

10. Y. Gu *et al.*, *Mol. Cell Biol.* **22**, 7645 (2002).
11. A. J. Ridley, H. F. Paterson, C. L. Johnston, D. Diekmann, A. Hall, *Cell* **70**, 401 (1992).
12. C. D. Nobes, A. Hall, *Cell* **81**, 53 (1995).
13. L. Van Aelst, C. D'Souza-Schorey, *Genes Dev.* **11**, 2295 (1997).
14. S. Etienne-Manneville, A. Hall, *Nature* **420**, 629 (2002).
15. K. Sugihara *et al.*, *Oncogene* **17**, 3427 (1998).
16. Materials and methods are available as supporting material on Science Online.
17. Rac3 was expressed by reverse transcriptase PCR in both primitive and differentiated hematopoietic cells.
18. R. Palacios, C. Bucana, X. Xie, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5247 (1996).
19. T. M. Dexter, T. D. Allen, L. G. Lajtha, *J. Cell. Physiol.* **91**, 335 (1976).
20. Mean fluorescence intensity (MFI) (mean \pm SD) for CD49d (a4 β 1): wild type, 122 \pm 15; *Rac1*^{-/-}, 117 \pm 19; *Rac2*^{-/-}, 111 \pm 11; *Rac1*^{-/-}*Rac2*^{-/-}, 107 \pm 7. MFI (mean \pm SD) for CD49e (a5 β 1): wild type, 125 \pm 12; *Rac1*^{-/-}, 120 \pm 8; *Rac2*^{-/-}, 133 \pm 16; *Rac1*^{-/-}*Rac2*^{-/-}, 108 \pm 9.
21. A. Peled *et al.*, *Science* **283**, 845 (1999).
22. Considerable F-actin polymerization at the cell periphery was seen in wild-type and *Rac1*^{-/-} HSC/PS stimulated with 100 ng/ml of SDF-1. In contrast, F-actin polymerization at the cell periphery was not detectable in *Rac2*^{-/-} and *Rac1*^{-/-}*Rac2*^{-/-} HSC/PS. Indeed, in the absence of both GTPases, no obvious F-actin was seen on immunofluorescence images, and the cytoplasm of these cells appeared to collapse toward the cell nucleus.
23. The well-defined rim of cortical actin present at the cell periphery of wild-type cells completely disappeared in *Rac2*^{-/-} cells but remained clearly visible in *Rac1*^{-/-} cells. In *Rac1*^{-/-}*Rac2*^{-/-} neutrophils, the F-actin structure appeared to collapse into a dense network immediately surrounding the nucleus, showing a complete loss of structural organization consistent with a lack of cell spreading (Fig. 4B). Whereas previous reports demonstrate an important role for Rac1 in F-actin assembly in many cells, most of these studies were done in nonhematopoietic cells, which do not express Rac2.
24. M. Glogauer *et al.*, *J. Immunol.* **170**, 5652 (2003).
25. A. Abo *et al.*, *Nature* **353**, 668 (1991).
26. U. G. Knaus, P. G. Heyworth, T. Evans, J. T. Curnutte, G. M. Bokoch, *Science* **254**, 1512 (1991).
27. T. Papayannopoulou, C. Craddock, *Acta Haematol.* **97**, 97 (1997).
28. D. Michaelson *et al.*, *J. Cell Biol.* **152**, 111 (2001).
29. J. C. van der Loo *et al.*, *J. Clin. Invest.* **102**, 1051 (1998).
30. H. Gu, J. D. Marth, P. C. Orban, H. Mossman, K. Rajewsky, *Science* **265**, 103 (1994).
31. Openlab software, Improvion, Andrew Waterfall, Lexington, MA.
32. We thank J. Towse for technical assistance, Y. Zheng for reviewing the manuscript, M. Hodgson and K. Steward for administrative assistance, D. Marmer and the Flow Core at Cincinnati Children's Hospital Medical Center for assistance with cell sorting, and Argen and Takara Bio for reagents. Supported by NIH DK62757 (D.A.W.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/302/5644/445/DC1

Materials and Methods

Figs. S1 to S16

References

Movies S1 and S2

26 June 2003; accepted 26 August 2003

A Bayesian Networks Approach for Predicting Protein-Protein Interactions from Genomic Data

Ronald Jansen,^{1*} Haiyuan Yu,¹ Dov Greenbaum,¹ Yuval Kluger,¹ Nevan J. Krogan,⁴ Sambath Chung,^{1,2} Andrew Emili,⁴ Michael Snyder,² Jack F. Greenblatt,⁴ Mark Gerstein^{1,3,†}

We have developed an approach using Bayesian networks to predict protein-protein interactions genome-wide in yeast. Our method naturally weights and combines into reliable predictions genomic features only weakly associated with interaction (e.g., messenger RNA coexpression, coessentiality, and colocalization). In addition to de novo predictions, it can integrate often noisy, experimental interaction data sets. We observe that at given levels of sensitivity, our predictions are more accurate than the existing high-throughput experimental data sets. We validate our predictions with TAP (tandem affinity purification) tagging experiments. Our analysis, which gives a comprehensive view of yeast interactions, is available at genecensus.org/intint.

Many fundamental biological processes involve protein-protein interactions, and comprehensively identifying them is important to systematically defining their cellular role. New experimental and computational methods have vastly increased the number of known or putative interactions, cataloged in databases (1–7). Much genomic information also relates to interactions indirectly: Interacting proteins are often significantly coexpressed (as shown by mi-

croarrays) and colocalized to the same subcellular compartment (8, 9).

Unfortunately, interaction data sets are often incomplete and contradictory (10–12). In the context of genome-wide analyses, these inaccuracies are greatly magnified because the protein pairs that do not interact (negatives) far outnumber those that do (positives). For instance, in yeast, the ~6000 proteins allow for ~18 million potential interactions, but the estimated number of actual interactions is <100,000 (10, 13, 14). Thus, even reliable techniques can generate many false positives when applied genome-wide. This is similar to a diagnostic with a 1% false-positive rate for a rare disease occurring in 0.1% of the population, which would roughly produce one true positive for every 10 false ones. Further information is necessary.

Consequently, when evaluating protein-protein interactions, one needs to integrate evidence from many different sources (15–17). Here, we propose a Bayesian approach for inte-

grating interaction information that allows for the probabilistic combination of multiple data sets and demonstrate its application to yeast (18). Our approach can be used for combining noisy interaction data sets and for predicting interactions de novo, from other genomic information. The basic idea is to assess each source of evidence for interactions by comparing it against samples of known positives and negatives (“gold-standards”), yielding a statistical reliability. Then, extrapolating genome-wide, we predict the chance of possible interactions for every protein pair by combining each independent evidence source according to its reliability. We verified our predictions by comparing them against existing experimental interaction data (not in the gold-standard) as well as new TAP (tandem affinity purification) tagging experiments.

Among the many possible machine-learning approaches that could be applied to predicting interactions (ranging from simple unions and intersections of data sets to neural networks, decision trees, and support-vector machines), Bayesian networks have several advantages (19): They allow for combining highly dissimilar types of data (i.e., numerical and categorical), converting them to a common probabilistic framework, without unnecessary simplification; they readily accommodate missing data; and they naturally weight each information source according to its reliability. In contrast to “black-box” predictors, Bayesian networks are readily interpretable as they represent conditional probability relationships among information sources.

The gold-standard data set on which we train (“parameterize”) the Bayesian network should ideally be (i) independent from the data sources serving as evidence, (ii) sufficiently large for reliable statistics, and (iii) free of systematic bias. We used the MIPS (Munich Information Center for Protein Sequences) complexes catalog as the gold-standard for positives (6). This hand-curated list of protein complexes is based on the literature [8250 pairs

¹Department of Molecular Biophysics and Biochemistry, ²Department of Molecular, Cellular and Developmental Biology, ³Department of Computer Science, Yale University, 266 Whitney Avenue, Post Office Box 208114, New Haven, CT 06520, USA. ⁴Banting and Best Department of Medical Research, Department of Molecular and Medical Research, University of Toronto, Toronto, M5G 1L6, Ontario, Canada.

*Present address: Computational Biology Center, Memorial Sloan-Kettering Cancer Center, 307 West 63rd Street, New York, NY 10021, USA.

†To whom correspondence should be addressed. E-mail: mark.gerstein@yale.edu

REPORTS

in our filtered version (19)]. A negatives gold-standard is harder to define, but essential for successful training. Thus, we synthesized negatives from lists of proteins in separate subcellular compartments (9). Our positive and negative gold-standards satisfy the first two criteria and provide a good practical solution for the third. Hence, our goal, precisely defined, was to predict whether two proteins are in the same complex, not whether they necessarily had direct physical contact.

As a measure of reliability, the overlap of information sources (i.e., “interaction data sets,” which could either be noisy experimental data or sets of genomic features) with the gold-standards can be expressed in terms of a “likelihood ratio.” For example, consider a genomic feature f expressed in binary terms (i.e., “present” or “absent”). The likelihood ratio $L(f)$ is then defined as the fraction of gold-standard positives having feature f divided by the fraction of negatives having f . For two features f_1 and f_2 with uncorrelated evidence, the likelihood ratio of the combined evidence is simply the product $L(f_1, f_2) = L(f_1)L(f_2)$. For correlated evidence, $L(f_1, f_2)$ cannot be factorized in this way. Bayesian networks are a formal representation of such relationships between features. The combined likelihood ratio is proportional to the estimated odds that two proteins are in the same complex, given multiple sources of information.

We predict a protein pair as positive if its combined likelihood ratio exceeds a particular cutoff ($L > L_{cut}$) (negative otherwise). To get an overall assessment of how the prediction performs, we segmented the gold-standard into

separate training and testing sets (using a sevenfold cross-validation protocol). Then we evaluated the number of true- (TP) and false-positive (FP) predictions in the testing set. Finally, we applied the Bayesian network beyond the testing set, computing likelihood ratios for all possible protein pairs in the genome.

Figure 1 schematically shows the information sources and results of our calculations. We term the results “probabilistic interactomes” (PIs), in which each protein pair is associated with a probability measure for being in the same complex (i.e., likelihood ratio L). Our procedure not only allows combining existing experimental interaction data sets (resulting in a PI-experimental or “PIE”), but also the de novo prediction of protein complexes from genomic data sets (when the input data are not interaction data sets per se, resulting in a PI-predicted or “PIP”).

We combined four interaction data sets from high-throughput experiments into the PIE (1–4) (Fig. 1B). The PIE represents a transformation of the individual binary-valued interaction sets into a data set where every protein pair is weighted according to the likelihood that it exists within a complex.

We computed the PIP from several genomic data sources: the correlation of mRNA amounts in two expression data sets (one with temporal profiles during the cell cycle, one of expression levels under 300 cellular conditions), two sets of information on biological function, and information about whether proteins are essential for survival (6, 20–22). Although none of these information sources are interaction data per se, they contain information weakly associated with in-

teraction: Two subunits of the same protein complex often have coregulated mRNA expression and similar biological functions and are more likely to be both essential or nonessential (8).

For computing the PIE and the PIP, we used two different types of Bayesian networks: a “naïve” network for the PIP and a fully connected one for the PIE (19). The naïve network is simpler to compute but requires information sources with essentially uncorrelated evidence. In contrast, the fully connected Bayesian network accommodates correlated evidence, which is the case for the four experimental interaction data sets.

Finally, we combined the PIP, PIE, and gold-standard into a total PI (PIT), which represents our most comprehensive view of the known and putative protein complexes in yeast (23). Because the PIP and PIE data provide essentially uncorrelated evidence for protein-protein interactions, we chose a naïve network to construct the PIT.

Figure 1C gives an overview of how we compared the PIP, PIE, gold-standard, and our new experiments. In particular, Fig. 2 shows the performance of the integration resulting in the PIT and PIE. When tested against the gold-standard, we observed that the ratio of true to false positives (TP/FP) increases monotonically with L_{cut} , confirming L as an appropriate measure of the odds of a real interaction. Conservatively estimated, protein pairs with $L > 600$ have a better than 50% chance of being in the same complex, suggesting $L_{cut} = 600$ as a useful threshold (19). Unless otherwise noted, we use this throughout our analysis. It gives 9897 predicted interactions from the PIP and

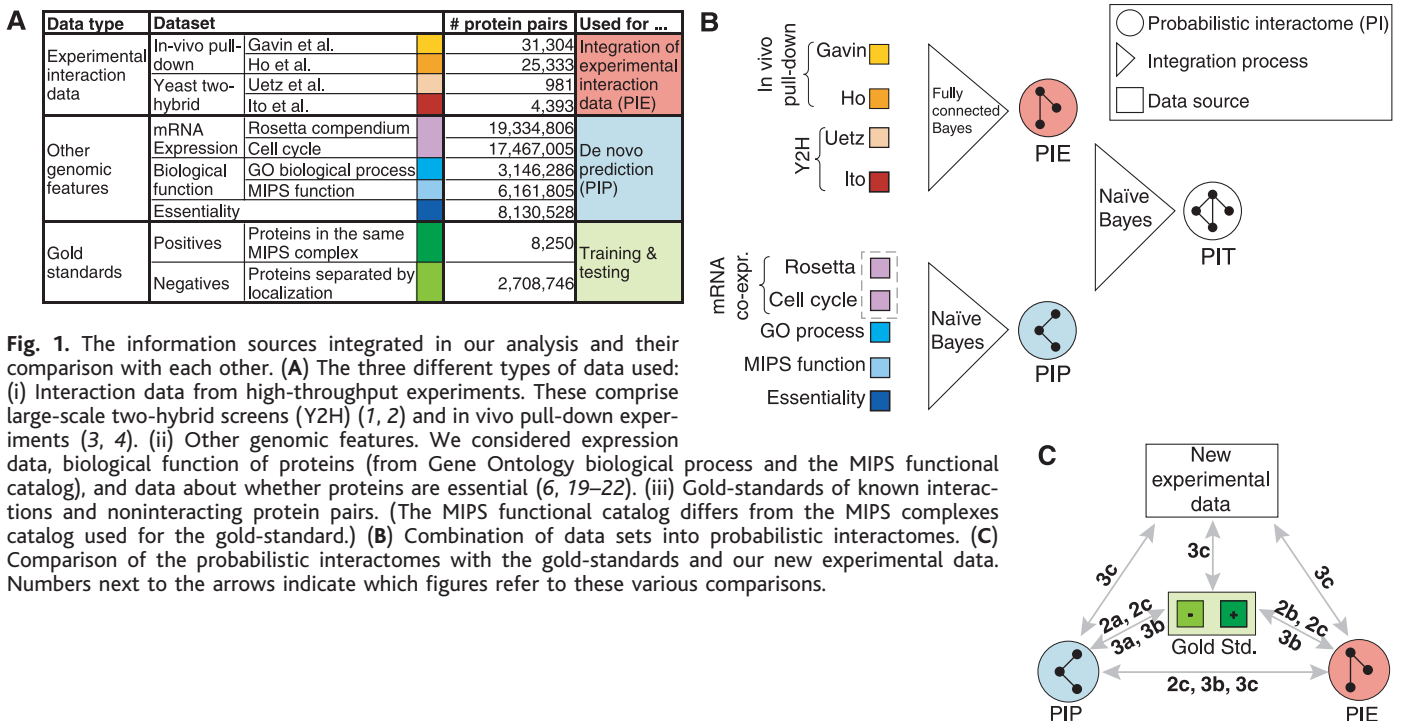


Fig. 1. The information sources integrated in our analysis and their comparison with each other. (A) The three different types of data used: (i) Interaction data from high-throughput experiments. These comprise large-scale two-hybrid screens (Y2H) (1, 2) and in vivo pull-down experiments (3, 4). (ii) Other genomic features. We considered expression data, biological function of proteins (from Gene Ontology biological process and the MIPS functional catalog), and data about whether proteins are essential (6, 19–22). (iii) Gold-standards of known interactions and noninteracting protein pairs. (The MIPS functional catalog differs from the MIPS complexes catalog used for the gold-standard.) (B) Combination of data sets into probabilistic interactomes. (C) Comparison of the probabilistic interactomes with the gold-standards and our new experimental data. Numbers next to the arrows indicate which figures refer to these various comparisons.

163 from the PIE. In contrast, likelihood ratios derived from single genomic features (e.g., mRNA coexpression) or from individual interaction experiments (e.g., the Ho data set) did not exceed the cutoff when used alone, with TP/FP values far below 1. This demonstrates that information sources that, taken alone, are only weak predictors of interactions can yield reliable predictions when combined.

The PIP had a higher sensitivity than the PIE for comparable TP/FP ratios (Fig. 2C). (“Sensitivity” measures coverage and is defined as TP/P , where P is the number of gold-standard positives.) Specifically, the sensitivity of the PIP is $\sim 27\%$ at our cutoff. This may seem low, but compares favorably with the PIE, which had a sensitivity of less than 1%. This means that we can predict, at comparable error levels, more complex interactions de novo than are present in the high-throughput experimental interaction data sets.

One might ask whether simpler voting procedures can match the performance of more

complicated machine-learning methods such as Bayesian networks. To test this hypothesis, we compared the PIP with a voting procedure where each of the four genomic features contributes an additive vote toward positive classification. We found that the Bayesian network achieved greater sensitivity for comparable TP/FP ratios (Fig. 2C) (19).

Figure 3 shows parts of the PIP and PIE graphs and how these compare with the gold-standard and our new experiments. First, to test whether the thresholded PIP was biased toward certain complexes, we looked at the distribution of predictions among gold-standard positives (Fig. 3A); they were roughly equally apportioned among the different complexes, suggesting a lack of bias.

We have thus far treated all interactions as independent. However, the joint distribution of interactions in the PIs can help identify large complexes: An ideal complex should be a “clique” in an interaction graph (i.e., a subgraph with $N(N - 1)/2$ links between N proteins).

Although this rarely happens in practice, because of incorrect or missing links, large complexes tend to have many interconnections within them, whereas false-positive links to outside proteins tend to occur randomly, without a coherent pattern (Fig. 4).

Figure 3B shows parts of the thresholded PIP that are restricted to proteins with ≥ 20 links (23), highlighting large complexes. Some predicted complexes overlap with the gold-standard positives (cytoplasmic ribosome) or the PIE (exosome, RNA polymerase I, 26S proteasome). Comparison with the gold-standard negatives showed where the PIP likely produced false complexes. Many protein associations only appear in the PIP and thus potentially represent new interactions and complexes. An interesting example is the mitochondrial ribosome; it has appreciable overlap with both gold-standard positives and the PIE and contains plausible, newly predicted interactions with three proteins (19).

To further test the predictions in the PIP, we conducted TAP-tagging experiments, in which a protein expressed at its normal intracellular concentration (“bait”) is tagged and used to “pull down” endogenous protein complexes. We picked 98 proteins as TAP-tagging baits. These produced 424 experimental interactions overlapping with the PIP thresholded at $L_{cut} = 300$. (Of these, 185, in turn, overlapped with gold-standard positives, and 16 with negatives, highlighting the reliability of our experiments.)

Figure 3C shows three examples of the overlap between the PIP and TAP-tagging. We predicted that the putative DEAD-box RNA helicase Dbp3 interacts with three other RNA helicases (Hca4, Mak5, and Dbp7), with proteins implicated in ribosomal RNA (rRNA) metabolism (e.g., Nop2, Rrp5, Mak5, and components of RNA polymerase I), and with Nsr1, the yeast homolog of mammalian Nucleolin and a GAR domain-containing protein (24). When Dbp3 was TAP-tagged and purified, we found previously unknown interactions with Nsr1, Hca4, and Nop1, connecting Dbp3 with known rRNA-processing proteins. Further purifications with TAP-tagged versions of Mak5, Rrp5, Dbp7, Dbp3, Nsr1, Hca4, and Nop2 verified the physical association.

The nucleosome, a fundamental unit within chromatin, provides a second example of overlap. It is composed of eight histones (two H2A, two H2B, two H3, and two H4), which can block RNA polymerase II progression. This blockage is relieved upon interaction with the FACT complex (also known as SPN or yFACT), which consists of Spt16 and Pob3 in yeast. Mammalian Pob3 has a high mobility group (HMG) domain for interaction with histones; however, yeast Pob3 lacks this domain. Instead, the HMG protein Nhp6 (with two virtually identical isoforms, Nhp6A and Nhp6B) binds histones (25–27).

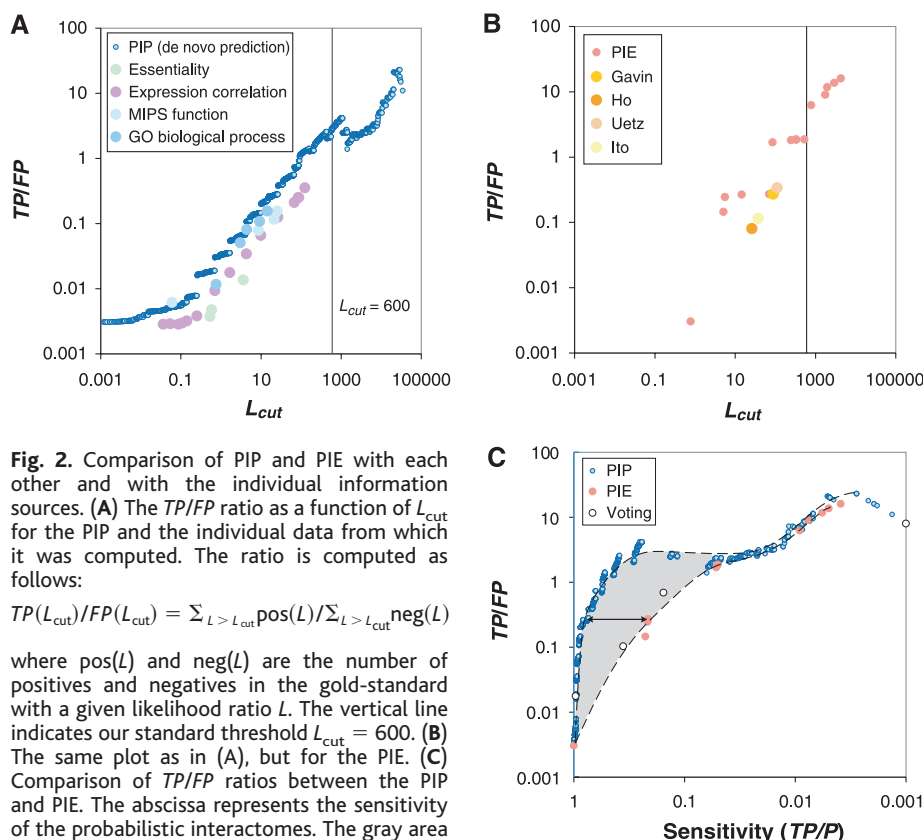


Fig. 2. Comparison of PIP and PIE with each other and with the individual information sources. (A) The TP/FP ratio as a function of L_{cut} for the PIP and the individual data from which it was computed. The ratio is computed as follows:

$$TP(L_{cut})/FP(L_{cut}) = \sum_{L > L_{cut}} \text{pos}(L) / \sum_{L > L_{cut}} \text{neg}(L)$$

where $\text{pos}(L)$ and $\text{neg}(L)$ are the number of positives and negatives in the gold-standard with a given likelihood ratio L . The vertical line indicates our standard threshold $L_{cut} = 600$. (B) The same plot as in (A), but for the PIE. (C) Comparison of TP/FP ratios between the PIP and PIE. The abscissa represents the sensitivity of the probabilistic interactomes. The gray area indicates the gain of sensitivity of the PIP over the PIE for equal TP/FP ratios. The arrow shows the difference in sensitivity at $TP/FP = 0.3$. At this level, the PIP contains 183,295 protein pairs, of which 6179 are gold-standard positives (75% sensitivity), whereas the PIE contains 31,511 protein pairs and 1758 gold-standard positives among these (21% sensitivity). This difference in sensitivity between PIE and PIP illustrates the value of the de novo prediction. It also reflects, to some degree, that the experiments were done only on subsets of the genome and may have been measuring different types of interactions than the complexes’ gold-standard, which we used to parameterize the PIP. The white circles show the performance of a voting procedure in which each of the four genomic features (from which we computed the PIP) contributed an additive vote. There are four possible outcomes in the additive voting procedure, depending on how many data sets contribute a positive vote (19).

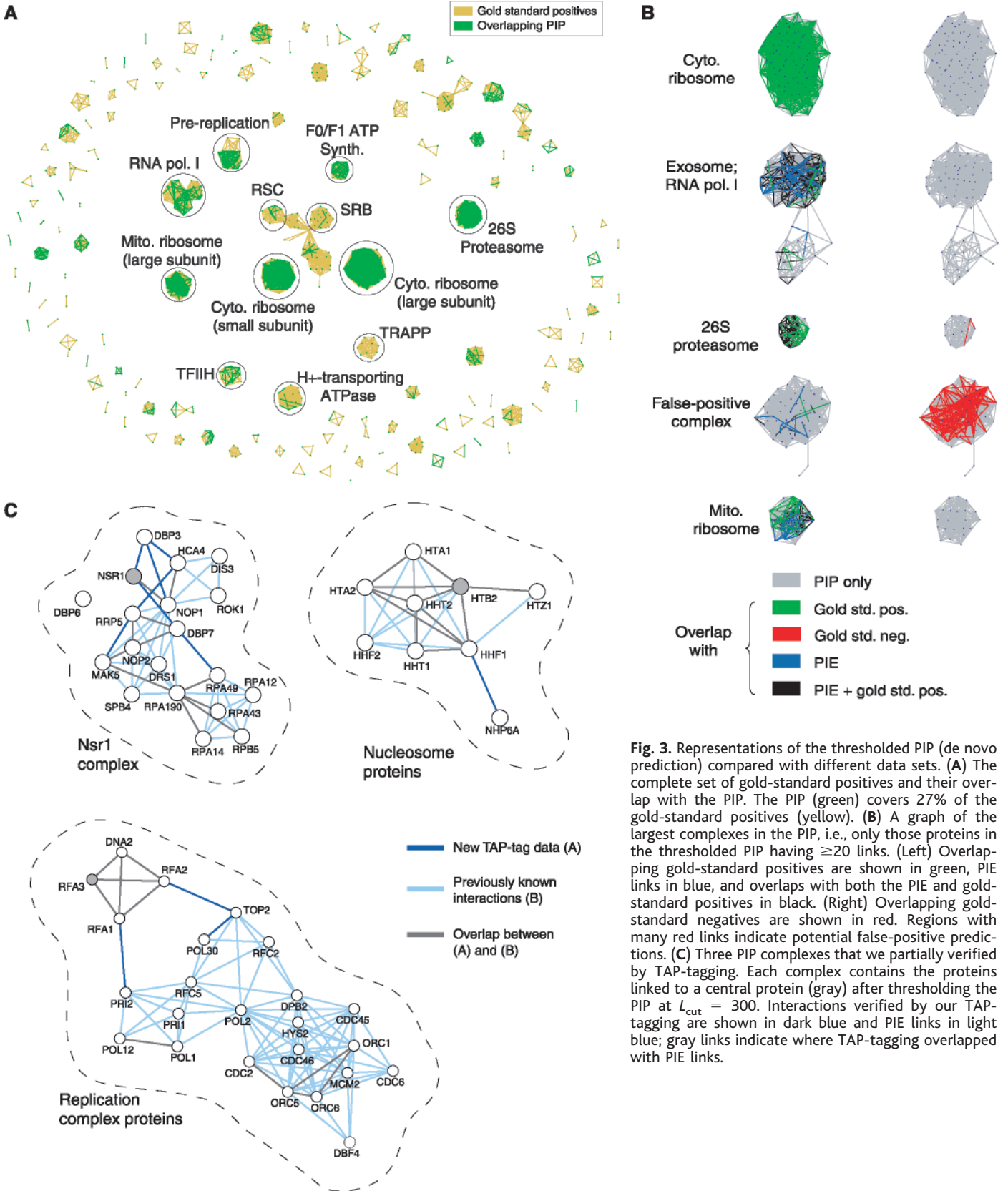


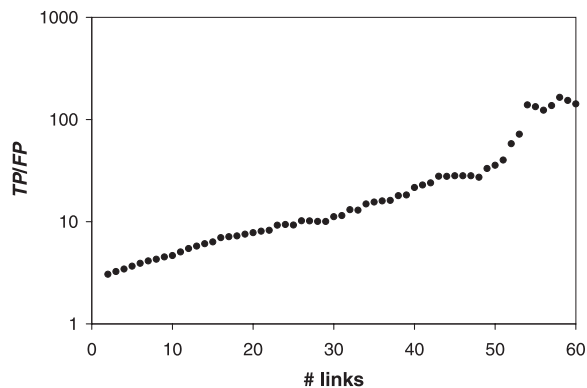
Fig. 3. Representations of the thresholded PIP (de novo prediction) compared with different data sets. (A) The complete set of gold-standard positives and their overlap with the PIP. The PIP (green) covers 27% of the gold-standard positives (yellow). (B) A graph of the largest complexes in the PIP, i.e., only those proteins in the thresholded PIP having ≥ 20 links. (Left) Overlapping gold-standard positives are shown in green, PIE links in blue, and overlaps with both the PIE and gold-standard positives in black. (Right) Overlapping gold-standard negatives are shown in red. Regions with many red links indicate potential false-positive predictions. (C) Three PIP complexes that we partially verified by TAP-tagging. Each complex contains the proteins linked to a central protein (gray) after thresholding the PIP at $L_{cut} = 300$. Interactions verified by our TAP-tagging are shown in dark blue and PIE links in light blue; gray links indicate where TAP-tagging overlapped with PIE links.

[Nhp6 also binds DNA in competition with the nucleosome (28).] Our thresholded PIP and experimental data document a specific interaction between Nhp6A and Hhf1 (H4),

pinpointing the contact between the nucleosome and Nhp6 to the H3-H4 heterodimer (Hhf1 and Hht1). This is plausible; because Nhp6 has been shown not to influence nu-

cleosome reassembly (29), it is unlikely that it binds with the H2A-H2B dimer, which needs to reassociate with the nucleosome after binding FACT.

Fig. 4. *TP/FP* for subsets of the thresholded PIP that only include proteins with a minimum number of links. Requiring a minimum number of links isolates large complexes in the thresholded PIP graph (Fig. 3B). Increasing the minimum number of links raises *TP/FP* by preserving the interactions among proteins in large complexes, while filtering out false-positive interactions with heterogeneous groups of proteins outside the complexes.



The replication complex, a third experimental validation of the PIP, assembles and disassembles from transiently interacting subcomplexes (e.g., MCM proteins, ORC, and polymerases) throughout the cell cycle (8, 30). Our predicted and experimentally verified interactions connect it, probably transiently, to another subcomplex, replication factor A (RFA, composed of Rfa1, Rfa2, and Rfa3). Specifically, we predicted and verified interactions between RFA and two proteins associated with other replication subcomplexes: Rfa2 with Top2 (a component of the nuclear synaptonemal complex) and Rfa1 with Pri2 (DNA polymerase α -primase subunit).

Finally, we predicted and verified by TAP-tagging that two proteins involved in translation elongation (Tef2 and Eft2) interact. This is plausible given that protein elongation is mediated by three factors in yeast: EF-1 α (Tef1, Tef2), EF-2 (Eft1, Eft2), and EF-3 (Hef3, Yef3); most other eukaryotes lack EF-3. Previous experimental data suggest an interaction between yeast EF-1 α and EF-3 (31). An interaction between EF-1 α and EF-2 had not been demonstrated, although this is reasonable given their similar roles in elongation and their overlapping binding sites on the ribosome (32).

In summary, we have developed a Bayesian approach for integrating weakly predictive genomic features into reliable predictions of protein-protein interactions. Our de novo prediction of complexes replicated interactions found in the gold-standard positives and PIE. In addition, we confirmed several of our predictions with new experiments. The accuracy of the PIP was comparable to that of the PIE while simultaneously achieving greater coverage.

Our procedure lends itself naturally to the addition of more features, possibly further improving results. We anticipate that protein-protein interactions in organisms other than yeast can be explored in similar ways.

References and Notes

1. P. Uetz et al., *Nature* **403**, 623 (2000).
2. T. Ito et al., *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4569 (2001).
3. A. C. Gavin et al., *Nature* **415**, 141 (2002).
4. Y. Ho et al., *Nature* **415**, 180 (2002).
5. I. Xenarios et al., *Nucleic Acids Res.* **30**, 303 (2002).

6. H. W. Mewes et al., *Nucleic Acids Res.* **30**, 31 (2002).
7. G. D. Bader et al., *Nucleic Acids Res.* **29**, 242 (2001).
8. R. Jansen, D. Greenbaum, M. Gerstein, *Genome Res.* **12**, 37 (2002).
9. A. Kumar et al., *Genes Dev.* **16**, 707 (2002).
10. C. von Mering et al., *Nature* **417**, 399 (2002).
11. A. M. Deane, L. Salwinski, I. Xenarios, D. Eisenberg, *Mol. Cell. Proteomics* **1**, 349 (2002).
12. A. M. Edwards et al., *Trends Genet.* **18**, 529 (2002).
13. G. D. Bader, C. W. Hogue, *Nature Biotechnol.* **20**, 991 (2002).
14. A. Kumar, M. Snyder, *Nature* **415**, 123 (2002).
15. A. M. Marcotte, M. Pellegrini, M. J. Thompson, T. O. Yeates, D. Eisenberg, *Nature* **402**, 83 (1999).
16. M. Steffen, A. Petti, J. Aach, P. D'Haeseleer, G. Church, *BMC Bioinformatics* **3**, 34 (2002).

17. R. Jansen, N. Lan, J. Qian, M. Gerstein, *J. Struct. Funct. Genomics* **2**, 71 (2002).
18. A. Drawid, M. Gerstein, *J. Mol. Biol.* **301**, 1059 (2000).
19. Materials and methods are available as supporting material on Science Online.
20. T. R. Hughes et al., *Cell* **102**, 109 (2000).
21. R. J. Cho et al., *Mol. Cell* **2**, 65 (1998).
22. M. Ashburner et al., *Nature Genet.* **25**, 25 (2000).
23. See <http://genecensus.org/intint>.
24. I. P. Girard et al., *EMBO J.* **11**, 673 (1992).
25. N. K. Brewster, G. C. Johnston, R. A. Singer, *Mol. Cell. Biol.* **21**, 3491 (2001).
26. A. A. Travers, *EMBO Rep.* **4**, 131 (2003).
27. T. Formosa et al., *Genetics* **162**, 1557 (2002).
28. Y. Yu, P. Eriksson, L. T. Bhoite, D. J. Stillman, *Mol. Cell. Biol.* **23**, 1910 (2003).
29. R. C. Bash, J. M. Vargason, S. Cornejo, P. S. Ho, D. Lohr, *J. Biol. Chem.* **276**, 861 (2001).
30. O. M. Aparicio, D. M. Weinstein, S. P. Bell, *Cell* **91**, 59 (1997).
31. M. Anand, K. Chakraborty, M. J. Marton, A. G. Hinnebusch, T. G. Kinzy, *J. Biol. Chem.* **278**, 6985 (2003).
32. O. Kovalchuk, R. Kambampati, E. Pladies, K. Chakraborty, *Eur. J. Biochem.* **258**, 986 (1998).
33. We thank C. Sander and G. Bader for critical discussions.

Supporting Online Material

www.sciencemag.org/cgi/content/full/302/5644/449/DC1

Materials and Methods

Figs. S1 to S3

Tables S1 and S2

References

29 May 2003; accepted 29 August 2003

Transcriptional Repression of Atherogenic Inflammation: Modulation by PPAR δ

Chih-Hao Lee,¹ Ajay Chawla,^{1*} Ned Urbiztondo,¹ Debbie Liao,¹ William A. Boisvert,² Ronald M. Evans^{1†}

The formation of an atherosclerotic lesion is mediated by lipid-laden macrophages (foam cells), which also establish chronic inflammation associated with lesion progression. The peroxisome proliferator-activated receptor (PPAR) γ promotes lipid uptake and efflux in these atherogenic cells. In contrast, we found that the closely related receptor PPAR δ controls the inflammatory status of the macrophage. Deletion of PPAR δ from foam cells increased the availability of inflammatory suppressors, which in turn reduced atherosclerotic lesion area by more than 50%. We propose an unconventional ligand-dependent transcriptional pathway in which PPAR δ controls an inflammatory switch through its association and disassociation with transcriptional repressors. PPAR δ and its ligands may thus serve as therapeutic targets to attenuate inflammation and slow the progression of atherosclerosis.

Lipid-accumulating macrophages, or foam cells, are the major component of the atherogenic lesion. Loss of either of the

nuclear receptors PPAR γ or liver X receptor (LXR) α in the macrophage disables lipid export and accelerates progression of the atherosclerotic lesion (1–3), establishing a transcriptional basis for lipid homeostasis within coronary arteries (4, 5).

Synthetic ligands of the nuclear receptor PPAR δ (also called PPAR β) modulate lipid transport in monocytic cell lines through regulated expression of the scavenger receptor CD36 and the efflux pump ABCA1 (6, 7). By analogy with PPAR γ , we hypothesized that loss of PPAR δ in macrophages may also disturb the dynamic balance of lipid uptake

¹Howard Hughes Medical Institute, Gene Expression Laboratory, Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA. ²Vascular Medicine Research, Brigham & Women's Hospital, 65 Landsdowne Street, Room 275, Cambridge, MA 02139, USA.

*Present address: Division of Endocrinology, Metabolism and Gerontology, Department of Medicine, Stanford University, Stanford, CA 94305, USA.

†To whom correspondence should be addressed. E-mail: evans@salk.edu