

# ROLE OF SYNTHETIC GENETIC INTERACTIONS IN UNDERSTANDING FUNCTIONAL INTERACTIONS AMONG PATHWAYS

SHAHIN MOHAMMADI\*, GIORGOS KOLLIAS, AND ANANTH GRAMA

*Department of Computer Science, Purdue University, West Lafayette, IN 47906, USA*

*\*E-mail: mohammas@purdue.edu*

Synthetic genetic interactions reveal buffering mechanisms in the cell against genetic perturbations. These interactions have been widely used by researchers to predict functional similarity of gene pairs. In this paper, we perform a comprehensive evaluation of various methods for predicting co-pathway membership of genes based on their neighborhood similarity in the genetic network. We clearly delineate the scope of these methods and use it to motivate a rigorous statistical framework for quantifying the contribution of each pathway to the functional similarity of gene pairs. We then use our model to infer interdependencies among KEGG pathways. The resulting KEGG crosstalk map yields significant insights into the high-level organization of the genetic network and is used to explain the effective scope of genetic interactions for predicting co-pathway membership of gene pairs. A direct byproduct of this effort is that we are able to identify subsets of genes in each pathway that act as ‘ports’ for interaction across pathways.

*Keywords:* Synthetic genetic interactions, Modified congruence score, Neighborhood overlap graph, KEGG pathways, Functional dependencies

## 1. Introduction

Systematic deletion of genes in *Saccharomyces cerevisiae* (yeast) reveals that only about 18% of its nearly 6000 genes are essential for cell growth on rich glucose media.<sup>1</sup> This observation suggests a strong underlying compensatory mechanism in the cell against genetic perturbations. Gene duplications and the existence of parallel (or alternative) pathways are well-known buffering mechanisms in the cell that are responsible for its robustness.<sup>2-4</sup> Gene duplications have been studied extensively. However, it is now believed that distributed compensatory mechanisms, such as parallel pathways, contribute the most towards cell robustness.<sup>2,5</sup> The underlying structure of these parallel pathways is determined by genetic interactions among genes having dependent, but not identical functions.<sup>6,7</sup> This is the rationale behind synthetic genetic interactions, in which double-mutant phenotypes significantly deviate from the expected (or typical) phenotype for gene pairs.

High-throughput technologies based on diploid synthetic lethal analysis by microarrays (dSLAM)<sup>8,9</sup> and/or synthetic genetic arrays (SGA) and its variants,<sup>10-13</sup> provide us with large datasets of genetic interactions in *Saccharomyces cerevisiae*. These interactions may be broadly classified into two main groups – negative (or aggravating/ synergistic) interactions, which describe double mutants that exhibit a more severe phenotype than expected; and positive (or alleviating/ epistatic) interactions, in which double mutants exhibit a less severe phenotype than expected. Synthetic lethal (SL) interactions correspond to a specific subset of negative interactions, in which the double-mutant is fatal while the individual mutations are vital. Analyzing these rapidly growing datasets of genetic interactions in order to understand the functional machinery of the cell is a challenging task. Three main models have been proposed

for the higher order organization of genetic interactions – the *between-pathway model* among redundant functions, the *within-pathway model* among genes with additive effect, and the *indirect effects model*.<sup>14</sup> These models have been used both to predict functional similarity of gene pairs, implicitly, and to shed light on the higher level organization of pathways.

The rationale behind function prediction methods is that gene pairs that share “many” neighbors are more likely to be functionally similar. This thesis has been used to construct a complementary network, orthogonal to the initial genetic network, in which edges code functional similarity of gene pairs. Tong et al.<sup>11</sup> use hierarchical clustering based on the similarity of genetic interaction patterns and show that sets of genes within the same pathway tend to cluster together. Schuldiner et al.<sup>12</sup> also use genetic interaction patterns, which they refer to as *phenotypic signatures*, on an E-MAP dataset that consists of both quantitative alleviating and aggravating interactions in early secretory pathway (ESP). They confirm that pairs of genes with intermediate levels of correlated phenotypic signatures have a higher probability of negatively interacting using aggravating links, while highly correlated pairs typically show no aggravating interactions, and in many cases, they are connected using alleviating interactions. Costanzo et al.<sup>13</sup> construct a complementary weighted network, called the ‘functional map’ of the cell, in which every edge weight encodes the strength of the Pearson correlation among interacting partners of gene pairs. After applying a force-directed layout on the constructed functional map, they find that genes participating in same pathway tend to be in close proximity to each other. Ye et al.<sup>15</sup> construct a complementary network based on the statistical significance of shared neighbors, called the congruence network. They show that there is a strong correlation between GO annotations and the congruence score of gene pairs. Another class of studies uses models of genetic interactions to identify the high-level organization of pathways, and provides insight into interactions among them. These methods either use a bottom-up approach, for e.g., Segre et al.<sup>16</sup> and Bandyopadhyay et al.,<sup>17</sup> to group genes into functional modules and then analyze functional enrichment of modules, or a top-down technique, for e.g., Zhang et al.,<sup>18</sup> to infer interactions among known pathways directly.

In spite of these research efforts, there remain significant unresolved questions: How accurate are genetic interactions in predicting different functional pathways? What is the scope of these predictions (in terms of functions)? What is the specific structure around each pathway that allows us to capture the relationships among its member genes? In order to answer these questions, we start with a comprehensive evaluation of existing methods for predicting functional similarity of gene pairs. We demonstrate that, surprisingly, all of the methods considered have limited scope of effectiveness. To study this shortcoming in terms of the specific structure around each pathway, we developed a novel statistical framework (along with efficient associated algorithms) to quantify the contribution of various pathways to the functional similarity of each gene pair. We use this framework to construct the neighborhood overlap graph (NOG) for each pair of pathways and assess their functional dependence. These dependencies result in the KEGG crosstalk map, which not only explains the performance limitations of genetic interactions in predicting functional similarity of genes, but also uncovers the high-level organization of the cellular machinery. An important consequence of our method is that it enables us to identify individual genes within pathways that act as ‘ports’ of interaction

across pathways. These genes can further be used to interpret the effective scope of functional similarity prediction methods.

## 2. Methods

### 2.1. Notations

Aggravating genetic interaction network, or *genetic interaction network (GIN)*, is a genetic map of the cell, represented as a graph  $G = (V, E)$ . Vertices  $V$  in the graph correspond to genes and edges encode aggravating interaction among gene pairs. A GIN of size  $n = |V|$  can be represented by its *adjacency matrix*,  $A$ , which is an  $n \times n$  matrix. The  $(i, j)$ th element in this matrix is 1 if gene  $v_i$  interacts with  $v_j$  and 0 otherwise. We denote the set of genes having aggravating interaction with a specific gene,  $v_i \in V$ , as  $N(v_i)$ . Given the set of *core genes*, that are annotated with at least one KEGG annotation, aggravating interactions are classified as *internal*, where both genes reside in the core set, *crossing*, in which one end of interaction is in the core set and the other is not, and *external*, in which both ends of the interaction reside outside of the core set.

Each KEGG pathway is represented as a subgraph of the GIN, represented by  $P_A = (V_A, E_A)$ , in which the set  $V_A$  consists of all genes in the pathway, and edge set  $E_A$  consists of all interactions incident on these genes. For each pathway pair  $\langle P_A, P_B \rangle$ , we define the *neighborhood overlap graph (NOG)*, represented by  $H_{A \rightarrow B}$ , as the graph defined over same vertices as  $V_A$ , with edges encoding the significance of common neighbors of gene pairs in  $P_A$  with respect to  $P_B$ . An *interaction port* is defined as a functionally coherent subset of genes in pathway  $P_A$  that have a (statistically) significant number of shared neighbors in  $P_B$ , which can be equivalently defined as a dense subgraph of  $H_{A \rightarrow B}$ .

We rely on standard statistical definitions – the binomial coefficient,  $C(r, s) = \frac{r!}{s!(r-s)!}$ , is the number of groups of  $s$  items selected from a population of  $r$  items when order does not matter. Hypergeometric distribution, represented by  $HG(k|N, s, t)$ , is the probability of having  $k$  successes in a sequence of  $s$  random draws without replacement, from a size  $N$ , where  $t$  items are classified as successes, and  $N - t$  items are classified as failures. The tail of hypergeometric distribution, denoted by  $HGT(k|N, s, t)$ , is the probability that a random variable  $X \sim HG(k|N, s, t)$  is greater than or equal to  $k$ .

### 2.2. Performance of local methods for predicting functional similarity of gene pairs

Functional similarity prediction of gene pairs is assessed by their co-pathway membership, as determined by the KEGG pathways. Each of the methods considered generates a symmetric matrix that codes predictions for co-pathway membership of genes. In order to evaluate these methods, we rank gene pairs based on the computed similarity matrix and use the *minimum HyperGeometric (mHG)*<sup>19</sup> score to evaluate each ranked-list. This method has been previously used to evaluate the enriched GO terms in a ranked list of genes.<sup>20</sup> Here, we extend this method to assess the significance of top-ranked *gene pairs*, as opposed to genes, that correctly predict co-pathway membership of genes with respect to a given pathway.

Given a ranked list of gene pairs and a cut-off value that partitions this list into *top-ranked* (target set) and *low-ranked* (background set) gene pairs, we are interested in evaluating relative density of functionally similar gene pairs in the target set versus background set. Denote the total number of gene-pairs by  $N = C(n, 2)$ , total number of gene-pairs in a given pathway  $P_A$  by  $N_A$ , and the size of target set as  $l$ . Also, let the random variable  $X$  denote the number of gene pairs in the target set that participate in  $P_A$ .  $X$  has hypergeometric distribution and its p-value can be computed as follows:

$$\begin{aligned} P - value(X = k) &= Prob(k \leq X) = HGT(k|N, N_A, l) \\ &= \sum_{x=k}^{\min(N_A, l)} \frac{C(l, x)C(N - l, N_A - x)}{C(N, N_A)} \end{aligned} \quad (1)$$

Without any prior information about the optimal cutoff, we need to automatically find  $l$  by finding the partition that minimizes the p-value. We encode the co-membership of gene-pairs in  $P_A$  using a binary vector  $\lambda = \lambda_1, \lambda_2, \dots, \lambda_N \in \{0, 1\}^N$ , having  $N_A$  ones and  $N - N_A$  zeros. Here  $\lambda_i$  is 1 if both of the genes in the  $i^{th}$  ranked gene pair are members of pathway  $P_A$ , and 0 otherwise. The minimum HyperGeometric (mHG) score is then defined as:

$$mHG(\lambda) = \min_{1 \leq l \leq N} HGT(b_l(\lambda); N, N_A, l), \quad (2)$$

where  $b_l(\lambda) = \sum_{i=1}^l \lambda_i$ . After correcting  $mHG$  scores for multiple comparison using the Bonferroni method,  $-\log(mHG)$  is used to quantify the performance of different methods in recapturing the co-pathway membership of different KEGG pathways.

### 2.3. Constructing the neighborhood overlap graph (NOG)

The neighborhood overlap graph (NOG) of a pathway  $P_A$  with respect to pathway  $P_B$  is an unweighted, undirected graph defined over the same vertices as  $P_A$ . In this graph, there is an edge between vertices  $v_i$  and  $v_j$  if their shared neighborhood, with respect to  $P_B$ , is statistically significant. Denote the degree of nodes  $v_i$  and  $v_j$  by  $D_i$  and  $D_j$ , respectively. Furthermore, denote the number of neighbors of nodes  $v_i$  and  $v_j$ , and their common neighbors that reside in a pathway  $P_B$  of size  $n_B$  nodes, using  $d_i^B$ ,  $d_j^B$ , and  $k_{ij}^B$ , respectively. Figure 1 illustrates these parameters in a sample case.

Let the random variable  $X$  denote the number of common neighbors of  $v_i$  and  $v_j$  that reside in  $P_B$ , if we were choosing their neighbors from  $P_B$  at random. Let the random variable  $Y_i$  denote the number of node  $v_i$ 's neighbors that reside in  $P_B$  if we were randomly selecting  $D_i$  neighbors of  $v_i$  from KEGG core genes ( $Y_j$  is defined similarly). It can be shown that  $Y_i$  and  $Y_j$  are independent and that all three variables follow the hypergeometric distribution. The significance of the neighborhood structure of  $v_i$  and  $v_j$  with respect to  $P_B$ , called the *modified congruence score (MCS)*, is defined as  $Prob(Y_i = d_i, Y_j = d_j, k_{ij}^B \leq X | n, n_B, D_i, D_j)$ , which can be computed using Bayes rule as follows:

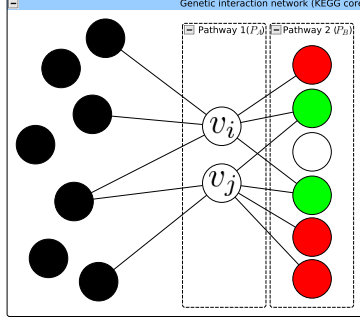


Fig. 1. A sample neighborhood configuration for  $v_i$  and  $v_j$ . Here  $n = 15$ ,  $D_i = 6$ ,  $D_j = 6$ ,  $n_B = 6$ ,  $d_i^B = 3$ ,  $d_j^B = 4$ , and  $k_{ij}^B = 2$

$$\begin{aligned}
MCS_B(i, j) &= \text{Prob}(Y_i = d_i, Y_j = d_j, k_{ij}^B \leq X | n, n_B, D_i, D_j) \\
&= \text{Prob}(k_{ij}^B \leq X | n_B, Y_i = d_i, Y_j = d_j) * \text{Prob}(Y_i = d_i, Y_j = d_j | n, n_B, D_i, D_j) \\
&= \text{HGT}(k_{ij}^B \leq X | n_B, Y_i = d_i, Y_j = d_j) \\
&\quad * \text{HG}(Y_i = d_i | n, n_B, D_i) * \text{HG}(Y_j = d_j | n, n_B, D_j).
\end{aligned} \tag{3}$$

For any given vertex pair  $\langle v_i, v_j \rangle \in V_H * V_H$ , there will be an edge in  $E_H$  if their MCS is less than or equal to a predefined threshold. Algorithm 2.1 computes the neighborhood overlap graph of a given pathway pair  $\langle P_A, P_B \rangle$  and threshold level  $\alpha$ . It may be noted here that under null model of random neighbor selection for gene pairs in  $V_A$ , the NOG follows the Erdos-Renyi random graph model with parameters  $|V_A|$  and  $\alpha$ ,  $ER(|V_A|, \alpha)$ . We use this property later to assess the density significance of each subgraph in NOG, as explained in the Section 2.4.

### Algorithm 2.1.

**Input:**  $P_A = (V_A, E_A)$ : Source pathway,  $P_B = (V_B, E_B)$ : Destination pathway,  $n$ : Network size ( $|V|$ ),  $\alpha$ : Significance threshold

**Output:**  $H = (V_H, E_H)$ : Neighborhood overlap graph

- 1: **Initialization:**  $n_B = |V_B|$ ,  $V_H = V_A$ ,  $E_H = \{\}$
- 2: **for each** vertex pair  $\langle v_i, v_j \rangle \in V_A * V_A$  **do**
- 3:  $D_i = |N(v_i)|$ ,  $D_j = |N(v_j)|$
- 4:  $d_i^B = |N(v_i) \cap V_B|$ ,  $d_j^B = |N(v_j) \cap V_B|$
- 5:  $k_{ij}^B = |N(v_i) \cap N(v_j) \cap V_B|$
- 6:  $MCS_B(i, j) = \text{HGT}(k_{ij}^B | n_B, d_i^B, d_j^B) * \text{HG}(d_i^B | n, n_B, D_i) * \text{HG}(d_j^B | n, n_B, D_j)$ .
- 7: **if**  $MCS_B(i, j) \leq \alpha$  **then**
- 8:  $E_H = E_H \cup \langle i, j \rangle$
- 9: **end if**
- 10: **end for**
- 11: **return**  $H$

## 2.4. Identifying interaction ports and inferring cross-pathway dependencies

If we decompose the neighborhood of gene pair  $(v_i, v_j) \in P_A$  with respect to different pathways, we can identify specific pathways (say  $P_B$ ) that contribute the most to the functional similarity of genes  $v_i$  and  $v_j$ , as predicted by local neighborhood similarity methods. In this case, we say that gene pair  $(v_i, v_j)$  depends on  $P_B$ . Similarly, we can identify a functionally coherent subset of genes in  $P_A$  that depends on  $P_B$ . This subset of genes in  $P_A$  is referred to as an *interaction port* between  $P_A$  and  $P_B$ . As defined in Section 2.3, the neighborhood overlap graph (NOG) represents the significance of the common neighbors of each gene pair  $\langle v_i, v_j \rangle \in P_A$  with respect to pathway  $P_B$ . This suggests that interaction ports can equivalently be defined in terms of dense subgraphs in the NOG.

In order to find interaction port(s), we use the concept of graph cores.<sup>21</sup> The  $k$ -core of a given graph  $G$  is its maximal subgraph (not necessarily connected) such that every vertex is connected to at least  $k$  other vertices in the subgraph. We use iterative peeling to find a cohesive subset of vertices in NOG and record connected components at each core. As mentioned in Section 2.3, the Erdos-Renyi (ER) random graph provides a good null model for comparison with NOG. Using this random graph model, we process connected components of each core separately based on their density and size, as proposed by Koyuturk et al.,<sup>22</sup> and assign p-values to each one of them. During this iterative process, we can find multiple ports and as we go to the inner cores, we find more refined ports with higher densities, but possibly fewer vertices (specialization of ports). We infer that pathway  $P_A$  depends on pathway  $P_B$  if its most significant interaction port has a smaller p-value than a predefined threshold.

## 3. Results

### 3.1. Datasets

#### 3.1.1. Genetic interaction network

We obtain the genome-scale quantitative genetic interactions from a recent SGA experiment in budding yeast, *Saccharomyces cerevisiae*.<sup>13</sup> This dataset consists of 1712 query genes – 1378 null alleles of non-essential genes and 334 hypomorphic or conditional alleles of essential genes (214 temperature-sensitive and 120 DAmP alleles), crossed over 3885 array strains, which spans a total of 5.4 million gene pairs that cover different biological processes. The aggravating genetic interaction network is constructed from the SGA dataset after applying the lenient cutoff.<sup>13</sup> Different hypomorphic alleles of each essential gene are merged together and represented by a single node, such that each node represents a unique gene in the SGA experiment. The final network consists of 363078 aggravating interactions among 4417 genes.

#### 3.1.2. Functional annotations

We downloaded KEGG<sup>23</sup> annotations for budding yeast from the KEGG FTP site on January 4th, 2011. This catalogue organizes genes in the epistatic map into 97 different pathways, ranging from metabolic pathways to genetic information processing pathways. In order to filter out pathways that are too general or too specific, we eliminate pathways with more

than 100 genes (*Metabolic pathways* and *Biosynthesis of secondary metabolites*, which are members of the Global Map) or less than 5 genes (15 pathways). The final dataset consists of 80 pathways, covering 1026 genes in the genetic interaction network.

### 3.1.3. Availability

All datasets and scripts used in this study can be downloaded from <http://www.cs.purdue.edu/homes/mohammas/PSB>.

## 3.2. Similarity of genetic neighborhood as a predictor of functional similarity

We evaluate three commonly used methods for predicting co-pathway membership of genes – sum of length-2 paths (or the number of shared neighbors), congruence score,<sup>15</sup> and the Pearson correlation of interaction profiles of gene pairs. After computing *mHG* score for each of these methods, we apply significance threshold  $\alpha = 0.01$  to identify the enriched pathways. Each pathway is considered enriched, if it is enriched in at least one of the tested methods. Figure 2 illustrates the performance of these methods in identifying co-pathway membership of genes in different KEGG pathways. Enriched pathways are sorted based on the overall performance of all three methods, and are labelled accordingly.

As depicted in the figure, there are only 42 pathways among 80 ( $\sim 50\%$ ) that are enriched. Further analysis of these enriched pathways indicates that most KEGG pathways that are involved in *genetic information processing* are highly enriched, while the bulk of the pathways that are *not* enriched are annotated as metabolic pathways. Among enriched pathways, *ribosome* had zero *mHG* score to machine precision (in spite of normalization before plotting log scores), which indicates high coverage of its member gene pairs among top ranked scores. One aspect of analysis that is not reflected in the figure is the coverage of each method at the point that minimizes *mHG* score, as well as the rank that caused the *mHG* to minimize. A detailed analysis of different methods indicates that congruence score and Pearson correlation methods have better precision with lower recall. This can be observed from the fact that the cut that minimizes *mHG* for these methods usually resides in top 1% of scores, while the same cut for number of shared neighbors method has better coverage of member genes, at the cost of reduced precision.

This observation suggests a specific structure around enriched pathways that boosts their similarity scores and allows us to recapture their co-pathway membership. In order to see how much of this signal comes from other KEGG pathways, we prune all crossing and external interactions in the genetic interaction network and recompute the enrichment scores only using KEGG core genes. Figure 3 illustrates the performance of local methods on the vertex induced subgraph of GIN using KEGG core genes. Most enriched pathways remain highly enriched even if we only use the internal interactions among KEGG pathways to compute functional similarity scores. This internal signal can be interpreted using the between-pathway model of genetic interactions and suggests a bi-cliquish structure around enriched pathways.





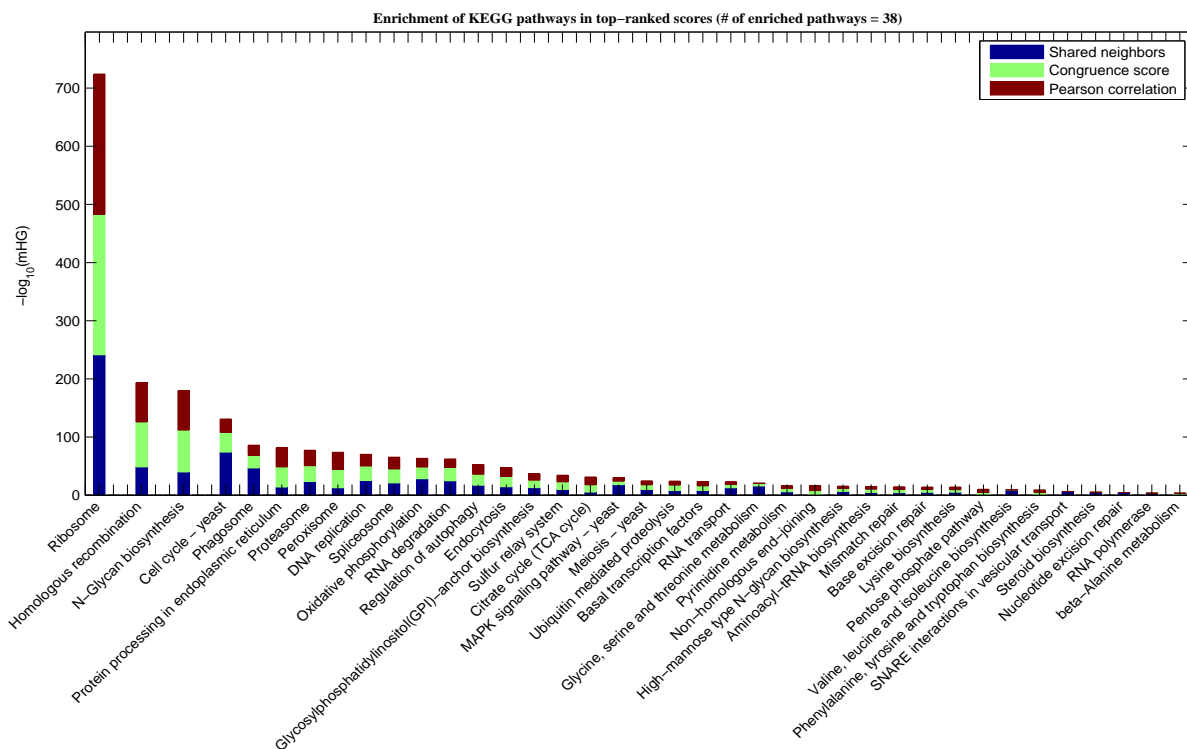


Fig. 3. Performance of local neighborhood methods in predicting co-pathway membership of gene pairs in different KEGG pathways, after pruning the crossing and external interactions.

Port genes are colored in red and are overlaid on the diagram of protein processing in ER (pathway diagram is adopted from KEGG).

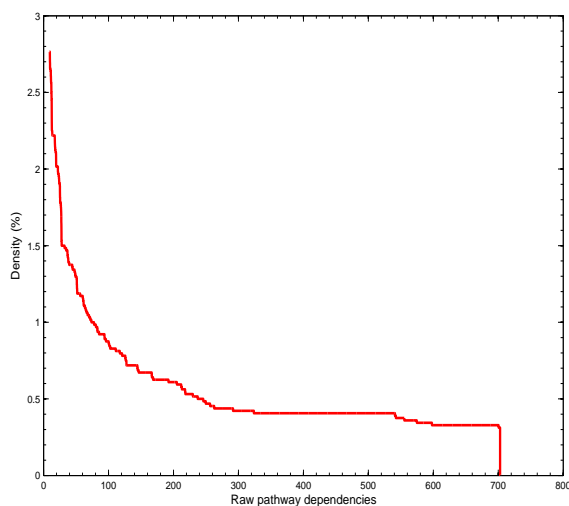


Fig. 4. Density of KEGG crosstalk map as a function of thresholding.

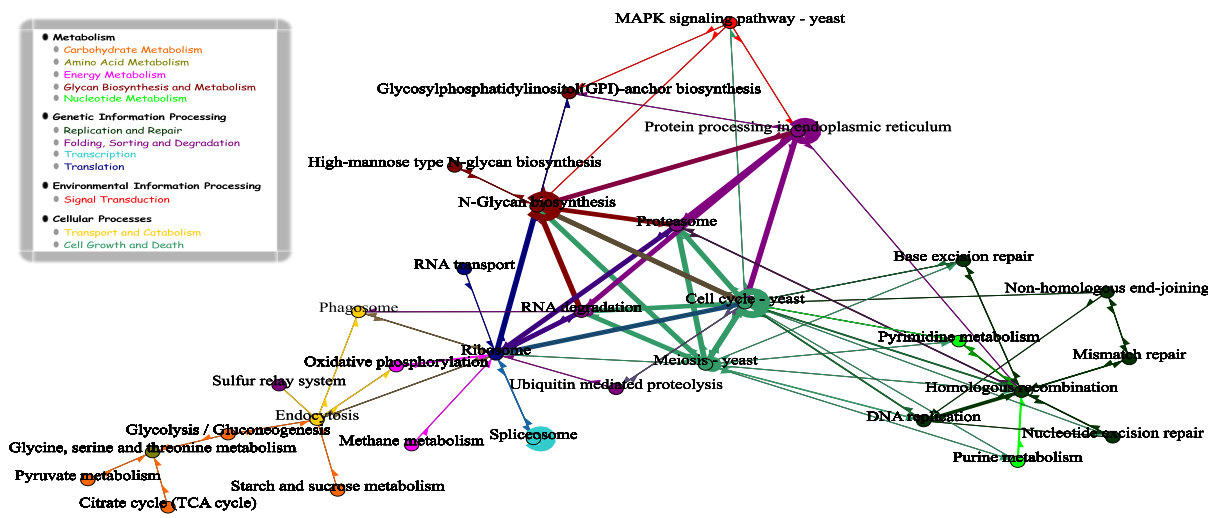


Fig. 5. Computed KEGG crosstalk map.

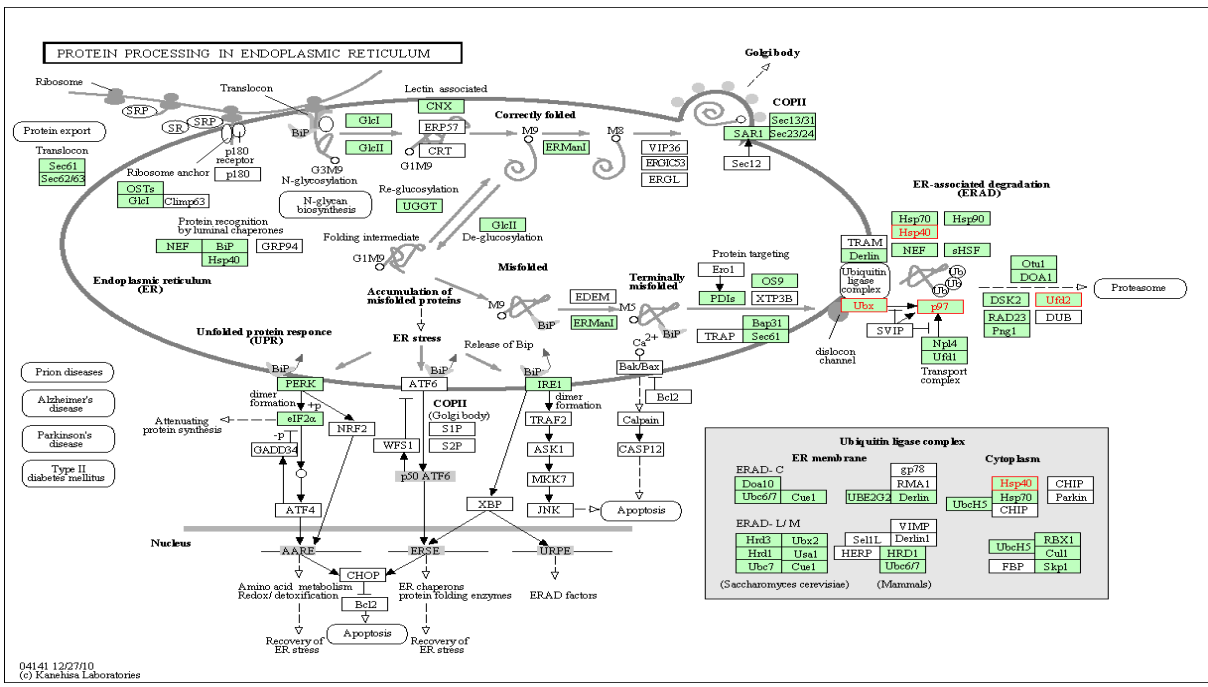


Fig. 6. Protein processing in ER, port genes interacting with proteasome pathway are colored in red (pathway diagram is adopted from KEGG website).

#### 4. Discussion

Figure 5 illustrates a strong internal buffering mechanism in *N-Glycan biosynthesis*, *cell cycle*, *protein processing in endoplasmic reticulum* and *spliceosome* pathways, which can be interpreted using the within-pathway model of genetic interactions. This buffering against genetic perturbations, preserved through evolution, hints at the important role of these pathways in maintaining the functional stability of the cell. Referring to Figure 3, we see that all of these

pathways are highly enriched, thus the internal redundancy of these pathways is a significant source of support for the functional similarity of their member genes.

Further analysis of the dependency map suggests that dependencies among functionally related pathways cluster them into coherent modules. The ribosome pathway resides at the heart of the map, having dependencies with both pre- and post-translational pathways. The former machinery can protect the cell against defective RNA constructed during transcription, through RNA degradation, and mis-spliced pre-mRNA constructed through Spliceosome. The latter mechanism regulates the concentration of specific proteins and degrades misfolded proteins by tagging them using ubiquitin mediated proteolysis and targets them for unfolding in proteasomes. This machinery is also known to be essential for many other cellular processes, including the cell cycle, the regulation of gene expression, and responses to oxidative stress. Perturbation of the oxidative phosphorylation pathway can also endanger cell integrity, since RNA molecules are known to be particularly vulnerable to oxidative stress.<sup>24</sup> The varied dependency of ribosome on a number of different pathways can explain its exceptionally high enrichment score, as measured with mHG and shown in Figures 2 and 3.

Most of the synthesized proteins in ribosome can fold themselves spontaneously into their active conformation. However, some of them need further post-translational modification, such as glycosylation, to aid their folding. The endoplasmic reticulum (ER) serves as the gateway for modification of newly synthesised proteins with the help of luminal chaperones. Terminally mis-folded proteins are degraded through the proteasome using a process called ER-associated degradation (ERAD). The high-level organization of the protein processing in ER is shown in Figure 6, with genes identified as the interaction port marked in red. It can be observed that the ERAD sub-function is faithfully recovered as the interaction port between protein processing in ER and the proteasome pathway. These functional dependencies can be also observed in the highly significant triangular dependencies between protein processing in ER, proteasome, and N-glycan biosynthesis in Figure 5.

Another observation relates to the densely connected module of DNA repair, clustered in bottom right part of the map. This module consists of the DNA replication pathway, together with both single strand repair pathways; including the base excision repair, mismatch repair, and nucleotide excision repair pathways; and double-strand repair pathways, namely the non-homologous end-joining and the homologous recombination pathways. The Purine and Pyrimidine metabolic pathways, both of which help in metabolising nucleotide, are also parts of this module. Referring back to the result of functional similarity prediction methods, we note that all of these pathways are highly enriched.

In this study, we demonstrated that synthetic genetic interactions provide powerful means for predicting functional similarity of gene pairs. However, their scope of effectiveness is limited to a specific subset of pathways. This subset includes most pathways involved in genetic information processing and cellular processes, while missing many of the metabolic pathways. To further study the specific structure around these pathways that boosts their functional similarity, we decomposed the neighborhood around each pathway and identified the dependencies among them, inferred from the observed topological patterns. We discovered that the strength of these dependencies, manifested as the significance of the interaction ports between

pathways, is highly correlated with the performance of functional prediction methods, and can be used to interpret their effective scope. In addition, observed dependencies reflect our previous understanding of higher level organization of pathways, which suggests that the KEGG crosstalk map by itself can be used to uncover the cellular machinery.

## 5. Acknowledgments

This work supported by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-0939370, and by NSF grants DBI 0835677 and 0641037.

## References

1. G. Giaever *et al.*, *Nature* **418**, 387 (July 2002).
2. A. Wagner, *Nature genetics* **24**, 355 (April 2000).
3. Z. Gu *et al.*, *Nature* **421**, 63 (January 2003).
4. K. Hannay, E. M. Marcotte and C. Vogel, *BMC genomics* **9**, p. 609 (December 2008).
5. A. Wagner, *BioEssays : news and reviews in molecular, cellular and developmental biology* **27**, 176 (February 2005).
6. Y. Qi *et al.*, *Genome Research* **18**, 1991 (December 2008).
7. S. J. Dixon *et al.*, *Annual review of genetics* **43**, 601 (August 2009).
8. X. Pan *et al.*, *Molecular cell* **16**, 487 (November 2004).
9. S. L. Ooi, D. D. Shoemaker and J. D. Boeke, *Nat Genet* **35**, 277 (October 2003).
10. A. H. Tong *et al.*, *Science* **294**, 2364 (December 2001).
11. A. H. Tong *et al.*, *Science (New York, N.Y.)* **303**, 808 (February 2004).
12. M. Schuldiner *et al.*, *Cell* **123**, 507 (November 2005).
13. M. Costanzo *et al.*, *Science* **327**, 425 (January 2010).
14. C. L. Tucker and S. Fields, *Nature Genetics* **35**, 204 (November 2003).
15. P. Ye *et al.*, *Molecular Systems Biology* **1**, msb4100034 (November 2005).
16. D. Segrè, A. Deluna, G. M. Church and R. Kishony, *Nature genetics* **37**, 77 (January 2005).
17. S. Bandyopadhyay, R. Kelley, N. J. Krogan and T. Ideker, *PLoS Comput Biol* **4**, p. e1000065 (April 2008).
18. L. V. Zhang, O. D. King, S. L. Wong, D. S. Goldberg, A. H. Tong, G. Lesage, B. Andrews, H. Bussey, C. Boone and F. P. Roth, *Journal of biology* **4**, 6+ (2005).
19. E. Eden, D. Lipson, S. Yogev and Z. Yakhini, *PLoS Comput Biol* **3**, p. e39 (March 2007).
20. E. Eden *et al.*, *BMC Bioinformatics* **10** (Feb 2009).
21. S. B. Seidman, *Social Networks* **5**, 269 (1983).
22. M. Koyutürk, W. Szpankowski and A. Grama, *Journal of Computational Biology* **14**, 747 (August 2007).
23. M. Kanehisa and S. Goto, *Nucl. Acids Res.* **28**, 27 (January 2000).
24. Q. Kong and C. L. Lin, *Cell Mol Life Sci* **67**, 1817 (June 2010).