

# Large-scale identification of yeast integral membrane protein interactions

John P. Miller\*<sup>†</sup>, Russell S. Lo<sup>‡</sup>, Asa Ben-Hur\*, Cynthia Desmarais\*, Igor Stagljar<sup>§¶</sup>, William Stafford Noble\*<sup>\*\*\*</sup>, and Stanley Fields\*<sup>††††</sup>

Departments of \*Genome Sciences, <sup>†</sup>Medicine and \*\*Computer Science and Engineering and <sup>‡</sup>Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195; and <sup>§</sup>Institute of Veterinary Biochemistry and Molecular Biology, University of Zurich-Irchel, CH-8057 Zurich, Switzerland

Contributed by Stanley Fields, June 29, 2005

We carried out a large-scale screen to identify interactions between integral membrane proteins of *Saccharomyces cerevisiae* by using a modified split-ubiquitin technique. Among 705 proteins annotated as integral membrane, we identified 1,985 putative interactions involving 536 proteins. To ascribe confidence levels to the interactions, we used a support vector machine algorithm to classify interactions based on the assay results and protein data derived from the literature. Previously identified and computationally supported interactions were used to train the support vector machine, which identified 131 interactions of highest confidence, 209 of the next highest confidence, 468 of the next highest, and the remaining 1,085 of low confidence. This study provides numerous putative interactions among a class of proteins that have been difficult to analyze on a high-throughput basis by other approaches. The results identify potential previously undescribed components of established biological processes and roles for integral membrane proteins of ascribed functions.

*Saccharomyces cerevisiae* | split-ubiquitin | support vector machine

Systematic studies of protein interactions in yeast have provided insights into the functions of many of the proteins encoded by this single-celled eukaryote. However, the roles of many integral membrane proteins remain poorly understood. Biochemical purifications require detergents to isolate proteins away from lipid molecules, and the large-scale nature of the affinity precipitation/mass spectrometry projects (1, 2) precluded adjusting the detergents for individual integral membrane proteins. Two-hybrid assays (3, 4) require that the two proteins localize to the nucleus; integral membrane proteins, targeted to an aqueous nuclear environment, may aggregate or misfold.

To increase the representation of integral membrane proteins in the protein-protein interaction network of *Saccharomyces cerevisiae*, we examined pair-wise interactions among 705 integral membrane proteins by the split-ubiquitin membrane yeast two-hybrid system (5). This modified form of the split-ubiquitin assay (6–8) is one of several hybrid protein approaches that detect interactions occurring at membranes. The split-ubiquitin membrane yeast two-hybrid system allows direct identification of yeast transformants that encode a pair of interacting proteins by use of a transcriptional reporter.

Analyses of previous large-scale interaction data sets revealed significant numbers of false negatives and false positives (9). False negatives may represent interactions unsuitable for detection by a particular technique and, thus, may not be easily remedied. False positives can potentially be identified by a failure to be validated through additional experiments. However, the large-scale nature of this study and the difficulties associated with biochemical analysis of integral membrane proteins preclude confirmation of these results by alternative experimental approaches. Therefore, we used a learning algorithm, the support vector machine (SVM) (10), to classify the interactions. The SVM algorithm has in recent years been applied to pattern-recognition problems in computational biology including pro-

tein remote homology detection, microarray gene expression analysis, and peptide identification from mass spectrometry data (11). We first trained the SVM on positive examples composed of interactions that were corroborated by independent studies. The SVM “learned” to distinguish these corroborated interactions from a selection of interactions chosen at random. Interactions consistently predicted as true by the SVM are considered highest confidence, whereas those infrequently or never classified as true are the most suspect.

## Materials and Methods

**Yeast Strains, Plasmids, and Stop Codon Removal rePCR.** The reporter strains used were L40 and AMR 70 (12). p415-Cub-PLV-WBP1 and p414-HA<sub>2</sub>-NubG-ADH1, the vectors used for homologous recombination of the ORFs, were constructed by standard methods. Using yeast ORFs with common 20-bp flanking sequences (13) as templates, we used PCR reactions to remove the stop codons. For more details, see *Supporting Methods*, which is published as supporting information on the PNAS web site.

**Selection of Integral Membrane Proteins from Yeast Proteome Database (YPD).** We chose the 642 proteins annotated with “integral membrane environment” on the YPD web page (14) (Dec. 10, 2001) and 63 proteins with similarity to these 642 proteins for a total of 705.

**Construction of an Array of Yeast Expressing ORF-HA<sub>2</sub>-NubG Fusions.** p414-HA<sub>2</sub>-NubG-ADH1 was linearized and cotransformed with PCR product in 96-well plates. Transformants were pinned onto 16 minimal medium (SD) (15) – Trp + Ade Omnitrays in a 96-spot format and incubated at 30°C for 2–3 days. Yeast were replica-pinned to a 384-spot format to generate four plates. Every ORF was present twice in adjacent spots. The array was alternately grown on SD-Trp and yeast extract peptone dextrose (15). Fusions of NubG-HA onto the amino termini of the set of 705 proteins were also generated but behaved promiscuously in the assay and were not used except as positive controls. See *Supporting Methods* for additional details.

Abbreviations: Cub-PLV, C-terminal half of ubiquitin/protein A/lexA/VP16; GO, Gene Ontology; HA<sub>2</sub>-NubG, two hemagglutinin tags fused to the N-terminal half of ubiquitin with an isoleucine to glycine mutation at position 13; Nubl, wild-type N-terminal half of ubiquitin; OST, oligosaccharyltransferase; SD, minimal medium; SVM, support vector machine.

<sup>†</sup>Present address: Department of Biochemistry and Department of Medical Genetics and Microbiology, The Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, 1 Kings College Circle, Toronto, ON, Canada M5S 1A8.

<sup>¶</sup>I.S. is a founder of Dualsystems Biotech in Zurich, which sells reagents and carries out screens using the split-ubiquitin method that is the basis for the work reported here. However, no funding from this company was used to support the work described in this work, and all of the reagents described in this paper will be made available on request by qualified researchers for their own use.

<sup>††</sup>To whom correspondence should be addressed. E-mail: fields@u.washington.edu.

© 2005 by The National Academy of Sciences of the USA

**Screening for Pairwise Interactions by Using the Split-Ubiquitin Membrane Yeast Two-Hybrid.** Screening was carried out as described in ref. 4. A single colony of a sequenced ORF-Cub-PLV fusion in the *MATa* strain L40 that passed a wild-type N-terminal half of ubiquitin (NubI)-binding test was inoculated into 25 ml of yeast extract peptone dextrose broth (15) with 0.5 mg of adenine. The cultures were grown overnight at 30°C, pelleted, resuspended in 8 ml of yeast extract peptone dextrose, and poured onto an Omnitray plate. A BioMek 2000 robot with an HDR-384 pinning tool transferred the suspension onto eight yeast extract peptone dextrose + Ade plates of 384 spots.

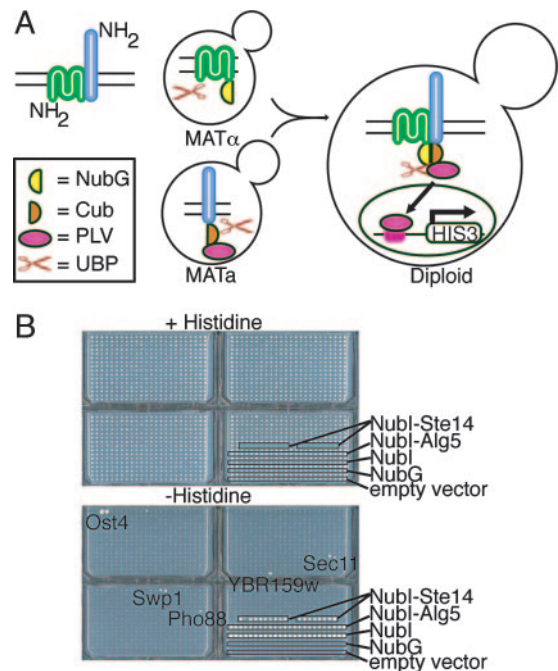
The four plates with the arrayed ORF- $HA_2$ -NubG fusions in the *MAT $\alpha$*  strain AMR70 were replica-pinned onto the ORF-Cub-PLV spots such that every pairwise screen was done in duplicate. Yeast were incubated overnight at 30°C, and spots were then replica-pinned to Omnitrays with SD – Trp – Leu + Ade (15) to select for diploids. These diploids were incubated at 30°C for 2–3 days, then replica-pinned to SD – Trp – Leu – His + Ade + 3 mM 3-amino-1,2,4-triazole and incubated for 1 week at 30°C. Interactions were scanned and scored by histidine phenotype.

**SVM Algorithm.** Full details of the SVM method are found in *Supporting Methods*. Briefly, the SVM was trained on a set of positive examples, and on 300 interactions selected at random as negative training examples. A set of negative examples was selected 100 times, and an SVM trained on each selection. Interactions were classified in each run, and each was scored according to the proportion of SVM runs in which it was classified as true (see Table 1, which is published as supporting information on the PNAS web site).

## Results

**Array Approach for the Split-Ubiquitin Membrane Yeast Two-Hybrid System.** In this approach, one integral membrane protein is fused at its C-terminal residue to two hemagglutinin tags fused to the N-terminal half of ubiquitin with an isoleucine to glycine mutation at position 13 ( $HA_2$ -NubG) (6). A second integral membrane protein is fused at its C terminus to the C-terminal half of ubiquitin and a protein A/lexA/VP16 (Cub-PLV) transcription factor. If the two membrane proteins interact, they reconstitute quasi-native ubiquitin, which is a target of ubiquitin-specific proteases that cleave after the C-terminal residue of ubiquitin. Cleavage releases the transcription factor to enter the nucleus and activate transcription of the *HIS3* reporter gene (Fig. 1A). Although these fusions leave N-terminal signal sequences intact, proteins with their carboxyl termini localized extracellularly or within the lumen of an organelle are not expected to meet the requirements of the assay. However, in some cases, the ubiquitin moiety may disrupt the topology of a membrane protein, such that the ubiquitin domain is exposed in the cytosol and available for detection in the assay. In support of this possibility, we observed interactions for a small number of proteins with characterized extracytosolic carboxyl termini. In addition, ubiquitin fusions may localize to other than the native environment of the integral membrane proteins, yet the assay may still be capable of yielding a signal when this occurs.

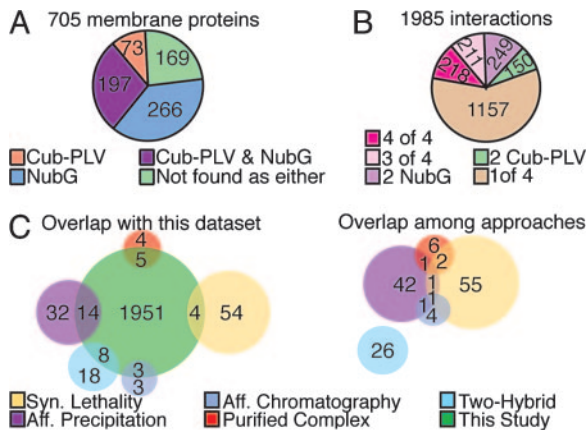
Each Cub-PLV fusion was screened twice against the array, and each  $HA_2$ -NubG fusion was present in two adjacent positions in the array. An example of a screen with the ORF YJR015W is shown (Fig. 1B). The most rapidly growing cells are those indicative of the interaction between Ost4 and YJR015w. Ost4 is a component of the oligosaccharyltransferase (OST) complex, a nine-protein enzyme that glycosylates proteins on asparagine in the sequence Asn-X-Ser/Thr as they translocate into the endoplasmic reticulum (16). Although growth is not as substantial as the Ost4-expressing yeast, yeast containing Swp1- $HA_2$ -NubG provide evidence that Swp1, a second component of



**Fig. 1.** The split-ubiquitin membrane yeast two-hybrid system. (A) Two integral membrane proteins (blue and green) are fused to the two halves of ubiquitin and expressed in cells of opposite mating type. Upon mating, the diploid coexpresses the proteins. An interaction brings the two halves of ubiquitin into close proximity, forming a reconstituted molecule that is cleaved by ubiquitin-specific proteases, releasing the transcription factor to enter the nucleus and activate *HIS3* gene transcription. (B) Yeast expressing YJR015w as a Cub-PLV fusion were mated to an array of yeast expressing the ORF- $HA_2$ -NubGs. *Upper* shows the diploids containing YJR015w-Cub-PLV and each of the  $HA_2$ -NubG fusions or controls on media that selects for the plasmids. *Lower* shows the growth on media selective for reporter gene activation. Putative YJR015w interactors are labeled by the name of the protein fused to the  $HA_2$ -NubG moiety. Black boxes show positive controls (NubI fusions onto the amino terminus of Ste14 and Alg5 and free NubI) and negative controls (NubG alone, empty vector).

the OST complex (17), also interacts with YJR015w. Potentially, this indicates the interaction between Ost4 and YJR015w is direct, whereas that between Swp1 and YJR015w is bridged by another protein, possibly Ost4. Both Ost4-Cub-PLV and Swp1-Cub-PLV, when screened against the array, interacted with YJR015w- $HA_2$ -NubG (data not shown). A third positive that shows growth for both positions represents the Sec11 and YJR015w interaction. Sec11 is a component of the signal peptidase complex, that, like OST, modifies nascent chains during translocation and is thought to be in close proximity to the translocation pore (18).

Before screening, we tested the ability of the ORF-Cub-PLV fusions to assemble with the NubI. Under the conditions of the assay, NubI retains enough affinity to assemble with the Cub moiety fusion proteins and, thus, *HIS3* transcription confirms the functionality of a Cub-PLV fusion. The mutant NubG is not able to bind Cub on its own and serves as a negative control. Based on this test, 365 of the 705 integral membrane proteins were competent for screening. The other 340 proteins may not have assembled with NubI because of failed construction of the fusions, instability, or mislocalization, or protein orientations such that the C terminus was not in the cytoplasm. We screened the array with the 365 functional Cub-PLV fusions and found a total of 1,985 interactions from 270 screens (Table 1). Another 10 of the Cub-PLV fusions appeared to interact with nearly all of the  $HA_2$ -NubG fusions and were therefore discarded, and 85,



**Fig. 2.** The integral membrane protein interaction data set. (A) The number of Cub-PLV and HA<sub>2</sub>-NubG fusions that identified interactions. (B) Reproducibility of the interactions. (C) Overlap of the interaction data with that found in the General Repository for Interaction Data and Database of Interacting Proteins databases of protein interactions. Two additional interactions, found in the literature but absent in the databases, were also included.

although able to confer growth in combination with NubI, did not support growth in combination with any HA<sub>2</sub>-NubG fusion. Overall, 270 Cub-PLV fusions identified 463 HA<sub>2</sub>-NubG fusions in at least one interaction. Of 705 proteins in the starting set, 536 identified at least one interaction partner (Fig. 2A). That some proteins were functional when fused to only one of the two ubiquitin moieties indicates the two fusions are not equivalent and may provide a rationale for why few interactions were observed reciprocally.

Each interaction may be detected up to four times in one orientation and up to eight times if the reciprocal orientation was tested. Of the 1,985 interactions, 38 were found in both orientations. For the rest, unidirectional interactions were observed from one to four times (Fig. 2B). Interactions may be detected fewer than four times because of replica-pinning errors or variations in growth of the yeast during multiple transfers. Positives that were observed only once likely include a fraction arising from stochastic activation of the *HIS3* reporter gene, not reflecting a true interaction.

In total, 34 of the 1,985 interactions (1.7%) had been reported in the literature (Fig. 2C). Among the 536 proteins identified in an interaction, previous experimental approaches had identified only 145 interactions that are present in the General Repository for Interaction Data (<http://biodata.mshri.on.ca/yeast-grid/servlet/SearchPage>). Of these 145 interactions, only 5 were detected by two methods and 1 by three methods. This paucity of overlap may result from the difficulty of applying experimental techniques to integral membrane proteins. However, synthetic lethality experiments should not be biased against membrane proteins, because it is the genes and not the proteins that are manipulated in such approaches. Nevertheless, of the 58 synthetic lethal interactions identified among any of the genes present in our data set, only 4 ( $\approx 7\%$ ) are recapitulated as physical interactions in our study. The fraction of interactions found in common between the split-ubiquitin approach and other physical interaction techniques is considerably higher: 4 of 9 interactions derived from analysis of purified complexes, 14 of 46 from affinity precipitations, 3 of 6 from affinity chromatography experiments, and 8 of 26 from two-hybrid experiments are also observed by using our approach.

**Support Vector Machine Analysis.** Given a data set in which each example is characterized by an  $n$ -dimensional vector, the SVM (11) “learns” a boundary between positive and negative inter-

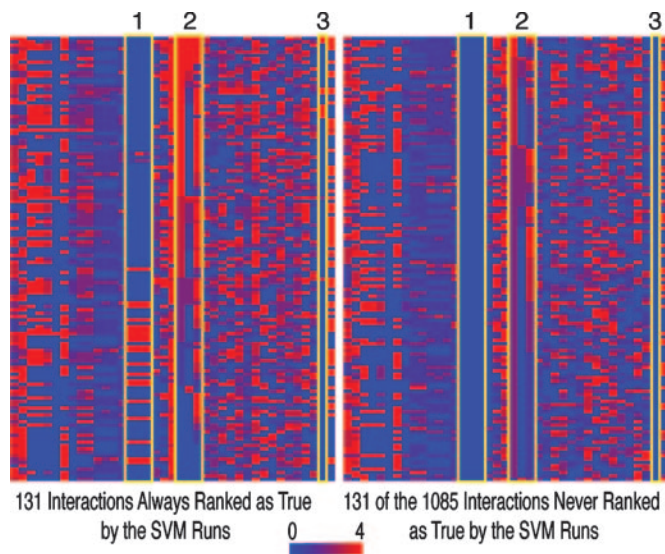
action examples with maximum margin. The remaining uncharacterized interactions in the data set are then classified according to their relationship with respect to the decision boundary. To carry out its rankings, the SVM used characteristics derived from our assay results and additional information that included the Gene Ontology (GO) annotations of biological process, molecular function, and cellular component (19); protein localization (20) (see Fig. 5, which is published as supporting information on the PNAS web site); transcriptional regulation (21); essentiality of the genes encoding the proteins (22); and protein expression level and codon enrichment correlation (23) (*Supporting Methods*). One type of interaction that cannot be supported in this manner is self-oligomerization because the evidence is identical for each of the partner proteins. These 36 interactions (Table 2, which is published as supporting information on the PNAS web site) were separated from the remaining data. Evidence from reproducibility, specificity, and corroboration by other experiments supports 21 of these interactions. Six proteins found to self-associate in this study also exhibit high confidence interactions with proteins highly homologous to them, suggesting the interactions may be mediated through similar domains of the proteins. That only 36 of 536 proteins (7%) yielded a signal for self-oligomerization argues against mere proximity of the two fusions to the same membrane compartment being sufficient for a signal in the assay.

To identify positive training examples, we searched two protein interaction databases, the Database of Interacting Proteins (<http://dip.doe-mbi.ucla.edu>) (24) and the General Repository for Interaction Data, and identified 27 non-self interactions (Table 3, which is published as supporting information on the PNAS web site). Two published interactions (25, 26) not present in either database were also included. We also made use of the paralogous verification method (PVM) (27) available from the Database of Interacting Proteins web site to support 25 additional interactions. The paralogous verification method scores an interaction as likely if paralogous proteins of the interacting pair are known to interact. Lastly, we included two interactions based on a synthetic lethality relationship (28). Mutations of genes for members of a physical complex may share a synthetic lethality relationship, as shown for Sec11 and Spc1, two components of the signal peptidase complex (25). In total, 56 corroborated interactions served as positive examples to train the SVM.

An examination of the features possessed by each of the positive training examples indicates that the examples differ significantly in their characteristics. For instance, the interaction of Emp24 with Erv25 (29) was observed only one of the possible four times, and the transcriptional regulation of the genes encoding these two proteins is highly similar across a number of experiments (21). In contrast, the interaction between Sec11 and Spc1 (30) was observed all four times, in each orientation, and has no significant similarity in the transcript profiles for the genes. The machine learning approach provides a way to integrate different forms of evidence into the prediction of an interaction’s validity.

A negative training set was derived from putative interactions randomly chosen from the data set. The SVM was trained on 100 selections of 300 random interactions and then used to classify the remaining interactions. Random interactions were chosen because we expected a sizable proportion of the total interactions were false positives and not physiologically relevant. An alternative method to choose negative examples is pairs of proteins that are annotated to different cellular compartments (31). However, because colocalization and participation in similar functions or processes are correlated, this method would preclude the use of GO annotations in making the predictions and can introduce additional biases as well (32).

The positive training set was held constant, and each SVM run



**Fig. 3.** Heat maps of SVM classified interactions. The 131 interactions classified as true by all of the SVM runs are compared with 131 (of 1,085) interactions not classified as true by any SVM run. The latter were selected based on the two proteins' GO biological process annotations being the most closely related. Interactions are presented as rows and features as columns (in the order presented in Table 1) in descending order of ranking by related GO annotation. The heat maps are normalized on a scale of zero to four, with the more desirable features for an interaction indicated by a four (red). Yellow boxes highlight: (1) the mutual clustering coefficients versus the whole yeast network, (2) "related GO," and (3) "Ihmels transcriptional modules," which are strongly enriched in the high confidence group relative to the low confidence group. The heat maps were generated by using `MATRIX2PNG` (33).

separated the positive interactions from a random sample of the remaining interactions. The overall ranking was generated as an average of 100 rankings. For a given interaction, this average was calculated by using only SVM rankings in which it was not chosen as a training example. In the absence of a defined set of negative examples, the performance of the method cannot be assessed by using cross-validation. Using a set of negative examples selected based on a lack of colocalization, we obtained an area under the receiver operating characteristic curve score of 0.94 by averaging over 10 runs of 10-fold cross-validation of an SVM with a Gaussian kernel (data not shown).

Using the trained algorithms, we found 131 interactions classified as true by all of the SVM runs; these interactions are our highest confidence set, excluding the 56 interactions that constitute the positive training set. We anticipate that even at this stringency, there may be false positive interactions in this group. There are 209 interactions classified as true in 50–99% of the SVM runs, and 468 interactions classified as true in 1–49% of the SVM runs. The remaining 1,085 interactions were never classified by any of the rankings as true. Some of these interactions may occur *in vivo*, but their features are not sufficiently similar to our positive examples for the algorithm to support them. We compared the 131 high confidence interactions to a set of 131 of the 1,085 lowest confidence interactions to determine which features correlated with the SVM rankings. Certain features, such as shared GO biological process and molecular function, mutual clustering coefficients, and inclusion of the transcripts for the proteins in one or more transcriptional modules (21), were enriched in the highest confidence set (Fig. 3). However, if we removed the feature of "shared GO biological process" from the SVM analysis, 121 of the 131 interactions were still classified as true in all 100 SVM runs. Thus, although shared GO biological process is the most informative feature, additional

features supported classification of most high confidence interactions.

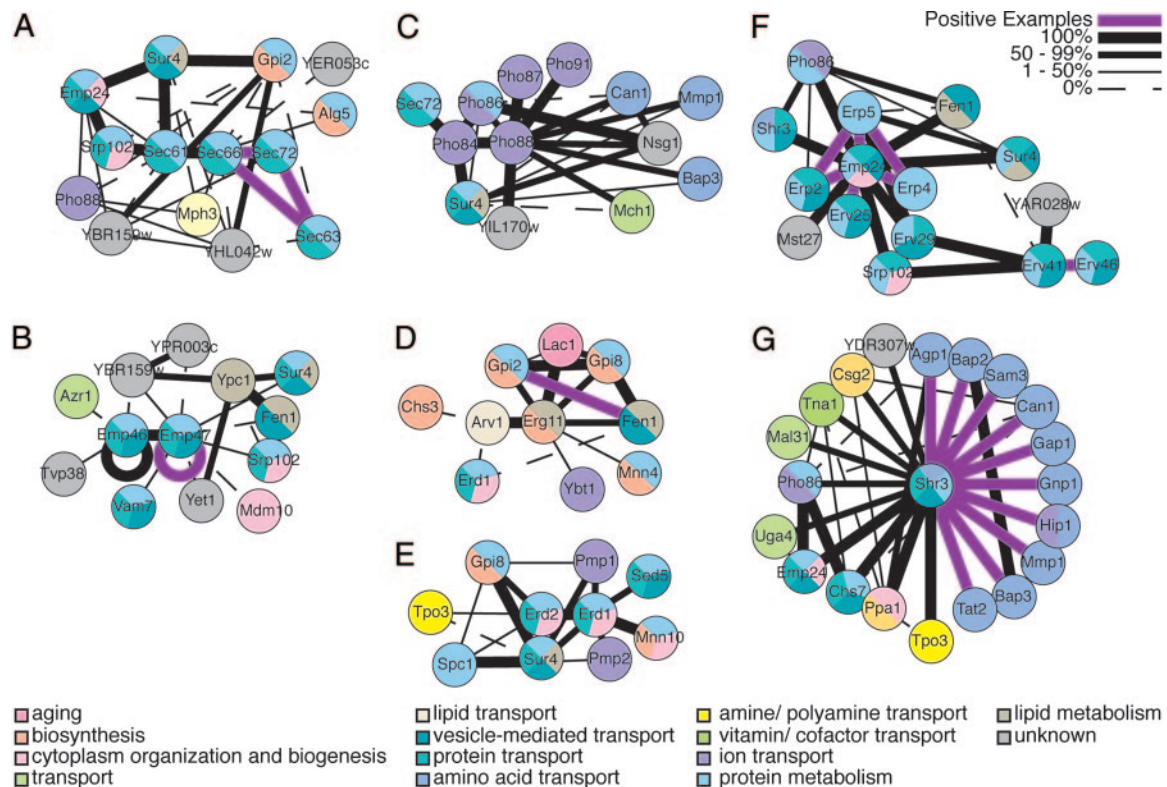
**Examples of Interactions Substantiated by the SVM.** Jansen *et al.* (31) used a Bayesian networks approach to predict protein–protein interactions. In total, 17 of the 1,985 interactions we identified were predicted to be "true-positive" interactions by this Bayesian analysis. Ten had been described previously by other experimental approaches and are thus in the positive training set. Of the seven interactions identified in this study and supported purely based on the computational Bayesian analysis, five are predicted to be true interactions by all of the SVM runs (Table 1). In addition to these five, several more interactions that are classified as true in all of the SVM runs are known from the literature but were not used as positive training examples. For example, Fig. 4A diagrams the high-confidence interactions for the central component of the Sec61 translocation complex that mediates insertion of secretory and membrane proteins into the ER. Sec61 has been copurified with Sec66 as part of a post-translational translocation complex that includes Sec63 and Sec72 (34). Srp102 is part of the receptor complex for the signal recognition particle that targets ribosomes harboring secreted or membrane proteins to the Sec61 complex, and independent lines of evidence indicate it is closely associated with Sec61 components (35, 36).

Emp47 and Emp46 are two proteins that share 46% similarity and have proposed roles in transporting cargo proteins from the ER into COPII vesicles. Emp46 and Emp47 physically interact, this interaction is necessary for the localization of Emp46, and Emp47 homooligomerizes (37). These two interactions are recapitulated in this study (Fig. 4B). Sato and Nakano (37) did not observe self-assembly of Emp46; the Emp46 self-interaction detected here may be due to bridging through a complex with Emp47.

Pho84, a phosphate transporter, identified Pho88 as an interactor (Fig. 4C). The function of Pho88 is unknown, but *pho88* yeast have a phenotype similar to that of *pho86* yeast; both mutants are defective in repression of *PHO81* expression in response to high phosphate in the media (38). Pho86 is necessary for trafficking Pho84 to the plasma membrane, and yeast with mutations in *PHO84* display a similar phenotype to *pho86* yeast (39). Interactions identified here suggest Pho88 and Pho86 may function by binding to Pho84 to promote its maturation or trafficking, suggesting that the absence of Pho88 inhibits Pho84-mediated transport of phosphate into the cell. Pho87 and Pho91, phosphate transporters that share homology, may be chaperoned by Pho88 to reach the plasma membrane in a manner similar to the interaction of Pho86 and Pho84.

An interaction between Arv1 and Erg11, both involved in sterol metabolism, was classified as true with high confidence (Fig. 4D). The genes for the two proteins share synthetic genetic and chemical genetic interactions (40): synthetic lethality between *erg11* and *arv1* strains and hypersensitivity of the *arv1* strain to fluconazole, an antimicrobial compound that targets Erg11. Yeast lacking Arv1 are defective in sphingolipid and sterol metabolism (41) and accumulate lanosterol, the substrate that Erg11 demethylates in the course of ergosterol biosynthesis (42). The identified interaction between these proteins may connect their functions in ergosterol metabolism.

ER retention of many soluble proteins involves their binding to Erd2, the HDEL receptor (43). Our results indicate an interaction between Erd2 and Erd1, suggesting that they may assemble for the purpose of restricting the transport of proteins out of the ER (Fig. 4E). Erd1 was originally isolated on the basis of its role in ER retention, but its specific mechanism of action has not been characterized (44). Our data suggest the role of Erd1 in ER retention is mediated by physical interaction with Erd2.



**Fig. 4.** Interactions validated by the SVM runs. Purple lines indicate interactions used in the positive training set. (A) Interactions of translocation channel proteins. (B) Interactions of Emp47 and Emp46. (C) Interactions of Pho88 and Pho84. For clarity, only the direct interaction partners of Pho84 and Pho88 that are identified as true by >50% of the SVM runs are presented; interconnections among proteins in the cluster of less confidence are included. (D) Interactions of sterol metabolism proteins. Only direct interactions of Erg11 found to be true by at least one SVM run are shown. (E) Interactions involved in HDL-mediated ER retention. Only interactions of Erd1 and Erd2 identified as true by at least one SVM run are presented. (F) Interactions of COPII vesicle proteins. Only interactions of Emp24 supported by all SVM runs are shown. (G) Interactions of Shr3. Only interactions predicted as true by at least half of the SVM runs are shown.

The SVM classified a cluster of interactions between proteins of COPII-coated vesicles as true. High confidence interactions were observed between Emp24 and Erv29, and between Erv29 and Erv41 (Fig. 4F). Erp5, a paralogue and paralogous verification method-supported interactor of Emp24, also interacted with Erv29. Mst27, a protein involved in transport of proteins mediated by COPII vesicles, was connected by a high confidence interaction to Emp24. These interactions are likely part of the assemblage of proteins involved in sorting proteins into vesicles, necessary perhaps for formation of the vesicles or for packaging the cargo proteins into them. In addition, a high confidence interaction linked Erv41 to YAR028w, a protein with similarity to Mst27, but of unknown function.

The Shr3 protein acts within the secretory pathway as a chaperone to promote the trafficking of amino acid transporters from the ER to the plasma membrane. Many of the interactions of Shr3 are present in the positive training examples because they have been observed by other approaches or because they fit the criteria for the paralogous verification method. Three other interactions of Shr3 with related transporters are classified as high confidence interactions (Fig. 4G). Uga4 is a GABA transporter (45), Tna1 is a nicotinamide transporter (46), and Tpo3 is a spermidine transporter (47). Our results lead to the idea that Shr3 may serve as a more general chaperone for transporter proteins. A high confidence interaction of Shr3 with the Emp24 protein suggests that the mechanism by which Shr3 mediates trafficking of transporters could involve its assembly with the p24 family of ER-derived vesicle proteins.

## Discussion

Studies aimed at identifying protein–protein interactions largely underrepresent interactions that involve integral membrane pro-

teins. This underrepresentation is due not only to the hydrophobic nature of these proteins but also reflects the fact that previous interaction studies have focused on soluble proteins. In this study, we applied the split-ubiquitin membrane yeast two-hybrid system, which can detect and report interactions between integral membrane proteins, on a large-scale basis. As anticipated, we identified many interactions that had not been observed by either systematic or small-scale approaches, which may reflect the ability of split-ubiquitin-based systems to detect transient interactions (7). Interactions identified between proteins annotated to unrelated processes may reflect this transient nature. Alternatively, annotations of proteins to biological processes may be incomplete or incorrect, or some interactions may lack physiological relevance.

The full data set contains both individual interactions and clusters of interactions that reflect endogenous assemblies in the cell and many likely false-positive interactions due to the nature of the assay. An interaction previously observed in another experiment can be considered of high confidence. However, novel interactions can be assigned confidence based on the behavior of the proteins in the assay itself or on other data such as protein colocalization, shared GO annotations, or shared mutant phenotypes. To combine the information from these different sources of data, we used an SVM classifier trained on a set of interactions considered to be biologically relevant based on their confirmation by other methods. The high confidence interaction set is enriched for pairs that share biological process and molecular function GO annotations and whose genes' transcription is coregulated across multiple microarray experiments, suggesting that many are physiologically relevant interactions. Additional relevant interactions

likely exist even in the lower confidence sets, and further functional studies are necessary to distinguish the true interactions from the false positives. This data set of integral membrane protein interactions should provide a complementary sampling of protein interaction space to that sampled by other experimental strategies.

We thank B. Byers, S. Emr, and A. Merz for critical reading of the manuscript; J. Hesselberth for assistance in the computational analysis;

S. Thamyin for unpublished reagents; and T. Hazbun and S. McCraith for helpful discussions. This work was funded by the National Center for Research Resources of the National Institutes of Health Grant P41 RR11823. W.S.N. was funded by National Institutes of Health Grant R33 HG003070, and I.S. was funded by grants from the Union Bank of Switzerland, Gebert R uf Foundation, the Swiss Cancer League, and the Swiss National Science Foundation. W.S.N. is an Alfred P. Sloan Research Fellow. S.F. is an investigator of the Howard Hughes Medical Institute.

- Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M., *et al.* (2002) *Nature* **415**, 141–147.
- Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S. L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., *et al.* (2002) *Nature* **415**, 180–183.
- Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M. & Sakaki, Y. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 4569–4574.
- Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., *et al.* (2000) *Nature* **403**, 623–627.
- Stagljar, I., Korostensky, C., Johnsson, N. & te Heesen, S. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 5187–5192.
- Johnsson, N. & Varshavsky, A. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 10340–10344.
- Johnsson, N. & Varshavsky, A. (1994) *EMBO J.* **13**, 2686–2698.
- Wittke, S., Lewke, N., Muller, S. & Johnsson, N. (1999) *Mol. Biol. Cell* **10**, 2519–2530.
- von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S. G., Fields, S. & Bork, P. (2002) *Nature* **417**, 399–403.
- Boser, B. E., Guyon, I. M. & Vapnik, V. N. (1992) in *Fifth Annual ACM Workshop on COLT* (ACM, Pittsburgh).
- Noble, W. S. (2004) in *Kernel Methods in Computational Biology*, eds. Schoelkopf, K. T. B. & Vert, J.-P. (MIT Press, Cambridge, MA), pp. 71–92.
- Hollenberg, S. M., Sternglanz, R., Cheng, P. F. & Weintraub, H. (1995) *Mol. Cell. Biol.* **15**, 3813–3822.
- Hudson, J. R., Jr., Dawson, E. P., Rushing, K. L., Jackson, C. H., Lockshon, D., Conover, D., Lanciault, C., Harris, J. R., Simmons, S. J., Rothstein, R. & Fields, S. (1997) *Genome Res.* **7**, 1169–1173.
- Costanzo, M. C., Crawford, M. E., Hirschman, J. E., Kranz, J. E., Olsen, P., Robertson, L. S., Skrzypek, M. S., Braun, B. R., Hopkins, K. L., Kondu, P., *et al.* (2001) *Nucleic Acids Res.* **29**, 75–79.
- Sherman, F., Fink, G. & Hicks, J. B. (1986) *Methods in Yeast Genetics: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY).
- Karaoglu, D., Kelleher, D. J. & Gilmore, R. (1997) *J. Biol. Chem.* **272**, 32513–32520.
- Kelleher, D. J. & Gilmore, R. (1994) *J. Biol. Chem.* **269**, 12908–12917.
- Johnson, A. E. & van Waes, M. A. (1999) *Annu. Rev. Cell Dev. Biol.* **15**, 799–842.
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., *et al.* (2000) *Nat. Genet.* **25**, 25–29.
- Huh, W. K., Falvo, J. V., Gerke, L. C., Carroll, A. S., Howson, R. W., Weissman, J. S. & O’Shea, E. K. (2003) *Nature* **425**, 686–691.
- Ihmels, J., Friedlander, G., Bergmann, S., Sarig, O., Ziv, Y. & Barkai, N. (2002) *Nat. Genet.* **31**, 370–377.
- Winzeler, E. A., Shoemaker, D. D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J. D., Bussey, H., *et al.* (1999) *Science* **285**, 901–906.
- Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O’Shea, E. K. & Weissman, J. S. (2003) *Nature* **425**, 737–741.
- Salwinski, L., Miller, C. S., Smith, A. J., Pettit, F. K., Bowie, J. U. & Eisenberg, D. (2004) *Nucleic Acids Res.* **32**, Suppl. 1, D449–D451.
- Mullins, C., Meyer, H. A., Hartmann, E., Green, N. & Fang, H. (1996) *J. Biol. Chem.* **271**, 29094–29099.
- Kane, P. M., Tarsio, M. & Liu, J. (1999) *J. Biol. Chem.* **274**, 17275–17283.
- Deane, C. M., Salwinski, L., Xenarios, I. & Eisenberg, D. (2002) *Mol. Cell Proteomics* **1**, 349–356.
- Beeler, T., Bacikova, D., Gable, K., Hopkins, L., Johnson, C., Slife, H. & Dunn, T. (1998) *J. Biol. Chem.* **273**, 30688–30694.
- Belden, W. J. & Barlowe, C. (1996) *J. Biol. Chem.* **271**, 26939–26946.
- Fang, H., Panzner, S., Mullins, C., Hartmann, E. & Green, N. (1996) *J. Biol. Chem.* **271**, 16460–16465.
- Jansen, R., Yu, H., Greenbaum, D., Kluger, Y., Krogan, N. J., Chung, S., Emili, A., Snyder, M., Greenblatt, J. F. & Gerstein, M. (2003) *Science* **302**, 449–453.
- Ben-Hur, A. and Noble, W. S. (2005) *Bioinformatics* **21**, Suppl. 1, i38–i46.
- Pavlidis, P. & Noble, W. S. (2003) *Bioinformatics* **19**, 295–296.
- Brodsky, J. L. & Schekman, R. (1993) *J. Cell Biol.* **123**, 1355–1363.
- Wittke, S., Dunnwald, M., Albertsen, M. & Johnsson, N. (2002) *Mol. Biol. Cell* **13**, 2223–2232.
- Helmers, J., Schmidt, D., Glavy, J. S., Blobel, G. & Schwartz, T. (2003) *J. Biol. Chem.* **278**, 23686–23690.
- Sato, K. & Nakano, A. (2003) *Mol. Biol. Cell* **14**, 3055–3063.
- Yompakdee, C., Ogawa, N., Harashima, S. & Oshima, Y. (1996) *Mol. Gen. Genet.* **251**, 580–590.
- Lau, W.-T. W., Howson, R. W., Malkus, P., Schekman, R. & O’Shea, E. K. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 1107–1112.
- Parsons, A. B., Brost, R. L., Ding, H., Li, Z., Zhang, C., Sheikh, B., Brown, G. W., Kane, P. M., Hughes, T. R. & Boone, C. (2004) *Nat. Biotechnol.* **22**, 62–69.
- Swain, E., Stuke, J., McDonough, V., Germann, M., Liu, Y., Sturley, S. L. & Nickels, J. T., Jr. (2002) *J. Biol. Chem.* **277**, 36152–36160.
- Aoyama, Y., Yoshida, Y., Sonoda, Y. & Sato, Y. (1989) *J. Biol. Chem.* **264**, 18502–18505.
- Semenza, J. C., Hardwick, K. G., Dean, N. & Pelham, H. R. (1990) *Cell* **61**, 1349–1357.
- Hardwick, K. G., Lewis, M. J., Semenza, J., Dean, N. & Pelham, H. R. (1990) *EMBO J.* **9**, 623–630.
- Andre, B., Hein, C., Grenson, M. & Jauniaux, J. C. (1993) *Mol. Gen. Genet.* **237**, 17–25.
- Llorente, B. & Dujon, B. (2000) *FEBS Lett.* **475**, 237–241.
- Tomitori, H., Kashiwagi, K., Asakawa, T., Kakinuma, Y., Michael, A. J. & Igarashi, K. (2001) *Biochem. J.* **353**, 681–688.