Functional Maps of Protein Complexes from Quantitative Genetic Interaction Data

Sourav Bandyopadhyay[1,2], Ryan Kelley[1,2], Nevan J. Krogan[3], Trey Ideker[1,2]*

1 Program in Bioinformatics, University of California San Diego, La Jolla, California, United States of America, 2 Department of Bioengineering, University of California San Diego, La Jolla, California, United States of America, 3 Department of Cellular and Molecular Pharmacology, University of California San Francisco, San Francisco, California, United States of America

Abstract

Recently, a number of advanced screening technologies have allowed for the comprehensive quantification of aggravating and alleviating genetic interactions among gene pairs. In parallel, TAP-MS studies (tandem affinity purification followed by mass spectroscopy) have been successful at identifying physical protein interactions that can indicate proteins participating in the same molecular complex. Here, we propose a method for the joint learning of protein complexes and their functional relationships by integration of quantitative genetic interactions and TAP-MS data. Using 3 independent benchmark datasets, we demonstrate that this method is >50% more accurate at identifying functionally related protein pairs than previous approaches. Application to genes involved in yeast chromosome organization identifies a functional map of 91 multimeric complexes, a number of which are novel or have been substantially expanded by addition of new subunits. Interestingly, we find that complexes that are enriched for aggravating genetic interactions (i.e., synthetic lethality) are more likely to contain essential genes, linking each of these interactions to an underlying mechanism. These results demonstrate the importance of both large-scale genetic and physical interaction data in mapping pathway architecture and function.

Introduction

Genetic interactions are logical relationships between genes that occur when mutating two or more genes in combination produces an unexpected phenotype [1–3]. Recently, rapid screening of genetic interactions has become feasible using Synthetic Genetic Arrays (SGA) or diploid Synthetic Lethality Analysis by Microarray (dSLAM) [4,5]. SGA pairs a gene deletion of interest against a deletion to every other gene in the genome (in turn). The growth/no growth phenotype measured over all pairings defines a *genetic interaction profile* for that gene, with no growth indicating a synthetic-lethal genetic interaction. Alternatively, all combinations of double deletions can be analyzed among a functionally-related group of genes [6–8]. A recent variant of SGA termed E-MAP [9] has made it possible to measure continuous rates of growth with varying degrees of epistasis (based on imaging of colony sizes). "Aggravating" interactions are indicated if the growth rate of the double gene deletion is slower than expected, while for "alleviating" interactions the opposite is true [10,11].

One popular method to analyze genetic interaction data has been to hierarchically cluster genes using the distance between their genetic interaction profiles. Clusters of genes with similar profiles are manually searched to identify the known pathways and complexes they contain as well as any genetic interactions between these complexes. This approach has been applied to several large-scale genetic interaction screens in yeast including genes involved in the secretory pathway [8] and chromosome organization [6]. Segré et al. [12] extended basic hierarchical clustering with the concept of "monochromaticity", in which genes were merged into the same cluster based on minimizing the number of interactions with other clusters that do not share the same classification (aggravating or alleviating).

Another set of methods has sought to interpret genetic relationships using physical protein-protein interactions [13]. Among these, Kelley and Ideker [14] used physical interactions to identify both "within-module" and "between-module" explanations for genetic interactions. In both cases, modules were detected as clusters of proteins that physically interact with each other more often than expected by chance. The "within-module" model predicts that these clusters directly overlap with clusters of genetic interactions. The "between-module" model predicts that genetic interactions run between two physical clusters that are functionally related. This approach was improved by Ulitsky et al. [15] using a relaxed definition of physical modules. In related work, Zhang et al. [16] screened known complexes annotated by the Munich Information Center for Protein Sequences (MIPS) to identify pairs of complexes with dense genetic interactions between them.

One concern with the above approaches, and the works by Kelley and Ulitsky in particular, is that they make assumptions about the density of interactions within and between modules which have not been justified biologically. Ideally, such parameters should be learned directly from the data. Second, between-module relationships are identified by separate, independent searches of the network seeded from each genetic interaction. This "local" search strategy can lead to a set of modules that are highly

Author Summary

Biologists are currently producing large amounts of data focused on physical and genetic protein interactions. Physical interactions dictate the architecture of the cell in terms of how direct associations between molecules constitute protein complexes, while genetic interactions define functional relationships through cause-and-effect relationships between genes. Both of these types of interactions can indicate shared protein functions; however, these two types of interactions are commonly non-overlapping, making their interpretation difficult. Along these lines, it has been noted that genetic interactions commonly occur between members of the same protein complex as well as between functionally related complexes. Here, we present an integrated framework that incorporates both types of interactions to generate large maps of protein complexes as well as highlight connections between related complexes. The ability to rapidly integrate these two types of data in an automated fashion can accelerate the discovery of new members of protein complexes as well as identify functionally related cellular components.

overlapping or even completely redundant with one another. Finally, genetic interactions are assumed to be binary growth/no growth events while E-MAP technology has now made it possible to measure continuous values of genetic interaction with varying degrees of epistasis. Here, we present a new approach for integrating quantitative genetic and physical interaction data which addresses several of these shortcomings. Interactions are analyzed to infer a set of modules and a set of inter-module links, in which a module represents a protein complex with a coherent cellular function, and inter-module links capture functional relationships between modules which can vary quantitatively in strength and sign. Our approach is supervised, in that the appropriate pattern of physical and genetic interactions is not predetermined but learned from examples of known complexes. Rather than identify each module in independent searches, all modules are identified simultaneously within a single unified map of modules and inter-module functional relationships. We show that this method outperforms a number of alternative approaches and that, when applied to analyze a recent EMAP study of yeast chromosome function, it identifies numerous new protein complexes and protein functional relationships.

Results

Characterization of Genetic and Physical Networks

We first sought to quantitatively confirm whether, and to what degree, physical and genetic interactions could indicate common membership in a protein complex. To provide genetic data for analysis, we obtained the previously-published results from a large E-MAP of yeast chromosomal biology [6]. This study consisted of genetic interactions measured among 743 genes (including 74 essential genes), yielding quantitative values for 182,669 gene pairs (66% of all possible pair-wise measurements). Each gene pair was assigned an S-score, where positive scores indicate protein pairs for which the double mutant grows better than expected (i.e., an alleviating interaction) and negative scores indicate pairs for which the double mutant grows worse than expected (i.e., a synthetic sick/lethal or aggravating interaction) where the expectation is that the double-deletion of unrelated proteins will have a growth rate equivalent to the multiplicative product of the two individual growth rates [17]. In all, 14,237 gene pairs (8%) showed strong genetic interactions with $|S| > 2.5$. Physical interactions were taken from a recent computational integration of two large datasets measured by co-immunoprecipitation followed by mass spectrometry [18]. This study assigned to each pairwise interaction a Purification Enrichment (PE) score, with larger values representing a greater likelihood of true binding.

Figure 1A confirms that protein pairs with higher PE-scores are more likely to operate in a known small-scale protein complex recorded in the MIPS database [19] (versus protein pairs chosen at
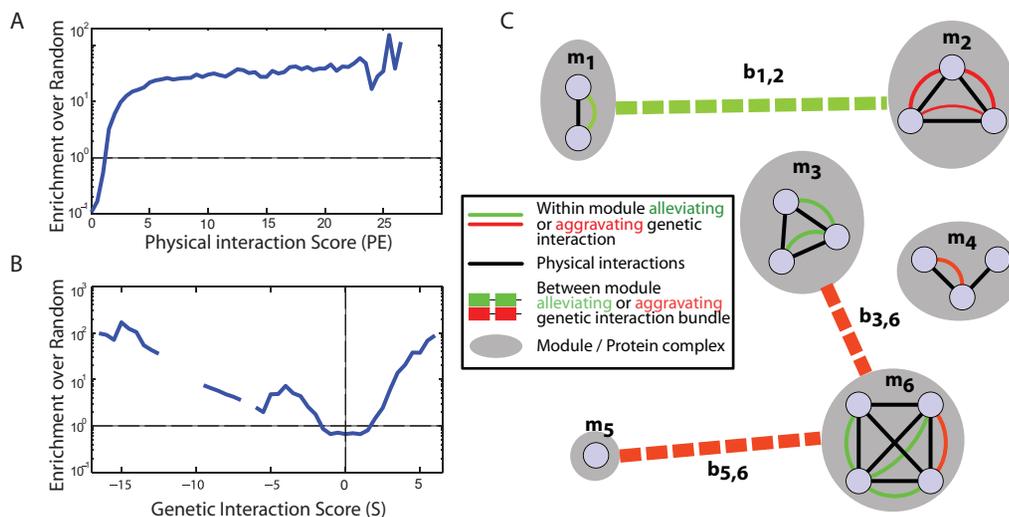
Figure 1. Combining physical and genetic interactions to define protein complexes. Correspondence of the physical interaction score (A) and the genetic interaction score (B) with the known small-scale, manually annotated protein complexes in MIPS. To compute the enrichment over random (y-axis), one first computes the fraction f of interactions at each score x that fall inside a MIPS small-scale complex (bin size of 1.5). The enrichment is the ratio f/r, where r is the fraction of random protein pairs within MIPS complexes. (C) Proteins are grouped into physically interacting sets called modules (gray ovals; $m_1 - m_6$). Pairs of modules may be linked to indicate a functional relationship (dotted lines; $b_1 - b_6$). The assignment of proteins to modules along with the list of inter-module links comprises the state of the system.
doi:10.1371/journal.pcbi.1000065.g001

random). This result is expected considering that PE-scores were trained based on these complexes [18]. Figure 1B shows that protein pairs with both positive and negative S-scores are more likely to operate within a known complex. Positive (alleviating) interactions are well-known to occur between subunits of a complex [6]. Negative (aggravating) interactions are to a lesser degree so, although the mechanism has not been as clear as for the alleviating case [20]. By comparing the magnitudes of enrichment between Figures 1A and 1B, it is apparent that extreme S-scores are at least as indicative of co-complex membership as strong PE-scores, if not more so (~100-fold enrichment versus ~50-fold enrichment, respectively). Together, these exploratory findings suggest that the physical and genetic information can indeed provide a basis for the identification of protein pairs involved in the same complex.

Functional Maps of Protein Complexes Involved in Yeast Chromosomal Biology

To capture these trends, we formulated an approach to classify a protein pair as operating either within the same module or between functionally related modules given its genetic and physical interaction scores. This approach seeks to categorize interactions supported by both strong genetic and physical evidence as operating within a module (i.e., complex). Interactions with a strong genetic but weak physical signal are better characterized as operating between two functionally related modules. Given within-module and between-module likelihoods for individual interactions, an agglomerative clustering procedure seeks to merge these interactions into increasingly larger modules and to identify pairs

of modules interconnected by bundles of many strong genetic interactions (Figure 1C). Full details are provided in Methods.

Applying this method, we identified 91 distinct modules with an average size of 4.1 proteins per module. Figure 2 gives an overview of a subset of the identified modules and inter-module links. Complete results are catalogued at http://www.cellcircuits.org/Bandyopadhyay2008/html/. Overall, these results suggest ten novel complexes not recorded in either the small-scale or high-throughput MIPS compendium, covering 23 proteins in total. The results also identify 84 new subunits of known complexes (Dataset S1). Through permutation testing, 19 versus 9 of the identified modules could be categorized as enriched for alleviating or aggravating genetic interactions, respectively. A total of 313 significant genetic relationships were identified between modules, 94 versus 219 of which were enriched for alleviating or aggravating interactions.

Comparison to Related Approaches

The method of choice for interpreting quantitative genetic interactions has been hierarchical clustering (HCL) of genes based on pair-wise distances between their genetic interaction profiles [6,8]. We compared the clusters obtained using HCL to the modules obtained with our present approach (Bandyopadhyay *et al.*) using three gold-standard metrics: gene co-expression (Figure 3A), co-functional annotation (Figure 3B), or membership in the same previously-identified complex (Figure 3C). To ensure a fair comparison between the two approaches, HCL and Bandyopadhyay *et al.* were evaluated across a range of coverages (number of gold-standard gene pairs recovered by the predicted

Figure 2. Global map of protein complexes involved in yeast chromosome biology. Each node represents a predicted multimeric protein complex, while each link represents a significantly alleviating or aggravating bundle of genetic interactions between complexes, indicative of an inter-complex functional relationship. Node colors indicate enrichment for alleviating or aggravating genetic interactions among members of the same complex. Node sizes are proportional to the number of proteins in the complex. When known, nodes are labeled with the common name of the complex. For complexes that are newly identified by our study and thus unnamed, the constituent proteins are listed. For clarity, the co-chaperone prefoldin complex (PFD1, PAC10, YKE2, GIM3, GIM4, GIM5, BUD27) and the 25 links associated with it have been removed.
doi:10.1371/journal.pcbi.1000065.g002

clusters/modules; see Methods). For all three benchmarks, our performance was substantially higher than that of the HCL-based approach at most levels of coverage (and at a level of coverage corresponding to the 91 modules reported above; dotted vertical line in Figure 3).

We considered that one reason why HCL performed less favorably might be that it was not given access to the same information (i.e., the physical network). This is especially true for the metric based on previously identified complexes, in which complexes were annotated based on the same high-throughput protein interactions used here. To investigate this possibility, we extended HCL to incorporate physical interactions in a straight-forward fashion, by merging only those clusters which share a physical interaction between them (HCL-PE). Although this approach outperformed hierarchical clustering without physical interactions, it was outperformed by the present approach by at least 50% across the three metrics. Finally, our method also shows improvement over the previous approach of Kelley and Ideker [14] for integrating genetic and physical interactions (Figure 3).

Aggravating Complexes Tend to be Essential

Nineteen versus nine of the learned modules had significant enrichment for alleviating versus aggravating genetic interactions, respectively. Identification of "alleviating" modules is expected, since subunits of a complex operate together and the phenotypic effect of removing any pair of proteins in a complex should be no worse than removing any single protein individually. The presence of aggravating interactions within modules was more intriguing. One way in which aggravating interactions could occur among the subunits of a complex is if its function is essential, i.e., the loss of the complex's function causes a lethal phenotype. In these cases, some protein subunits should be encoded by essential genes, while other subunits might be redundant and thus essential in pairwise combinations [20].

To test the hypothesis that essential genes are more likely in aggravating modules, we analyzed both MIPS small-scale complexes and our learned modules for the presence of essential genes (Figure 4). We found that 80% of aggravating MIPS complexes contained an essential gene, compared to only 20% of alleviating MIPS complexes (a four-fold increase). Similarly, of the aggravating modules determined by our approach, 55% contained an essential gene compared to only 21% of alleviating modules (a 2.6-fold increase). These results are not correlated with module size, as the median size of aggravating learned modules is less than the median size of alleviating learned modules. They suggest that, regardless of the technique for identifying complexes, those containing essential genes tend to be composed of primarily aggravating genetic interactions. Mechanistically, this might occur through a variety of means, including proteins with separate but functionally redundant roles in maintaining complex integrity, or subunit substitution by paralogous proteins.

Discussion

Figure 5 presents detailed diagrams of example functional relationships elucidated by our module mapping method. Figure 5A shows the alleviating relationship between the RTT109-VPS75 histone acetyltransferase complex [6,21,22] and Elongator, a complex that is associated with RNA Polymerase II and is involved in transcriptional elongation [23]. Since several subunits both of Elongator and RTT109/VPS75 have been shown to be involved in histone acetylation levels [22,24], these two complexes may operate together to effectively clear histones from actively transcribed regions. To identify further mechanisms of
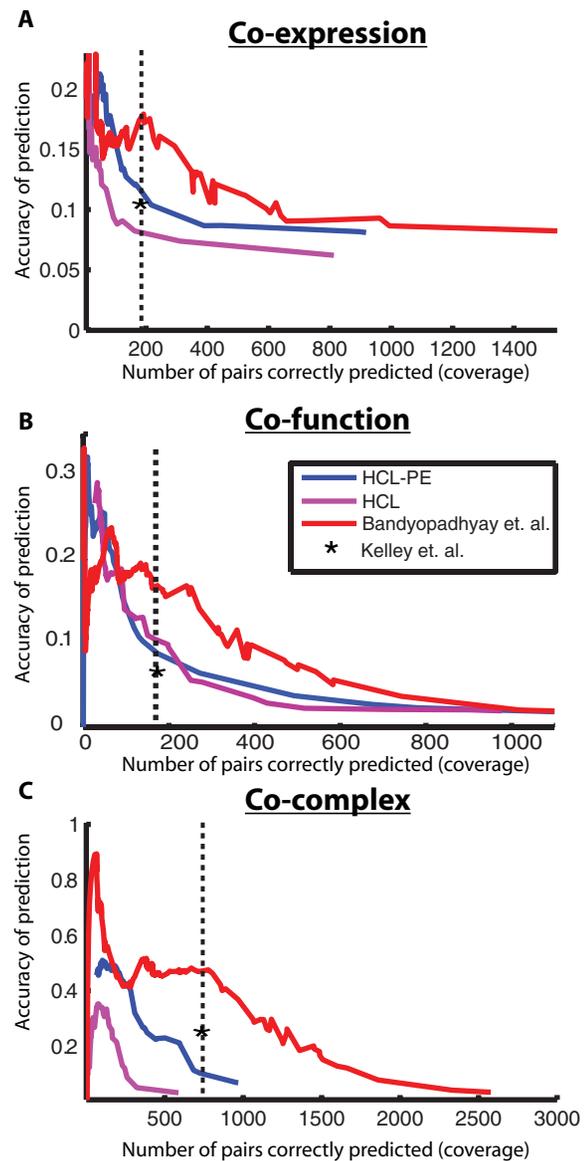

Figure 3. Performance of complex identification. The proposed approach is compared to several competing methods of discovering protein complexes within genetic interaction networks: HCL implements hierarchical clustering with a distance measure computed from the genetic interaction profiles only (S-scores), while HCL-PE extends HCL by merging clusters only if there is a physical interaction between them (PE-score>1). For the modules defined by each method, accuracy versus coverage is plotted over a range of values for tuning the module size (see Methods). Accuracy is estimated as the fraction of protein pairs in a predicted module that are in a gold-standard set; coverage is estimated as the number of gold-standard pairs that fall in the same predicted module. Gold-standard sets are defined by protein pairs that are either (A) co-expressed, (B) functionally-related, or (C) assigned to the same complex in high-throughput data sets (as annotated in MIPS). The performance at the chosen parameter setting ($\alpha = 1.6$) is indicated by the dotted vertical line. The performance of the method of Kelley et al. is reported for the same level of coverage as the present approach (asterisk). Since it operates on binary interaction data, we converted quantitative genetic and physical interaction scores to binary values based on a threshold of $|S|>2.5$ and PE>1.
doi:10.1371/journal.pcbi.1000065.g003

their cooperation, future studies may search for specific residues of histone H3 whose acetylation levels are modulated by both complexes. This example highlights the utility of an integrated
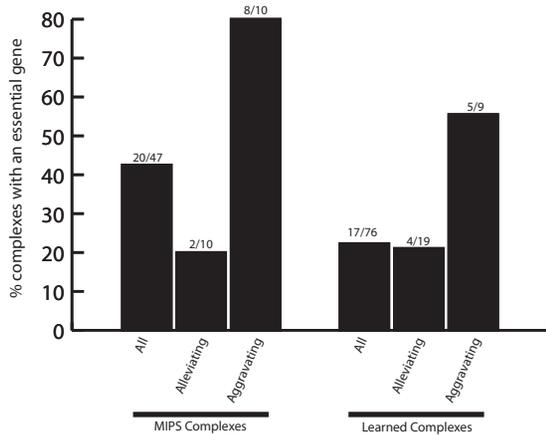
Figure 4. Aggravating complexes are more likely to contain essential genes. The percentage of complexes that contain at least one essential gene is shown, for various groups of complexes defined within small-scale complexes in MIPS (left three bars) or complexes identified in this study (right three bars). In MIPS, approximately 80% of "aggravating" complexes (see text) contain an essential gene, versus 20% for "alleviating" complexes. The trend is similar for the complexes reported in this study, with 55% versus 22% of aggravating versus alleviating complexes containing an essential gene. The list of all essential genes was taken from (http://www-sequence.stanford.edu/group/yeast_deletion_project/deletions3.html).
doi:10.1371/journal.pcbi.1000065.g004

approach, since although RTT109 and VPS75 are known to form a complex their genetic interaction profiles are not congruent (correlation of profiles of −0.1) and had been missed by hierarchical clustering. Figure 5B highlights non-essential components (LRP1 and RRP6) of the exosome, which contributes to the quality-control system that retains and degrades aberrant mRNAs in the nucleus [25]. These components have alleviating interactions with a complex composed of Lsm proteins involved in mRNA decay.

Figure 5C centers on BRE1/LGE1, subunits of the Rad6 Histone Ubiquitination Complex (RAD6-C; the Rad6 protein itself was not covered by the original E-MAP screen) [26,27]. RAD6-C is functionally connected with two other complexes, SWR-C and COMPASS. SWR-C functions to regulate gene expression through the incorporation of transcriptionally active histone variant H2AZ [28–30], while COMPASS is involved in mediating transcriptional elongation and silencing at telomeres through methylation of histone H3 [31]. Interactions between RAD6-C and SWR are aggravating, suggesting synergy or redundancy towards an essential cellular function. Interactions between RAD6-C and COMPASS are alleviating, suggesting they operate in a potentially serial fashion. Consistent with this analysis, it has been shown that histone H2B ubiquitination by RAD6-C is a prerequisite for histone H3 methylation by COMPASS [32,33].

Several trends emerge from the performance analysis in Figure 3. First, genetic interaction data alone can yield substantial information about molecular pathways. Functionally similar proteins often
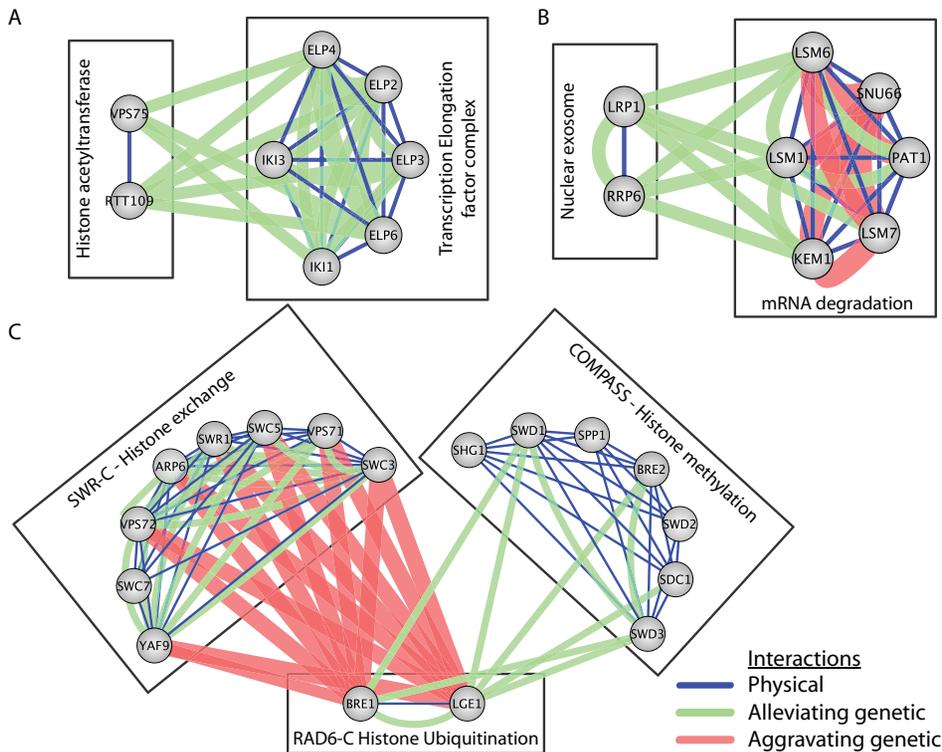

Figure 5. Pathway models identify novel functional associations among cellular machinery. Each panel represents complexes and between-complex links taken from Figure 2. Physical interactions with PE>1 are shown and strong genetic interactions (|S|>2.5) are shown with increased thicknesses corresponding to stronger genetic interactions. (A) Histone acetyltransferase complex RTT109 – VPS75 showing strong alleviating interactions with the Elongator transcription elongation factor complex. (B) Between-complex model highlighting alleviating interactions between the LRP1 – RRP6 nuclear exosome complex and an mRNA degradation complex. (C) Complexes associated with the RAD6-C histone ubiquitination complex (BRE1/LGE1).
doi:10.1371/journal.pcbi.1000065.g005

have similar profiles of genetic interaction, a feature we have previously exploited to identify functional interactions between complexes as well as to identify new members of complexes based on a combination of weak physical and genetic data [14]. On the other hand, the ability to detect complexes can be greatly improved by adding information about protein physical interactions. Even the straightforward HCL-PE method was able to greatly improve the accuracy and coverage according to most metrics, while the greatest performance was achieved by the improved probabilistic framework we have presented in this study. This framework has led to the inclusion of YKL023W as a potential new member of the SKI complex and YGR071C in a complex with VID22/TBF1 (Figure 2), for a total of 84 novel protein subunit assignments to complexes (Dataset S1). Both of these examples have both physical and genetic support and would have been missed by an approach based on either type of interaction alone.

Future work may seek to incorporate yet additional types of linkages such as protein-DNA interactions [34,35], kinase-substrate phosphorylations [36], or other genetic perturbation data such as eQTLs [37]. There are also opportunities to refine the modeling framework further. Here, a gold-standard set of complexes was used to explicitly learn the relationship between physical interactions, genetic interactions, and module membership. This supervised approach could be extended to also learn which features best indicate the inter-module functional relationships, perhaps through curation of a gold-standard set of interacting complexes.

Methods

Problem Definition

We analyze the interaction data to infer *a set of protein modules* and *a set of inter-module links* (Figure 1C). A protein module is defined as a set of proteins that are connected through protein-protein interactions and are likely to represent a protein complex with a coherent cellular function. Inter-module links capture functional relationships between modules and may be of two types, aggravating or alleviating. The complete state of the system is described by a set M of modules, each module defining a set of proteins, and a set N of pairs of modules that are functionally linked.

Scoring Module Co-Membership

For each pair of proteins (a,b) we compute a log ratio W of the likelihood that a and b fall *within* the same module versus the likelihood that they are unrelated (i.e., occur in the background). The function uses two sources of information that are indicative of protein complex co-membership: the strength of protein-protein physical interaction (PE) and the strength of genetic interaction (S):

$$W(a,b) = LLR_{PE}(a,b) + LLR_S(a,b) \qquad (1)$$

For a given data type (PE or S) the log likelihood ratio (LLR) is defined as:

$$LLR(a,b) = \log \frac{P_{within}(a,b)}{P_{background}(a,b)} \qquad (2)$$

The probability P_{within} is determined using logistic regression training on 217 complexes curated from small-scale studies in MIPS [19]. $P_{background}$ is the probability of randomly observing the observed value (PE or S) for the pair (a,b) in the background of all gene pairs. As shown in Figure 1A and 1B, it is clear that higher

values of PE are predictive of MIPS complex membership. As both positive and negative values of S are predictive, $LLR_S(a,b)$ is trained on the absolute value of S. A third predictor based on the correlation of genetic interaction profiles was also evaluated but did not result in any gain in performance (Figure S1).

Scoring Inter-Module Links

A similar function $B()$ is formulated to assess the likelihood that two proteins fall *between* modules that are functionally linked. The function inputs the same two sources of information on protein-protein and genetic interactions (PE and S). Unfortunately, there is no curated set of functionally related complexes that can be used as positive training examples for regression. Instead, $B()$ is derived from the within-module LLRs, assuming that between-module interactions have a similar pattern of genetic interactions but lack physical interactions:

$$B(a,b) = -LLR_{PE}(a,b) + LLR_S(a,b) \qquad (3)$$

This function captures both aggravating and alleviating genetic interactions between two functionally-related modules. It also ensures such modules are physically separate—if not, they would be better considered as a single module.

Global Optimization of Module Memberships and Links

Given the above functions $W()$ and $B()$, we compute the likelihood of the complete system (i.e., given a particular choice M of modules and N of inter-module links):

$$L = \left(\sum_{m \in M} \sum_{(a,b) \in m \times m} W(a,b) \right) + \left(\sum_{(m_1,m_2) \in N} \sum_{(a,b) \in m_1 \times m_2} B(a,b) \right) + \left(\sum_{m \in M} |m|^{\alpha} \right) \qquad (4)$$

The first term accumulates the within-module scores among gene pairs assigned to the same module. The second term accumulates the inter-module scores for gene pairs spanning any two modules. Gene pairs spanning unlinked modules do not contribute to L. The final term is a tunable reward which scales with module size. Larger values of α result in fewer, larger complexes. The final module map shown in Figure 2 was generated using $\alpha = 1.6$, based on its good coverage and performance across all three metrics in Figure 3.

Module Search

Assignment of gene to modules and of inter-module links is performed using a simple variant of UPGMA hierarchical clustering [38]: (a) Initially, each gene is assigned to a separate module; (b) Each pair of modules (m_1, m_2) is evaluated for merging into a single module $m = m_1 \cup m_2$; the pair-wise merging that results in the largest increase in L is chosen; (c) Repeat step b until no module merge operation increases L.

At each iteration of step b, L is optimized over all possible ways of assigning inter-module links (i.e., module pairs are linked whenever the second term in Equation 4 is positive). Because each inter-module link is scored independently, additions or deletions of links from the system need only be evaluated for modules that are under evaluation for merging.

Subsequent to the above procedure, each between-module link is evaluated to assess its significance and whether it represents

predominantly aggravating or alleviating genetic interactions. A two-tailed p-value is computed by indexing the sum of S-scores for gene pairs falling across the two modules against a distribution of 10^6 sums of equal numbers of S-scores drawn from random gene pairs. To account for multiple testing, we use the distribution of between-module p-values to compute a local false discovery rate (FDR) [39]. All reported between-module links have an inferred FDR of <10% with the global map in Figure 2 constrained to links with an FDR of <1%. Module maps in Figure 2 and Figure 5 are visualized using the Cytoscape package [40,41].

To label modules as "aggravating" or "alleviating" (Figure 2), the sum of S-scores for gene pairs assigned to the same module is compared to a distribution of sums of equal numbers of randomly drawn S-scores. Modules with a two-tailed p-value<0.05 are labeled as either alleviating (right tail) or aggravating (left tail).

Validation Using Co-Expression, Co-Function, or Co-Complex Annotations

Co-expressed gene pairs were defined using gene expression datasets culled from the Stanford Microarray Database covering ~790 conditions [42]. The validation set was taken as the top 5% (13,014) of pairs ranked by Pearson correlation coefficient. The co-function set was based on yeast Gene Ontology annotations from November 2005 which predates the publication of large scale TAP-MS studies that were used to generate the PE-score [43]. This set was taken as the top 5% (13,052) most functionally similar gene pairs covered in the E-MAP. Functional similarity was determined by comparison to the background probability of picking two genes with the same shared functional annotation from the entire yeast genome (via a hypergeometric test). Similar analysis using current Gene Ontology annotation was also performed (Figure S2). The co-complex validation set was defined as gene pairs from 846 MIPS complexes annotated using high-throughput approaches (with interactions also appearing in small-scale studies removed) for a total of 2,885 gold-standard pairs.

The size and number of final modules was varied by altering the α parameter (see above). To assess performance at low coverage we ran the method with no reward contribution (remove the third term in Equation 4 by setting $\alpha = -\infty$) and plotted the performance of the algorithm at each merge step, which ultimately connects with the performance of the method as α is increased. For HCL and HCL-PE methods, the size and number of modules were varied by changing the level at which the hierarchy was cut.

Supporting Information

Figure S1 Addition of congruence as a predictor of pathway membership.
Found at: doi:10.1371/journal.pcbi.1000065.s001 (0.10 MB DOC)

Figure S2 A current version of the Gene Ontology shows similar performance.
Found at: doi:10.1371/journal.pcbi.1000065.s002 (0.09 MB DOC)

Dataset S1 Results tables in Excel format.
Found at: doi:10.1371/journal.pcbi.1000065.s003 (0.06 MB XLS)

Acknowledgments

The authors thank Sean Collins for his useful comments and suggestions.

Author Contributions

Conceived and designed the experiments: SB RK NK TI. Performed the experiments: SB RK. Analyzed the data: SB. Wrote the paper: SB RK NK TI.

References

1. Avery L, Wasserman S (1992) Ordering gene function: the interpretation of epistasis in regulatory hierarchies. Trends Genet 8: 312–316.
2. Carter GW, Prinz S, Neou C, Shelby JP, Marzolf B, et al. (2007) Prediction of phenotype and gene expression for combinations of mutations. Mol Syst Biol 3: 96.
3. Hereford LM, Hartwell LH (1974) Sequential gene function in the initiation of Saccharomyces cerevisiae DNA synthesis. J Mol Biol 84: 445–461.
4. Ooi SL, Shoemaker DD, Boeke JD (2003) DNA helicase gene interaction network defined using synthetic lethality analyzed by microarray. Nat Genet 35: 277–286.
5. Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, et al. (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science 294: 2364–2368.
6. Collins SR, Miller KM, Maas NL, Roguev A, Fillingham J, et al. (2007) Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. Nature 446: 806–810.
7. Collins SR, Schuldiner M, Krogan NJ, Weissman JS (2006) A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. Genome Biol 7: R63.
8. Schuldiner M, Collins SR, Thompson NJ, Denic V, Bhamidipati A, et al. (2005) Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. Cell 123: 507–519.
9. Schuldiner M, Collins SR, Weissman JS, Krogan NJ (2006) Quantitative genetic analysis in *Saccharomyces cerevisiae* using epistatic miniarray profiles (E-MAPs) and its application to chromatin functions. Methods 40: 344–352.
10. Drees BL, Thorsson V, Carter GW, Rives AW, Raymond MZ, et al. (2005) Derivation of genetic interaction networks from quantitative phenotype data. Genome Biol 6: R38.
11. St Onge RP, Mani R, Oh J, Proctor M, Fung E, et al. (2007) Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. Nat Genet 39: 199–206.
12. Segre D, Deluna A, Church GM, Kishony R (2005) Modular epistasis in yeast metabolism. Nat Genet 37: 77–83.
13. Beyer A, Bandyopadhyay S, Ideker T (2007) Integrating physical and genetic maps: from genomes to interaction networks. Nat Rev Genet 8: 699–710.
14. Kelley R, Ideker T (2005) Systematic interpretation of genetic interactions using protein networks. Nat Biotechnol 23: 561–566.
15. Ulitsky I, Shamir R (2007) Pathway redundancy and protein essentiality revealed in the Saccharomyces cerevisiae interaction networks. Mol Syst Biol 3: 104.
16. Zhang LV, King OD, Wong SL, Goldberg DS, Tong AH, et al. (2005) Motifs, themes and thematic maps of an integrated Saccharomyces cerevisiae interaction network. J Biol 4: 6.
17. Phillips PC, Otto SP, Whitlock MC (2000) Beyond the average: the evolutionary importance of gene interactions and variability of epistatic effects in epistasis and evolutionary process. New York: Oxford University Press.
18. Collins SR, Kemmeren P, Zhao XC, Greenblatt JF, Spencer F, et al. (2007) Toward a comprehensive atlas of the physical interactome of Saccharomyces cerevisiae. Mol Cell Proteomics 6: 439–450.
19. Guldener U, Munsterkotter M, Oesterheld M, Pagel P, Ruepp A, et al. (2006) MPact: the MIPS protein interaction resource on yeast. Nucleic Acids Res 34: D436–D441.
20. Boone C, Bussey H, Andrews BJ (2007) Exploring genetic interactions and networks with yeast. Nat Rev Genet 8: 437–449.
21. Driscoll R, Hudson A, Jackson SP (2007) Yeast Rtt109 promotes genome stability by acetylating histone H3 on lysine 56. Science 315: 649–652.
22. Han J, Zhou H, Horazdovsky B, Zhang K, Xu RM, et al. (2007) Rtt109 acetylates histone H3 lysine 56 and functions in DNA replication. Science 315: 653–655.
23. Otero G, Fellows J, Li Y, de Bizemont T, Dirac AM, et al. (1999) Elongator, a multisubunit component of a novel RNA polymerase II holoenzyme for transcriptional elongation. Mol Cell 3: 109–118.
24. Winkler GS, Kristjuhan A, Erdjument-Bromage H, Tempst P, Svejstrup JQ (2002) Elongator is a histone H3 and H4 acetyltransferase important for normal histone acetylation levels in vivo. Proc Natl Acad Sci U S A 99: 3517–3522.
25. Mitchell P, Petfalski E, Shevchenko A, Mann M, Tollervey D (1997) The exosome: a conserved eukaryotic RNA processing complex containing multiple 3′→5′ exoribonucleases. Cell 91: 457–466.
26. Hwang WW, Venkatasubrahmanyam S, Ianculescu AG, Tong A, Boone C, et al. (2003) A conserved RING finger protein required for histone H2B monoubiquitination and cell size control. Mol Cell 11: 261–266.
27. Wood A, Krogan NJ, Dover J, Schneider J, Heidt J, et al. (2003) Bre1, an E3 ubiquitin ligase required for recruitment and substrate selection of Rad6 at a promoter. Mol Cell 11: 267–274.

28. Kobor MS, Venkatasubrahmanyam S, Meneghini MD, Gin JW, Jennings JL, et al. (2004) A protein complex containing the conserved Swi2/Snf2-related ATPase Swr1p deposits histone variant H2A.Z into euchromatin. PLoS Biol 2: e131.

29. Krogan NJ, Dover J, Wood A, Schneider J, Heidt J, et al. (2003) The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. Mol Cell 11: 721–729.

30. Mizuguchi G, Shen X, Landry J, Wu WH, Sen S, et al. (2004) ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. Science 303: 343–348.

31. Li B, Carey M, Workman JL (2007) The role of chromatin during transcription. Cell 128: 707–719.

32. Dover J, Schneider J, Tawiah-Boateng MA, Wood A, Dean K, et al. (2002) Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6. J Biol Chem 277: 28368–28371.

33. Sun ZW, Allis CD (2002) Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. Nature 418: 104–108.

34. Berger MF, Philippakis AA, Qureshi AM, He FS, Estep PW 3rd, et al. (2006) Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. Nat Biotechnol 24: 1429–1435.

35. Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, et al. (2004) Transcriptional regulatory code of a eukaryotic genome. Nature 431: 99–104.

36. Ptacek J, Devgan G, Michaud G, Zhu H, Zhu X, et al. (2005) Global analysis of protein phosphorylation in yeast. Nature 438: 679–684.

37. Brem RB, Storey JD, Whittle J, Kruglyak L (2005) Genetic interactions between polymorphisms that affect gene expression in yeast. Nature 436: 701–703.

38. Sokal RR, Michener CD (1958) A statistical method for evaluating systematic relationships. University of Kansas Sci Bull 28: 1409–1438.

39. Benjamini Y, Hochberg, Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. JRSSB 57: 289–300.

40. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, et al. (2007) Integration of biological networks and gene expression data using Cytoscape. Nat Protoc 2: 2366–2382.

41. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13: 2498–2504.

42. Demeter J, Beauheim C, Gollub J, Hernandez-Boussard T, Jin H, et al. (2007) The Stanford Microarray Database: implementation of new analysis tools and open source release of software. Nucleic Acids Res 35: D766–D770.

43. Gene Ontology (November 2005) CVS log for go/gene. Available at: http://cvsweb.geneontology.org/cgi-bin/cvsweb.cgi/go/gene-associations/gene_association.sgd.gz. Accessed 26 March 2008.