A network of protein–protein interactions in yeast

Benno Schwikowski[1,2]*, Peter Uetz[3], and Stanley Fields[3,4]

*[1]The Institute for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA 98105. [2]Department of Computer Science and Engineering, University of Washington, Box 352350, Seattle, WA 98195. [3]Departments of Genetics and Medicine, [4]Howard Hughes Medical Institute, University of Washington, Box 357360, Seattle, WA 98195. *Corresponding authors (benno@systemsbiology.org, uetz@u.washington.edu, fields@u.washington.edu).*

Received 27 June 2000; accepted 13 October 2000

A global analysis of 2,709 published interactions between proteins of the yeast *Saccharomyces cerevisiae* has been performed, enabling the establishment of a single large network of 2,358 interactions among 1,548 proteins. Proteins of known function and cellular location tend to cluster together, with 63% of the interactions occurring between proteins with a common functional assignment and 76% occurring between proteins found in the same subcellular compartment. Possible functions can be assigned to a protein based on the known functions of its interacting partners. This approach correctly predicts a functional category for 72% of the 1,393 characterized proteins with at least one partner of known function, and has been applied to predict functions for 364 previously uncharacterized proteins.

Keywords: Protein interactions, *Saccharomyces cerevisiae*, functional genomics, proteomics, bioinformatics

Functional genomics strategies are of increasing importance in characterizing proteins newly discovered by genome sequencing projects. In one approach, large sets of mutants created systematically[1] or randomly[2] are analyzed by physiological tests. In a second approach, DNA arrays are interrogated in order to cluster genes the transcripts of which are co-regulated with genes of known functions[3–5]. A third approach uses computational strategies to assign functions based on the coevolution of sets of proteins[6] or on the existence of a fusion protein in one organism corresponding to two separate proteins in another organism[7,8]. As a fourth approach, we use protein–protein interaction data (for example, ref. 9) to classify proteins based on the properties of their interacting partners.

The identification of protein interactions has escalated in scale from the analysis of small numbers of proteins[10–12] to more comprehensive analyses covering more than 1,000 proteins[13]. In particular, the ~6,000 predicted proteins of the yeast *S. cerevisiae* have been used extensively in two-hybrid searches to detect interacting partners[14,15]. When the data from these studies are combined with previously published interactions from the entire community of yeast researchers, they total over 2,709 putative interactions encompassing 2,039 different proteins[16,17].

Here we analyze all of these *S. cerevisiae* interactions in an effort to diagram the set of links within a large protein network. This network and the accompanying computational approaches can be used to view interactions among proteins within any defined functional category or cellular localization. The interaction data can be used to predict function for those uncharacterized yeast proteins that have partners of known function. Such approaches will find additional utility when applied to other organisms with increasing numbers of interactions, and as uncharacterized yeast proteins placed on the interaction map are used to predict possible functions for their orthologs in other species.

Results

1,548 proteins can be linked by protein interactions. We analyzed 2,709 published interactions involving 2,039 yeast proteins, available from public databases[16,17] and from two recent large-scale studies[14,15] (see Experimental Protocol). Our analysis includes only direct interactions identified by biochemical experiments or two-hybrid studies, but not protein complexes for which specific protein contacts are unknown. We sought to determine whether comprehensive interaction maps could be assembled from these data, and if so, whether such maps afforded insight into functional relationships among characterized and uncharacterized proteins. To visualize interactions, we developed a software program based on the graph-drawing library "AGD" (http://www.mpi-sb.mpg.de/AGD). Surprisingly, only a single large network of protein interactions was obtained, containing 2,358 links among 1,548 individual proteins (Fig. 1A). The next largest network contained only 19 proteins; 9 networks contained between 5 and 11 proteins; and the remaining 193 networks contained 4 or fewer proteins (data not shown), for a total of 204 independent networks (each interaction appears in only a single network).

Proteins have been assigned to 42 "cellular role" categories in the Yeast Protein Database (YPD)[16]. The term "function" is used here as in YPD to mean the cellular role a protein is engaged in and not its precise biochemical activity. The YPD categories are broad, and 39% of the 1,485 characterized proteins in the networks are assigned more than one cellular role. Most proteins within one of these functional classes cluster in a specific region of the large network, if a "cluster" includes any three or more proteins of the same function separated by no more than two other proteins. For example, in Figure 1, proteins involved in chromatin structure are highlighted in gray (89% of all such annotated proteins within clusters); in red, cytokinesis proteins (75% in clusters); in blue, membrane fusion proteins (94% in clusters); in green, cell structure proteins (90% in clusters); and in yellow, lipid metabolism proteins (58% in clusters).

A major concern in the delineation of these 204 networks is the quality of both the data and the database annotations. For example, interactions derived from genomic two-hybrid approaches are generally uncorroborated by additional experiments, and false positives are commonplace with this technique. Annotations of "cellular role" sometimes do not match experimentally determined properties, and other annotation information often has been transferred from

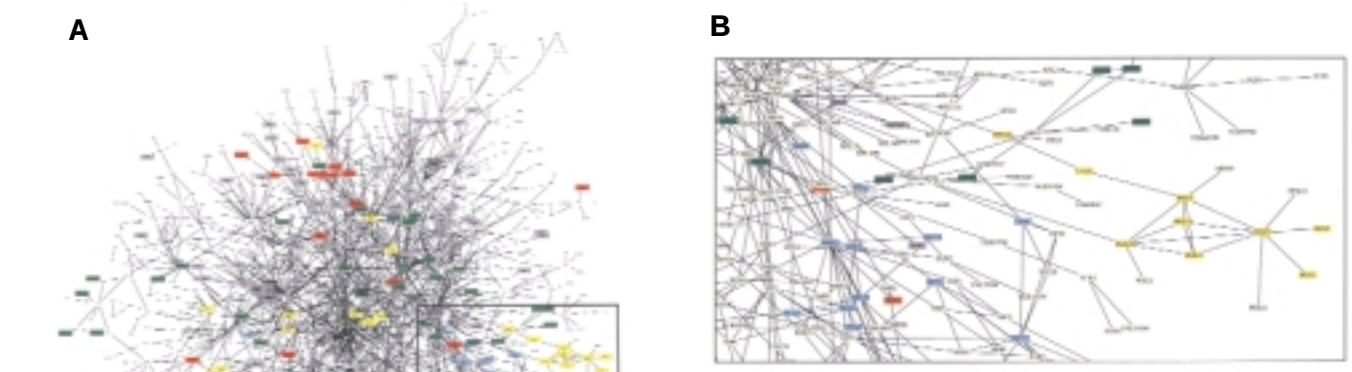

Figure 1. (A) An interaction map of the yeast proteome assembled from published interactions. The map contains 1,548 proteins and 2,358 interactions. Proteins are colored according to their functional role as defined by the Yeast Protein Database[16]; proteins involved in membrane fusion (blue), chromatin structure (gray), cell structure (green), lipid metabolism (yellow), and cytokinesis (red). For other maps with different functional groups highlighted, see http://depts.washington.edu/sfields/. On-line maps can also be zoomed and searched for protein names. (B) Section of part A showing the clustering of proteins involved in membrane fusion (blue), lipid metabolism (yellow), and cell structure (green).

homologous proteins without experimental confirmation. To assess the reliability of these networks, we determined how well they could be used to predict function for characterized proteins. For each such protein, we counted how many times, if at all, each of the 42 functions occurs as annotation in the direct interaction partners of the protein. If the protein has at least one characterized interaction partner, this leads to a list of functions for the partner(s) that can be sorted by frequency. We took the most frequent functions (at most three, if the list was longer) of the partners as indicators of the function of each characterized protein, and, if the known function occurred among these indicators, counted the protein as a case of a correct prediction. By this method, the networks allowed a correct prediction for 72% of the 1,393 characterized proteins with at least one characterized partner. (We also performed 100 trials with the same set of proteins and annotations but scrambled links between proteins; this randomization yielded on average only 12% correct predictions.) The approximately one-quarter of links between characterized proteins that do not have matching functions are likely due to false positives, incomplete annotations, or unknown biological connections. Although this considerable minority of possibly suspect links in the networks necessitates caution in the interpretation of new inferences from these data, nevertheless the overall validity of the networks analysis seems substantiated.

Crosstalk between and within functional groups. In addition to connections within a broad functional group, relationships between groups can be plotted to reveal connections that may be biologically meaningful. Figure 2 shows the number of interactions (if 15 or greater) of proteins from each functional group with proteins of their own and other groups. As expected, 65% of the interactions in the complete set of networks occur among proteins with at least one common functional assignment. However, numerous cross-connections are observed. For example, proteins involved in protein folding had 16 connections to proteins involved in protein translocation. Note that the graph-drawing algorithm that generated Figure 2 tends to place related functional groups next to each other. Cell cycle control proteins, which show the most interactions with other classes, are thereby placed near the center of the map, a positioning that reflects their central role in regulating other cellular processes.

Although many cross-connections are known, some revealed by Figure 2 seem surprising. For instance, proteins assigned a role in RNA processing not only show many connections to RNA splicing, RNA turnover, and RNA polymerase II transcription, but also to unexpected groups such as mitosis, chromatin, and protein synthe-

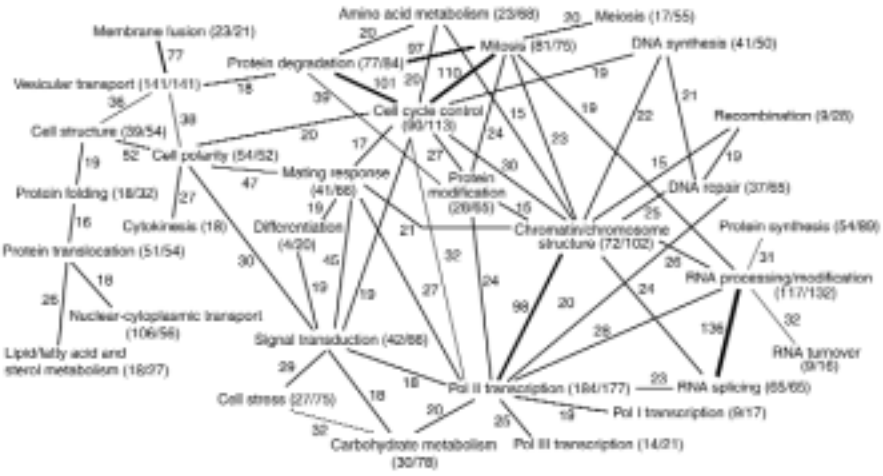

Figure 2. Interactions between functional groups. Numbers in parentheses indicate, first, the number of interactions within a group, and second, the number of proteins in a group. Numbers near connecting lines indicate the number of interactions between proteins of the two connected groups. For example, there are 77 interactions between the 21 proteins involved in membrane fusion and the 141 proteins involved in vesicular transport (upper left corner); 23 protein interactions connect the 21 proteins involved in membrane fusion. Only connections with 15 or more interactions are included here. Note that only proteins with known function are shown (many of these have several functions). The sum of all interactions in this diagram is therefore smaller than the number of all interactions.

sis. Such data are derived from studies, for example, showing that RNA processing factors interact with proteins of the ribosome[14,18].

Crosstalk between and within subcellular compartments. The logic of depicting interactions based on functional roles can also be applied to protein localizations. The localization of 1,203 of the proteins in the complete data set of interactions is known[16,19], and these localization data can be used to visualize potential crosstalk between cellular compartments (Fig. 3). In ~78% of the 1,432 interactions between proteins of known localization, the proteins share one or more annotated compartments. Interactions of membrane proteins are likely to be underrepresented in Figure 3 because of technical difficulties; most interactions described here were found by the two-hybrid assay, which requires nuclear localization. Interestingly, 31 of the 109 interactions (28%) of membrane proteins involved proteins of unknown function, which constitute targets for additional study.

The large number of interactions between nuclear and cytoplasmic proteins is expected, with many reports of proteins moving between these compartments through the nuclear pore[20]. Interactions between proteins located in less clearly connected compartments, like the nucleus and mitochondrion, are more difficult to explain. Some may represent nonphysiological interactions that occur in the two-hybrid or other assays, and others may be false positives.

Prediction of functions. Of the 2,039 proteins in the data set, 554 have no annotation in YPD for "cellular role." An obvious approach to functional prediction is to identify the most common function among the partners of an uncharacterized protein and to assume that the protein in question shares the same or a related function. However, only 364 proteins of unknown function have at least one partner of known function, and only 69 have two or more partners of known function (see Supplementary Table 2 in the Web Extras page of *Nature Biotechnology* online). Furthermore, only 29 of these 69 have two or more interacting proteins with at least one function in common (Table 1), which strengthens the predictive assignments.

One example to illustrate these functional predictions (other cases in Table 1 and Supplementary Table 2) is YNR053C, which interacts with five proteins implicated in splicing. Whereas these interactions allow an obvious prediction, this protein has also been implicated in nuclear-cytoplasmic transport because it is associated with the nuclear pore complex[21]. However, because there is no experimental evidence implicating the protein in one or the other of these functions, the protein has not been assigned a function.

One benefit of the network view of interactions is the ability to discern local patterns that are more informative than those in simpler datasets. For example, experiments on an uncharacterized protein "B" may detect links to proteins "A" and "C," and experiments on the uncharacterized protein "C" may detect a link to protein "D," to yield the simple network A–B–C–D. The validity of the functional assignments for B and C is considerably strengthened in those cases in which A and D have been annotated to be in the same functional class. An example is YHR105W, which interacts with one protein involved in vesicular transport, Akr2, and with YGL161C, an uncharacterized protein that interacts with two transport proteins, Yip1 and

Table 1. Prediction of protein function from protein–protein interactions compared to predictions by a combined algorithm.[22]

Protein	Prediction from protein interactions[a,b]	Prediction after Marcotte et al.[b,c]
YNR053C	RNA-processing/modification (5/5); **RNA-splicing** (5/5)	Transcription; Cell organization; nuclear organization; mRNA transcription; **mRNA splicing** (e: 3, p: 1)
BIR1	Cell-cycle control (2/4)	–
CNS1	Cell stress (2/3); Protein folding (2/3)	Cell organization; nuclear organization (p: 33)
YGL161C	Vesicular transport (2/3); membrane fusion (2/3)	–
YIP3	Vesicular-transport (3/3)	–
YMR322C	Cell stress (3/3)	–
YPR105C	Vesicular transport (2/3)	–
AIP2	Cell polarity (2/2)	Cell organization; carbohydrate utilization; carbohydrate metabolism; metabolism; fermentation; energy (e: 2, p: 4)
DUO1	Mitosis (2/4)	–
EBS1	Carbohydrate metabolism (2/5)	–
FPR4	Protein synthesis (2/3)	Cell organization; transcription; nuclear organization (e: 1, two: 2)
GIP2	Meiosis (2/2)	-
ISA1	Protein synthesis (2/2)	Nitrogen and sulfur utilization; metabolism; nitrogen and sulfur metabolism (p: 2)
RDI1	Cell polarity (2/3)	G-protein (e: 1)
SMY2	**RNA splicing** (2/2)	Transcription; cell organization; **mRNA splicing**; nuclear organization; mRNA transcription (e: 1)
YBR270C	Protein synthesis (2/4)	–
YDR100W	Vesicular transport (2/2); membrane fusion (2/2)	–
YEL015W	**RNA processing**/modification (2/6)	Transcription; cell organization; **mRNA splicing**; nuclear organization; mRNA transcription (e: 1)
YER079W	Signal transduction (2/2)	–
YGL096W	RNA processing/modification (2/2); **RNA splicing** (2/2)	Transcription; cell organization; **mRNA splicing**; nuclear organization; mRNA transcription (e: 1)
YJL019W	Chromatin/chromosome structure (2/2)	–
YKR030W	Vesicular transport (2/2)	–
YLR128W	Cell polarity (2/4)	–
YLR269C	RNA processing/modification (2/2);RNA splicing (2/2)	–
YLR368W	Protein degradation (2/5); amino acid metabolism (2/5); Cell cycle control (2/5)	–
YLR435W	Protein synthesis (2/2)	–
YLR456W	**RNA processing**/modification (2/4)	Transcription; mRNA transcription; **mRNA processing** (e: 1)
YNL311C	Amino acid metabolism (2/2)	–
YPL105C	RNA splicing (2/2)	–

[a]Prediction from protein interaction lists only proteins with two or more interaction partners of known function. Numbers in parentheses denote number of neighbors with the given function/number of interaction partners with known functions. Predictions from protein interactions are classified according to YPD.
[b]For complete data set, see Supplementary Table 2.
[c]Prediction after Marcotte et al. include only "high-quality predictions," with the method used for the prediction indicated in parentheses (e, experimental evidence; p, phylogenetic profile; two, two or more methods; the number is the number of links, see Marcotte et al.[22] for details). Identical or similar predictions are highlighted in bold typeface. Marcotte's predictions use terminology of MIPS.

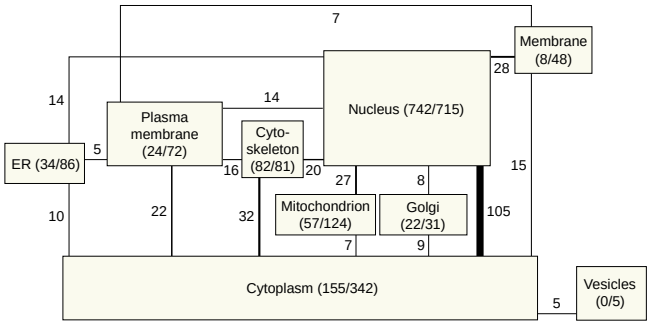

Figure 3. Interactions between proteins of different compartments. Numbers in parentheses indicate, first, the number of interactions among the proteins of this compartment, and second, the number of proteins that occur in this compartment. Numbers near connecting lines indicate the number of interactions between proteins of the two connected locations. For example, there are 9 interactions between the 31 Golgi proteins and the 342 cytoplasmic proteins. Note that only intercompartmental interactions that involve 5 or more interactions are shown. ER, endoplasmic reticulum.

Pep12 (Fig. 4). YHR105W also interacts with YPL246C, another uncharacterized protein that interacts with Ypt1 and Vam7, proteins implicated in vesicular transport and membrane fusion, respectively. Thus, interaction data can allow clear predictions even if the partners of a protein are largely of unknown function, as long as these partners interact with proteins of known function. Overall, we found 22 cases in which such indirect interactions strengthen functional predictions.

Discussion

Our analysis of protein–protein interactions in *S. cerevisiae* allows us to place fully one-quarter of the proteins predicted from the genome sequence within a single large network. This network reveals global patterns of interactions of proteins within functional classes or localization assignments, as well as many possible cross-connections. The interaction data can be used to make functional predictions for uncharacterized proteins, with the validity of this methodology established by an analysis showing that 72% of the characterized proteins with characterized partners could be assigned a correct functional category. Although the large network is not, of course, a wholly accurate view of cellular connections, it retains its utility for analyzing protein function even when local regions of the network are examined.

Marcotte et al.[22] predicted the functions of yeast proteins based on RNA expression, coevolution, protein fusions, and keyword searches of protein databases. They provided "high-quality" predictions for 90 of the 364 uncharacterized proteins with partners of known function. However, 27 of these predictions included the use of experimental evidence such as protein interactions. "Low-quality" predictions were made for 236 of the 364 proteins, based mainly on RNA expression patterns and protein fusions. We compared predictions for the 29 uncharacterized proteins with two or more known partners having at least one function in common (Table 1; Supplementary Table 2). Of the 10 Marcotte et al.[22] predictions for this set, 5 agree with those made from protein interactions (YNR053C, Smy2, YEL015W, YGL096W, YLR456W), although the combined algorithm includes experimental evidence in all these cases, usually interaction data. Three other predictions from Marcotte et al.[22] are based on experimental evidence (Aip2, Fpr4, Rdi1) and the other two on phylogenetic profiles (Cns1, Isa1). For the 63 proteins (out of 364 uncharacterized proteins with characterized partners) for which Marcotte et al.[22] make "high-quality" predictions without experimental evidence, the comparison with interaction data indicates that only 9 (15%) agree (see Supplementary Table 2). These comparisons indicate that

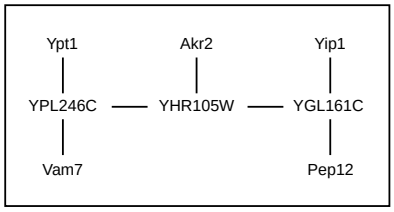

Figure 4. Prediction of function by direct and indirect protein interactions. YHR105W, YPL246C, and YGL161C are proteins of unknown function. Akr2 is a protein involved in endocytosis and therefore suggests a function for YHR105W. This potential function is supported by indirect interactions with Ypt1, Vam7, Yip1, and Pep12, which have been also implicated in vesicular transport and/or membrane fusion (data from refs 14 and 15).

whereas both purely computational methods and genome-wide interaction approaches can predict functions for unknown proteins, uncertainties remain that will require additional experimentation to resolve.

Annotations of proteins in databases are incomplete and therefore introduce additional uncertainties. A number of proteins with well-characterized function or at least well-supported functional prediction are annotated as "unknown function." For example, Tra1, a component of the transcriptional SAGA complex, is said to be of "unknown cellular role" in YPD (as of April 2000). This annotation problem can limit the usefulness of computational predictions that are heavily dependent on the accuracy and completeness of databases.

The protein network presented here demonstrates the complexity of cellular processes, particularly in light of the fact that only a fraction of all protein interactions are known. In addition to the identification of the remaining interactions, much remains to be done in terms of modeling and visualization. Continued development of bioinformatics tools is needed in order to cope with the emerging data sets and manually prepared wiring diagrams (as in ref. 23). Yeast gives us just a glimpse of what challenges lie ahead for the analysis of more complex, multicellular organisms.

Experimental protocol

Data sources. Of the 2,709 interactions analyzed, 1,183 are available from the MIPS site (http://www.mips.biochem.mpg.de/proj/yeast/tables/interaction/physical_interact.html); however, this site continues to be updated and contains additional interactions not available at the time of our analysis. Another 808 interactions are from Uetz et al.[14], 134 are from Ito et al.[15], 37 are from the Database of Interacting Proteins[9] (http://dip.doe-mbi.ucla.edu/), and 76 are newly generated as part of our continuing two-hybrid analyses. We obtained an additional 471 interactions from YPD as part of a larger compilation; this list of interactions can be requested from Proteome, Inc., Beverly, MA. Our complete data set, excluding those from YPD, together with sources, can be found in Supplementary List of Interactions in the Web Extras page of Nature Biotechnology online. Note that protein names, functional assignments, localizations, and interaction lists in other databases are subject to change and may not be identical to the names and lists used here. Cellular roles for proteins in this study are available at http://www.proteome.com/databases/YPD/YPDcategories/Cellular_Role.html. Subcellular localizations are available at http://www.proteome.com/databases/YPD/YPDcategories/Subcellular_Localization.html.

Generation of graphic representations. Graphs were generated using the AGD software library (http://www.mpi-sb.mpg.de/AGD). For Figure 1 we used the algorithm by Sugiyama et al.[24] to generate an initial layout and then applied a spring embedder algorithm[25] for the final placement of the proteins. In Figure 2, an interaction between two functional groups A and B was counted whenever our data set contained an interaction between a protein with an annotated function A and another protein with an annotated function B. Notice that an interaction between a protein with x annotated functions and a protein with y annotated functions was counted (x)(y) times in Figure 2. This calculation led to each interaction being counted an average of 2.74 times. Interactions between compartments in Figure 3 were counted analogously whenever the sets of annotated compartments for A and B were

nonoverlapping. However, when the sets of compartments intersected, only one internal interaction was counted for each of the compartments occurring in the intersection. This procedure led to each interaction in Figure 3 being counted an average of 1.07 times.

Prediction of function. We predicted the function of a protein P as follows. The annotated functions of all neighbors of P are ordered in a list, from the most frequent to the least frequent. Functions that occur the same number of times are ordered arbitrarily. Everything after the third entry in the list is discarded, and the remaining three or fewer functions are declared as predictions for the function of P.

Reliability of the generated networks. We were able to evaluate the quality of the networks on those 1,393 of the 2,039 proteins that were annotated with some function and had at least one neighbor annotated with a function, so that the prediction could be compared to the actual annotation. In 1,005 of the 1,393 cases (72.1%), at least one annotated function was predicted correctly by the above method. To evaluate the significance of this success rate, we performed the same prediction algorithm 100 times on the basis of randomly generated interactions. Each time, a set of 2,709 interactions among the 2,039 proteins was drawn uniformly at random from all possible such sets. On the average, only 12.2% of the predictions yielded a prediction that agreed with the known annotation.

Note: Supplementary information can be found on the Nature Biotechnology *website in Web Extras (http://www.biotech.nature.com/web_extras).*

Acknowledgments

We thank Phil Green, Maynard Olson, Peer Bork, David Eisenberg, Trey Ideker, and members of the laboratory for helpful discussions and comments on the manuscript. This work has been supported by fellowships from the German Academic Exchange Service DAAD (to B.S. and P.U.) and grants from the National Institutes of Health (P41 RR11823) and The Institute for Systems Biology. S.F. is an investigator of the Howard Hughes Medical Institute.

1. Winzeler, E.A. et al. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
2. Ross-Macdonald, P. et al. Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* **402**, 413–418 (1999).
3. Zhang, M.Q. Promoter analysis of co-regulated genes in the yeast genome. *Comput. Chem.* **23**, 233–250 (1999).
4. Harrington, C.A., Rosenow, C. & Retief, J. Monitoring gene expression using DNA microarrays. *Curr. Opin. Microbiol.* **3**, 285–291 (2000).
5. Granjeaud, S., Bertucci, F. & Jordan, B.R. Expression profiling: DNA arrays in many guises. *Bioessays* **21**, 781–790 (1999).
6. Pellegrini, M., Marcotte, E.M., Thompson, M.J., Eisenberg, D. & Yeates, T.O. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl. Acad. Sci. USA* **96**, 4285–4288 (1999).
7. Marcotte, E.M. et al. Detecting protein function and protein–protein interactions from genome sequences. *Science* **285**, 751–753 (1999).
8. Enright, A.J., Iliopoulos, I., Kyrpides, N.C. & Ouzounis, C.A. Protein interaction maps for complete genomes based on gene fusion events. *Nature* **402**, 86–90 (1999).
9. Xenarios, I. et al. DIP: the database of interacting proteins. *Nucleic Acids Res.* **28**, 289–291 (2000).
10. Bartel, P.L., Roecklein, J.A., SenGupta, D. & Fields, S. A protein linkage map of *Escherichia coli* bacteriophage T7. *Nat. Genet* .**12**, 72–77 (1996).
11. Fromont-Racine, M., Rain, J.C. & Legrain, P. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nat. Genet* .**16**, 277–282 (1997).
12. Walhout, A.J. et al. Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* **287**, 116–122 (2000).
13. Uetz, P. & Hughes, R.E. Systematic and large-scale two-hybrid screens. *Curr. Opin. Microbiol.* **3**, 304–309 (2000).
14. Uetz, P. et al. A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627 (2000).
15. Ito, T. et al. Toward a protein–protein interaction map of the budding yeast: a comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proc. Natl. Acad. Sci. USA* **97**, 1143–1147 (2000).
16. Costanzo, M.C. et al. The yeast proteome database (YPD) and *Caenorhabditis elegans* proteome database (WormPD): comprehensive resources for the organization and comparison of model organism protein information. *Nucleic Acids Res.* **28**, 73–76 (2000).
17. Mewes, H.W. et al. MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.* **28**, 37–40 (2000).
18. Neubauer, G. et al. Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nat. Genet* .**20**, 46–50 (1998).
19. Kumar, A. et al. TRIPLES: a database of gene function in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **28**, 81–84 (2000).
20. Michael, W.M. Nucleocytoplasmic shuttling signals: two for the price of one. *Trends Cell Biol.* **10**, 46–50 (2000).
21. Rout, M.P. et al. The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J. Cell Biol.* **148**, 635–651 (2000).
22. Marcotte, E.M., Pellegrini, M., Thompson, M.J., Yeates, T.O. & Eisenberg, D. A combined algorithm for genome-wide prediction of protein function. *Nature* **402**, 83–86 (1999).
23. Kohn, K.W. Molecular interaction map of the mammalian cell cycle control and DNA repair systems. *Mol. Biol. Cell* **10**, 2703–2734 (1999).
24. Sugiyama, K., Tagawa, S. & Toda, M. Methods for visual understanding of hierarchical systems. *IEEE Trans. Syst. Man Cybern.* **SMC-11**, 109–125 (1981).
25. Fruchterman, T. & Reingold, E. Graph drawing by force-directed placement. *Software - Pract. Exp.* **21**, 1129–1164 (1991).