



Predicting genetic modifier loci using functional gene networks

Insuk Lee, Ben Lehner, Tanya Vavouri, et al.

Genome Res. published online June 9, 2010

Access the most recent version at doi:[10.1101/gr.102749.109](https://doi.org/10.1101/gr.102749.109)

Supplemental Material <http://genome.cshlp.org/content/suppl/2010/05/10/gr.102749.109.DC1.html>
<http://genome.cshlp.org/content/suppl/2010/05/10/gr.102749.109.DC2.html>

P<P Published online June 9, 2010 in advance of the print journal.

Email alerting service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#)

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

To subscribe to *Genome Research* go to:
<http://genome.cshlp.org/subscriptions>

Method

Predicting genetic modifier loci using functional gene networks

Insuk Lee,^{1,2,7} Ben Lehner,^{3,4,7} Tanya Vavouri,³ Junha Shin,¹ Andrew G. Fraser,^{5,7} and Edward M. Marcotte^{2,6,7}

¹Department of Biotechnology, College of Life science and Biotechnology, Yonsei University, Seodaemun-ku, Seoul 120-749, South Korea; ²Center for Systems and Synthetic Biology, Institute for Cellular and Molecular Biology, University of Texas, Austin, Texas 78712, USA; ³EMBL-CRG Systems Biology Research Unit, Centre for Genomic Regulation, Pompeu Fabra University (UPF), Barcelona 08003, Spain; ⁴ICREA, Centre for Genomic Regulation, UPF, Barcelona 08003, Spain; ⁵Donnelly CCB, University of Toronto, Toronto M5S 3E1, Canada; ⁶Department of Chemistry and Biochemistry, Institute for Cellular and Molecular Biology, University of Texas, Austin, Texas 78712, USA

Most phenotypes are genetically complex, with contributions from mutations in many different genes. Mutations in more than one gene can combine synergistically to cause phenotypic change, and systematic studies in model organisms show that these genetic interactions are pervasive. However, in human association studies such nonadditive genetic interactions are very difficult to identify because of a lack of statistical power—simply put, the number of potential interactions is too vast. One approach to resolve this is to predict candidate modifier interactions between loci, and then to specifically test these for associations with the phenotype. Here, we describe a general method for predicting genetic interactions based on the use of integrated functional gene networks. We show that in both *Saccharomyces cerevisiae* and *Caenorhabditis elegans* a single high-coverage, high-quality functional network can successfully predict genetic modifiers for the majority of genes. For *C. elegans* we also describe the construction of a new, improved, and expanded functional network, WormNet 2. Using this network we demonstrate how it is possible to rapidly expand the number of modifier loci known for a gene, predicting and validating new genetic interactions for each of three signal transduction genes. We propose that this approach, termed network-guided modifier screening, provides a general strategy for predicting genetic interactions. This work thus suggests that a high-quality integrated human gene network will provide a powerful resource for modifier locus discovery in many different diseases.

[Supplemental material is available online at <http://www.genome.org>.]

Most diseases in humans are known to be genetically complex, resulting from mutations in many different genes (Bonetta 2008). In model organisms, systematic studies have shown that for nearly all traits studied, the traits are not only affected by multiple genes, but also by nonadditive interactions between mutations in these genes (Flint and Mackay 2009). Combined with the results of synthetic lethal screens (Boone et al. 2007; Lehner 2007), this shows that nonadditive (epistatic) interactions between mutations are pervasive in biology. For essentially all phenotypes (Flint and Mackay 2009) and nearly all genes (Tong et al. 2004) the outcomes of mutations are affected by the alleles carried by an organism at other genomic loci. These synthetic interactions between mutations have been suggested as one explanation for why genome-wide association studies have failed to identify loci that explain more than a minority of the known genetic contribution to most common diseases in humans (Maher 2008; Flint and Mackay 2009).

With the aim of understanding how mutations combine to cause disease, large-scale genetic interaction screens have been performed in model organisms (Tong et al. 2004; Schuldiner et al. 2005; Lehner et al. 2006a; Pan et al. 2006; Byrne et al. 2007; Collins

et al. 2007; Lin et al. 2008). In these screens, pairs of genes are systematically inhibited and the effects on viability phenotypes assayed (for review, see Boone et al. 2007; Lehner 2007). These screens have highlighted the enormous extent to which mutations in one gene alter the phenotypic outcome of mutations in a second locus. Furthermore, the data from these screens show that most genetic interactions do not represent simple cases of “redundancy” between genes encoding similar biochemical functions (Tong et al. 2004; Wong et al. 2004; Kelley and Ideker 2005; Ulitsky and Shamir 2007).

Interactions between mutations are extremely difficult to identify in human population studies. This is because of the large number of statistical tests required to identify interactions between pairs of alleles, limited disease population sizes, and the low frequency at which any particular combination of alleles is present in a population (Hartman et al. 2001; Flint and Mackay 2009). This makes identifying modifier loci in human disease a very challenging prospect, and, indeed, only a few such loci are known (Bochdanovits et al. 2008; Flint and Mackay 2009; Mackay et al. 2009).

An alternative approach to identify interactions between mutations would be to first predict them and then to validate these predictions in population studies. This would reduce the number of tests being performed, therefore increasing power to detect interactions. Analysis of genetic interaction data from model organisms has shown that genes that participate in common biological processes often show similar patterns of genetic interactions (Tong et al. 2004; Kelley and Ideker 2005; Ye et al. 2005; Ulitsky and Shamir 2007). This means that genetic interactions can, in part, be

⁷Corresponding authors.

E-mail insuksysbio@gmail.com.

E-mail ben.lehner@crg.es.

E-mail andyfraser.utoronto@gmail.com.

E-mail marcotte@icmb.utexas.edu.

Article published online before print. Article and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.102749.109>.

predicted using physical interaction data to identify proteins that act in common functional pathways (Wong et al. 2004; Le Meur and Gentleman 2008).

In biology, however, functional relationships among proteins often transcend direct physical interactions (Fraser and Marcotte 2004). Many proteins can be important for common biological processes without physically interacting or associating. For example, proteins functioning in the same biosynthesis pathway, but at different biochemical steps, may never physically contact each other but are functionally associated because they act in the same biological process. This insight led to the concept of “functional gene networks” (Marcotte et al. 1999). In these networks genes are coupled if they are predicted to participate in a common biological process. Further, the strength of interaction between any two genes indicates the confidence in the functional coupling between the two genes (Troyanskaya et al. 2003; Lee et al. 2004; Rhodes et al. 2005; Jensen et al. 2009).

Many different types of biological data can be used to predict functional interactions between genes. Indeed, this is an important advantage of the approach, as it provides a generic framework for integrating disparate data types into a common predictive network. To date, functional networks have been constructed for organisms ranging from unicellular yeast (Mellor et al. 2002; Troyanskaya et al. 2003; Karaoz et al. 2004; Lee et al. 2004, 2007; Tian et al. 2008), through invertebrate model organisms (Gunsalus et al. 2005; Lee et al. 2008), to mammals (Rhodes et al. 2005; Guan et al. 2008; Kim et al. 2008; Peña-Castillo et al. 2008; Linghu et al. 2009).

These studies have demonstrated that a single functional network can be used to predict gene loss-of-function phenotypes across a majority of genes in an organism (McGary et al. 2007; Lee et al. 2008; Linghu et al. 2009). These networks successfully predict loss-of-function phenotypes, because genes that act in a common process tend to cause similar phenotypic effects when they are mutated, as has been known for many years in genetics and has been quantitatively confirmed for gene networks (Fraser and Plotkin 2007; Hart et al. 2007; Lage et al. 2007; Lee et al. 2008). Thus, given knowledge of a few genes associated with a process or phenotype of interest, it is possible to successfully predict more genes that will alter this process or phenotype when inhibited. We also observed, for a few tested cases, that a functional network was also predictive for genetic modifiers of a mutation (Lee et al. 2008).

In this study, we have expanded and tested the general validity of the approach of using functional networks for predicting specific genetic modifiers. First, we describe the construction of a new and improved functional network for the model animal *Caenorhabditis elegans*. This network has both higher coverage and greater predictive power for single gene loss-of-function phenotypes than our first generation network. We then show that this network is also highly predictive for identifying interactions between mutations, considering all previously identified systematic interaction datasets. Indeed, the network is similarly predictive for identifying genetic interactions as for single gene loss-of-function phenotypes. Moreover, we demonstrate the advantage over purely physical interaction networks. We further validate the approach in yeast, the organism for which most genetic and functional data is currently available. Finally, we demonstrate how it is possible to rapidly expand the set of modifier loci known for a mutation using the examples of three signal transduction genes. Taken together, our work proposes that a single, high-quality functional network for human genes will provide a powerful resource for predicting modifier genes in human disease.

Results

WormNet version 2: An extended functional network for *C. elegans*

To test the ability of functional networks to predict interactions between different genes, we first constructed an improved and expanded functional network for the animal *C. elegans*. Currently, *C. elegans* is the only animal in which genetic interactions have been systematically identified in vivo (Lehner et al. 2006a; Byrne et al. 2007) and so it provides the best animal system in which to test the predictive power of integrated networks. Our new network incorporates many improvements compared with our first generation network (see Methods and Supplemental Table S1) and extends the number of functional couplings between genes over 2.5-fold to 999,367 links covering 15,139 genes (75.5% of 20,081 protein-encoding genes).

Important improvements derive from multiple sources (Supplemental material and Supplemental Table S1). These include the more stringent treatment of coexpression datasets, the incorporation of information from other organisms as individual rather than preintegrated data, and the addition of many new and expanded datasets (see Methods). These additional data types include linkages derived from protein domain co-occurrence patterns, linkages based on mRNA coexpression relationships of human orthologs (I Lee, M Blom, PI Wang, J Shim, and EM Marcotte, in prep.), and additional human protein interaction data (Kim et al. 2005; Stelzl et al. 2005; Ewing et al. 2007). In total, 21 data sets (Supplemental Table S2) from four different species (yeast, fly, worm, and human) were integrated to achieve a high-coverage, high-accuracy network model (Fig. 1A). About 58% of links modeled by the previous network are recapitulated by the new network (Fig. 1B).

WormNet 2 provides improved predictive power for loss-of-function phenotypes

To evaluate the predictive power of the new network we used a compilation of data from genome-wide RNAi screens covering 43 different loss-of-function phenotypes as listed in Supplemental Table S3. These phenotypes range from defects in cellular processes to gross morphological phenotypes and physiological processes such as aging. The phenotype datasets are independent of any data used to construct either versions of our network and so provide a common benchmarking set to compare the performance of both networks. We evaluated predictability of the networks for each phenotype by receiver operating characteristic (ROC) curve analysis as described in Methods. This method provides a measure of the recovery of true-positive genes (those known to be associated with a phenotype) compared with false-positive genes (others) when all genes are ranked by their network connections to the known phenotypic genes.

Example ROC curves for two phenotypes (PTEN synthetic lethality and ruptured) for WormNet versions 1 and 2 are illustrated in Figure 1C. For both phenotypes the newer version of WormNet shows improved ROC curve performance. These ROC curve behaviors can be summarized as a simple score, the area under the ROC curve (AUC), which would be close to 0.5 for a random network, and approaching 1 for a perfect predictor. As shown in Figure 1D, for nearly all phenotypes the predictive power of the new version of WormNet is greater than that of the old version ($P = 3 \times 10^{-4}$, Wilcoxon signed rank test, unless noted otherwise). This improvement stems from links that are

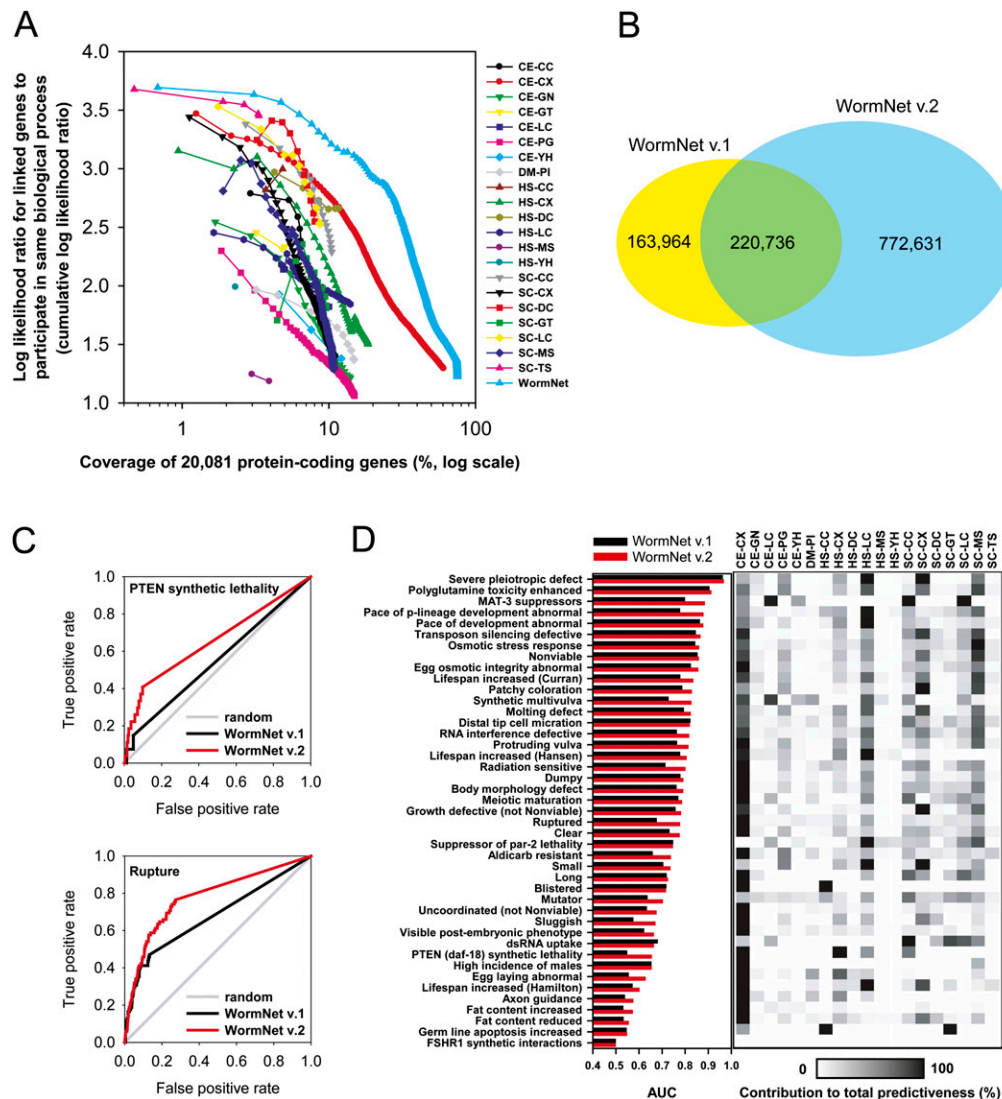


Figure 1. Improved performance of the updated gene network for *C. elegans*, WormNet version 2. (A) Benchmarking results of network by each 21 data sets and one by their integration (WormNet). Each data set is indicated by XX-YY code, in which XX represents origin species of data—CE, *C. elegans*; DM, *Drosophila melanogaster*, HS, *Homo sapiens*, SC, *Saccharomyces cerevisiae*—and YY represents data type—CC, cocitation; CX, coexpression; GN, gene neighbors; GT, genetic interaction; LC, literature curated protein–protein interaction; PG, phylogenetic profiling; YH, high-throughput yeast two-hybrid interaction; PI, protein interaction; DC, domain co-occurrence; MS, mass spectrometry analysis; TS, inferred interaction from protein tertiary structure. The x-axis represents coverage of 20,081 protein-coding worm genes by different components of WormNet version 2 (log scaled); the y-axis represents predictive performance of the network and components, measured as the cumulative log likelihood for linked genes to participate in same Gene Ontology biological process. (B) Venn diagram between old and new WormNet linkages. The new WormNet more than doubles the number of linkages, with >57% (220,736 of 384,700) of the WormNet v.1 linkages recapitulated in v.2. (C) Improved predictability for RNAi phenotypes by WormNet v.2 is illustrated by ROC curve analysis for two RNAi phenotypes. A version of WormNet omitting literature-based genetic interaction datasets (CE-CC and CE-GT) was used for all ROC analyses in Figures 1–4, to minimize the possibility of circular reasoning when predicting published genetic interactions. (D) The improved predictability of WormNet v.2 (red bars) over v.1 (black bars) is illustrated by a comparison of AUC scores for 43 different RNAi phenotypes. For each RNAi phenotype (rows), the supporting power of each data set (columns, labeled as in A) to the prediction was measured as a fractional score—the sum of log likelihood scores of all supporting evidence for that data set divided by the sum of log likelihood scores across all data sets. The degree of contribution is indicated by grayscale, where the higher support the darker indicator. Various data sets including those from other species make significant contribution for predictability for worm RNAi phenotypes.

derived from heterogeneous data types and species (Fig. 1D). We conclude that WormNet 2 has increased predictability for loss-of-function phenotypes compared with the previous network model.

We extended the comparison of predictability to other large-scale gene networks including STRING and FunCoup (Supplemental Fig. 1). The STRING version 8.2 (Jensen et al. 2009)

network includes 150,462 linkages among 8570 genes (42.7%), and FunCoup version 1 (Alexeyenko and Sonnhammer 2009) includes 1,868,676 linkages among 13,415 genes (66.8%). WormNet v. 2 shows a comparable or higher combination of predictability (measured as AUC) and phenotypic gene coverage than other networks, including WormNet v. 1. More than 50% of RNAi phenotypes (22/43) were best predicted by WormNet v. 2.

Predicting genetic interactions in *C. elegans* using a functional network

The example ROC curve shown in Figure 1C for PTEN synthetic lethality suggests that a functional network may also provide a general method for predicting interactions between mutations in addition to single gene loss-of-function phenotypes. That is, given a few genes that are known to interact genetically with a mutation, it may be possible to identify new modifier loci by their functional coupling to these known “seed” genes (Fig. 2A).

In *C. elegans* genetic interactions between mutations have been systematically identified in two large-scale interaction screens (Lehner et al. 2006a; Byrne et al. 2007). In each screen, modifier loci that enhance the effects of a mutation in one gene were identified using RNAi to inhibit the expression of other genes. In a few cases, RNAi was used to inhibit two genes simultaneously.

We used the same ROC AUC analysis to assess the ability of WormNet 2 to predict these genetic interactions between loci. As

shown in Figure 2B, we observe a high AUC for the modifier genes identified in nearly all of the screens. Indeed, the network is similarly predictive for identifying modifier loci (Fig. 2B) as it is for identifying single gene loss-of-function phenotypes (Fig. 1D). We conclude that genetic interactions can be predicted with similar performance as loss-of-function phenotypes.

Predicting genetic interactions in yeast

To further assess the generality of our approach, we tested whether a functional network could predict genetic interactions in a second organism. We focused on the budding yeast *Saccharomyces cerevisiae* because it is the species for which the most genetic interactions have been systematically identified (Boone et al. 2007) and also the organism for which the highest coverage functional gene networks are currently available (Lee et al. 2007; Myers and Troyanskaya 2007; Tian et al. 2008). We considered two large-scale genetic interaction screens: one testing for interactions with nonessential genes (Tong et al. 2004) and the other with essential loci (Davierwala et al. 2005). As a functional network, we used YeastNet version 2 (Lee et al. 2007), which is similar in construction to the *C. elegans* network described in this work.

As for *C. elegans*, we observed high AUC scores for the majority of interaction screens, both for nonessential and essential genes (Fig. 2C). We conclude that a single integrated functional network can successfully predict modifier loci in both unicellular and multicellular organisms. Interestingly, we note that the predictive power of genetic interactions with nonessential genes is higher than for essential genes (Fig. 2C). Essential genes are probably required for buffering many cellular processes (Davierwala et al. 2005), while nonessential genes, in general, function as specific modifiers for a few pathways (Lehner et al. 2006a). Thus, the differences in predictive power for these two gene classes may represent a general dichotomy between two types of genetic modifiers—general buffers and specific modifiers (Lehner et al. 2006a), with interactions for specific modifiers being more straightforward to predict using functional networks.

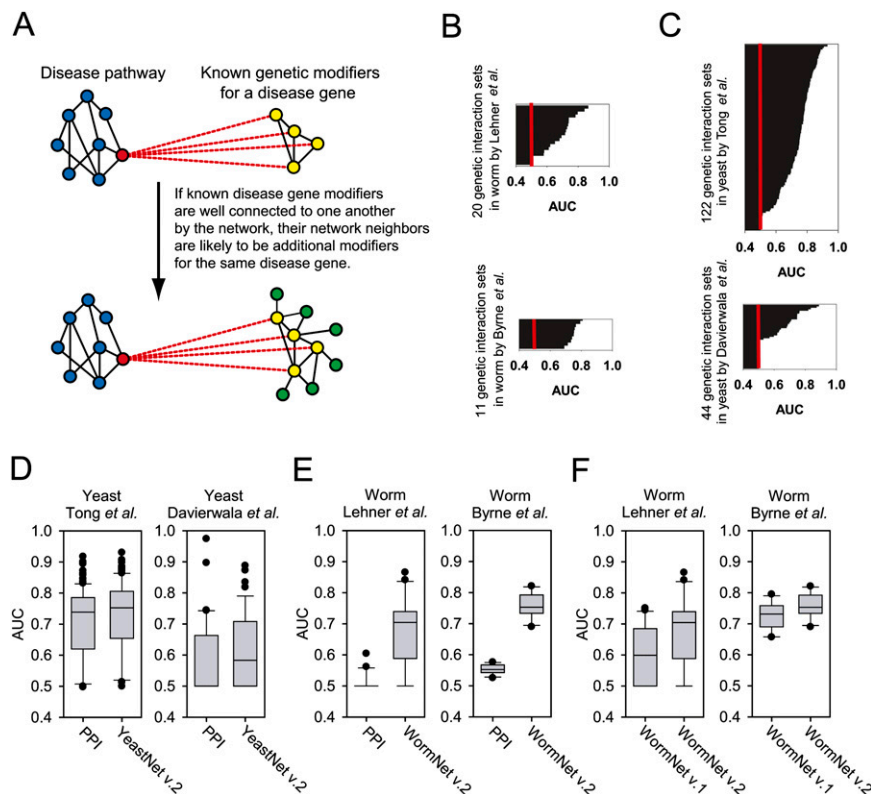


Figure 2. Network-guided prediction of genetic modifiers. (A) A schematic figure of network-guided prediction of genetic modifiers. Assuming there are some known disease gene modifiers (yellow nodes) for a disease gene (red node), we can predict additional candidate genetic modifiers (green nodes) connecting to known genetic modifiers in the functional gene network, because genes with similar functions tend to share genetic interaction partners (here, the disease gene). (B) AUC of known genetic modifiers for worm genes from two independent screens. The majority of groups of congruent genes show high AUC score (e.g., AUC > 0.6). AUC for random expectation (AUC = 0.5) is indicated by red line. (C) AUC of known synthetic lethal partners for yeast genes, from two independent studies—one for nonessential genes and the other for essential genes. These also show high predictability, indicated by high AUC scores for the majority of groups (red line for AUC = 0.5). (D) Predictability for synthetic lethal partners for yeast nonessential genes or essential genes by protein–protein interaction or functional gene network. In bar-and-whiskers plots, the central horizontal line in the box indicates the median AUC, and the boundaries of the box indicate the first and third quartiles of the AUC distribution, whiskers indicate the 10th and 90th percentiles, and filled circles indicate individual outliers. (E) Same analysis as D for genetic modifiers of worm orthologs of human disease genes. (F) Comparison of predictability between two different versions of WormNet.

Functional networks are more predictive than current protein interaction networks

Previous work has shown that it is possible to predict genetic interactions using physical protein–protein interaction networks (Tong et al. 2004; Wong et al. 2004; Kelley and Ideker 2005; Le Meur and Gentleman 2008). We therefore asked whether using functional interactions—which extend beyond physical interactions—improves the predictive power of a network.

In yeast, in addition to the numerous small- or medium-scale protein–protein interaction studies, multiple genome-wide screens have been performed for protein–protein interactions. As a consequence, a consolidated yeast protein–protein interaction network is expected to be fairly complete. We constructed a yeast protein interaction network by consolidating protein–protein interactions derived from various sources (Supplemental Table S4). The resulting interactome covers 5275 yeast genes (91% of predicted coding genes) with 65,033 interactions. We compared the prediction power of this high coverage protein interactome with that of the functional gene network (YeastNet version 2; Lee et al. 2007) using the data from two large-scale genetic interaction screens (Tong et al. 2004; Davierwala et al. 2005). For both datasets YeastNet 2 is more predictive than the physical interactome ($P = 7 \times 10^{-14}$ for Tong et al. 2004; $P = 2 \times 10^{-7}$ for Davierwala et al. 2005) (Fig. 2D), suggesting that even for an organism with a very extensive physical interaction data set, considering nonphysical functional interactions enhances predictive performance.

In *C. elegans* we observed an even greater benefit of using a functional network compared with a physical network. The current physical interactome for *C. elegans*—Worm Interactome version 8 (WI8) (Simonis et al. 2009) covers only 15% of protein-coding genes with 4402 interactions. With this limited coverage, most pathways are out of view, and consequently, most genetic modifier groups show an AUC < 0.55 (Fig. 2E). In contrast, WormNet version 2 covers >75% of genes and achieves a median AUC > 0.7 for genetic modifiers identified in both large-scale studies. We conclude that functional networks can provide better predictive performance for genetic modifiers than protein–interaction networks, particularly for species with less complete physical interactomes.

WormNet version 2 has improved predictive power for genetic modifiers

Next we tested whether the updated functional gene network for *C. elegans* improves predictability for known genetic modifiers compared with the previous version (Lee et al. 2008). We observed improvements of the AUC scores for most of the groups of genetic interaction partners ($P = 5 \times 10^{-5}$ for Lehner et al. 2006a; $P = 4 \times 10^{-3}$ for Byrne et al. 2007) (Fig. 2F). Together with the results for the 43 single gene loss-of-function phenotypes (Fig. 1C,D), we conclude that WormNet version 2 is improved in its general predictive power. We expect that as new datasets become available we will be able to further improve this predictive power as the network model is updated.

Data from diverse sources is used to predict genetic modifiers

We found that the high predictability of WormNet 2 does not depend on only a few dominant types of data. Rather, most of the data types integrated into the final gene network contribute to the predictions (Fig. 3A,B). This shows that data integration was crucial to achieve the high predictive performance of the network model. The ability to predict novel genetic modifiers depends on the connectivity among known genetic modifiers. Therefore, we might expect a correlation between the number of known genetic modifiers and predictability. However, this correlation is low (Fig. 3C), suggesting that the number of known interaction partners does not strongly affect prediction power, provided there are sufficient known cases to seed the prediction.

It is noteworthy that of the functional gene linkage information derived from other species, the majority comes from yeast orthologs. In Figure 3, A and B, yeast data account for 33% of predictability for the Lehner et al. (2006a) data set, while human

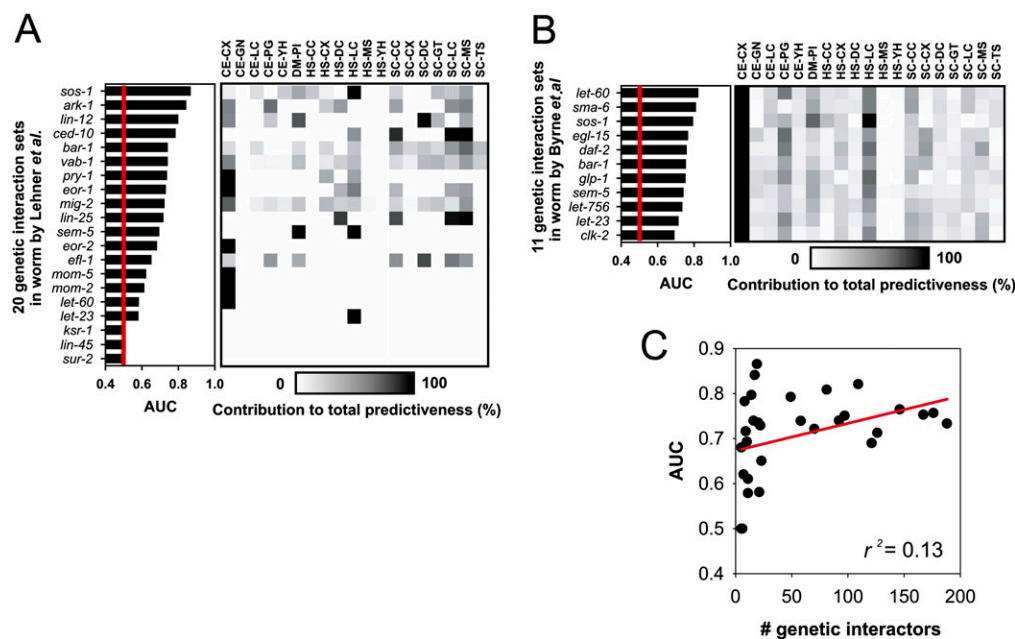


Figure 3. High predictive power for genetic modifiers stems from a wide variety of data types integrated into WormNet. Both independent screens for genetic modifier for worm orthologs of human disease genes by Lehner et al. (2006a) (A) and by Byrne et al. (2007) (B) shows that groups of genetic modifiers (labeled by sharing genetic interaction partner disease gene names at y-axis) with high AUC scores (random expectation is indicated by red line where AUC = 0.5) are supported by contribution (degree of contribution is measured and indicated as in Fig. 1D) of diverse data types (listed at x-axis with same code scheme as Fig. 1A). (C) Predictability does not depend on seed set size, seen by a low correlation between the number of known genetic modifiers (number of seed genes) and predictability (indicated by AUC).

and fly data account for 21% and 6.7%, respectively (compared with 39% from worm). For the Byrne et al. (2007) interactions, yeast data account for 22%, human 19%, and fly 6.4% of predictability (compared with 52.7% from worm). It has been shown previously that genetic interactions are relatively poorly conserved between species (Dixon et al. 2008; Roguev et al. 2008; Tischler et al. 2008), which might suggest that genetic interactions could not be easily predicted using cross-species data. This is not what we find, since yeast data make large contributions to the predictability of genetic interactions in the worm. While individual genetic interactions themselves may be poorly conserved, functional modules and pathways do tend to be highly conserved between species, and this means that functional couplings derived from one species can be highly predictive of genetic interactions in another species with the method proposed here (Fig. 2). Thus, functional genomic datasets from yeast and other model organisms—including genetic interaction screens—can be highly informative for predicting modifier loci in higher organisms, including humans.

Network-guided modifier screening: Verifying new interactions for three signaling genes

Based on our analysis of existing genetic interaction data in worms and yeast, we propose a general strategy for discovering disease gene modifiers using integrated functional networks. Starting with a small number of known modifiers, a functional network can be used to predict new candidate modifiers. These candidates can then be experimentally validated in a cost-effective and statistically powerful way (Fig. 2A).

To illustrate this approach, we used the network to predict new genetic modifiers for three signal transduction genes in *C. elegans* and tested these predictions experimentally. Previously, we performed focused screens testing the effects of inhibiting ~8% of *C. elegans* genes on the phenotypic outcome of these three mutations. These ~8% of genes were chosen because of their predicted functions in signaling, transcription regulation, or chromatin remodeling, which was designed to increase the recovery of interactions in the screen. We used the modifiers identified in each of these previous screens as seed genes in WormNet 2, ranking all genes by their functional coupling to these genes (Fig. 2A).

The genes that we considered encode an ephrin receptor (*vab-1*) and the son-of-sevenless (*sos-1*) and ACK kinase (*ark-1*) orthologs that function in growth-factor receptor/MAP kinase signaling. For each gene, we tested approximately 90 additional loci ranked as most closely functionally coupled to the previously identified modifiers, focusing only on genes that had not been tested in the previous screen (Lehner et al. 2006a) and for which RNAi clones were available in the Ahringer feeding library (Kamath et al. 2003).

In total we identified 31 novel modifier interactions (11% of tested genes). This represents a 7.3-fold enrichment compared with our previous focused screens ($P = 8 \times 10^{-16}$, Fisher's exact test). For the individual genes, the validation rates were: 4% (*ark-1*), 14% (*sos-1*), and 15% (*vab-1*), all significantly enriched compared with the previous screens (1%, 1%, and 2.4%, respectively, $P < 0.05$ in all cases). In Figure 4, we list the new and previously identified modifiers of each mutation together with the functional couplings that connect them in WormNet 2. Example interactions are illustrated in Figure 5. It can be seen that the functional couplings derive from multiple data types and multiple species, clearly demonstrating the benefits of data integration. Together with the analysis of previous screens, we conclude from this that integrated networks provide a powerful method for predicting genetic modifiers in animals.

Discussion

Functional networks provide a general strategy for predicting genetic modifier loci

We have demonstrated here that integrated, functional networks provide a general method for predicting genetic modifier loci in both unicellular and multicellular organisms. The basis for this is the principle that genes that act in a common pathway or processes generally act as modifier loci for the same mutations.

Our method extends beyond previous work that has used physical interaction networks to predict genetic interactions (Wong et al. 2004; Kelley and Ideker 2005; Le Meur and Gentleman 2008) and we demonstrated that considering functional interactions offers improved predictive power. Two reasons for this are the integration of many more diverse datasets and the fact that functional couplings transcend physical interactions. Our method also differs from previous efforts to predict genetic interactions by training a network on known genetic interactions (Zhong and Sternberg 2006) or inferring new interactions from large datasets of genetic interactions (Qi et al. 2008). Such approaches would be very difficult in higher organisms given the very small number of previously identified interactions (Flint and Mackay 2009). In contrast, human functional networks can be constructed using existing data, meaning that it will be possible to easily extend our approach to our own species.

The main caveat of our method is that it requires knowledge of a set of “seed” genes that are known to modify any mutation of interest. These seed genes may be identified from previous knowledge, an unbiased screen, or from testing candidate genes.

Network-guided modifier screening

In yeast and *C. elegans*, genetic interaction screens are quite straightforward. However, given the enormous number of gene pair combinations that need to be tested, they are still highly labor intensive. For example, the two largest studies to date in these species tested only ~3% and <0.02% of possible interactions (Tong et al. 2004; Lehner et al. 2006a).

Our method offers an alternative approach. We showed how it is possible to first screen a subset of the genome, and then to use the interactions identified in this first screen to predict more interactions with genes in the rest of the genome. Using interactions derived from screening ~8% of the genome with three mutations in signal transduction genes, and testing approximately 90 predicted interactions for each gene, we were able to identify a total of 31 new genetic modifiers. Thus, we achieved a four- to 12-fold enhancement of interaction discovery compared with our previous screens.

A new improved functional gene network for *C. elegans*

The accuracy and coverage of a network model is critical for its predictability. To increase predictability we improved and extended an earlier version of a functional gene network for *C. elegans* (Lee et al. 2008) by incorporating a large number of modifications and new datasets as summarized in Supplemental Table S1. We found that the updated network, WormNet version 2, significantly improves not only prediction of loss-of-function phenotypes (Fig. 1C,D) but also prediction of genetic modifiers (Fig. 2F). This new network can be accessed through a web interface (<http://www.functionalnet.org/wormnet>). Using this interface, researchers

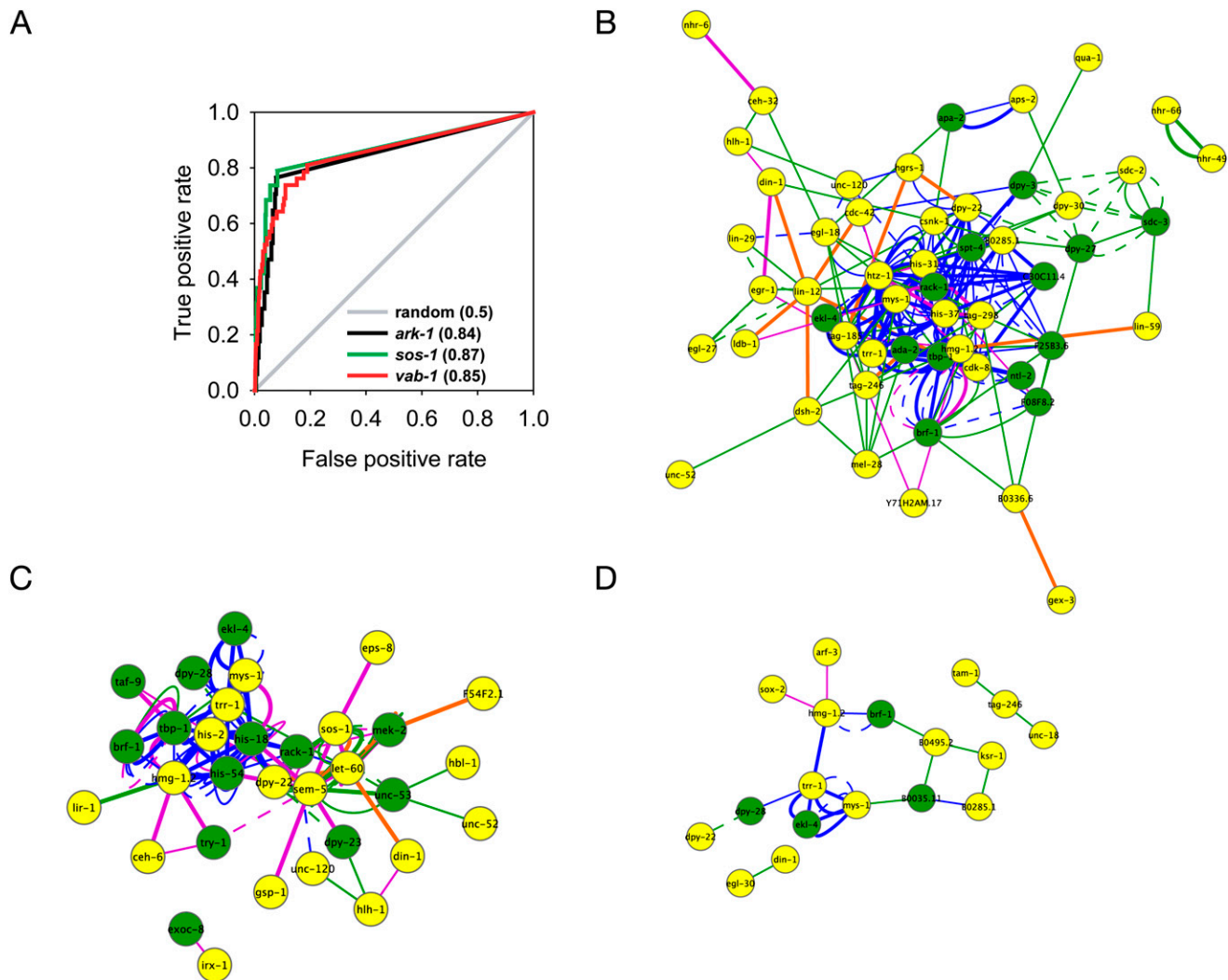


Figure 4. Predicting and validating novel genetic modifiers for three signaling genes in *C. elegans*. New candidate genetic modifiers were predicted for mutations in three signal transduction genes by ranking all genes in the genome by their functional coupling to the known genetic modifiers for each gene as shown schematically in Figure 2A. The top ranked approximately 90 genes that had not been previously tested for their ability to interact were then tested. (A) Cross-validated ROC curves for each of the three genes' previously known genetic modifiers. AUCs are indicated after each gene name in key. (B–D) Previously known (yellow nodes) and newly verified (green nodes) genetic modifiers of mutations in *vab-1* (B), *sos-1* (C), and *ark-1* (D) are plotted as networks, showing the high network connectivity among the genes that led to their prediction as genetic modifiers. Edges show WormNet 2 predicted functional couplings between the modifier loci, derived from evidence in *C. elegans* (green edges), *D. melanogaster* (orange edges), *S. cerevisiae* (blue edges), and *H. sapiens* (purple edges). Physical interaction evidence is shown as thick lines, cocitation evidence as dashed lines, and all other evidence as solid lines. Gene networks were plotted using Cytoscape 2.5.2. (Cline et al. 2007).

can easily search the network using a set of “seed” genes of interest. The interface returns a list of genes ranked according to their connections to the seed genes together with the evidence used to identify each coupling. The interactions and evidence can be downloaded, and a network visualization tool has been incorporated.

A general strategy for predicting genetic modifier loci in human disease

The success of our approach suggests that a similar strategy should also work for identifying modifier loci in humans. Most diseases in humans are known to be genetically complex, but systematically identifying interactions between mutations in population studies is extremely difficult because of a lack of statistical power. Our re-

sults suggest an alternative approach: Starting from a small seed set of interactions, an integrated functional network can be used to predict more candidate modifier loci. For example, a few modifier loci might be known from previous studies or a genome-wide association study. Using this set and an integrated network, new modifiers could be predicted. These candidates could then be verified in a population study, which would obtain much greater statistical power because of the smaller number of genes being considered.

Concluding remarks

In summary, the approach outlined here provides a general method for predicting genetic modifier loci. This will be useful in experimental model organisms, but also in species of clinical or

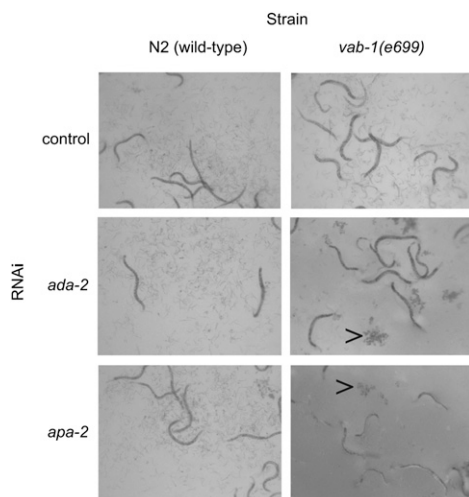


Figure 5. Examples of novel genetic modifier interactions in *C. elegans*. Here, the phenotypic consequences of inhibiting each of two genes *ada-2* and *apa-2* in wild-type (Bristol N2) and *vab-1(e699)* mutant animals are compared. In wild-type worms, inhibiting either of these genes produces minimal phenotypic change. In the images, adult worms can be observed together with their larval progeny. In contrast, in *vab-1(e699)* mutants, *ada-2(RNAi)* and *apa-2(RNAi)* produce embryonic lethality (example clusters of unhatched embryos are marked by arrowheads), a reduced brood size, and for *apa-2(RNAi)* delayed growth of the first generation worms.

agricultural importance. Indeed, our results suggest that a single high-quality functional network for humans will provide a powerful resource for the systematic identification of modifier loci in multiple genetic diseases.

Methods

Caenorhabditis elegans gene annotation reference sets and network integration

WormNet version 2 is based on the 20,081 *C. elegans* protein-coding genes annotated by WormBase 170 (downloaded from the WormBase ftp site in January 2007) (Uren et al. 2008). All linkages and calculations of genome coverage are based on this gene set. A reference set for gene functional associations was assembled from gene pairs sharing any Gene Ontology (GO) biological process annotations (downloaded from WormBase ftp site in January 2007). We excluded the following annotations: (1) seven overdominant terms—embryonic development (sensu Metazoa) (GO no. 0009792), larval development (sensu Nematoda) (GO no. 0002119), positive regulation of growth rate (GO no. 0040010), locomotory behavior (GO no. 0007626), regulation of transcription, DNA dependent (GO no. 0006355), gametogenesis (GO no. 0007276), transport (GO no. 0006810)—and three terms about general post-translational modification that labeled highly diverse pathways—G-protein coupled receptor protein signaling pathway (GO no. 0007186), protein amino acid phosphorylation (GO no. 0006468), protein amino acid dephosphorylation (GO no. 0006470). If not excluded, these 10 terms would account for 90% of the total positive training gene pairs. (2) All terms at the first level in the GO hierarchy (taking the parental term “biological process” as level zero)—growth (GO no. 0040007), reproduction (GO no. 0000003), metabolic process (GO no. 0008152), locomotion (GO no. 0040011), localization (GO no. 0051179). The resulting set of reference gene annotations contained 626,342 pairs covering 5178 *C. elegans* genes (~25.8% of 20,081 WormBase170 genes encoding proteins).

Functional associations were calculated using the log likelihood scoring (LLS) scheme:

$$LLS = \ln \left(\frac{P(L|E)/P(-L|E)}{P(L)/P(-L)} \right),$$

where $P(L|E)$ and $P(-L|E)$ are the frequencies of gold standard linkages (L) observed in the given experiment (E) between annotated *C. elegans* genes operating in the same pathway (indicated by L , positive instances) and in different pathways (indicated by $-L$, negative instances), respectively, while $P(L)$ and $P(-L)$ represent the prior expectations (i.e., the total frequency of linkages between all annotated *C. elegans* genes operating in the same pathway and operating in different pathways, respectively). To avoid circular training and monitor overtraining, we used 0.632 bootstrapping constructs training sets from data sampled with replacement and test sets from the remaining data that weren't sampled from the original training set. Each linkage has a probability of $1 - 1/n$ of not being sampled, resulting in ~63.2% of the original training data in the training set and ~36.8% in the test set. The overall LLS is the weighted average of results on the two sets, equal to $0.632 \times LLS_{\text{test}} + (1 - 0.632) \times LLS_{\text{train}}$. Log likelihood scores from each contributing data set were integrated using the weighted sum (WS) method.

$$WS = L_0 + \sum_{i=1}^n \frac{L_i}{D \times i}, \text{ for all } L \geq T,$$

where L_0 is the best LLS score among all LLS s for that gene pair, D is a free parameter for the overall degree of dependence among the data sets, T is a LLS threshold for all data sets being integrated, and i is the order index of the data sets after rank-ordering LLS scores according to descending magnitude. The values of two free parameters (D and T) are systematically chosen to maximize overall performance (LLS and gene coverage) on the benchmark.

Inferring functional associations from *C. elegans* transcript expression data

mRNA coexpression associations were inferred from public DNA microarray datasets as described in Lee et al. (2004) (referred to as data set CE-CX) using data listed in Supplemental Table S5 with the following modification: For WormNet version 2, we removed gene pairs likely to cross-hybridize to each other's DNA microarray probes based on observing significant DNA sequence homology, defined by a BLASTN E -value $\leq 10^{-4}$ and nucleotide sequence identity $\geq 70\%$, as established by Ramani et al. (2008). We integrated log likelihood scores from distinct sets of microarray experiments using the weighted sum method described above.

Functional associations inferred from physical and genetic interactions

WormNet version 2 incorporates physical protein interaction datasets from high and medium-throughput yeast two-hybrid analyses of *C. elegans* genes (CE-YH) reported in the Worm Interactome Version 5 (WI5) (Li et al. 2004). We treated subsets of the WI5 (literature, scaffold, core1, core2, noncore) separately, providing different confidence scores for the different data subsets, rather than a single averaged confidence score across all interactions of the set. The WI5 literature-curated subset was combined with literature-based protein physical interactions from BIND (Alfarano et al. 2005), IntAct (Kerrien et al. 2007), and MINT (Chatr-aryamontri et al. 2007) to construct the literature-curated protein physical interaction set (CE-LC). Genetic interactions (CE-GT) (~2200 interactions among ~1000 genes) were included from

WormBase170 (Uren et al. 2008) derived from more than 1000 primary publications. Additional linkages were identified based upon cocitation of gene names in Medline abstracts downloaded on December 2004). We analyzed a set of $n = 7732$ Medline abstracts that included the word “*elegans*” in the abstract for perfect matches to either the systematic names or common names of 20,081 genes of *C. elegans*, scoring gene pairs according to the scheme of Lee et al. (2004) (CE-CC). We excluded all previously predicted genetic interactions (Zhong and Sternberg 2006) as well as those from two genome-scale screens (Lehner et al. 2006a; Byrne et al. 2007).

Inferring functional associations from phylogenetic profiles and genomic context

To discover functional associations between genes on the basis of the genomic context of orthologs of *C. elegans* genes, we used the phylogenetic profile method (Pellegrini et al. 1999; Huynen et al. 2000; Wolf et al. 2001) and the gene neighbors method (Dandekar et al. 1998; Overbeek et al. 1999; Bowers et al. 2004). For both methods, we analyzed 424 bacterial genome sequences (31 archaeobacteria and 393 eubacteria, downloaded from NCBI in December 2006). *C. elegans* protein sequences were aligned to protein sequences encoded by the 424 bacterial genomes, using the program BLASTP with default settings (Altschul et al. 1997) and alignment scores analyzed as in Date and Marcotte (2003) with the modification of using discretized BLASTP *E*-values when calculating mutual information between phylogenetic profiles. We used bins of equal numbers of *E*-values rather than equal intervals of *E*-values, accounting for the nonuniform *E*-value distribution. We benchmarked linkages inferred from three subsets of the genomes—the complete set of 424 genomes, the set of 311 genomes representing unique species, and the set of 181 genomes representing unique genes, selecting the representative species or genus with the maximum number of BLASTP hits to *C. elegans* proteins. Using recall/precision analysis, we found that the 181 genome set maximized performance for the gene neighbor method (CE-GN) and the 424 genome set performed best for the phylogenetic profiling method (CE-PG). We further limited the phylogenetic profile analysis to *C. elegans* proteins with fewer than 19 InterPro domains (Hunter et al. 2009), which performed considerably better as measured by recall-precision analysis, due to a tendency for larger proteins to show promiscuous functional associations arising from misestimated coinheritance patterns. Log likelihood scores were assigned as described above.

Inferring gene functional associations using associologs

In addition to *C. elegans* data, we also analyzed datasets collected for yeast, fly, and human. Linkages between *C. elegans* gene pairs were identified based on the data associated with the orthologs of the *C. elegans* genes (“associologs”) (Lee et al. 2008). A total of 14 linkage sets were analyzed based on datasets previously incorporated into gene networks for yeast (seven sets from YeastNet; Lee et al. 2007), human (six sets from HumanNet; I Lee, M Blom, PI Wang, J Shim, and EM Marcotte, in prep.), and fly (fly protein-protein interactions derived from BIOGRID [Breitkreutz et al. 2008], IntAct [Kerrien et al. 2007], and MINT [Chatr-aryamontri et al. 2007], downloaded in March 2007) and summarized in Supplemental Table S2. Orthologs were defined between *C. elegans* proteins and other organisms using Inparanoid (Remm et al. 2001). Transferred linkages were weighted by Inparanoid-assigned confidence scores in the orthology assignments using the following scheme: We defined an Inparanoid weighted log likelihood score (IWLLS) equal to the LLS from the network of origin + log(Inparanoid score for gene A) + log(Inparanoid score for gene B). *C. elegans* gene pairs were ranked by the IWLLS scores, then log likelihood scores calculated using the *C. elegans* reference annotation set, as for *C. elegans* datasets.

noid score for gene A) + log(Inparanoid score for gene B). *C. elegans* gene pairs were ranked by the IWLLS scores, then log likelihood scores calculated using the *C. elegans* reference annotation set, as for *C. elegans* datasets.

ROC analysis of predictability of genetic modifier identification

A network’s predictive power for inferring loss-of-function phenotypes or genetic modifiers was tested using leave-one-out cross-validation on known phenotypic genes or known genetic modifiers (termed “seed” genes) by scoring all genes in the genome by each gene’s associated sum of LLS scores to the seed genes, omitting each seed gene in turn from the seed set for purposes of its own evaluation. A ROC curve was calculated by plotting the true-positive rate (TP/[TP + FN]) versus false-positive rate (FP/[FP + TN]) for all genes scored above a sliding score threshold. (A set of equally scoring genes is thus evaluated as a single step in the ROC plot.) Seed genes (positives) well connected to one another will thus score higher than nonseed genes (negatives), resulting in a ROC curve above the diagonal. Each ROC analysis was summarized by the AUC, where AUC = 0.5 is expected for a random predictor and AUC = 1 for a perfect predictor, arising in the case of all of the seed genes being tightly interconnected in the network.

Predicting new modifier loci for signal transduction genes in *C. elegans*

We predicted new modifier loci for mutations in three *C. elegans* genes by their connectivity to the previously identified genetic interaction partners of these genes (Lehner et al. 2006a). For each gene in the genome, the log likelihood scores of their WormNet 2 interactions to the previously identified modifiers were summed. All genes were then ranked using these summed scores. For each mutation we aimed to test the 96 highest-ranked predictions that had not been tested in our previous screen (Lehner et al. 2006a) and for which RNAi clones were available in the Ahringer feeding library (Kamath et al. 2003). Excluding poorly growing bacterial clones that were not tested, these RNAi library clones are listed in Supplemental Tables S6–S8.

Experimental validation of novel interactions in *C. elegans*

We performed RNAi screens in liquid culture in 96-well plates as described (Lehner et al. 2006a,b). In brief, synchronized L₁ stage animals were added at a density of approximately 15 animals per well to a total volume of 50 μ L of media containing each RNAi feeding strain or control bacteria and incubated with shaking at 20°C. Viability phenotypes (sterility, lethality) were scored by visual inspection after 4 and 5 d of incubation, directly comparing the phenotype of each RNAi treatment in mutant and wild-type (Bristol N2) animals. To be considered an interaction, the RNAi-treated mutant strain had to show a phenotype considerably stronger than both that of the mutant strain and that resulting from the RNAi treatment in wild-type (N2) animals, as described (Lehner et al. 2006a,b). All feeding experiments and controls were repeated four times in each experiment, and the complete experiment was replicated three times. An interaction had to be observed in at least two wells of each individual experiment, and in at least two of three repeats. The identities of all positive RNAi feeding clones were confirmed by sequencing. The following *C. elegans* strains were used: PS1461 [*ark-1(sy247)*] (Hopper et al. 2000), UP604 [*sos-1(cs41)*] (Rocheleau et al. 2002), CZ414 [*vab-1(e1699)*] (George et al. 1998).

Acknowledgments

This work was supported by grants from the National Research Foundation of Korea (NRF) funded by the Korea government (MEST) (nos. 2009-0063342, 2009-0070968, 2009-0087951) and Yonsei University (no. 2008-7-0284, 2008-1-0018) (I.L.), from the NSF, NIH, and Welch (F1515) and Packard Foundations (E.M.M.), and from the ERC, MICINN, ICREA, AGAUR, and the EMBL-CRG Systems Biology Program (B.L.); T.V. is supported by a Marie Curie intra-European fellowship. Some nematode strains used in this work were provided by the *Caenorhabditis* Genetics Center, which is funded by the NIH National Center for Research Resources (NCRR).

Author Contributions: I.L., E.M.M., A.G.F., and B.L. conceived the project; I.L. constructed WormNet using approaches developed with E.M.M.; B.L. and T.V. performed the experimental validation; J.S. assisted in construction of the web-based prediction server; I.L. and B.L. analyzed data and wrote the manuscript. I.L., E.M.M., A.G.F., and B.L. edited the manuscript.

References

- Alexeyenko A, Sonhammer EL. 2009. Global networks of functional coupling in eukaryotes from comprehensive data integration. *Genome Res* **19**: 1107–1116.
- Alfarano C, Andrade CE, Anthony K, Bahroos N, Bajec M, Bantoft K, Betel D, Bobechko B, Boutilier K, Burgess E, et al. 2005. The Biomolecular Interaction Network Database and related tools 2005 update. *Nucleic Acids Res* **33**: D418–D424.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.
- Bochdanovits Z, Sondervan D, Perillous S, van Beijsterveldt T, Boomsma D, Heutink P. 2008. Genome-wide prediction of functional gene-gene interactions inferred from patterns of genetic differentiation in mice and men. *PLoS One* **3**: e1593. doi: 10.1371/journal.pone.0001593.
- Bonetta L. 2008. Getting up close and personal with your genome. *Cell* **133**: 753–756.
- Boone C, Bussey H, Andrews BJ. 2007. Exploring genetic interactions and networks with yeast. *Nat Rev Genet* **8**: 437–449.
- Bowers PM, Pellegrini M, Thompson MJ, Fierro J, Yeates TO, Eisenberg D. 2004. Prolinks: A database of protein functional linkages derived from coevolution. *Genome Biol* **5**: R35. <http://genomebiology.com/2004/5/5/R35>.
- Breitkreutz BJ, Stark C, Reguly T, Boucher L, Breitkreutz A, Livstone M, Oughtred R, Lackner DH, Bahler J, Wood V, et al. 2008. The BioGRID Interaction Database: 2008 Update. *Nucleic Acids Res* **36**: D637–D640.
- Byrne AB, Weirauch MT, Wong V, Koeva M, Dixon SJ, Stuart JM, Roy PJ. 2007. A global analysis of genetic interactions in *Caenorhabditis elegans*. *J Biol* **6**: 8.
- Chatr-aryamontri A, Ceol A, Palazzi LM, Nardelli G, Schneider MV, Castagnoli L, Cesareni G. 2007. MINT: The Molecular Interaction database. *Nucleic Acids Res* **35**: D572–D574.
- Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, et al. 2007. Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc* **2**: 2366–2382.
- Collins SR, Miller KM, Maas NL, Roguev A, Fillingham J, Chu CS, Schuldiner M, Gebbia M, Recht J, Shales M, et al. 2007. Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* **446**: 806–810.
- Dandekar T, Snel B, Huynen M, Bork P. 1998. Conservation of gene order: A fingerprint of proteins that physically interact. *Trends Biochem Sci* **23**: 324–328.
- Date SV, Marcotte EM. 2003. Discovery of uncharacterized cellular systems by genome-wide analysis of functional linkages. *Nat Biotechnol* **21**: 1055–1062.
- Davierwala AP, Haynes J, Li Z, Brost RL, Robinson MD, Yu L, Mnaimneh S, Ding H, Zhu H, Chen Y, et al. 2005. The synthetic genetic interaction spectrum of essential genes. *Nat Genet* **37**: 1147–1152.
- Dixon SJ, Fedysyn Y, Koh JL, Prasad TS, Chahwan C, Chua G, Toufighi K, Baryshnikova A, Hayles J, Hoe KL, et al. 2008. Significant conservation of synthetic lethal genetic interaction networks between distantly related eukaryotes. *Proc Natl Acad Sci* **105**: 16653–16658.
- Ewing RM, Chu P, Elisma F, Li H, Taylor P, Climie S, McBroom-Cerajewski L, Robinson MD, O'Connor L, Li M, et al. 2007. Large-scale mapping of human protein-protein interactions by mass spectrometry. *Mol Syst Biol* **3**: 89. doi: 10.1038/msb4100134.
- Flint J, Mackay TF. 2009. Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res* **19**: 723–733.
- Fraser AG, Marcotte EM. 2004. A probabilistic view of gene function. *Nat Genet* **36**: 559–564.
- Fraser HB, Plotkin JB. 2007. Using protein complexes to predict phenotypic effects of gene mutation. *Genome Biol* **8**: R252. doi: 10.1186/gb-2007-8-11-r252.
- George SE, Simokat K, Hardin J, Chisholm AD. 1998. The VAB-1 Eph receptor tyrosine kinase functions in neural and epithelial morphogenesis in *C. elegans*. *Cell* **92**: 633–643.
- Guan Y, Myers CL, Lu R, Lemischka IR, Bult CJ, Troyanskaya OG. 2008. A genome-wide functional network for the laboratory mouse. *PLoS Comput Biol* **4**: e1000165. doi: 10.1371/journal.pcbi.1000165.
- Gunsalus KC, Ge H, Schetter AJ, Goldberg DS, Han JD, Hao T, Berriz GF, Bertin N, Huang J, Chuang LS, et al. 2005. Predictive models of molecular machines involved in *Caenorhabditis elegans* early embryogenesis. *Nature* **436**: 861–865.
- Hart GT, Lee I, Marcotte EM. 2007. A high-accuracy consensus map of yeast protein complexes reveals modular nature of gene essentiality. *BMC Bioinformatics* **8**: 236. doi: 10.1186/1471-2105-8-236.
- Hartman J, Garvik B, Hartwell L. 2001. Principles for the buffering of genetic variation. *Science* **291**: 1001–1004.
- Hopper NA, Lee J, Sternberg PW. 2000. ARK-1 inhibits EGFR signaling in *C. elegans*. *Mol Cell* **6**: 65–75.
- Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, et al. 2009. InterPro: The integrative protein signature database. *Nucleic Acids Res* **37**: D211–D215.
- Huynen M, Snel B, Lathe W III, Bork P. 2000. Predicting protein function by genomic context: Quantitative evaluation and qualitative inferences. *Genome Res* **10**: 1204–1210.
- Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, et al. 2009. STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* **37**: D412–D416.
- Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, Kanapin A, Le Bot N, Moreno S, Sohrmann M, et al. 2003. Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**: 231–237.
- Karaoz U, Murali TM, Letovsky S, Zheng Y, Ding C, Cantor CR, Kasif S. 2004. Whole-genome annotation by using evidence integration in functional-linkage networks. *Proc Natl Acad Sci* **101**: 2888–2893.
- Kelley R, Ideker T. 2005. Systematic interpretation of genetic interactions using protein networks. *Nat Biotechnol* **23**: 561–566.
- Kerrien S, Alam-Faruque Y, Aranda B, Bancarz I, Bridge A, Derow C, Dimmer E, Feuermann M, Friedrichsen A, Huntley R, et al. 2007. IntAct—open source resource for molecular interaction data. *Nucleic Acids Res* **35**: D561–D565.
- Kim JK, Gabel HW, Kamath RS, Tewari M, Pasquinelli A, Rual JF, Kennedy S, Dybbs M, Bertin N, Kaplan JM, et al. 2005. Functional genomic analysis of RNA interference in *C. elegans*. *Science* **308**: 1164–1167.
- Kim WK, Krumpelman C, Marcotte EM. 2008. Inferring mouse gene functions from genomic-scale data using a combined functional network/classification strategy. *Genome Biol* **9**: S5. doi: 10.1186/gb-2008-9-s1-s5.
- Lage K, Karlberg EO, Stirling ZM, Olason PI, Pedersen AG, Rigina O, Hinsby AM, Tumer Z, Pociot F, Tommerup N, et al. 2007. A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat Biotechnol* **25**: 309–316.
- Le Meur N, Gentleman R. 2008. Modeling synthetic lethality. *Genome Biol* **9**: R135. doi: 10.1186/gb-2009-9-9-r135.
- Lee I, Date SV, Adai AT, Marcotte EM. 2004. A probabilistic functional network of yeast genes. *Science* **306**: 1555–1558.
- Lee I, Li Z, Marcotte EM. 2007. An improved, bias-reduced probabilistic functional gene network of baker's yeast, *Saccharomyces cerevisiae*. *PLoS One* **2**: e988. doi: 10.1371/journal.pone.0000988.
- Lee I, Lehner B, Crombie C, Wong W, Fraser AG, Marcotte EM. 2008. A single gene network accurately predicts phenotypic effects of gene perturbation in *Caenorhabditis elegans*. *Nat Genet* **40**: 181–188.
- Lehner B. 2007. Modelling genotype-phenotype relationships and human disease with genetic interaction networks. *J Exp Biol* **210**: 1559–1566.
- Lehner B, Crombie C, Tischler J, Fortunato A, Fraser AG. 2006a. Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. *Nat Genet* **38**: 896–903.
- Lehner B, Tischler J, Fraser AG. 2006b. RNAi screens in *Caenorhabditis elegans* in a 96-well liquid format and their application to the systematic identification of genetic interactions. *Nat Protoc* **1**: 1617–1620.

- Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, Vidalain PO, Han JD, Chesneau A, Hao T, et al. 2004. A map of the interactome network of the metazoan *C. elegans*. *Science* **303**: 540–543.
- Lin YY, Qi Y, Lu JY, Pan X, Yuan DS, Zhao Y, Bader JS, Boeke JD. 2008. A comprehensive synthetic genetic interaction network governing yeast histone acetylation and deacetylation. *Genes Dev* **22**: 2062–2074.
- Linghu B, Snitkin ES, Hu Z, Xia Y, Delisi C. 2009. Genome-wide prioritization of disease genes and identification of disease-disease associations from an integrated human functional linkage network. *Genome Biol* **10**: R91. doi: 10.1186/gb-2009-10-9-r91.
- Mackay TF, Stone EA, Ayroles JF. 2009. The genetics of quantitative traits: Challenges and prospects. *Nat Rev Genet* **10**: 565–577.
- Maher B. 2008. Personal genomes: The case of the missing heritability. *Nature* **456**: 18–21.
- Marcotte EM, Pellegrini M, Thompson MJ, Yeates TO, Eisenberg D. 1999. A combined algorithm for genome-wide prediction of protein function. *Nature* **402**: 83–86.
- McGary KL, Lee I, Marcotte EM. 2007. Broad network-based predictability of *Saccharomyces cerevisiae* gene loss-of-function phenotypes. *Genome Biol* **8**: R258. doi: 10.1186/gb-2007-8-12-r258.
- Mellor JC, Yanai I, Clodfelter KH, Mintseris J, DeLisi C. 2002. Predictome: A database of putative functional links between proteins. *Nucleic Acids Res* **30**: 306–309.
- Myers CL, Troyanskaya OG. 2007. Context-sensitive data integration and prediction of biological networks. *Bioinformatics* **23**: 2322–2330.
- Overbeek R, Fonstein M, D'Souza M, Pusch GD, Maltsev N. 1999. The use of gene clusters to infer functional coupling. *Proc Natl Acad Sci* **96**: 2896–2901.
- Pan X, Ye P, Yuan DS, Wang X, Bader JS, Boeke JD. 2006. A DNA integrity network in the yeast *Saccharomyces cerevisiae*. *Cell* **124**: 1069–1081.
- Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO. 1999. Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. *Proc Natl Acad Sci* **96**: 4285–4288.
- Peña-Castillo L, Tasan M, Myers CL, Lee H, Joshi T, Zhang C, Guan Y, Leone M, Pagnani A, Kim WK, et al. 2008. A critical assessment of *Mus musculus* gene function prediction using integrated genomic evidence. *Genome Biol* **9**: S2. doi: 10.1186/gb-2008-9-s1-s2.
- Qi Y, Suhail Y, Lin YY, Boeke JD, Bader JS. 2008. Finding friends and enemies in an enemies-only network: A graph diffusion kernel for predicting novel genetic interactions and co-complex membership from yeast genetic interactions. *Genome Res* **18**: 1991–2004.
- Ramani AK, Li Z, Hart GT, Carlson MW, Boutz DR, Marcotte EM. 2008. A map of human protein interactions derived from co-expression of human mRNAs and their orthologs. *Mol Syst Biol* **4**: 180. doi: 10.1038/msb.2008.19.
- Remm M, Storm CE, Sonnhammer EL. 2001. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J Mol Biol* **314**: 1041–1052.
- Rhodes DR, Tomlins SA, Varambally S, Mahavisno V, Barrette T, Kalyana-Sundaram S, Ghosh D, Pandey A, Chinnaiyan AM. 2005. Probabilistic model of the human protein-protein interaction network. *Nat Biotechnol* **23**: 951–959.
- Rocheleau CE, Howard RM, Goldman AP, Volk ML, Girard LJ, Sundaram MV. 2002. A *lin-45 raf* enhancer screen identifies *cor-1*, *cor-2* and unusual alleles of Ras pathway genes in *Caenorhabditis elegans*. *Genetics* **161**: 121–131.
- Roguev A, Bandyopadhyay S, Zofall M, Zhang K, Fischer T, Collins SR, Qu H, Shales M, Park HO, Hayles J, et al. 2008. Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. *Science* **322**: 405–410.
- Schuldiner M, Collins SR, Thompson NJ, Denic V, Bhamidipati A, Punna T, Ihmels J, Andrews B, Boone C, Greenblatt JF, et al. 2005. Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* **123**: 507–519.
- Simonis N, Rual JF, Carvunis AR, Tasan M, Lemmens I, Hirozane-Kishikawa T, Hao T, Sahalie JM, Venkatesan K, Gebreab F, et al. 2009. Empirically controlled mapping of the *Caenorhabditis elegans* protein-protein interactome network. *Nat Methods* **6**: 47–54.
- Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S, et al. 2005. A human protein-protein interaction network: A resource for annotating the proteome. *Cell* **122**: 957–968.
- Tian W, Zhang LV, Tasan M, Gibbons FD, King OD, Park J, Wunderlich Z, Cherry JM, Roth FP. 2008. Combining guilt-by-association and guilt-by-profiling to predict *Saccharomyces cerevisiae* gene function. *Genome Biol* **9**: S7. doi: 10.1186/gb-2008-9-s1-s7.
- Tischler J, Lehner B, Fraser AG. 2008. Evolutionary plasticity of genetic interaction networks. *Nat Genet* **40**: 390–391.
- Tong AH, Lesage G, Bader GD, Ding H, Xu H, Xin X, Young J, Berriz GF, Brost RL, Chang M, et al. 2004. Global mapping of the yeast genetic interaction network. *Science* **303**: 808–813.
- Troyanskaya OG, Dolinski K, Owen AB, Altman RB, Botstein D. 2003. A Bayesian framework for combining heterogeneous data sources for gene function prediction (in *Saccharomyces cerevisiae*). *Proc Natl Acad Sci* **100**: 8348–8353.
- Ulitsky I, Shamir R. 2007. Pathway redundancy and protein essentiality revealed in the *Saccharomyces cerevisiae* interaction networks. *Mol Syst Biol* **3**: 104.
- Uren AG, Kool J, Matentzoglou K, de Ridder J, Mattison J, van Uiter M, Lagcher W, Sie D, Tanger E, Cox T, et al. 2008. Large-scale mutagenesis in p19(ARF)- and p53-deficient mice identifies cancer genes and their collaborative networks. *Cell* **133**: 727–741.
- Wolf YI, Rogozin IB, Kondrashov AS, Koonin EV. 2001. Genome alignment, evolution of prokaryotic genome organization, and prediction of gene function using genomic context. *Genome Res* **11**: 356–372.
- Wong SL, Zhang LV, Tong AH, Li Z, Goldberg DS, King OD, Lesage G, Vidal M, Andrews B, Bussey H, et al. 2004. Combining biological networks to predict genetic interactions. *Proc Natl Acad Sci* **101**: 15682–15687.
- Ye P, Peysers BD, Pan X, Boeke JD, Spencer FA, Bader JS. 2005. Gene function prediction from congruent synthetic lethal interactions in yeast. *Mol Syst Biol* **1**: 2005.0026. doi: 10.1038/msb4200034.
- Zhong W, Sternberg PW. 2006. Genome-wide prediction of *C. elegans* genetic interactions. *Science* **311**: 1481–1484.

Received November 1, 2009; accepted in revised form May 4, 2010.