Title
Systematic identification of functional orthologs based on protein network comparison.

Permalink
https://escholarship.org/uc/item/4tm7d73p

Journal
Genome Research, 16(3)

ISSN
1088-9051

Authors
Bandyopadhyay, Sourav
Sharan, Roded
Ideker, Trey

Publication Date
2006-03-01

Peer reviewed
Systematic identification of functional orthologs based on protein network comparison

Sourav Bandyopadhyay, Roded Sharan and Trey Ideker

Genome Res. published online Jan 27, 2006; doi:10.1101/gr.4526003

Published online January 27, 2006 in advance of the print journal.

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

Notes

Online First contains unedited articles in manuscript form that 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). Online First articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Online First articles must include the digital object identifier (DOIs) and date of initial publication.

To subscribe to Genome Research go to: http://www.genome.org/subscriptions/

© 2006 Cold Spring Harbor Laboratory Press
Systematic identification of functional orthologs based on protein network comparison

Sourav Bandyopadhyay,1,2 Roded Sharan,3,4 and Trey Ideker1,2,4

1Program in Bioinformatics, 2Department of Bioengineering, University of California at San Diego, La Jolla, California 92093, USA; 3School of Computer Science, Tel-Aviv University, Tel-Aviv 69978, Israel

Annotating protein function across species is an important task that is often complicated by the presence of large paralogous gene families. Here, we report a novel strategy for identifying functionally related proteins that supplements sequence-based comparisons with information on conserved protein–protein interactions. First, the protein interaction networks of two species are aligned by assigning proteins to sequence homology clusters using the Inparanoid algorithm. Next, probabilistic inference is performed on the aligned networks to identify pairs of proteins, one from each species, that are likely to retain the same function based on conservation of their interacting partners. Applying this method to Drosophila melanogaster and Saccharomyces cerevisiae, we analyze 121 cases for which functional orthology assignment is ambiguous when sequence similarity is used alone. In 61 of these cases, the network supports a different protein pair than that favored by sequence comparisons. These results suggest that network analysis can be used to provide a key source of information for refining sequence-based homology searches.

[Supplemental material is available online at www.genome.org and http://www.cellcircuits.org/Bandyopadhyay2006/.]
ancestry). The remaining 692 clusters contained multiple pro-
thologs we take to be functionally equivalent because of direct
represent unambiguous or
tained only a single yeast/fly protein pair and were assumed to
yeast and 3881 fly proteins overall. Of these, 1552 clusters con-
clusters. A total of 2244 clusters were generated, covering 2834

Motivation: Interaction conservation is related to orthology

Protein–protein interaction networks for yeast and fly were ob-
tained from the Database of Interacting Proteins (December 2004
download) ( Xenarios et al. 2002 ). These contained 14,319 inter-
actions among 4389 proteins in yeast and 20,720 interactions
among 7038 proteins in fly. First, we applied the Inparanoid
( Remm et al. 2001 ) algorithm to the complete sets of proteins
from Saccharomyces cerevisiae and the fruit fly Drosophila melan-
gaster. Not only are these species among the most important
model eukaryotes, they are also associated with the largest
numbers of experimentally measured protein interactions to
date.

Results

Network-based identification of functional orthologs

To capture these trends to identify functional orthologs, we for-
mulated a procedure to estimate the likelihood of functional or-
thology for each ambiguous functional ortholog given its con-
servation index. By this method, the probability of functional

![Figure 1](https://www.genome.org)

| Figure 1. | (A) Network neighborhood conservation for definite orthologs vs. other yeast/fly protein pairs. The distribution of the conservation index \( c \) is shown for definite functional orthologs (sole members of an Inparanoid cluster), ambiguous functional orthologs (in a cluster with multiple members), homology (different clusters but similar sequences), and random protein pairs. Definite functional orthologs show a shift toward higher conservation of protein interactions between the yeast and fly protein networks. Mean \( c = 0.1512, 0.1171, 0.0870, 0.0615 \) for definite functional orthologs, ambiguous functional orthologs, homologs, and random pairs, respectively. (B) Logistic function relating conservation index to probability of functional orthology. Logistic regression was performed by using the “definite functional ortholog” and “homolog” pairs as positive vs. negative training data, respectively. The resulting function is shown. |
orthology for a pair of proteins is influenced by the probabilities of functional orthology for their network neighbors, which in turn depend on their network neighbors, and so on. This type of probabilistic model is known as a Markov random field (Besag 1974). Exact inference in this model is not tractable because of the complex interrelationships between network nodes. Rather, approaches such as clustering, conditioning, and stochastic simulation have been used to derive estimates for the posterior probabilities of node properties. Here, we implemented a method based on Gibbs sampling for its computational tractability and accuracy in densely connected networks (Pearl 1988). An overview of the approach is given in Figure 2, with full details provided in the Methods.

Application to yeast and fly identifies new putative functionally orthologous pairs

We applied this approach to resolve ambiguous functional orthology relationships in the yeast and fly protein networks. Of the 692 ambiguous Inparanoid clusters, 121 contained protein pairs for which at least one pair had conserved interactions between networks. Application of our Gibbs sampling procedure yielded estimates of probability of functional orthology for each protein pair in these 121 ambiguous clusters. In 60 of these clusters, the highest probability was assigned to the protein pair that was also the most sequence-similar via BLAST. These cases reinforced the intuition that the best sequence matches are also the most functionally similar. The remaining 61 clusters showed the opposite behavior; that is, the highest probability pair was not the most sequence similar pair. Of these 61 cases, 15 were supported by two or more conserved interactions (Table 1). Because the yeast and fly networks are incomplete (i.e., they contain false negatives), in some of these cases we cannot rule out the possibility that conserved interactions with the best BLAST matches have been missed (see Discussion). A complete listing of the results can be found on the Supplemental Web site (http://www.cellcircuits.org/Bandyopadhyay2006/).

Validation

A straightforward validation of the approach would be to analyze its accuracy in recapitulating a gold standard set of protein functional annotations. However, databases of functional annotations are based directly on sequence similarity, and they typically lack the specificity to discriminate among subtle functional differences across large gene families. As an alternative, we used the technique of cross-validation to test the ability of the approach to reclassify protein pairs in the definite functional ortholog set (positive test data) versus the nonorthologous homolog set (negative test data). In each cross-validation trial, 1% of these assignments were hidden (declassified) and monitored during Gibbs sampling to obtain probabilities of functional orthology for positive and negative examples. Reclassification was judged successful if the probability of functional orthology exceeded a particular cutoff value. These statistics were compiled over 100 trials. Figure 3A charts cross-validation performance over a range of probability cutoffs. At a probability cutoff of 0.5, we observed a 50% true-positive rate and a 15% false-positive rate. This shows marked improvement over a random predictor, where we would expect to see the same true-positive rate as false-positive rate.

Declassifying 1% of the known functional orthologous and nonorthologous pairs reduces the amount of information available to the algorithm and, thus, can reduce its predictive ability. To assess the severity of this effect, we repeated the cross-validation analysis at varying percentages of declassification of positive and negative data (ranging from 1%–100%) (Fig. 3B). For instance, changing the amount of declassification of available training data from 1% to 25% reduced the maximum precision from 83% to 75%. Further declassification yielded more marked reductions in precision and recall.
Discussion

Specific examples of yeast/fly functional orthologs resolved by the network-based approach are shown graphically in Figure 4. In Figure 4A, yeast transportin (Kap104) is orthologous to both Trn and CG8219 in fly with highly significant sequence homology (BLAST E-values $9 \times 10^{-128}$ and $7 \times 10^{-96}$, respectively). Transportin is a member of a complex responsible for the nuclear import of mRNA binding proteins and is known to be highly conserved among diverse organisms (Aitchison et al. 1996). D. melanogaster Trn was identified by using sequence homology based on human transportin1 (Siomi et al. 1998). Both Trn and CG8219 in fly interact with orthologs of members of the Kap104-associated complex in yeast (Ho et al. 2002) suggesting that both of these fly proteins may participate in the functionally similar complex in fly. Our analysis suggests that CG8219 retains more of the original functions of Kap104 (probability of functional orthology of 47% vs. 41% for the Kap104/Trn pairing) due to their conserved interactions with members of the spliceosome complex (Yju2, Ssa1, and Ssa2 in yeast). This result is a case in which the most sequence-similar protein does not appear to be the most functionally related protein in an orthologous cluster given the current network data.

The cluster in Figure 4C contains two alternative catalytic subunits of the protein phosphatase type 2A family (yeast Pph21 and Pph22). Both alternatives interact with a member of the Rts1 subunit (Rts1) and have high sequence similarity to the fly Mts protein (75% identity for Pph21, 76% for Pph22). Since Pph21 and Pph22 are at least partially redundant (disruption of both genes in combination is synthetic lethal) (Ronne et al. 1991), it appears that the array of interactions carried out by Mts is conserved across the two yeast orthologs. Nonetheless, based on the available protein interaction data, Pph22 alone has conserved interactions with the proteasome (Pre2/Pros and Pre4/CG12000), which has been shown to be important for the role of the Pph21/22 complex in degradation of Sve1p (Yang et al. 2000).

As a final example, Figure 4D shows evidence that the yeast Calmodulin (Cmd1) protein is functionally orthologous to fly Androcam (And) rather than to the more sequence-similar fly Calmodulin (Cam1; 60% identity vs. 51% for And). The existence of many conserved interactions for the Cmd1/And pair, compared with only one for Cmd1/Cam1, does not appear to be a result of incomplete coverage: Cmd1 has a total of 61 interactions in the yeast network, and Cam1 and And have 19 and 26 interactions, respectively, in the fly network (most of these do

### Table 1. Inparanoid clusters for which the network suggests different functional pairings than BLAST

<table>
<thead>
<tr>
<th>Inparanoid cluster</th>
<th>Yeast/fly pairings in cluster</th>
<th>Total protein interactions in yeast/fly</th>
<th>BLAST E-value</th>
<th>No. of conserved interactions</th>
<th>P (z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>Ssa3/Hsc70-4</td>
<td>3/29</td>
<td>1E-277</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Ssa2/Hsc70-4</td>
<td>10/29</td>
<td>7E-275</td>
<td>4</td>
<td>53.22%</td>
</tr>
<tr>
<td></td>
<td>Ssa1/Hsc70-4</td>
<td>13/29</td>
<td>2E-275</td>
<td>4</td>
<td>48.10%</td>
</tr>
<tr>
<td>94</td>
<td>Act1/Act5c</td>
<td>38/48</td>
<td>9E-201</td>
<td>3</td>
<td>39.56%</td>
</tr>
<tr>
<td></td>
<td>Act1/Act42a</td>
<td>38/3</td>
<td>3E-200</td>
<td>1</td>
<td>39.24%</td>
</tr>
<tr>
<td></td>
<td>Act1/Act87e</td>
<td>38/11</td>
<td>1E-199</td>
<td>3</td>
<td>43.53%</td>
</tr>
<tr>
<td></td>
<td>Act1/CG10067</td>
<td>38/9</td>
<td>1E-198</td>
<td>2</td>
<td>38.20%</td>
</tr>
<tr>
<td></td>
<td>Act1/Act88f</td>
<td>38/2</td>
<td>9E-198</td>
<td>2</td>
<td>40.17%</td>
</tr>
<tr>
<td>126</td>
<td>Vph1/CG7678</td>
<td>12/0</td>
<td>2E-174</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Vph1/CG18617</td>
<td>12/13</td>
<td>3E-170</td>
<td>2</td>
<td>41.87%</td>
</tr>
<tr>
<td></td>
<td>Stv1/CG18617</td>
<td>11/13</td>
<td>1E-148</td>
<td>1</td>
<td>38.44%</td>
</tr>
<tr>
<td>246</td>
<td>Kap104/Trn</td>
<td>47/7</td>
<td>9E-128</td>
<td>2</td>
<td>40.64%</td>
</tr>
<tr>
<td></td>
<td>Kap104/CG8219</td>
<td>47/20</td>
<td>7E-96</td>
<td>5</td>
<td>46.78%</td>
</tr>
<tr>
<td>376</td>
<td>Pda1/CG7024</td>
<td>8/1</td>
<td>9E-101</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Pda1/L(T)g0334</td>
<td>8/12</td>
<td>6E-99</td>
<td>2</td>
<td>57.90%</td>
</tr>
<tr>
<td>425</td>
<td>Gpa1/G-isa65a</td>
<td>14/2</td>
<td>1E-90</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Gpa1/G-oa47a</td>
<td>14/12</td>
<td>5E-67</td>
<td>2</td>
<td>41.53%</td>
</tr>
<tr>
<td>707</td>
<td>Rp112/Rp112</td>
<td>0/11</td>
<td>2E-63</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Rp112a/Rp112</td>
<td>6/11</td>
<td>2E-63</td>
<td>2</td>
<td>48.39%</td>
</tr>
<tr>
<td>917</td>
<td>Cmd1/Cam</td>
<td>61/19</td>
<td>1E-49</td>
<td>1</td>
<td>35.90%</td>
</tr>
<tr>
<td></td>
<td>Cmd1/And</td>
<td>61/26</td>
<td>4E-40</td>
<td>6</td>
<td>44.39%</td>
</tr>
<tr>
<td>1236</td>
<td>Fkh2/CG11799</td>
<td>5/14</td>
<td>4E-31</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Fkh1/CG11799</td>
<td>29/14</td>
<td>3E-18</td>
<td>2</td>
<td>42.34%</td>
</tr>
<tr>
<td>1550</td>
<td>Kelz/CG12081</td>
<td>3/10</td>
<td>3E-19</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kelz/CG12081</td>
<td>16/16</td>
<td>1E-17</td>
<td>2</td>
<td>45.41%</td>
</tr>
<tr>
<td>1562</td>
<td>Egd1/Bcd</td>
<td>3/16</td>
<td>2E-19</td>
<td>1</td>
<td>47.19%</td>
</tr>
<tr>
<td></td>
<td>Btt1/Bcd</td>
<td>3/16</td>
<td>2E-13</td>
<td>1</td>
<td>40.86%</td>
</tr>
<tr>
<td></td>
<td>Btt1/CG11835</td>
<td>3/2</td>
<td>2E-09</td>
<td>2</td>
<td>70.50%</td>
</tr>
<tr>
<td>1643</td>
<td>Ngr1/CG12478</td>
<td>1/1</td>
<td>6E-16</td>
<td>0</td>
<td>45.06%</td>
</tr>
<tr>
<td></td>
<td>Nam8/Aret</td>
<td>22/10</td>
<td>7E-06</td>
<td>2</td>
<td>45.06%</td>
</tr>
<tr>
<td>1687</td>
<td>Tpm2/Tm1</td>
<td>1/7</td>
<td>3E-15</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Tpm1/Tm2</td>
<td>3/17</td>
<td>2E-14</td>
<td>2</td>
<td>43.98%</td>
</tr>
<tr>
<td>1740</td>
<td>Mig2/Vpa</td>
<td>0/31</td>
<td>5E-13</td>
<td>0</td>
<td>40.42%</td>
</tr>
<tr>
<td></td>
<td>Mig3/Vpa</td>
<td>2/31</td>
<td>1E-09</td>
<td>2</td>
<td>40.42%</td>
</tr>
<tr>
<td>2037</td>
<td>Gld8/CG18467</td>
<td>3/0</td>
<td>8E-03</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Gld8/CG6617</td>
<td>3/8</td>
<td>0.001</td>
<td>2</td>
<td>76.51%</td>
</tr>
</tbody>
</table>

*aFor brevity, only pairings with conserved interactions, or with the best BLAST E-value, are shown.*
not appear in Fig. 4 because the network alignment only shows interactions that are conserved). Furthermore, multiple sequence alignment and phylogenetic analysis of these genes over a larger number of organisms, including worms and mammals, indicates a closer phylogenetic relationship for yeast Cmd1 and fly And, supporting our hypothesis that they are the true functional orthologs (Supplemental Fig. 1). This apparent discrepancy between functional and sequence similarity is probably a result of the large amount of sequence variability among the calmodulin family of proteins (Tombes et al. 2003) and would have been difficult to probe without protein network information.

In future work, it is possible that incorporating yet other types of conserved linkages, such as transcriptional interactions (Harbison et al. 2004), synthetic-lethal interactions (Guarente 1993), and coexpression relations (Stuart et al. 2003) will allow for a more complete and multifaceted view of protein function. The proposed method would also benefit from a more accurate understanding of network evolution. At the core of our approach is a model for measuring the divergence of orthologous proteins by means of a network “conservation index.” It encapsulates the notion that shorter evolutionary distances correspond to greater relative numbers of conserved interactions. However, a more sophisticated metric might represent explicit evolutionary mechanisms, such as formation of new interactions through gene mutation or duplication (Wagner 2003). It should also be noted that comparative methods rely on the conservation of function between evolutionarily related proteins, and that this functional similarity may be lost among orthologs due to large evolutionary distance; thus, network-based methods that search for the absence of a functional ortholog may also be useful. Finally, further work is needed to analyze the impact of data quality, that is, numbers of false-positive and false-negative interactions (Sprinzak et al. 2003). False positives are largely mitigated by the focus on only those interactions that are conserved across species, because spurious interactions are typically not reproducible (Sharan et al. 2005). False negatives are a larger concern, because they might cause a functionally orthologous pair to be wrongly rejected due to lack of conserved interactions. Certainly, a preponderance of conserved interactions for one particular pair of proteins versus others provides evidence that these proteins are indeed functional orthologs. Although the expected number of false-negative interactions will decrease with forthcoming interaction data sets, future approaches may explicitly incorporate the false-negative rate into the probabilistic model.

In summary, we have presented an algorithm that uses protein interaction measurements to achieve more specific discrimination of functional orthologs than is possible with sequence-based methods alone. It is built on the concept that conserved proteins typically do not function independently but rely on interactions with other proteins to form conserved pathways, and that the specific patterns of conservation of these pathways are informative for determining which cross-species protein pairs have similar functional roles. As these methods mature and as ever greater numbers of protein interactions become available across species, comparative network analysis is likely to play an increasingly central role as a bridge among protein sequence, evolution, and function.

Methods
Inparanoid clusters generation
The complete sets of 5878 yeast and 18,746 fly protein sequences were downloaded from the Saccharomyces Genome Database (Christie et al. 2004) and Flybase (Drysdale et al. 2005), respectively. Protein sequences for both species were grouped together into orthology clusters by using the Inparanoid algorithm (Remm et al. 2001) with default parameters (overlap threshold = 0.5; confidence = 0.05). Inparanoid optionally allows a third genome to be used as an outgroup, which can detect missing sequences and thus improve ortholog detection. However, use of Escherichia coli as an outgroup had a negligible impact on our analysis.

Network alignment
A global network alignment between yeast and fly was constructed as described in Kelley et al. (2004), with the difference that Inparanoid clusters were used instead of BLAST E-values for pairing proteins between the two networks. Briefly, the network alignment is represented as a graph of nodes and links (Fig. 2b). Each node denotes a pair of putatively orthologous proteins $a$ and $a'$. Each link between a pair of nodes $a/a'$ and $b/b'$ denotes a conserved protein interaction, that is, an interaction observed for both $(a, b)$ and $(a', b')$. To tolerate a certain amount of missing interaction data, “indirect” links are also defined if a pair of pro-
We define the conservation index \( c(i) \) of node \( i \) (representing protein pair \( a/a' \)) as twice its number of strongly conserved interactions divided by its total number of interactions over both species:

\[
c(i) = \frac{2d(i)}{d(a) + d(a')}
\]

where \( d(i) \) denotes the number of strongly conserved links involving node \( i \), while \( d(a) \) and \( d(a') \) denote the degrees (numbers of interactions) of proteins \( a \) and \( a' \) in their respective single-species networks.

**Probabilistic model**

We model the orthology relations for two species by using a Markov random field (Besag 1974). This model is specified by an undirected graph \( G = (V, E) \) corresponding to a network alignment, and conditional probability distributions that relate the event that a given node represents a functionally orthologous pair with those events for its neighbors. A Markov random field model is specified in terms of potential functions on the cliques in the graph

\[
P(z) = \frac{1}{Z} \exp(-U(z))
\]

where \( z \) is some assignment to the states of all nodes in the graph, \( U(z) \) is an "energy" function that integrates the potentials over all cliques in the graph, and \( Z \) is a normalizing constant. It is not necessary to compute the normalization constant, since all that is required are the conditional probabilities for each node given its neighbors (rather than the joint distribution). For computational efficiency, we used the common auto-logistic model (Besag 1974), which assigns zero potential to cliques of size greater than two. Under this model, the energy takes the form

\[
U(z) = \sum_i \alpha_i z_i - \sum_{(i,j) \in E} \beta_{ij} z_i z_j
\]

which, when substituted into the equation for \( P(z) \) above, reduces to a logistic function. Based on our initial observation that the functional orthology of a node is a function of its conservation index (well approximated by a logistic function) (see Fig. 1A; Results), we set \( \alpha_i = \alpha \) and \( \beta_{ij} = \beta_{i} = 2\beta/|d(a_i) + d(a'_j)| \) to obtain the following:

\[
P(z_i | z_{N(i)}) = \frac{1}{1 + \exp[-\alpha_i - \sum_{j \in N(i)} \beta_{ij} z_j]} = \frac{1}{1 + \exp[-\alpha_i - \beta c(i)]}
\]

where \( N(i) \) is the set of neighbors of node \( i \), and \( z_{N(i)} \) denotes the set of all \( z_j \) such that \( j \in N(i) \). Note that \( \alpha_i \) and \( \beta_{ij} \) could be set to accommodate other equations for conservation index, as long as they are linear in the number of strongly conserved neighbors \( d(i) \).

**Fitting the logistic function**

To provide a set of training data for fitting the parameters \( \alpha \) and \( \beta \) of the logistic function, 100 of the 212 definite functional orthologs having at least one conserved interaction were randomly selected as positive examples, and their states were set to \( z = 1 \). Negative examples of "nonorthologs" were generated by randomly selecting 100 yeast proteins and pairing each with its best BLAST matching fly protein not in the same cluster; their states were set to \( z = 0 \) (ideally, the negative training data would consist of orthologs that are not functional orthologs, but few
such examples exist). Parameters $\alpha$ and $\beta$ were optimized by maximizing the product of $P(z_i | z_{\text{neg}})$ over all positive and $[1 - P(z_i | z_{\text{neg}})]$ over all negative training data using the method of conjugate gradients (Press 1992). The optimal logistic function is shown in Figure 1B. Note that the equal numbers of positive and negative training data assume a prior probability of 0.5 of observing a true functional ortholog. Although the actual prior is unknown and may differ from this value, $P(z_i | z_{\text{neg}})$ remains monotonically related to the true probability of functional orthology.

Orthology inference

We used the above model to estimate the final posterior probabilities ($P(z_i)$) by using the method of Gibbs sampling (Smith and Roberts 1993). In this approach, nodes representing ambiguous functional orthologs are each assigned a temporary state, $z_i$ of zero or one, initially at random. At each iteration, a node $i$ is sampled (with replacement) and its value of $z_i$ is updated given the states of its neighbors, $z_{\text{neg}}$. The new value of $z_i$ is set to one with probability $P(z_i | z_{\text{neg}})$, else zero. Over all iterations, the nodes designated as definite functional orthologs and “nonorthologs” are forced to states of one and zero, respectively. This process is illustrated in Figure 2, c and d.

The Gibbs sampling procedure was carried out for an initial period of $2 \times 10^7$ “burn-in” iterations. From this point onward, $2 \times 10^7$ additional iterations were performed and statistics computed on the fraction of iterations in which each node acquires a “functionally orthologous” $z = 1$ state. The final probabilities of functional orthology for each node, $P(z_i)$, were estimated as this fraction. The above numbers of iterations were chosen to ensure that results were stable across multiple runs of the algorithm and mean probabilities reported.

Acknowledgments

We thank Tomer Shlomi and Ryan Kelley for critical reading of the manuscript, as well as Silpa Suthram for providing interaction data. We gratefully acknowledge the following sources of funding for this project: award R01-GM070743-01 from the NIGMS (T.I.), an Alon fellowship (R.S.), and a Quantitative Systems Biology grant from the NSF (S.B.). T.I. is a fellow of the David and Lucille Packard Foundation.

References


Received August 3, 2005; accepted in revised form November 21, 2005.