanations are better than within-pathway explanations at interpreting genetic interactions identified in systematic screens. This study provides a road map for how genetic and physical interactions can be integrated to reveal pathway organization and function.

---

A major biological challenge is to interpret observed genetic interactions in a physical cellular context<sup>1-3</sup>. There are several major types of genetic interactions: synthetic-lethal interactions, in which mutations in two nonessential genes are lethal when combined; suppressor interactions, in which one mutation is lethal but when combined with a second, cell viability is restored; and an array of other effects such as enhancement and epistasis. Genetic interactions have been used extensively to shed light on pathway organization in model organisms<sup>1-4</sup>. In humans, genetic interactions are critical in linkage analysis of complex diseases<sup>5</sup> and in discovery of new pharmaceuticals<sup>6</sup>. Although genetic interactions are classically identified by mutant screens<sup>7</sup>, recent studies have applied systematic 'reverse' methods such as synthetic genetic arrays (SGA)<sup>8</sup> or synthetic lethal analysis by microarrays (SLAM)<sup>9</sup> to catalog ~4,000 synthetic-lethal and synthetic-sick interactions in *Saccharomyces cerevisiae*.

Because of the high-throughput nature of SGA, discovery of new genetic interactions is largely automated. However, interpreting the functional significance of each result remains a relatively slow process. The problem is compounded by the large number of genetic interactions measured when screening one gene versus all others (~34 on average<sup>10</sup>) as well as possible false positives if the interactions are not confirmed by tetrad or random spore analysis. Thus,

---

Correspondence should be addressed to T.I. (trey@bioeng.ucsd.edu).

Note: Supplementary information is available on the Nature Biotechnology website.

### COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

without further methods to aid in characterizing synthetic lethals, large-scale interpretation is a daunting prospect.

A promising solution may be to integrate synthetic lethals with other types of high-throughput interactions. For instance, direct physical interactions among proteins are being mapped by systematic two-hybrid<sup>11–15</sup> or immunoprecipitation studies<sup>16,17</sup>, whereas physical interactions between transcription factors and promoter sites are determined using chromatin-immunoprecipitation in conjunction with DNA microarrays<sup>18,19</sup>. These interactions comprise a physical network, which correlates with the network of genetic interactions and provides potential clues as to the mechanisms behind particular synthetic-lethal effects. Previous studies have demonstrated this correlated structure in yeast, by showing that two proteins in the same region of the genetic network are likely to also physically interact<sup>8,10</sup>, that genes with similar patterns of genetic interactions often occur within the same protein complex<sup>10</sup> and that a protein with many interactions in the physical network typically has many interactions in the genetic network also<sup>20</sup>.

These studies suggest that it may be possible to interpret observed synthetic-lethal relationships explicitly using physical interactions. In this regard, previous authors<sup>1,21</sup> have noted that synthetic-lethal interactions are typically associated with one of three types of physical interpretations: between-pathway models, within-pathway models and indirect effects (Box 1).

Here, we demonstrate a computational framework for assembling genetic and physical interactions into models corresponding to between- versus within-pathway interpretations. Regions of the physical network that correspond to each type of model are identified using a probabilistic scoring scheme. These models predict new protein functions and suggest that genetic interactions are more likely to bridge redundant or complementary processes than to combine additively within the same process.

## Construction of genetic and physical networks

We assembled a genetic interaction network from two primary data sources (Fig. 1). The first was generated by SGA, a large-scale screen<sup>10</sup> crossing 132 yeast gene deletion strains versus each of the ~4,700 available deletion strains<sup>22</sup> and resulting in 2,012 observed synthetic-lethal interactions and 2,113 synthetic-sick interactions. The second data source consisted of an additional 687 synthetic-lethal interactions culled from the literature and catalogued at the Munich Information Center for Protein Sequences (MIPS)<sup>23</sup>. The combined genetic network synthesizing these data consisted of 1,434 proteins (genes) linked by 4,812 synthetic-lethal interactions.

We also assembled a physical network of 5,993 yeast proteins connected by physical interactions of three types: 15,429 protein-protein interactions (the two proteins *a* and *b* display physical binding); 5,869 protein-DNA regulatory interactions (*a* binds upstream of the gene encoding *b*) and 6,306 shared-reaction metabolic relationships (*a* and *b* are enzymes that operate on at least one metabolite in common). The protein-protein interactions were downloaded from the DIP database<sup>24</sup> as of July 2004 and predominantly included data from large-scale experiments<sup>13,15–17</sup>. The protein-DNA interactions were obtained from a large-scale chromatin-immunoprecipitation study of 106 transcription factors<sup>18</sup> (interactions with  $P = 0.001$ ). Enzymatic reactions linked by common metabolites were obtained from KEGG<sup>25</sup>, excluding metabolite cofactors such as ATP or H<sub>2</sub>O (listed in Supplementary Table 1 online). The combined physical network covered 94.4% of all proteins in the genetic network. Both networks are provided at <http://www.cellcircuits.org/Kelley2005/> in Cytoscape<sup>26</sup> (SIF) format.

## Between-pathway interpretations for genetic interactions

Preliminary statistical analyses confirm a limited relationship between genetic and physical interactions (see Supplementary Fig. 1 online and Tong *et al.*<sup>8,10</sup>), but demonstrate a need for structured models to efficiently separate signal from noise. Towards this goal, we implemented a probabilistic modeling procedure to capture the between-pathway interpretation of genetic interactions. This procedure involved a search for pairs of physical pathways that were densely connected by genetic interactions, in which a ‘pathway’ was loosely defined as any densely connected set of proteins in the physical network (this definition generically covers many network structures, including protein complexes). Pairs of pathways (constituting a single network model; see Fig. 1) were assigned a score proportional to the density of physical interactions falling within each pathway and the density of genetic interactions bridging between pathways (Box 2). This search generated 360 significant models covering 401 pathways and incorporating a total of 1,573 genetic interactions (196 MIPS, 687 SGA synthetic lethal, 690 SGA synthetic sick) and 1,931 physical interactions (1,248 protein binding, 77 regulatory, 606 shared reaction). Significance of these models was assessed by comparison to random genetic and physical networks. Detailed information for all models is provided in Supplementary Tables 2 and 3 online and at <http://www.cellcircuits.org/Kelley2005/>.

Pooling diverse genetic and physical interaction data sets widens the search but also has the potential to decrease the coverage of network models, because not all data sets may be equally predictive and high-scoring network models are more likely to arise at random in large networks. To investigate the effect of data pooling, we repeated the search on a smaller network comprising large-scale synthetic-lethal (SGA) and protein-binding (DIP) interactions only. This reduced search identified 20 models containing a total of 137 synthetic-lethal and 120 protein-binding interactions (Fig. 2). In comparison to the complete search, fewer protein-binding and SGA synthetic-lethal interactions were incorporated into models, demonstrating the synergy obtained by data pooling (although models generated by the restricted search performed somewhat better in validation). Supplementary Table 4 online analyzes the impact of removing each physical and genetic data set from the modeling procedure.

## Within-pathway interpretations

We next searched the physical and genetic networks for within-pathway explanations. This procedure assigned a high score to single sets of proteins that were densely connected by both physical and genetic interactions (see Fig. 1, Box 2 and Supplementary Fig. 2 online). This search yielded 91 significant models. In all, these contained 272 MIPS, 225 SGA synthetic lethal and 169 SGA synthetic-sick interactions associated with 318 protein-binding, 37 regulatory and 36 shared-reaction interactions. Four representative within-pathway models are shown in Figure 3.

## Functional enrichment of models

As initial validation of the between- and within-pathway models, we found that both types were significantly enriched for particular functional annotations recorded in the Gene Ontology database<sup>27</sup>. Two-hundred and fifty-one out of 401 pathways in between-pathway models were enriched for proteins with a common Molecular Function, Biological Process or Cellular Component annotation using the hypergeometric test ( $P = 0.05$ ; Bonferroni-corrected for multiple testing)<sup>28</sup>. Similarly, 52 of the 91 within-pathway models were enriched for Gene Ontology annotations. Moreover, these functional enrichments were higher than expected based on the physical interaction network alone (see Supplementary Table 5 online).

## Prediction of new protein functions

Having established that proteins in many of the between- and within-pathway models were enriched for specific annotations, we used this concept to predict new protein functions. Specifically, for physical pathways in which a majority of proteins were already assigned a common significant annotation, we predicted this term for the remaining proteins in the pathway. To eliminate overly general predictions, significance was assigned only to those terms that were enriched at a level of  $P = 0.05$  and were associated with fewer than 100 yeast proteins overall.

For between-pathway models, this approach predicted 745 molecular function, biological process or cellular component annotations among 282 proteins. In comparison, the within-pathway models predicted 285 annotations involving 127 proteins, bringing the total to 973 annotations for 343 proteins accounting for repeated predictions. A list of novel functional predictions is provided in Supplementary Table 6 online. Less than a quarter of these predictions were attainable using a similar approach based on the physical network only (Supplementary Table 7 online).

Accuracy of these predictions was estimated using cross validation<sup>29</sup>. Using a standard five-way procedure, the set of yeast proteins was partitioned such that annotations were hidden for one-fifth of the proteins and annotations for the remaining four-fifths of proteins were used to predict the hidden information. Each prediction for a protein in the 'hidden set' was scored as a success or failure depending on whether it recovered a hidden annotation. Using this approach, the success rate was estimated to be 63% for between-pathway models, 69% for within-pathway models.

## Prediction of new genetic interactions

Finally, we investigated whether the network models could predict the existence of new genetic interactions (Fig. 4). According to the between-pathway model, proteins in one pathway genetically interact with many of the same partners in a second pathway. This leads to the occurrence of 'complete bipartite motifs' in the genetic interaction network, defined as four-protein subnetworks in which the first two proteins are connected to the second two proteins by all four possible genetic interactions (Fig. 4a; see Milo *et al.*<sup>30</sup> for an introduction to network motifs). When an incomplete motif (IM) is observed, for which only three of the four genetic interactions are present, the motif implies that the remaining interaction is true. Physical network information is incorporated by requiring that valid incomplete motifs fall within (i.e., are subgraphs of) a between-pathway model.

We applied the technique of five-way cross-validation to estimate the accuracy of genetic interaction prediction versus the minimum number of required incomplete motifs (Fig. 5). In each of five cross-validation trials, approximately one-fifth of the genetic interaction data were withheld, including both positive and negative interactions measured for each genetic 'bait' in SGA. These positive and negative interactions were subsequently used to test prediction accuracy. For instance, at a prediction threshold of eight or more incomplete motifs, the between-pathway models predicted 43 new genetic interactions with 87% estimated accuracy (Fig. 5). To assess the contribution of the physical models in the prediction process, we also predicted 'naive' genetic interactions by relaxing the requirement that incomplete motifs fall in a between-pathway model. The estimated accuracy fell to 5% for these naive predictions, evaluated at the same threshold of eight incomplete motifs.

For the within-pathway models, genetic interactions were implied between proteins that had genetic interactions with one or more common neighbors (Fig. 4b). The physical network was incorporated by restricting the proteins and neighbors to fall into a single within-pathway

model. The number of common neighbors was used as a measure of confidence in the implied genetic interaction, and cross validation was used to estimate the prediction accuracy as a function of this number. The maximal prediction accuracy was 38%, achieved at a prediction threshold of three or more common neighbors (Supplementary Fig. 3 online). The corresponding success rate for naive predictions, made without constraining the proteins to occur in within-pathway models, was 15%. Thus, both types of models enhance the accuracy of prediction of genetic interactions, but between-pathway models appear to be better predictors than within-pathway models.

## Preponderance of between-pathway interactions

Given a systematic approach for associating genetic interactions with physical interpretations, it is of interest to ask which type of interpretation is most common. Focusing on large-scale SGA measurements, roughly three-and-a-half times as many genetic interactions are associated with between- as opposed to within-pathway models (1,377 versus 394 SGA interactions). These figures can be viewed as an a priori expectation that a newly determined SGA interaction will fall between versus within pathways, suggesting that SGA interactions typically span between multiple physical network regions instead of occurring within a single complex or pathway. One reason for the preference towards between-pathway models may be that SGA interactions are mainly targeted to nonessential genes (due to their use of complete gene deletions as opposed to, *e.g.*, point mutations made by classical techniques).

Using physical models, it is possible to characterize approximately 40% of the genetic interactions as occurring between or within pathways. Whether the remaining interactions belong to between-pathway models, within-pathway models or are best characterized as 'indirect' (Box 1) cannot be reliably determined at this stage. For example, consider the case of two related pathways, each with only one protein required for pathway function. In this case, only the required proteins would be connected by a (single) genetic interaction across the pathways, making it difficult for the between-pathway model to achieve statistical significance.

Further examination of the between-pathway models reveals that many of the genetically linked pathways have clear interdependent functional relationships. For example, pathway M contains members of the prefoldin complex, which have synthetic-lethal interactions with members of pathways N and T forming parts of the dynactin complex and kinetochore, respectively (Fig. 2a). The prefoldin complex promotes folding of  $\alpha$ - and  $\beta$ -tubulin into functional microtubules<sup>31</sup>. These are important for the function of dynactin, an adaptor complex involved in translocating the spindle and other molecular cargos along microtubules<sup>32</sup>, as well as the kinetochore, which anchors chromosomes to spindle microtubules during metaphase<sup>33</sup>. Apparently, deletion of proteins in the prefoldin complex reduces micro-tubule stability, leading to synthetic-lethal interactions with pathways that are directly dependent on microtubule function.

These pathways also predict a new function for the uncharacterized protein YII049w (pathway N). This protein binds Jnm1, a dynactin protein which is required for spindle partitioning in anaphase<sup>32</sup>. In addition, it has synthetic-lethal interactions with members of the prefoldin complex in a manner similar to dynactin genes. Together, these relationships suggest that YII049w is associated with dynactin during spindle partitioning. However, because Jnm1 has 12 physical interactions overall, and YII049w has a total of 14 interactions in the genetic network, this prediction would have been difficult to make without an integrated approach.

Pathways O, U and Y provide another example of synergistic pathways linked by genetic interactions (Fig. 2a). Pathways U and Y mediate retrograde transport of proteins to the Golgi apparatus<sup>34,35</sup>. Pathway O (Bre1, Lge1) is involved in histone ubiquitination and cell size control, where cell size is influenced by the histone ubiquitination activity by an unknown

process<sup>36</sup>. The abundant genetic interactions between pathways O and U indicate a possible role for retrograde transport in histone ubiquitination, or reciprocally, for histone ubiquitination in retrograde transport. Moreover, the uncharacterized protein Yel043w is physically associated with Bre1 and Lge1 and also has the same pattern of genetic interactions, suggesting that the three proteins may function together.

In summary, we have presented a methodology for integrating large-scale genetic and physical networks to capture the physical context behind observed genetic interactions. Approximately 40% of yeast synthetic-lethal genetic interactions can be incorporated into high-level physical pathway models and are approximately three and a half times as likely to span pairs of pathways than to occur within pathways. Further studies will be needed to address other types of genetic effects to extend this approach from yeast to the growing number of other organisms for which protein networks are now available. As systematic approaches generate ever larger databases of interactions across a variety of species, integrative modeling approaches such as the one proposed here will be indispensable for selecting and organizing the information into predictive models.

### Box 1

#### Interpretations of genetic interactions

**Between-pathway interpretations.** The genetic interaction bridges genes operating in two pathways with redundant or complementary functions. Deletion of either gene is expected to abrogate the function of one but not both pathways.

**Within-pathway interpretations.** The genetic interaction occurs between protein subunits within a single pathway. A single gene is dispensable for the function of the overall pathway, but the additive effects of several gene deletions are lethal.

**Indirect effects.** The synthetic lethal phenotype is not mediated by a localized mechanism in the physical network. Indirect effects can occur because a deletion phenotype represents not just the absence of one particular gene, but also the response of the cell to its absence, involving many diverse pathways<sup>21</sup>.

### Box 2

#### Scoring the models

**Scoring within-pathway explanations.** The within-pathway model implies dense interactions within a single group of proteins in both the physical and genetic networks. We adopt a previously described log-odds score<sup>37</sup> to assess the likelihood that a group of proteins is more densely connected than would be expected at random:

$$S_{within}(V, E) = \log \frac{P(V, E | Model_{dense})}{P(V, E | Model_{random})}$$

$$= \log \frac{\prod_{(a,b) \in V \times V} \beta^{I_E(a,b) + (1-\beta)(1-I_E(a,b))}}{\prod_{(a,b) \in V \times V} r_{a,b}^{I_E(a,b) + (1-r_{a,b})(1-I_E(a,b))}}$$

where  $V$  is a set of proteins and  $E$  a set of interactions among those proteins (genetic or physical).  $I_E(a,b)$  is an indicator function which equals 1 if and only if the interaction  $(a,b)$  occurs in  $E$  and otherwise 0. For  $Model_{dense}$ , interactions are expected to occur with high probability ( $\beta$ ) for every pair of proteins in  $V$ . In this work,  $\beta$  is set to 0.9

(Supplementary Fig. 2 shows how the results depend on choice of  $\beta$ ). For *Model<sub>random</sub>*, the probability of observing each interaction ( $r_{a,b}$ ) is determined by estimating the fraction of all networks with identical degree distribution which also contain that interaction.

Comparable random networks are generated by ‘crossing’ pairs of edges in a process similar to that described by Milo *et al.*<sup>30</sup> In this randomization, only edges of the same type are allowed to be crossed. In addition, for undirected types, either interacting node is allowed to serve as the ‘source’ in crossing the edges. Such randomization generates a family of random networks which resemble the original network and corrects for the presence of highly connected proteins, which score highly under both models. The interaction density is evaluated independently for the physical and genetic networks, yielding an overall score for the within-pathway model:

$$S = S_{\text{within}}(V, E_{\text{physical}}) + S_{\text{within}}(V, E_{\text{genetic}}).$$

**Scoring between-pathway explanations.** The between-pathway model implies dense genetic interactions connecting two separate, nonoverlapping groups of proteins, where each group is densely connected by physical interactions. The density of physical interactions is scored independently within two sets of proteins  $V_1$  and  $V_2$  using the above function  $S$ . A related log-odds score is used to evaluate the probability that the genetic interactions  $E_{\text{genetic}}$  bridging these sets are denser than random:

$$S_{\text{between}}(V_1, V_2, E_{\text{genetic}}) = \log \frac{\prod_{(a,b) \in V_1 \times V_2} \beta I_{E_{\text{genetic}}}(a,b) + (1-\beta)(1 - I_{E_{\text{genetic}}}(a,b))}{\prod_{(a,b) \in V_1 \times V_2} r_{a,b} I_{E_{\text{genetic}}}(a,b) + (1 - r_{a,b})(1 - I_{E_{\text{genetic}}}(a,b))}$$

The final scoring function for the between-pathway model is then:

$$S(V_1, V_2, E_{\text{all}}) = \sum_{i=1,2} S_{\text{within}}(V_i, E_{\text{physical}}) + S_{\text{between}}(V_1, V_2, E_{\text{genetic}})$$

**Search and Significance.** Sets of proteins that are well explained by either the within-pathway or between-pathway models are identified using a greedy network search procedure. The search is as previously described by Sharan *et al.*<sup>37</sup> except that it is seeded from each pair of genetically interacting proteins. Pathways that share more than 50% of genetic interactions with a higher-scoring result are discarded. To determine the significance threshold, identical searches are performed over 100 random trials in which both the genetic and physical networks are randomized as described above. Models that score higher than the maximal-scoring models in 95% of random trials are reported as significant.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank Jonathan Wang, Owen Ozier and Gopal Ramachandran for preliminary investigations and Vineet Bafna, Ben Raphael and Vikas Bansal for insightful commentary. Craig Mak, Silpa Suthram and Taylor Sittler provided

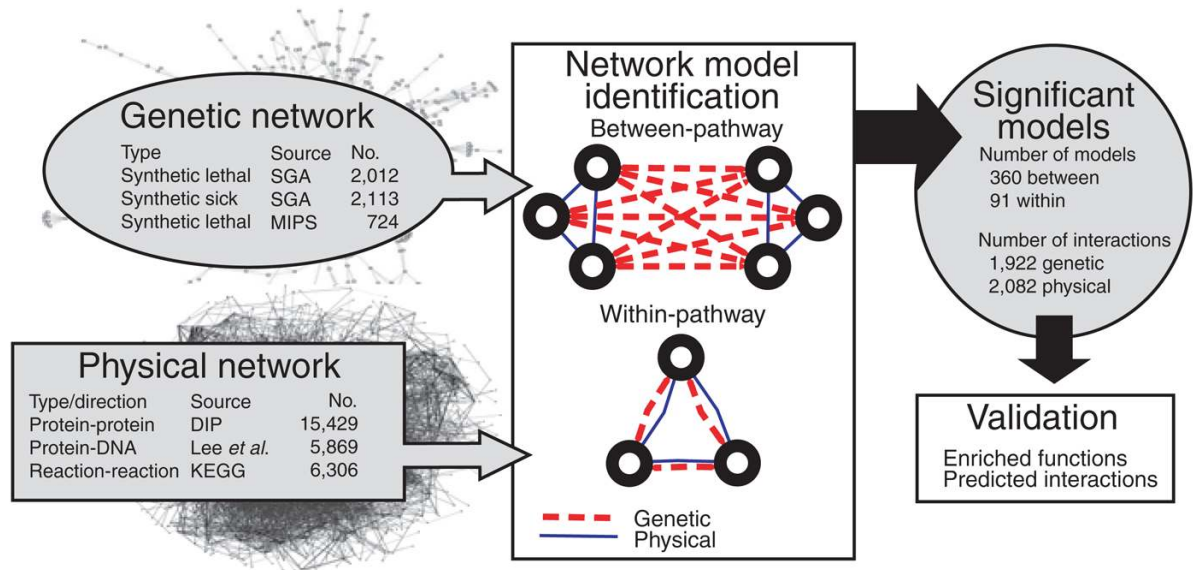
helpful reviews of the text. Funding was provided by the National Institute of General Medical Sciences (GM070743-01) and the National Science Foundation (NSF 0425926).

## References

1. Guarente L. Synthetic enhancement in gene interaction: a genetic tool come of age. *Trends Genet* 1993;9:362–366. [PubMed: 8273152]
2. Avery L, Wasserman S. Ordering gene function: the interpretation of epistasis in regulatory hierarchies. *Trends Genet* 1992;8:312–316. [PubMed: 1365397]
3. Thomas JH. Thinking about genetic redundancy. *Trends Genet* 1993;9:395–399. [PubMed: 8310537]
4. Hartman JL, Garvik B, Hartwell L. Principles for the buffering of genetic variation. *Science* 2001;291:1001–1004. [PubMed: 11232561]
5. Sham P. Shifting paradigms in gene-mapping methodology for complex traits. *Pharmacogenomics* 2001;2:195–202. [PubMed: 11535109]
6. Dolma S, Lessnick SL, Hahn WC, Stockwell BR. Identification of genotype-selective antitumor agents using synthetic lethal chemical screening in engineered human tumor cells. *Cancer Cell* 2003;3:285–296. [PubMed: 12676586]
7. Forsburg SL. The art and design of genetic screens: yeast. *Nat Rev Genet* 2001;2:659–668. [PubMed: 11533715]
8. Tong AH, et al. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* 2001;294:2364–2368. [PubMed: 11743205]
9. Ooi SL, Shoemaker DD, Boeke JD. DNA helicase gene interaction network defined using synthetic lethality analyzed by microarray. *Nat Genet* 2003;35:277–286. [PubMed: 14566339]
10. Tong AH, et al. Global mapping of the yeast genetic interaction network. *Science* 2004;303:808–813. [PubMed: 14764870]
11. Li S, et al. A map of the interactome network of the metazoan *C. elegans*. *Science* 2004;303:540–543. [PubMed: 14704431]
12. Rain JC, et al. The protein-protein interaction map of *Helicobacter pylori*. *Nature* 2001;409:211–215. [PubMed: 11196647]
13. Uetz P, et al. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 2000;403:623–627. [PubMed: 10688190]
14. Giot L, et al. A protein interaction map of *Drosophila melanogaster*. *Science* 2003;302:1727–1736. [PubMed: 14605208]
15. Ito T, et al. Toward a protein-protein interaction map of the budding yeast: a comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proc Natl Acad Sci USA* 2000;97:1143–1147. [PubMed: 10655498]
16. Ho Y, et al. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 2002;415:180–183. [PubMed: 11805837]
17. Gavin A-C, Bösch M, Krause R, Grandi P, Marzioch M. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 2002;415:141–147. [PubMed: 11805826]
18. Lee TI, et al. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 2002;298:799–804. [PubMed: 12399584]
19. Harbison CT, et al. Transcriptional regulatory code of a eukaryotic genome. *Nature* 2004;431:99–104. [PubMed: 15343339]
20. Ozier O, Amin N, Ideker T. Global architecture of genetic interactions on the protein network. *Nat Biotechnol* 2003;21:490–491. [PubMed: 12721566]
21. Tucker CL, Fields S. Lethal combinations. *Nat Genet* 2003;35:204–205. [PubMed: 14593402]
22. Winzler EA, et al. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* 1999;285:901–906. [PubMed: 10436161]
23. Mewes HW, et al. MIPS: a database for genomes and protein sequences. *Nucleic Acids Res* 2002;30:31–34. [PubMed: 11752246]

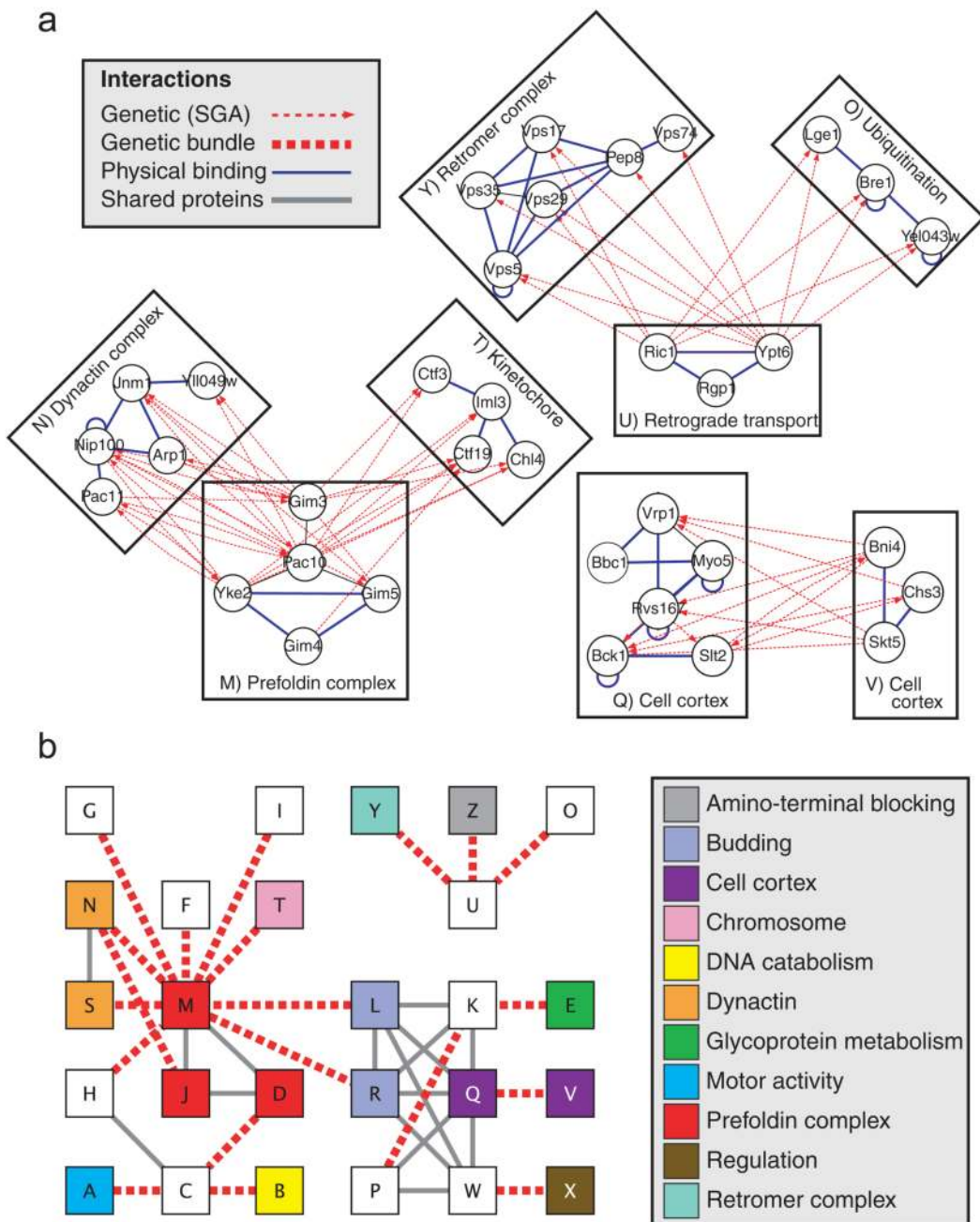


24. Xenarios I, et al. DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res* 2002;30:303–305. [PubMed: 11752321]
25. Kanehisa M, Goto S, Kawashima S, Nakaya A. The KEGG databases at GenomeNet. *Nucleic Acids Res* 2002;30:42–46. [PubMed: 11752249]
26. Shannon P, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498–2504. [PubMed: 14597658]
27. Gene Ontology Consortium. Creating the gene ontology resource: design and implementation. *Genome Res* 2001;11:1425–1433. [PubMed: 11483584]
28. Zar, JH. *Biostatistical Analysis*. 3. Prentice Hall; New Jersey: 1996.
29. Kendall, SM.; Stuart, A.; Ord, JK. *Kendall's Advanced Theory of Statistics*. 5. Oxford University Press; NY: 1987.
30. Milo R, et al. Network motifs: simple building blocks of complex networks. *Science* 2002;298:824–827. [PubMed: 12399590]
31. Geissler S, Siegers K, Schiebel E. A novel protein complex promoting formation of functional alpha- and gamma-tubulin. *EMBO J* 1998;17:952–966. [PubMed: 9463374]
32. Kahana JA, et al. The yeast dynactin complex is involved in partitioning the mitotic spindle between mother and daughter cells during anaphase B. *Mol Biol Cell* 1998;9:1741–1756. [PubMed: 9658168]
33. Pidoux AL, Allshire RC. Centromeres: getting a grip of chromosomes. *Curr Opin Cell Biol* 2000;12:308–319. [PubMed: 10801468]
34. Pfeffer SR. Membrane transport: retromer to the rescue. *Curr Biol* 2001;11:R109–R111. [PubMed: 11231171]
35. Siniossoglou S, Peak-Chew SY, Pelham HR. Ric1p and Rgp1p form a complex that catalyses nucleotide exchange on Ypt6p. *EMBO J* 2000;19:4885–4894. [PubMed: 10990452]
36. Hwang WW, et al. A conserved RING finger protein required for histone H2B monoubiquitination and cell size control. *Mol Cell* 2003;11:261–266. [PubMed: 12535538]
37. Sharan, R.; Ideker, T.; Kelley, BP.; Shamir, R.; Karp, RM. Identification of protein complexes by comparative analysis of yeast and bacterial protein interaction data. *Proceedings of the Eighth Annual International Conference on Research in Computational Molecular Biology–RECOMB*; New York: ACM Press; 2004. p. 282–289.

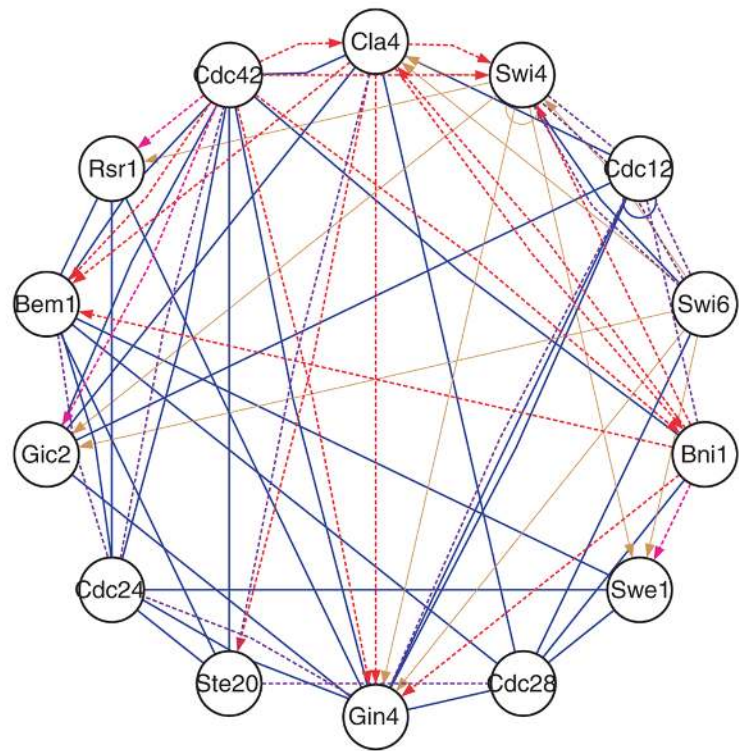
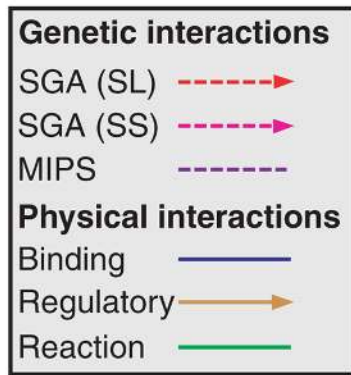


**Figure 1.**

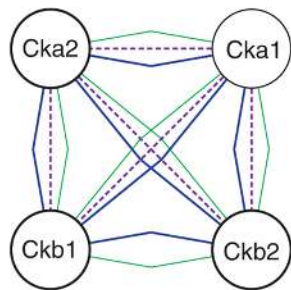
Method overview. A combined physical and genetic network is searched to identify between- or within-pathway models of genetic interactions. The between-pathway model implies two groups of proteins (pathways) with many physical connections within each pathway (solid blue links) and genetic interactions spanning between pathways (dotted red links). The within-pathway model implies many physical and genetic interactions within the same group of proteins. In the search, 360 and 91 network models were identified that correspond to between- or within-pathway searches, respectively.



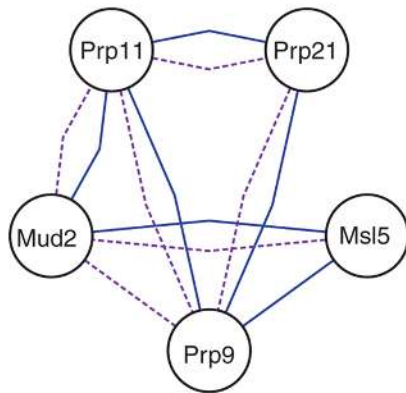
**Figure 2.** Between-pathway explanations for genetic interactions. **(a)** Several high-scoring models are shown (M,N,T; Q,V; O,U,Y). Blue solid and red dotted links indicate physical and genetic interactions, respectively. **(b)** Bird's-eye view of all between-pathway models obtained from a search on a reduced network composed of SGA and DIP interactions. Each node [A]–[Z] represents a physical pathway; groups of genetic interactions between pathways are condensed into a single link called a 'bundle.' Node colors indicate significant Gene Ontology annotations. Solid gray lines connect pathways that share one or more proteins; such pathways may represent different components of a larger mechanism.



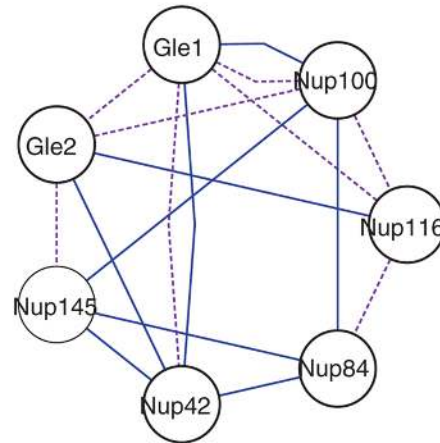
Cellular morphogenesis



CK2 complex

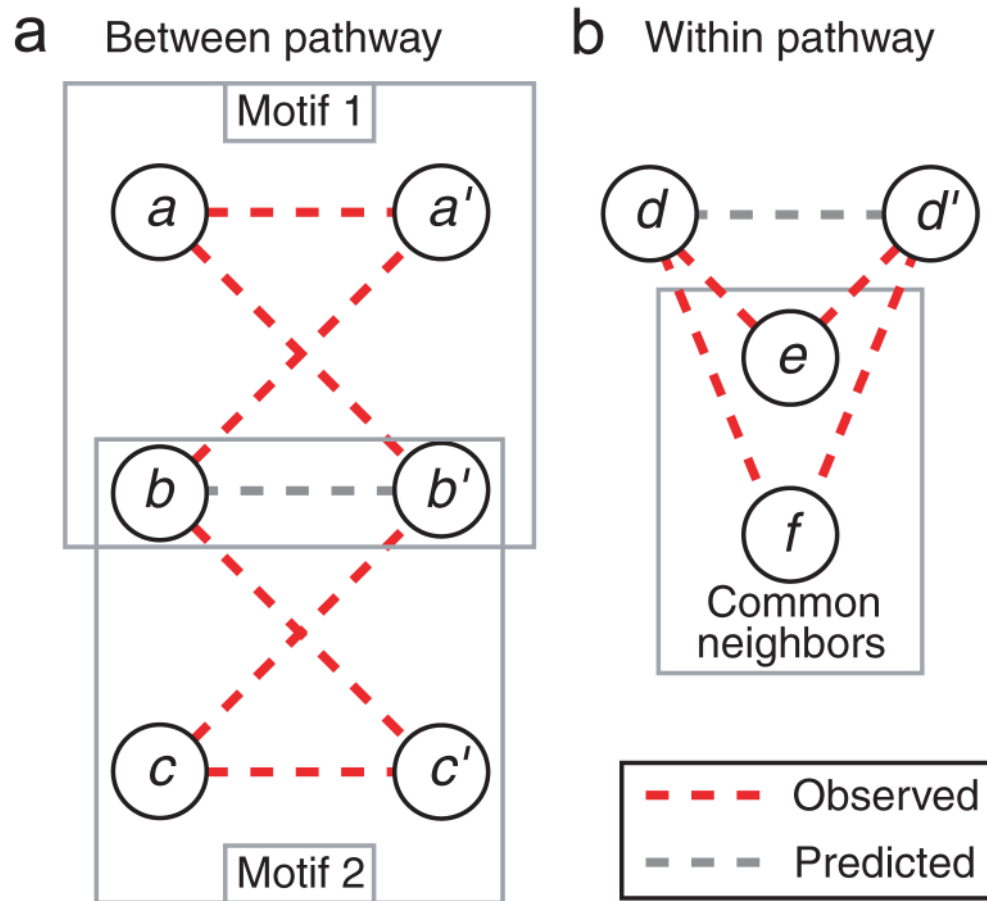


Spliceosome

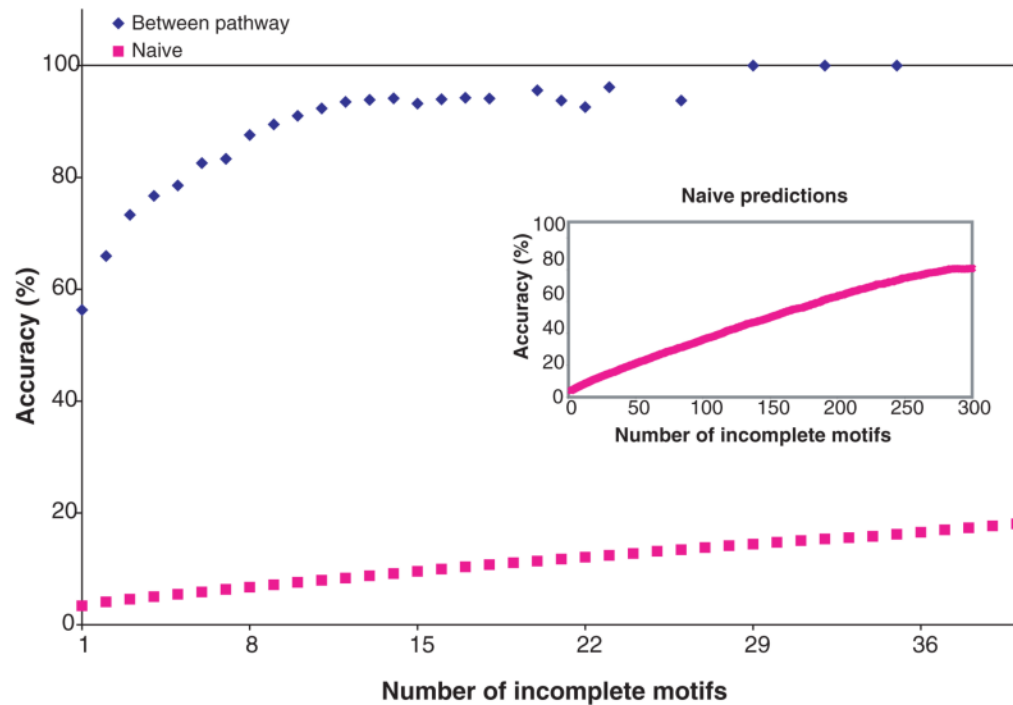


Nucleic acid and related transport

**Figure 3.** Within-pathway explanations for genetic interactions. A total of 91 pathways were identified, of which four examples are displayed. Color is used to indicate the data set from which each interaction was drawn.



**Figure 4.** Genetic interaction prediction schemes. Two different schemes are proposed for predicting genetic interactions, depending on the underlying network model. Observed genetic interactions are shown in red, while the corresponding predicted genetic interactions are shown in gray. **(a)** Under the between-pathway model, two incomplete bipartite motifs are shown which predict a genetic interaction between genes  $b$  and  $b'$ . **(b)** Under the within-pathway model, common genetic neighbors are used to predict a genetic interaction between genes  $d$  and  $d'$ . Note that these diagrams contain additional incomplete motifs which have been omitted for clarity: the motifs in **a** can be rearranged to predict genetic interactions ( $a$  to  $c'$ ) and ( $c$  to  $a'$ ); the motifs in **b** can be rearranged to predict ( $e$  to  $f$ ).



**Figure 5.**

Success rate of genetic interaction prediction versus the stringency of prediction. Success rate is measured through cross-validation as (predicted positives)/(predicted positives and negatives). Stringency is defined by the minimum number of incomplete bipartite motifs required for prediction. Blue diamonds mark the success rate for predictions in which incomplete motifs must occur in a between-pathway model. The success rate is dramatically higher than for naive predictions (magenta) which predict interactions in the same manner, but are not constrained by the physical network. Even for much more stringent prediction criteria, the success rate of naive predictions fails to exceed that of the between-pathway predictions (inset).