Cell, Vol. 118, 31–44, July 9, 2004, Copyright ©2004 by Cell Press

# Exploration of Essential Gene Functions via Titratable Promoter Alleles

**Sanie Mnaimneh,[1,5] Armaity P. Davierwala,[1,5] Jennifer Haynes,[2,5] Jason Moffat,[2] Wen-Tao Peng,[1] Wen Zhang,[1,2] Xueqi Yang,[1] Jeff Pootoolal,[1] Gordon Chua,[1] Andres Lopez,[1] Miles Trochesset,[1] Darcy Morse,[3] Nevan J. Krogan,[1,2] Shawna L. Hiley,[1] Zhijian Li,[1,2] Quaid Morris,[1] Jörg Grigull,[1] Nicholas Mitsakakis,[1] Christopher J. Roberts,[4] Jack F. Greenblatt,[1,2] Charles Boone,[1,2] Chris A. Kaiser,[3] Brenda J. Andrews,[2] and Timothy R. Hughes[1,2,*]**

[1]Banting and Best Department of Medical Research
University of Toronto
112 College Street
Toronto, ON M5G 1L6
Canada

[2]Department of Medical Genetics and Microbiology
University of Toronto
1 Kings College Circle
Toronto, ON
Canada

[3]Department of Biology
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

[4]Rosetta Inpharmatics LLC, a wholly owned subsidiary of Merck & Co., Inc.
401 Terry Avenue North
Seattle, Washington 98109

## Summary

**Nearly 20% of yeast genes are required for viability, hindering genetic analysis with knockouts. We created promoter-shutoff strains for over two-thirds of all essential yeast genes and subjected them to morphological analysis, size profiling, drug sensitivity screening, and microarray expression profiling. We then used this compendium of data to ask which phenotypic features characterized different functional classes and used these to infer potential functions for uncharacterized genes. We identified genes involved in ribosome biogenesis (*HAS1*, *URB1*, and *URB2*), protein secretion (*SEC39*), mitochondrial import (*MIM1*), and tRNA charging (*GSN1*). In addition, apparent negative feedback transcriptional regulation of both ribosome biogenesis and the proteasome was observed. We furthermore show that these strains are compatible with automated genetic analysis. This study underscores the importance of analyzing mutant phenotypes and provides a resource to complement the yeast knockout collection.**

## Introduction

The budding yeast *Saccharomyces cerevisiae* is one of the most thoroughly characterized organisms in molecular biology, due in part to the ease of making targeted integrations in the genome. Recently, a consortium of laboratories created a collection of isogenic knockouts for virtually all yeast genes, enabling an unparalleled degree of unbiased functional analysis (Giaever et al., 2002). We previously utilized this collection to generate a "compendium" of microarray expression profiles (Hughes et al., 2000). Application of statistical classification algorithms to this compendium showed that patterns of transcriptional changes could be used to distinguish perturbation of diverse cellular functions and thereby assign gene function or link bioactive compounds to their intracellular target pathways (Hughes et al., 2000).

A major outcome of the deletions consortium effort was the demonstration that 1,105 (18.7%) of the ~5,800 protein-coding yeast genes (Cliften et al., 2003; Kellis et al., 2003) are essential for haploid viability under standard laboratory conditions (growth at 30°C in rich medium with glucose as the carbon source). Transcription, splicing, ribosome biosynthesis, translation, cell wall and membrane biogenesis, DNA replication, nuclear transport, and basic cytoskeletal functions are all required for cell proliferation, and genes involved in these processes tend to be essential. Specific aspects of other cellular functions are also required for viability. For example, although the vast majority of mitochondrial proteins are nonessential, some components involved in mitochondrial import and iron-sulfur complex assembly are essential, presumably because of the role of mitochondria in cellular iron metabolism (Baker and Schatz, 1991; Lill and Kispal, 2000). Essential genes tend to be more highly conserved in humans; 38% of essential yeast proteins have counterparts in humans, versus 20% for nonessential genes (with a Blastp E value of E-50; Hughes, 2002).

The precise molecular and genetic functions of many essential yeast proteins have not been studied in detail, at least in part because it is difficult to study essential genes using deletion mutants. Heterozygous diploids have proven to be useful in screens for haploinsufficiency in the presence of specific insults (Giaever et al., 1999, 2004; Lum et al., 2004), but in the absence of an insult most of them have little or no growth defect in rich medium and do not manifest the full range of phenotypes associated with alleles that cause growth inhibition (C.R. and T.R.H., unpublished data). Genetic analysis of essential proteins has traditionally relied on conditional mutants, typically temperature-sensitive (ts) alleles (e.g., Hartwell et al., 1970). A systematic method for constructing ts alleles involves the construction of a gene fusion of the coding sequence to create a heat-inducible-degron domain, which modulates the stability of the protein (Dohmen et al., 1994; Kanemaki et al., 2003). Another systematic approach involves replacement of the native promoter with one that can be rapidly repressed. We and other researchers (Hughes et al., 2000; Peng et al., 2003; Zhang et al., 2003) have employed the tetracycline (tet)-regulatable promoter (Gari et al., 1997) to create mutant alleles of essential genes.

*Correspondence: t.hughes@utoronto.ca
[5]These authors contributed equally to this work.

A major advantage of promoter-replacement systems is that the native open reading frame of the gene is maintained. Moreover, with the tet system, repression is controlled by addition of doxycycline to the growth medium, which has little effect on yeast physiology and virtually no effect on global gene expression at concentrations used for promoter shutoff, which is critical for analysis using microarray expression profiling (Hughes et al. 2000; Peng et al., 2003).

To facilitate uniform and systematic analysis of yeast essential genes, we have now created $TetO_7$-promoter alleles for over two-thirds of all yeast essential genes. We describe an initial phenotypic characterization of 602 of these strains that encompasses morphological characterization, size profiling, drug sensitivity analysis, and expansion of our previous "compendium" of expression profiles to include 215 of the $TetO_7$-promoter mutants. We show that precise statistical inferences regarding gene functions can be drawn from patterns within all of these data types, and we present in-depth examination of several mutants in previously uncharacterized essential genes. These strains and the data presented here are a critical counterpart to sequence features, localization, and protein-protein interactions that have been compiled on a subset of yeast essential genes (Rout et al., 2000; Uetz et al., 2000; Ito et al., 2001; Gavin et al., 2002; Ho et al., 2002; Kumar et al., 2002; Huh et al., 2003; Hazbun et al., 2003; Sickmann et al., 2003; Krogan et al., 2004). In some cases, however, we obtained dramatic phenotypes from mutants in genes for which there is little or no functional information from sequence or other large-scale studies. These strains are publicly available, and since they are derived from the deletions consortium wild-type MATa strain, they are compatible with most screens designed to use haploid yeast cells. To illustrate this, we show that the array of $TetO_7$-promoter mutants is compatible with synthetic genetic array analysis (Tong et al., 2001, 2004).

## Results

### Construction of 602 $TetO_7$-Promoter Strains

We constructed 602 yeast strains, each with a $kan^R$-$tetO_7$-TATA cassette integrated into the promoter of a different essential gene in strain R1158 (Hughes et al., 2000) (Figure 1A). In the presence of doxycycline (10 μg/ml), 74% (447) yielded colonies smaller than wild-type upon visual inspection (Figures 1B and 1C). Among these, 20% (123) yielded smaller colonies even in the absence of doxycycline (Figure 1B, boxed in red; Figure 1C), suggesting that these genes are highly sensitive to perturbation of promoter activity. Colony size reflected growth rate in liquid (Figure 1D).

### Systematic Phenotypic Analysis

We applied four general purpose phenotypic assays to initially characterize the 602 strains. First, we examined the morphology of each strain in the presence and absence of 10 μg/ml doxycycline by manually scoring each strain with one or more of 16 morphological descriptors. This confirmed expected phenotypes for many previously characterized genes: for example, the $TetO_7$-*CDC53* strain displayed an "elongated" morphology, as previously described (Chun and Goebl, 1997) (Figure 2A), and several mutants in the secretion pathway displayed a "wide bud neck" ($TetO_7$-*SEC8* is shown in Figure 2A). Second, we measured the size distribution of individual cells in cultures of each strain using a Coulter Channelizer (Figure 2B). Third, we measured colony size under 15 diverse growth conditions, using a colony arraying robot (Tong et al., 2001) (Figure 2C). Fourth, we examined 215 of the $TetO_7$-promoter strains by microarray expression profiling (Figure 2D). These 215 strains were hand selected to represent a wide variety of functional classes and to encompass all mutants in uncharacterized genes that displayed a growth defect.

### Phenotypic Features that Characterize Mutants in Specific Functional Categories

We next sought to objectively determine whether specific features or patterns in these data correlated with known functions of the mutated genes, and whether these patterns could be used as diagnostics to identify new genes with related functions. While clustering analysis (Eisen et al., 1998) (e.g., Figure 2D) is useful for gaining an initial overview of the data, it is not ideal for surveying all relationships among gene classes and features in the data set (Wu et al., 2002; Gasch and Eisen, 2002) and it does not automatically determine how well patterns in the data correspond to distinct functional classes. Therefore, we applied Support Vector Machines (SVM) (Brown et al., 2000), a machine learning algorithm (i.e., a computer program) that seeks features or patterns of features in the data that separate two classes, e.g., genes that are known to be in a functional category and those that are known to not be in the category. SVM has previously been established to work well for predicting gene functions in yeast on the basis of microarray expression data, and it can also be applied to other types of data that can be represented in a spreadsheet, such as the size, drug sensitivity, and morphology data that we collected. SVM considers each functional category separately and outputs a single "discriminant value" for each gene that represents relative confidence that the gene is in the given category. These values can be processed to obtain an estimate of the probability that the prediction for each gene in each category is correct (i.e., "precision"), on the basis of how well previously annotated genes in the given category can be distinguished from previously annotated genes that are not in the category. This is accomplished by performing the analysis several times, on each repetition "masking" the known functions of a portion of the genes to ask how well they are classified. In this way, SVM can identify which gene categories are distinguishable in the data and can also provide a relative confidence that an uncharacterized gene is in any given category, on the basis of the phenotype displayed by the mutant in that gene.

For each of 643 gene ontology (Ashburner et al., 2000) "biological process" categories represented among the 602 mutants, we trained separate SVMs for each data set. We also trained SVMs for each data set in which the mutants considered were restricted to the 215 mutants that were included in the microarray analysis. After excluding eleven broad categories in which predictions

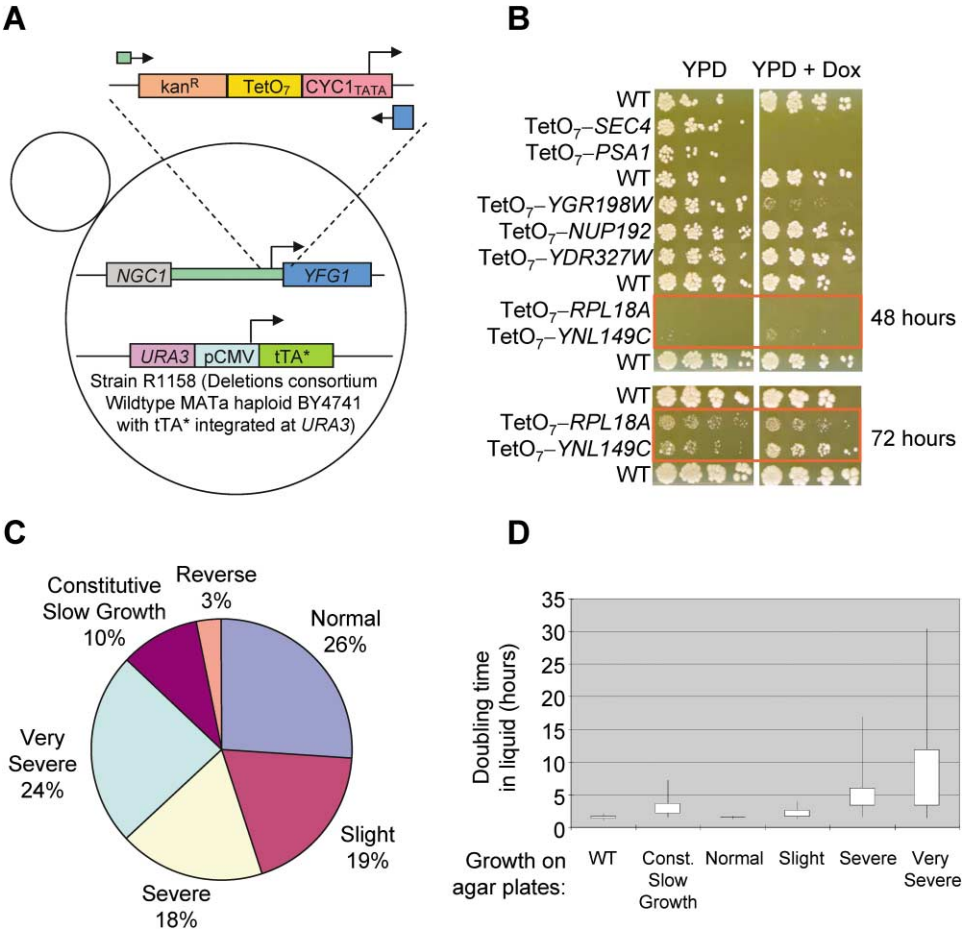


Figure 1. Construction and Growth Characteristics of tet-Promoter Strains

(A) Strain construction by one-step homologous replacement of native promoters with a $TetO_7$-containing cassette. *YFG1*, your favorite gene; *NGC1*, next gene on chromosome.

(B) Examples of different growth phenotypes on YPD-agar plates. $TetO_7$-*SEC4* and $TetO_7$-*PSA1,* very severe; $TetO_7$-*YGR198W*, severe; $TetO_7$-*NUP192,* normal; $TetO_7$-*YDR327W,* reverse (i.e., growth slightly better in presence of doxycycline); $TetO_7$-*RPL18A,* $TetO_7$-*YNL149C*, constitutive slow growth.

(C) Proportion of the 602 strains with different growth phenotypes. A subset of strains in the slight, severe, and very severe categories also exhibit constitutive slow growth that is accentuated by growth on doxycycline.

(D) Correspondence between growth on YPD-agar plates and doubling time measured in liquid SC medium (2% dextrose) between 18 and 24 hr after addition of doxycycline (10 μg/ml).

do not lead to testable hypotheses (listed in Supplemental Data at http://www.cell.com/cgi/content/full/118/1/31/DC1), we asked how many categories and how many genes achieved precision values of over 50% (i.e., in a blind test, among previously annotated genes identified by the algorithm as being in the class, at least 50% are indeed in that class). This was accomplished for 61 categories in the morphology data, 43 categories in the size data, 64 categories in the drugs data, and 79 categories in the microarray expression data. Using the same SVMs, a total of 75 unannotated genes were classified in at least one category with at least 50% precision by at least one assay. Frequently, these classifications were supported by multiple data types, particularly if lower confidence thresholds were allowed. All of the predictions with precision values above 5%, for both characterized and uncharacterized genes, are found in Supplemental Table S1 on the *Cell* website. Figure 3 shows phenotypic data for mutants in selected categories that were identified in this analysis as being "predictable" by one or more data types. These mutants represent genes that were previously known to be in the category as well as those that are not known to be in the category. The drug/media sensitivity profiles were less informative for these categories and are not shown.

Since it is difficult to determine exactly which parts of the data the SVM uses to make each prediction, we used Fisher's score to ask which of the individual features in the morphology, drugs, and expression data distinguish mutants in each of the functional categories (values for all measurements in all categories are listed in Supplemental Data online). For example, among the mutants in Figure 3, the morphology category "amorphous" was characteristic of mutants in ubiquitin-dependent protein catabolism, whereas the categories "large" and "round" were characteristic of mutants in

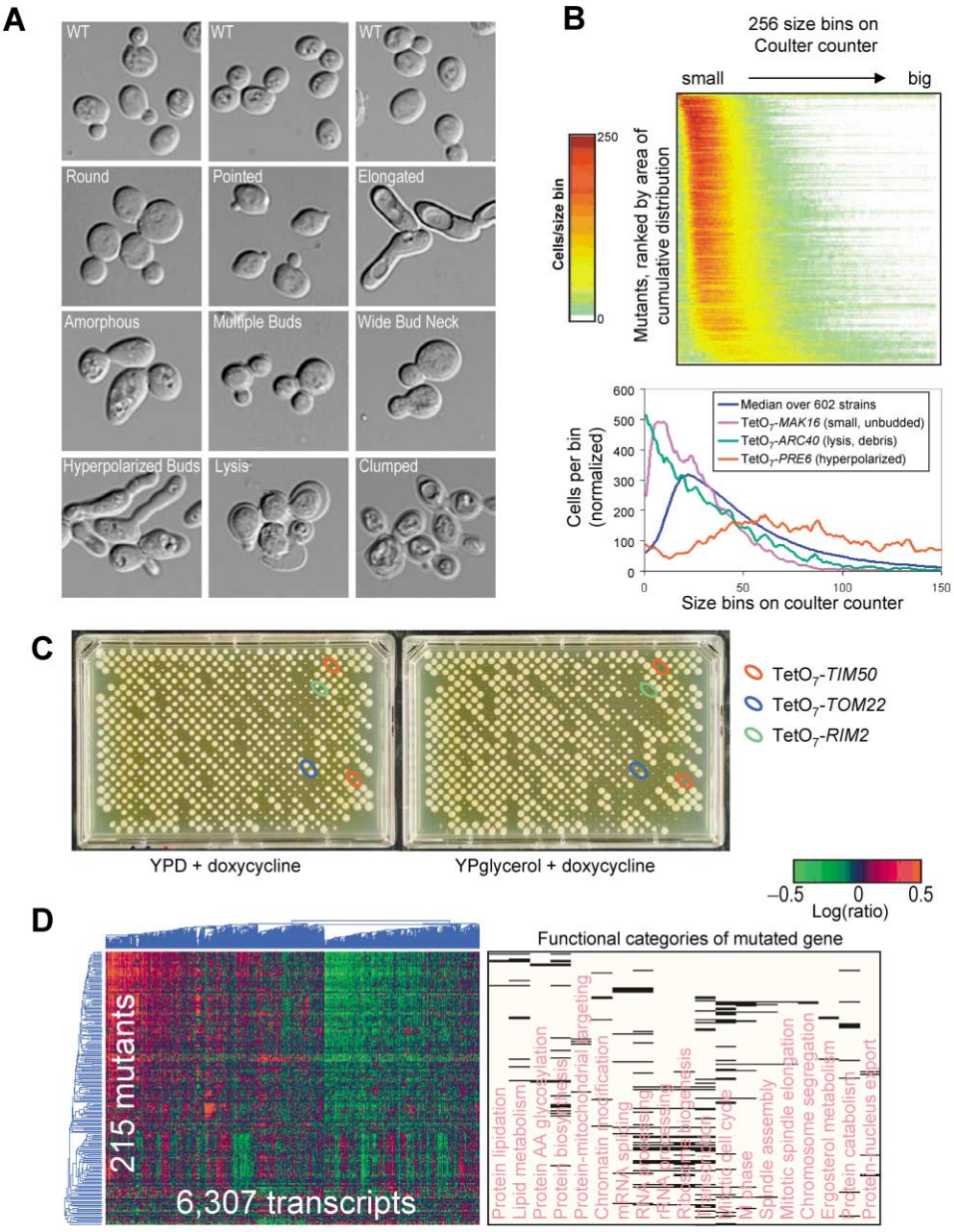


Figure 2. Four Phenotypic Assays Applied to the $TetO_7$-Promoter Strains

(A) Morphological characteristics, assayed by microscopy. The strains shown are round, $TetO_7$-*YIP1*; pointed, $TetO_7$-*NUP192*; elongated, $TetO_7$-*CDC53*; amorphous, $TetO_7$-*RPT6*; multiple buds, $TetO_7$-*COG4*; wide bud neck, $TetO_7$-*SEC8*; hyperpolarized buds, $TetO_7$-*PRE6*; lysis, $TetO_7$-*ARC40*; clumped, $TetO_7$-*HRR25*.

(B) *Top*, cell size distributions, from smallest to largest overall distribution. *Bottom*, examples of cell size distributions from individual strains, in comparison to the median cell size distribution over all experiments (blue line).

(C) Growth phenotypes assayed by robotically depositing colonies on plates.

(D) 2D hierarchical clustering analysis (Eisen et al., 1998) of microarray expression data for 6,307 yeast genes in 215 different $TetO_7$-promoter strains. The full data matrix is given in the Supplemental Data. Representation of nineteen GO categories among the mutants, selected on the basis of being "learnable" by SVM, are shown at right.

the secretory pathway. The microarray measurements, however, were able to correctly distinguish the largest number of distinct categories (Figure 3 and Supplemental Data). In several cases, the identities of the genes whose transcript levels increased are consistent with what is known about the functional categorization of the mutants that caused their induction. For example, mutants in genes involved in amino acid activation featured transcriptional induction of genes involved in amino acid biogenesis; these are most likely regulated by Gcn4p, a transcription factor that is known to regulate amino acid biogenesis and whose translation is coupled to tRNA charging (Vazquez de Aldana et al., 1994) (Figure 3). Mutants in genes involved in protein-mitochondrial

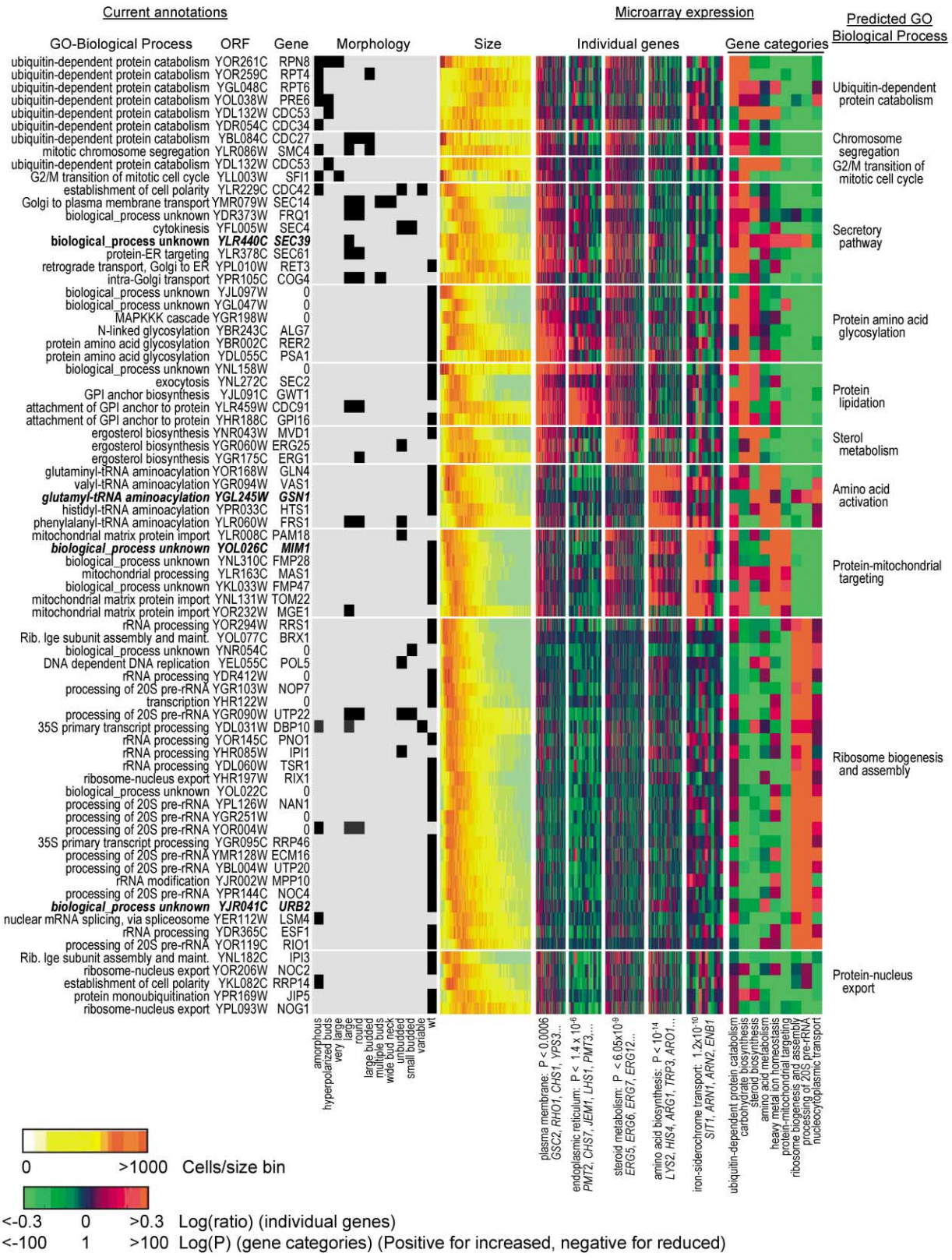


**Figure 3. Relationships among Gene Functions and Phenotypes Allows Systematic Prediction of the Functions of Uncharacterized Genes**

Phenotypic data (morphology, size profiles, and microarray expression) are shown for mutants in functional classes that can be distinguished by SVM in at least one data type. Mutants in bold at left are further characterized in Figures 5 and 6. The microarray measurements shown were selected on the basis of Fisher's score (see Experimental Procedures). The labels at bottom indicate gene properties enriched among the separate groups and representative genes. The rightmost panel of microarray data represents the behavior of functional classes of genes where the values indicated are p values output by the Kolmogorov-Smirnov test (Sachs, 1982), which estimates the likelihood (i.e., assigns a p value) that the distribution of a subset of ranks is nonrandom. Not all mutants are shown for all categories.

transport specifically featured transcriptional induction of genes involved in iron transport, which is consistent with the idea that iron regulation is the essential mitochondrial function (Baker and Schatz, 1991; Lill and Kispal, 2000) (Figure 3). Genes specifically induced upon perturbation of sterol metabolism included many members of the sterol metabolic pathway (Figure 3), as previously described (Daum et al., 1998; Hughes et al., 2000; Vik and Rine, 2001), reconfirming what appears to be negative feedback regulation of this essential pathway.

### Negative Feedback Regulation of *Trans*-Acting Ribosome Biogenesis Factors

To more systematically identify examples of feedback transcriptional regulation, we examined whether the overall induction or repression of genes in entire functional categories were characteristic of mutants in any given category. The Supplemental Data contain the full analysis. Selected results are shown at the right in Figure 3 and expose characteristic trends that are not evident from looking at transcriptional responses of individual genes. For example, the proteasome appears to be subject to negative feedback regulation at the level of transcription, consistent with results obtained previously using chemical inhibition (Fleming et al., 2002). Another prominent example is that the majority of mutants affecting ribosome biogenesis and assembly displayed a specific induction of genes encoding *trans*-acting factors involved in ribosome biogenesis and assembly. This suggests that, like many other cellular activities, this process is controlled by a negative feedback loop that regulates transcript levels. Although autoregulation of the ribosome itself has been previously observed (Zhao et al., 2003), to our knowledge feedback control of transcripts encoding the *trans*-acting factors has not been previously noted. Figure 4 shows in detail the identities of relevant mutants and individual induced genes in these two categories.

### Verification of Inferred Gene Functions

As with any screening approach, directed experimentation (preferably coupled with independently obtained supporting observations) is required to characterize gene and protein functions with confidence. To show explicitly that patterns and features in the data are predictors of new gene functions, we further pursued de novo predictions made regarding the functions of uncharacterized genes in several different functional categories.

### HAS1 Is Required for 18S rRNA Biogenesis

In addition to the genes encoding the 78 ribosomal proteins, over 100 essential genes are already known to be involved in ribosome biogenesis (Fatica and Tollervey, 2002). In our analyses, depletion of these factors was not only characterized by the aforementioned transcriptional response, but also by generally smaller cell size and absence of other morphological defects (Figure 3), as has been previously described for nonessential genes (Jorgensen et al., 2002). These phenotypes were shared by $TetO_7$-promoter mutants in several poorly characterized genes, which we then examined by other assays. In some cases, our data confirmed gene functions that have been suspected from sequence features, physical associations, and/or localization. For example, Has1p ("helicase associated with Set1") has been found in physical association with a variety of nucleolar rRNA processing complexes (Gavin et al. 2002; Ho et al., 2002; Fatica et al., 2002) and is localized to the nucleus, nuclear membrane, and/or nucleolus (Rout et al., 2000; Kumar et al., 2002). In the $TetO_7$-*HAS1* mutant, we observed a striking loss of 18S rRNA and 20S pre-rRNA, as well as accumulation of the 35S pre-rRNA and the aberrant 23S species ($A_3$-$B_1$), which is indicative of a defect in cleavage of the pre-rRNA at sites $A_0$, $A_1$, or $A_2$ (Fatica and Tollervey, 2002) (Figures 5A and 5B). Virtually identical results were published recently in an independent study (Emery et al., 2004).

### YKL014C (URB1) and YJR041C (URB2) Are Required for Normal Ribosome Biogenesis

We also observed a reduction in levels of both 18S and 25S rRNAs in $TetO_7$-promoter mutants of *YKL014C* and *YJR041C* (Figure 5C). We confirmed by pulse-chase analysis that there is a delay in rRNA synthesis in both of these mutants (Figure 5D; $TetO_7$-*YJR041C* has a more severe growth defect and displays a more prominent delay). Both of the encoded proteins have been localized to the nucleolus (Huh et al., 2003), and although they have been detected in physical association with rRNA processing proteins (Gavin et al., 2002; Ho et al., 2002), they have also been reported to be associated with proteins in other processes (Ho et al., 2002). Using a high-stringency affinity purification procedure (Krogan et al., 2004), we found that Ykl014cp copurified with Yjr041cp almost exclusively (Figure 5E). Although their precise role in ribosome biogenesis remains to be determined, these observations together indicate that these two proteins share a close functional relationship and are directly required for ribosome biogenesis and/or rRNA processing. We have named the genes *URB1 (YKL014C)* and *URB2 (YJR041C)* (URB = *u*nhealthy *r*ibosome *b*iogenesis).

### YGL245W (GSN1) Displays a tRNA Charging-like Defect and Associates with tRNA-Glu In Vivo

Among the mutants that displayed transcriptional induction of genes involved in amino acid biosynthesis was $TetO_7$-*YGL254W* (Figure 3), suggesting that the product of this gene is required for amino acid activation. In fact, the Ygl254w protein sequence indicates that it is a tRNA-Glu synthetase, and the encoded protein has previously been shown to physically associate with both methionyl-tRNA synthetase (Mes1p) and Arc1p (Simos et al., 1996). However, to our knowledge it has not been previously shown that *YGL245W* functions as a tRNA synthetase in vivo, and the gene has not been assigned a standard name. To provide further support that Ygl254wp is a bona fide tRNA-Glu synthetase in vivo, we epitope-tagged and purified Ygl254wp, phenol extracted any associated RNAs, labeled them with fluors, and hybridized them to a microarray containing oligonucleotides complementary to all known yeast noncoding RNAs (Xing et al., 2004). This revealed that the most highly

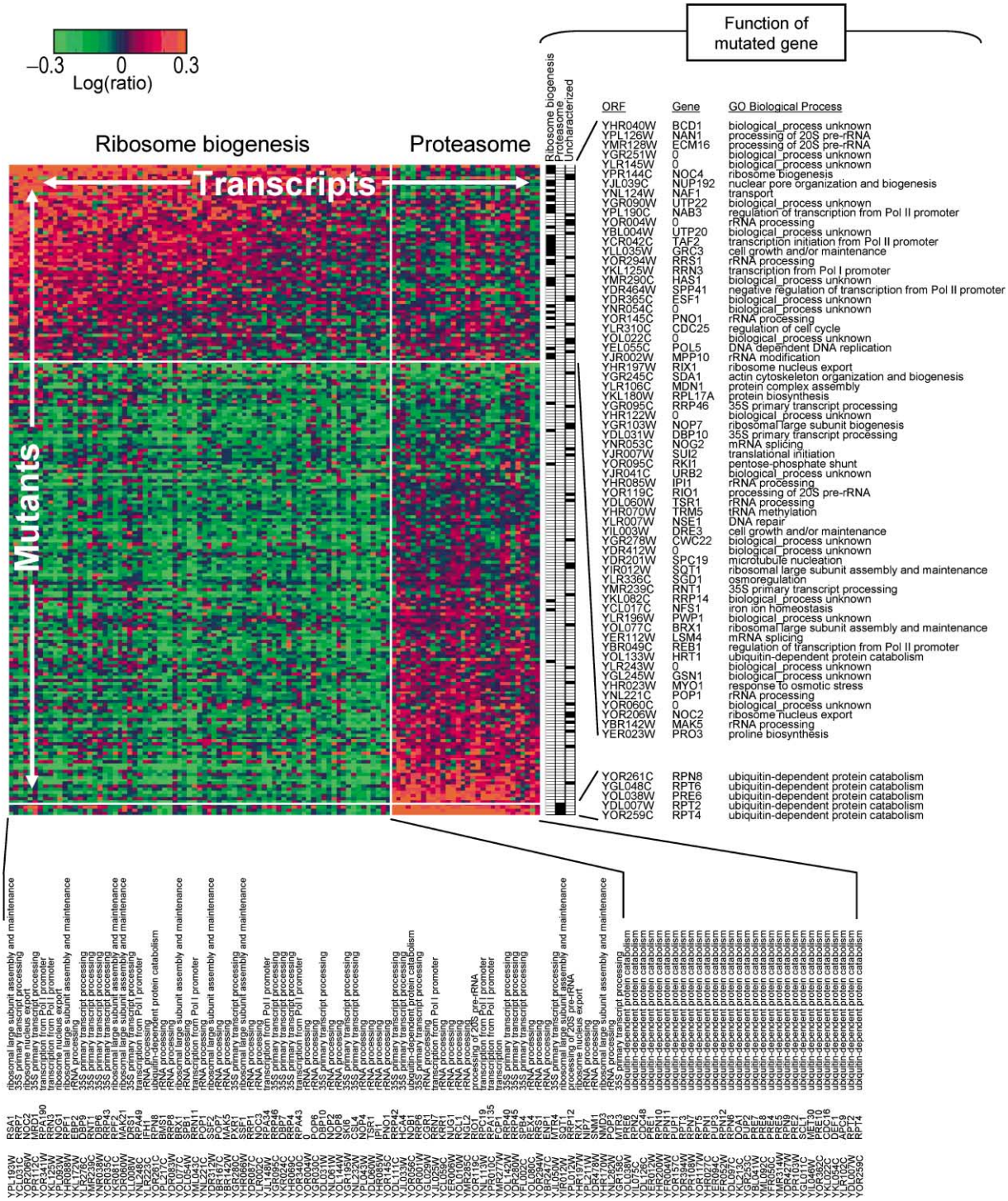


Figure 4. Negative Feedback Regulation of Ribosome and Proteasome Biogenesis

Induction or repression of individual genes (horizontal axis) in individual mutants (vertical axis) are shown. The mutants are sorted such that those that induce transcript levels of ribosome biogenesis factors (not ribosomal proteins) are shown at the top, those that induce transcript levels of proteasome components most highly are shown at the bottom. The remainder are shown in the middle, sorted by the overall proteasome induction level. The transcripts are sorted on their level of induction; i.e., the ribosome biogenesis transcripts that are induced most highly are farthest to the left.

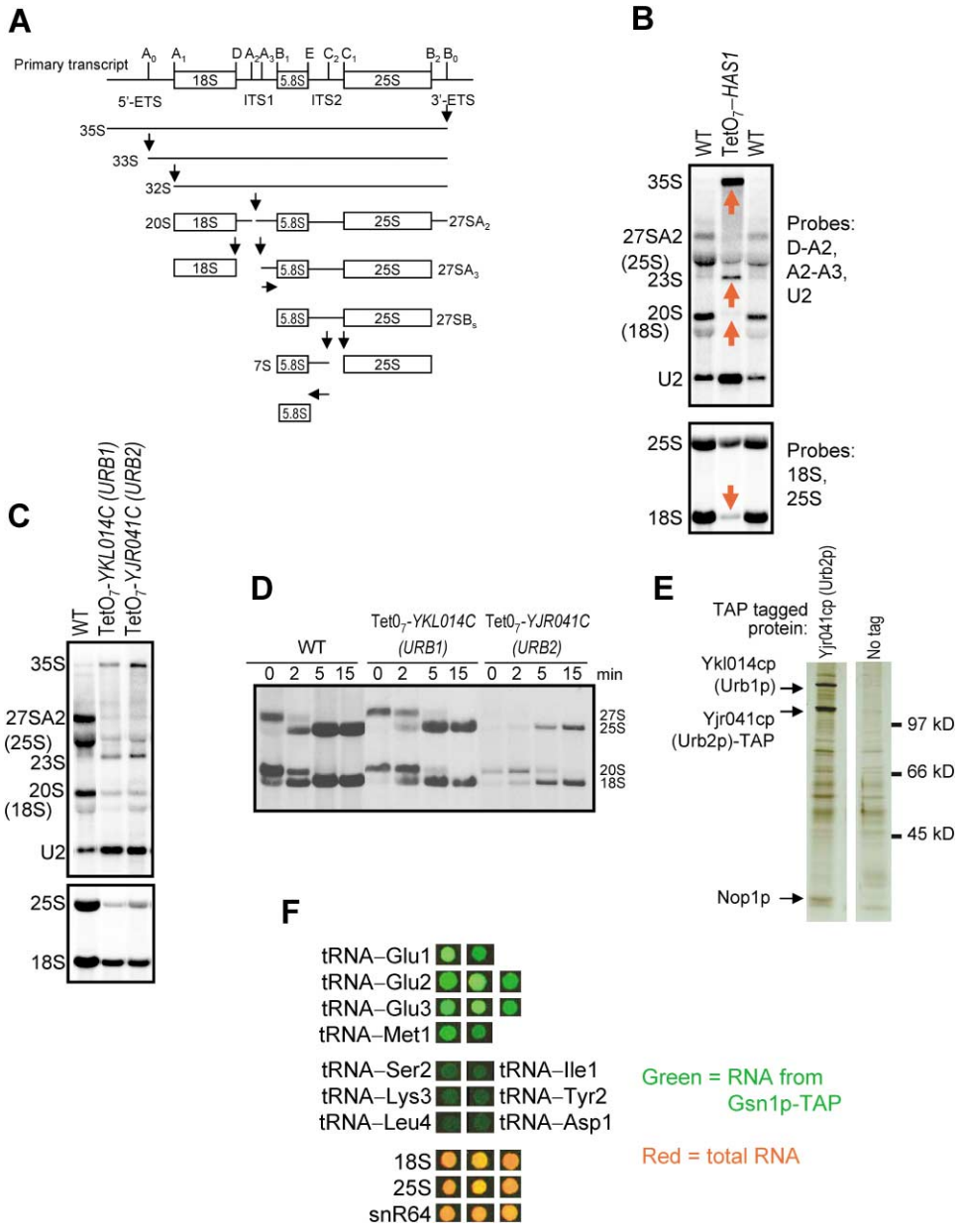


Figure 5. Essential Genes Involved in RNA Metabolism

(A) Schematic representation of rRNA processing pathway.

(B) Northern analysis of rRNA and pre-rRNA in TetO$_7$-*HAS1*. The blot in the top panel was probed simultaneously with oligonucleotides complementary to D-A2, A2-A3, and U2 snRNA (which serves as a loading control). The blot was then stripped and probed with oligonucleotides complementary to the 18S and 25S rRNAs.

(C) Northern analysis of rRNA and pre-rRNA in TetO$_7$-promoter mutants of *YKL014C* (*URB1*) and *YJR041C* (*URB2*), as in (B).

(D) Pulse-chase analysis verifying delayed rRNA processing in TetO$_7$-*URB1* and TetO$_7$-*URB2*.

(E) TAP-tagged Urb2p primarily recovers Urb1p.

(F) Ygr245w (Gsn1)-TAP associates with tRNA-Glu in vivo. A preparation of purified Gsn1-TAP was phenol extracted, and the resulting RNA was DNase I-treated, labeled with fluorescent dyes, and hybridized to a microarray containing probes complementary to tRNA, rRNA, and other yeast noncoding RNAs. In the image shown, the Gsn1-TAP-purified RNA was labeled with Cy3 (green), yeast total RNA was labeled with Cy5 (red), and the two were hybridized simultaneously.

enriched sequences (relative to total yeast RNA that was hybridized in the other channel on the array) were tRNA-Glu, followed by tRNA-Met (Figure 5F). We have named the *YGL245W* gene *GSN1* (GSN = tRNA *g*lutamine *syn*thetase).

## YLR440C (SEC39) Is Required for Protein Secretion

In our analysis, mutants in the secretory pathway were characterized by a large, round morphology, often with a wide bud neck and/or large size distribution, and a

microarray expression phenotype that is similar to (but distinguishable from) mutants in glycosylation. Previously characterized genes that shared one or more of these distinctive phenotypes included *SEC4*, *SEC14*, *SEC61*, *COG4* (*SEC38*), and *RET3* (a vesicle coat component; Cosson et al., 1996). This group also included the $TetO_7$-promoter mutant of *YLR440C*, a gene whose product has been localized to the endoplasmic reticulum (Huh et al., 2003) and for which a two-hybrid interaction has been observed with Dsl1p (Ito et al., 2001), a protein required for Golgi-to-ER retrograde traffic (Andag et al., 2001; Reilly et al., 2001). We further characterized the phenotype of $TetO_7$-*YLR440C* by pulse-chase analysis, which confirmed a striking defect in processing of Carboxypeptidase Y (CPY), a marker of protein secretion (Conibear and Stevens, 1998). On the basis of these observations, we named the *YLR440C* gene *SEC39*.

### YOL026C (MIM1) Is Required for Mitochondrial Import

Mutants in four genes known to be involved in mitochondrial import (*MAS1*, *TOM22*, *MGE1*, and *PAM18*) displayed similar phenotypic features to one another, most notably induction of genes involved in iron homeostasis (Figure 3). Mutants in three additional genes also displayed these features: *FMP28*, *FMP47*, and *YOL026C*. Fmp28p and Fmp47p have recently been localized to mitochondria (Sickmann et al., 2003; FMP = *f*ound in *m*itochondrial *p*roteome). Aside from cytoplasmic localization (Kumar et al., 2002), there is currently no supporting data regarding properties of Yol026cp. A growth defect of the $TetO_7$-*YOL026C* strain on medium with glycerol as the sole carbon source was dependent on addition of doxycycline to the medium, supporting a role in mitochondrial function (Figure 6B). Moreover, Western blotting showed that the $TetO_7$-*YOL026C* mutant accumulates uncleaved Atp2p precursor, a classical mitochondrial import defect shared by mutants in other mitochondrial import proteins (Figure 6C). We have named the *YOL026C* gene *MIM1* (for *m*itochondrial *im*port).

### Synthetic Genetic Array Analysis using $TetO_7$-Promoter Alleles

As a final illustration of the utility of the $TetO_7$-promoter strain collection for exploring essential gene functions, we conducted synthetic genetic array (SGA) analysis (Tong et al., 2001) by crossing a *cdc40-1*$^{ts}$ strain to the arrayed $TetO_7$-promoter strains (Figure 2C). Consistent with the known functions of *CDC40* in both DNA replication and pre-mRNA splicing (Ben-Yehuda et al., 2000), we identified synthetic genetic interactions with DNA replication factors (*PSF2*, *RFA2*) and splicing factors (*PRP16*, *PRP24*, *PRP38*) (Figure 7); in fact, the genetic interaction we observed between $TetO_7$-*PRP16* and *cdc40-1*$^{ts}$ recapitulates a previous result (Ben-Yehuda et al., 2000). Genetic interactions were obtained with genes in other pathways, including protein targeting (*HSP60*, *KAP95*) and membrane lipid biosynthesis (*GPI10*, *GWT1, TSC11*), indicating that *CDC40* function is also linked to these processes.

## Discussion

The analysis of mutant phenotypes is a singularly powerful means of defining gene functions. The "compendium" approach we previously described employed microarray expression profiles as a one-size-fits-all phenotypic assay (Hughes et al., 2000). This approach was the most fruitful for mutants that had a growth defect since these consistently exhibited gene expression changes relative to wild-type. This suggests that the systematic analysis of conditional alleles of essential genes should yield highly informative data. Here, we show that the majority of a set of ~700 different $TetO_7$-promoter strains display dramatic alterations in expression profiles. Furthermore, we show that similar "compendia" of other phenotypes provide complementary perspectives on the mutant phenotype and can also be used independently to generate functional inferences.

Although all four data types were informative (Figure 2 and Supplemental Data online), microarray expression profiles appear to be able to discern the largest number of distinct functional classes (Figure 3 and Supplemental Data), perhaps due to the large number of independent measurements made by microarrays; the yeast genome encodes at least 100 different transcription factors, implying that microarrays can measure at least 100 different variables. The drug/media sensitivity profiles did not correlate as tightly with distinct functional classes as might have been expected on the basis of how extensively such tests have previously been used as screening tools (Supplemental Data). This may be due to the fact that the robotic arraying system we used deposited a large number of cells on the test plates, possibly leading to false negatives (data not shown). We did, however, detect a number of expected sensitivities; for example, our $TetO_7$-*ALG7* mutant was sensitive to tunicamycin, consistent with previous observations (Barnes et al., 1984; Giaever et al., 1999), and our $TetO_7$-*POL30* mutant was sensitive to hydroxyurea, consistent with its role in DNA replication (Pol30p is the yeast homolog of PCNA) (Bauer and Burgers, 1990) (Supplemental Data). Figures 2C and 6B show that several $TetO_7$-promoter mutants in genes involved in mitochondrial function (*TIM50*, *TOM22*, *RIM2*, *MIM1*) are compromised for growth on glycerol.

The microarray expression data were particularly effective for discerning perturbation of biosynthetic pathways. This appears to be due at least in part to negative feedback loops in which perturbation of a pathway results in transcriptional induction of the same pathway or a related pathway. Previous analyses in which much smaller numbers of experiments were examined (Birrell et al., 2002; Giaever et al., 2002) suggested that there is a poor correlation between the genes that are required to survive a condition and the genes that are transcriptionally induced under the same condition. Our results (Figures 3 and 4) suggest that there is often a meaningful correspondence between the perturbed pathway and the genes that are specifically induced, although a large number of experiments may be required to determine this specificity. In particular, we observed an apparent negative feedback loop controlling *trans*-acting factors involved in ribosome biogenesis. To our knowledge, this

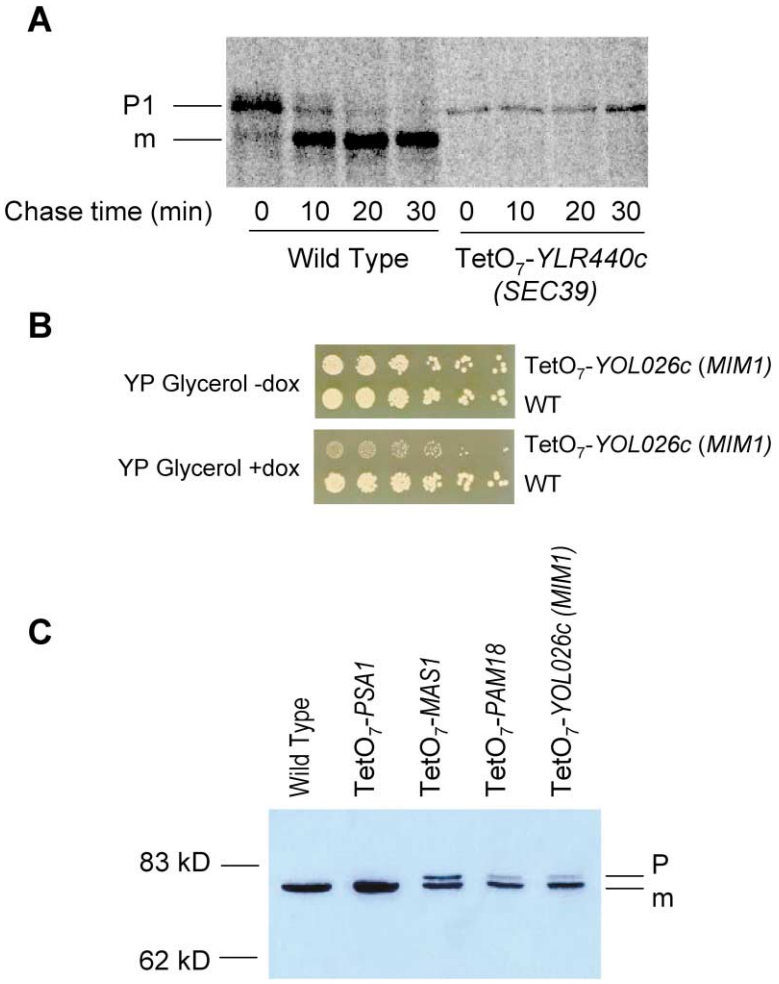


Figure 6. Mutants with Defects in Protein Processing

(A) SDS-PAGE and autoradiography of anti-CPY immunoprecipitates showing that $TetO_7$-*YLR440C (SEC39)* accumulates CPY precursor. The ER precursor (P1) and mature vacuolar (m) forms of CPY are indicated.

(B) Dilution spot assays, on agar plates containing glycerol as the sole carbon source, showing that the $TetO_7$-*YOL026C (MIM1)* growth defect on glycerol is dependent upon addition of doxycycline to the medium.

(C) Western blot showing that the $TetO_7$-*YOL026C (MIM1)* strain accumulates Atp2p precursor. $TetO_7$-*PSA1* is shown as a negative control (*PSA1* is involved in cell wall function); $TetO_7$-*MAS1* and $TetO_7$-*PAM18* are positive controls.

feedback loop has not been previously described; this shows that microarray expression profiling can reveal new regulatory pathways governing even the most fundamental cellular processes.

In many cases, our data are complementary to data from other large-scale studies. The development and implementation of techniques for systematic characterization of protein localizations and complexes have led to the generation of some type of functional information for a majority of yeast genes, including many that are essential for viability (Rout et al., 2000; Uetz et al., 2000; Ito et al., 2001; Gavin et al., 2002; Ho et al., 2002; Kumar et al., 2002; Huh et al., 2003; Hazbun et al., 2003; Sickmann et al., 2003; Krogan et al., 2004). In particular, a recent study described systematic analysis of 100 uncharacterized essential yeast genes using four such assays (affinity purification and mass spectrometry, two-hybrid, GFP localization, and PSI-BLAST) (Hazbun et al., 2003). Among the 602 mutants we analyzed, 49 overlapped with the 77 for which Hazbun et al. (2003) reported results. Among these, our inferences are in agreement on twelve genes, which is a higher level of concordance than would be expected by chance. However, we are in disagreement on fifteen genes, which will require further study. There are also 19 genes in which our SVM analysis produced predictions with expected precision of 30% or higher, but for which Hazbun et al. made no functional assignments. These include

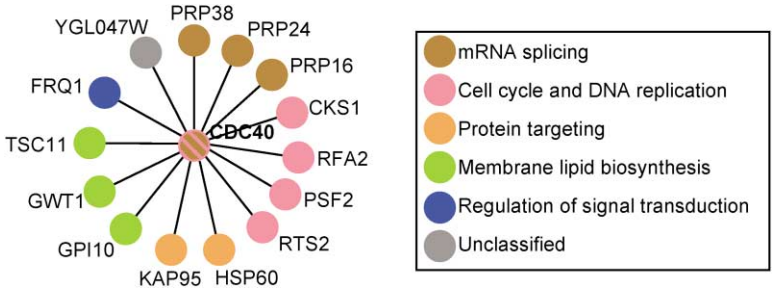


Figure 7. Network Diagram Summarizing Synthetic Genetic Interactions between *cdc40-1*$^{ts}$ and Strains in the $TetO_7$-Promoter Array

correct predictions for *YHR040W* (*BCD1*), *YHR085W* (*IPI1*), *YHR197W* (*RIX1*), *YNL182C* (*IPI3*), and *YOR004W*; although these proteins were still unannotated at the time of our analysis, we have recently shown in independent work that all of them are involved in ribosome biogenesis (Wu et al., 2002; Peng et al., 2003; Krogan et al., 2004). Similarly, among the six genes whose functions we analyzed here using specific assays, only four (*HAS1*, *URB1*, *GSN1*, and *SEC39*) could have been predicted with confidence (i.e., on the basis of agreement among more than one data set) from any of the functional genomics or proteomics data currently in the literature. For *URB2* and *MIM1*, previous evidence was either conflicting or insufficiently informative.

The strains described here are derived from the deletions consortium wild-type MATa strain (BY4741) (Giaever et al., 2002) and consequently are compatible with any assay developed for the deletion strains that does not require the strains to be *ura3* (since *URA3* marks the tet activator). For example, the $TetO_7$ strains are compatible with synthetic genetic array analysis (Tong et al., 2001, 2004) (Figure 7), and this may provide a means for mapping large-scale interaction networks among essential genes. We plan to ultimately generate $TetO_7$ strains encompassing all essential yeast genes, and we have already created an additional 100 strains, bringing the current collection to over 700 strains. In addition, an effort is also underway to integrate molecular barcodes into these strains in order to facilitate genetic analysis in pools using a microarray readout (Giaever et al., 2002), as is already possible with the deletions consortium strains (C. Nislow and G. Giaever, personal communication).

## Experimental Procedures

### Strain Construction

$TetO_7$-promoter alleles were constructed in strain R1158 (Hughes et al., 2000), which expresses the tet "off" activator (tTA*), by replacing the 100 bases upstream of the start codon with a cassette ($kan^R$-$TetO_7$-$TATA_{CYC1}$) from plasmid RP188 via one-step integration. Colony-purified G418-resistant transformants were frozen as glycerol stocks and confirmed by PCR. The Supplemental Data on the *Cell* website includes a table with all of the primer sequences, as well as a table of strains. The full collection is available from Open Biosystems (Huntsville, Alabama).

### Growth Assays

Growth rates on solid medium (YPD) were scored by manual inspection of spotted dilution series in the presence and absence of 10 μg/ml doxycyline. Growth rates in liquid (Figure 1D) were calculated from the cell densities (at 18 and 24 hr after the addition of doxycycline) of cultures used for microarray expression profiling.

### Morphological Characterization

Strains were grown 12–16 hr to mid-log phase at 30°C in YPD with and without 10 μg/ml doxycycline, fixed by addition of formaldehyde (3.7%), and sonicated gently for 8 s to disperse aggregated cells. At least 100 cells of each strain were examined using phase-contrast microscopy. Mutants were scored for 16 categories: an increase of unbudded, small-budded, or large-budded cells; small, large, or variable size; round, pointed, elongated, or amorphous shape; hyperpolarized buds; multiple buds; wide bud neck; lysis and clumped. Elongated and amorphous categories were subclassified as slight, moderate, or severe. Mutants that did not have a phenotype were classified as wild-type.

### Size Analysis

Cultures were grown in YPD ± 10 μg/ml doxycycline at 30°C to 0.6 to 3 × $10^7$ cells/ml, a density range in which wild-type cell size distributions do not vary. Cells were sonicated as above and size distributions were obtained with a Coulter Channelizer Z2 (Beckman-Coulter) as described previously (Jorgensen et al., 2002). Batches of up to 84 mutants at a time were analyzed, with multiple wild-type cultures interspersed among the mutants. Batches were normalized by translating the data across the 256 size bins and adjusting the spread by rescaling so that each batch had an identical overall distribution. The counts for each profile were then renormalized to an identical sum. Empty leading and trailing bins were filled in by linear extrapolation. We confirmed that this transform brought the wild-types into register.

### Drug Sensitivity Analysis and Synthetic Genetic Array Analysis

Strains were transferred onto agar plates at a density of 768 colonies per plate (384 strains in duplicate, as in Figure 2C) using a VersArray Colony Arrayer (BioRad) (Tong et al., 2001). Drug concentrations were hydroxyurea, 100 mM; benomyl, 15 μg/ml; tunicamycin, 1 μg/ml; rapamycin, 50 ng/ml; 6-azauracil, 100 μg/ml; cycloheximide, 0.1 μg/ml; caffeine, 8 mM; 5-fluorouracil, 20 mM, 40 mM, 80 mM; glucosamine, 3%; sorbitol, 1.5 M; all with and without doxycycline, 10 μg/ml. The strains were also arrayed on YPglycerol (3% glycerol) and YPgalactose media. Drug sensitivity screens were performed in duplicate; SGA analysis was performed in triplicate. Sensitivities were scored using an automated system that measures the colony area from digital images of the plates and reports the probability (p value) that the colony area is different between reference and treatment plates (C.B. and H. Ding, unpublished data). For SGA analysis, the *CDC40* gene was replaced with a $nat^R$-marked *cdc40-1*$^{ts}$ allele in the haploid MATα strain Y3656 (Tong et al., 2004) to create strain Y5065. Y5056 was crossed to the $TetO_7$-promoter array following Tong et al. (2001) with modifications. The Supplemental Data contain a full protocol. Interactions were confirmed by random spore analysis following Tong et al. (2004). Supplemental Figure S1 shows random spore analyses for each of the interactions shown in Figure 7.

### Microarray Analysis

Mutant cultures were grown in SC + 2% glucose + 10 μg/ml doxycycline for 24 hr, harvested, and processed in parallel with corresponding wild-type control cultures (R1158) as described (Hughes et al., 2000). Samples were hybridized to spotted arrays containing 70-mer oligonucleotides complimentary to 6,307 yeast genes (Operon), with spotting and hybridizations following Hegde et al., 2000. Arrays were scanned, images were quantified, and physical artifacts (dust and salt residue) edited as described previously (Hughes et al., 2000) and normalized by loess smoothing (Yang et al., 2002). Spatial trends were removed by applying a high-pass filter with a Gaussian window in the frequency domain (FFT transformed slides with missing values set to the slide mean). Remaining spatial trends were manually flagged. Replicate slides (dye-swaps) were then compared, and spots with poor correlation were manually flagged on both arrays. Dye-swaps were combined by averaging the ratio of the two measurements. The data were assembled into a 215 × 6,307 data matrix and missing values (9.8%) were imputed using BPCAfill (Oba et al., 2003). The final data are in the Supplemental Data online and individual microarray experiments have been submitted to the GEO.

### Northern Blotting and Pulse-Chase Analysis

RNA extraction and Northern blotting were performed as described (Peng et al., 2003). For pulse-chase analysis, strains were transformed with pRS411, grown to 0.8 × $10^7$ cells/ml in SD-Met+10 μg/ml doxycycline, and labeled following Kressler et al. (1999). Aliquots (20,000 cpm) were run on a 1% agarose-glyoxal gel, transferred to Hybond-N+ (Amersham) by overnight downward capillary transfer, dried, and visualized by autoradiography.

### TAP Tagging and Mass Spectrometry

Four liters of TAP strains for *URB1*, *URB2*, and *GSN1* (Ghaemmaghami et al., 2003) were grown in YPD to 1.5 × $10^7$ cells/ml. Proteins

were extracted with high-speed clarification, TAP-purified, separated by 10% SDS-PAGE gels, visualized by silver staining, and identified by MALDI-TOF mass spectrometry as previously described (Krogan et al., 2004).

### Secretion Assays

Strains were grown to the exponential phase in selective minimal medium (with doxycycline) at 24°C, suspended at OD600 = 5 in selective minimal medium without methionine, pulse-labeled with 35S-methionine for 7 min, and chased for 0–30 min as indicated. The pulse and chase were done at room temperature. Extracts prepared by lysis in SDS were diluted into Triton X-100 and immunoprecipitated with anti-CPY.

### Mitochondrial Import Assay

*ATP2* was TAP-tagged in each of the strains shown in Figure 6C by transformation with C-terminal *ATP2* PCR products from an *ATP2*-TAP strain (Ghaemmaghami et al., 2003). Strains were grown in SC-His+10 μg/ml doxycycline for 24 hr, extracted, and precipitated by vortexing in 20% TCA in the presence of glass beads and subjected to Western blotting with rabbit IgG to recognize the protein A component of the TAP tag.

### Statistical Analysis

For SVM, we used Gist (http://svm.sdsc.edu) version 2.0.5 for Linux with default parameters. Precision and recall were established by 3-fold cross validation. GO annotations were downloaded from *Saccharomyces* Genome Database (Issel-Tarver et al., 2002) and all categories were "propagated up" the GO graph. Morphology and drug sensitivity data were input into SVM verbatim. To avoid overfitting, the size profile for each mutant used by the SVM was a six-dimensional vector consisting of the projections of the original 256-dimensional profile on to the first six principal components of the covariance matrix of the normalized size profiles. The expression profile for each mutant used by the SVM was the projection of the initial profile onto the first 20 principal components of the expression covariance matrix. A separate SVM for each mutant utilized a 65-dimensional vector composed of 65 groups of coregulated genes identified manually from clustering diagrams.

Fisher's score (Bishop, 1995) was computed for all features in each data type for all GO-BP categories. Matrices of all Fisher's scores are given in the Supplemental Data. Microarray data shown in Figure 3 scored above the 99$^{th}$ percentile for at least one of the indicated categories (corresponding to a Fisher's score of 1.8).

The Kolmogorov-Smirnov test and hypergeometric p value calculations used tools at http://kstest.med.utoronto.ca/ and http://funspec.med.utoronto.ca/ (Robinson et al., 2002).

### Acknowledgments

We thank HuiMing Ding, Amy Tong, Renee Brost, Diana Ho, Nina Enriques, Darlene Ellenor, Mariana Kekis, Natalie Gabovic, Melissa Ballantine, and Jialin Wu for technical contributions; Dawn Richards and Victoria Canadien for assistance with mass spectrometry; Michael Yaffe, Howard Bussey, Roberto J. Rodriguez-Suarez, and Arnis Kuksis for advice and assistance in analyzing specific strains; and Grant Brown for evaluation of the manuscript. This work was supported by grants from the CIHR and Genome Canada to T.R.H. and B.J.A. A.D. was supported by a Charles H. Best fellowship and J.H. was supported by an Estate of Betty Irene West/CIHR doctoral research award.

Received: April 3, 2004
Revised: May 13, 2004
Accepted: May 17, 2004
Published: July 8, 2004

### References

Andag, U., Neumann, T., and Schmitt, H.D. (2001). The coatomer-interacting protein Dsl1p is required for Golgi-to-endoplasmic reticulum retrieval in yeast. J. Biol. Chem. *276*, 39150–39160.

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). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. *25*, 25–29.

Baker, K.P., and Schatz, G. (1991). Mitochondrial proteins essential for viability mediate protein import into yeast mitochondria. Nature *349*, 205–208.

Barnes, G., Hansen, W.J., Holcomb, C.L., and Rine, J. (1984). Asparagine-linked glycosylation in Saccharomyces cerevisiae: genetic analysis of an early step. Mol. Cell. Biol. *4*, 2381–2388.

Bauer, G.A., and Burgers, P.M. (1990). Molecular cloning, structure and expression of the yeast proliferating cell nuclear antigen gene. Nucleic Acids Res. *18*, 261–265.

Ben-Yehuda, S., Dix, I., Russell, C.S., McGarvey, M., Beggs, J.D., and Kupiec, M. (2000). Genetic and physical interactions between factors involved in both cell cycle progression and pre-mRNA splicing in Saccharomyces cerevisiae. Genetics *156*, 1503–1517.

Birrell, G.W., Brown, J.A., Wu, H.I., Giaever, G., Chu, A.M., Davis, R.W., and Brown, J.M. (2002). Transcriptional response of Saccharomyces cerevisiae to DNA-damaging agents does not identify the genes that protect against these agents. Proc. Natl. Acad. Sci. USA *99*, 8778–8783.

Bishop, C.M. (1995). Neural Networks for Pattern Recognition (New York: Oxford University Press).

Brown, M.P., Grundy, W.N., Lin, D., Cristianini, N., Sugnet, C.W., Furey, T.S., Ares, M., Jr., and Haussler, D. (2000). Knowledge-based analysis of microarray gene expression data by using support vector machines. Proc. Natl. Acad. Sci. USA *97*, 262–267.

Chun, K.T., and Goebl, M.G. (1997). Mutational analysis of Cak1p, an essential protein kinase that regulates cell cycle progression. Mol. Gen. Genet. *256*, 365–375.

Cliften, P., Sudarsanam, P., Desikan, A., Fulton, L., Fulton, B., Majors, J., Waterston, R., Cohen, B.A., and Johnston, M. (2003). Finding functional features in Saccharomyces genomes by phylogenetic footprinting. Science *301*, 71–76.

Conibear, E., and Stevens, T.H. (1998). Multiple sorting pathways between the late Golgi and the vacuole in yeast. Biochim. Biophys. Acta *1404*, 211–230.

Cosson, P., Demolliere, C., Hennecke, S., Duden, R., and Letourneur, F. (1996). Delta- and zeta-COP, two coatomer subunits homologous to clathrin-associated proteins, are involved in ER retrieval. EMBO J. *15*, 1792–1798.

Daum, G., Lees, N.D., Bard, M., and Dickson, R. (1998). Biochemistry, cell biology and molecular biology of lipids of Saccharomyces cerevisiae. Yeast *14*, 1471–1510.

Dohmen, R.J., Wu, P., and Varshavsky, A. (1994). Heat-inducible degron: a method for constructing temperature-sensitive mutants. Science *263*, 1273–1276.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA *95*, 14863–14868.

Emery, B., de la Cruz, J., Rocak, S., Deloche, O., and Linder, P. (2004). Has1p, a member of the DEAD-box family, is required for 40S ribosomal subunit biogenesis in Saccharomyces cerevisiae. Mol. Microbiol. *52*, 141–158.

Fatica, A., and Tollervey, D. (2002). Making ribosomes. Curr. Opin. Cell Biol. *14*, 313–318.

Fatica, A., Cronshaw, A.D., Dlakic, M., and Tollervey, D. (2002). Ssf1p prevents premature processing of an early pre-60S ribosomal particle. Mol. Cell *9*, 341–351.

Fleming, J.A., Lightcap, E.S., Sadis, S., Thoroddsen, V., Bulawa, C.E., and Blackman, R.K. (2002). Complementary whole-genome technologies reveal the cellular response to proteasome inhibition by PS-341. Proc. Natl. Acad. Sci. USA *99*, 1461–1466.

Gari, E., Piedrafita, L., Aldea, M., and Herrero, E. (1997). A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in Saccharomyces cerevisiae. Yeast *13*, 837–848.

Gasch, A.P., and Eisen, M.B. (2002). Exploring the conditional co-

regulation of yeast gene expression through fuzzy k-means clustering. Genome Biol. *3*, RESEARCH0059.

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). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature *415*, 141–147.

Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737–741.

Giaever, G., Shoemaker, D.D., Jones, T.W., Liang, H., Winzeler, E.A., Astromoff, A., and Davis, R.W. (1999). Genomic profiling of drug sensitivities via induced haploinsufficiency. Nat. Genet. *21*, 278–283.

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B., et al. (2002). Functional profiling of the Saccharomyces cerevisiae genome. Nature *418*, 387–391.

Giaever, G., Flaherty, P., Kumm, J., Proctor, M., Nislow, C., Jaramillo, D.F., Chu, A.M., Jordan, M.I., Arkin, A.P., and Davis, R.W. (2004). Chemogenomic profiling: identifying the functional interactions of small molecules in yeast. Proc. Natl. Acad. Sci. USA *101*, 793–798.

Hartwell, L.H., Culotti, J., and Reid, B. (1970). Genetic control of the cell-division cycle in yeast. I. Detection of mutants. Proc. Natl. Acad. Sci. USA *66*, 352–359.

Hazbun, T.R., Malmstrom, L., Anderson, S., Graczyk, B.J., Fox, B., Riffle, M., Sundin, B.A., Aranda, J.D., McDonald, W.H., Chiu, C.H., et al. (2003). Assigning function to yeast proteins by integration of technologies. Mol. Cell *12*, 1353–1365.

Hegde, P., Qi, R., Abernathy, K., Gay, C., Dharap, S., Gaspard, R., Hughes, J.E., Snesrud, E., Lee, N., and Quackenbush, J. (2000). A concise guide to cDNA microarray analysis. Biotechniques *29*, 548–556.

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). Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature *415*, 180–183.

Hughes, T.R. (2002). Yeast and drug discovery. Funct. Integr. Genomics *2*, 199–211.

Hughes, T.R., Marton, M.J., Jones, A.R., Roberts, C.J., Stoughton, R., Armour, C.D., Bennett, H.A., Coffey, E., Dai, H., He, Y.D., et al. (2000). Functional discovery via a compendium of expression profiles. Cell *102*, 109–126.

Huh, W.K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., and O'Shea, E.K. (2003). Global analysis of protein localization in budding yeast. Nature *425*, 686–691.

Issel-Tarver, L., Christie, K.R., Dolinski, K., Andrada, R., Balakrishnan, R., Ball, C.A., Binkley, G., Dong, S., Dwight, S.S., Fisk, D.G., et al. (2002). Saccharomyces Genome Database. Methods Enzymol. *350*, 329–346.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc. Natl. Acad. Sci. USA *98*, 4569–4574.

Jorgensen, P., Nishikawa, J.L., Breitkreutz, B.J., and Tyers, M. (2002). Systematic identification of pathways that couple cell growth and division in yeast. Science *297*, 395–400.

Kanemaki, M., Sanchez-Diaz, A., Gambus, A., and Labib, K. (2003). Functional proteomic identification of DNA replication proteins by induced proteolysis in vivo. Nature *423*, 720–724.

Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E.S. (2003). Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature *423*, 241–254.

Kressler, D., Doere, M., Rojo, M., and Linder, P. (1999). Synthetic lethality with conditional dbp6 alleles identifies rsa1p, a nucleoplasmic protein involved in the assembly of 60S ribosomal subunits. Mol. Cell. Biol. *19*, 8633–8645.

Krogan, N.J., Peng, W.T., Cagney, G., Robinson, M.D., Haw, R., Zhong, G., Guo, X., Zhang, X., Canadien, V., Richards, D.P., et al. (2004). High-definition macromolecular composition of yeast RNA-processing complexes. Mol. Cell *13*, 225–239.

Kumar, A., Agarwal, S., Heyman, J.A., Matson, S., Heidtman, M., Piccirillo, S., Umansky, L., Drawid, A., Jansen, R., Liu, Y., et al. (2002). Subcellular localization of the yeast proteome. Genes Dev. *16*, 707–719.

Lill, R., and Kispal, G. (2000). Maturation of cellular Fe-S proteins: an essential function of mitochondria. Trends Biochem. Sci. *25*, 352–356.

Lum, P.Y., Armour, C.D., Stepaniants, S.B., Cavet, G., Wolf, M.K., Butler, J.S., Hinshaw, J.C., Garnier, P., Prestwich, G.D., Leonardson, A., et al. (2004). Discovering modes of action for therapeutic compounds using a genome-wide screen of yeast heterozygotes. Cell *116*, 121–137.

Oba, S., Sato, M., Takemasa, I., Monden, M., Matsubara, K., and Ishii, S. (2003). A Bayesian missing value estimation method. Bioinformatics *19*, 2088–2096.

Peng, W.T., Robinson, M.D., Mnaimneh, S., Krogan, N.J., Cagney, G., Morris, Q., Davierwala, A.P., Grigull, J., Yang, X., Zhang, W., et al. (2003). A panoramic view of yeast noncoding RNA processing. Cell *113*, 919–933.

Reilly, B.A., Kraynack, B.A., VanRheenen, S.M., and Waters, M.G. (2001). Golgi-to-endoplasmic reticulum (ER) retrograde traffic in yeast requires Dsl1p, a component of the ER target site that interacts with a COPI coat subunit. Mol. Biol. Cell *12*, 3783–3796.

Robinson, M.D., Grigull, J., Mohammad, N., and Hughes, T.R. (2002). FunSpec: a web-based cluster interpreter for yeast. BMC Bioinformatics *3*, 35.

Rout, M.P., Aitchison, J.D., Suprapto, A., Hjertaas, K., Zhao, Y., and Chait, B.T. (2000). The yeast nuclear pore complex: composition, architecture, and transport mechanism. J. Cell Biol. *148*, 635–651.

Sachs, L. (1982). Applied Statistics (Berlin: Springer-Verlag).

Sickmann, A., Reinders, J., Wagner, Y., Joppich, C., Zahedi, R., Meyer, H.E., Schonfisch, B., Perschil, I., Chacinska, A., Guiard, B., et al. (2003). The proteome of Saccharomyces cerevisiae mitochondria. Proc. Natl. Acad. Sci. USA *100*, 13207–13212.

Simos, G., Segref, A., Fasiolo, F., Hellmuth, K., Shevchenko, A., Mann, M., and Hurt, E.C. (1996). The yeast protein Arc1p binds to tRNA and functions as a cofactor for the methionyl- and glutamyl-tRNA synthetases. EMBO J. *15*, 5437–5448.

Tong, A.H., Evangelista, M., Parsons, A.B., Xu, H., Bader, G.D., Page, N., Robinson, M., Raghibizadeh, S., Hogue, C.W., Bussey, H., et al. (2001). Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science *294*, 2364–2368.

Tong, A.H., Lesage, G., Bader, G.D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G.F., Brost, R.L., Chang, M., et al. (2004). Global mapping of the yeast genetic interaction network. Science *303*, 808–813.

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). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature *403*, 623–627.

Vazquez de Aldana, C.R., Wek, R.C., Segundo, P.S., Truesdell, A.G., and Hinnebusch, A.G. (1994). Multicopy tRNA genes functionally suppress mutations in yeast eIF-2 alpha kinase GCN2: evidence for separate pathways coupling GCN4 expression to unchanged tRNA. Mol. Cell. Biol. *14*, 7920–7932.

Vik, A., and Rine, J. (2001). Upc2p and Ecm22p, dual regulators of sterol biosynthesis in Saccharomyces cerevisiae. Mol. Cell. Biol. *21*, 6395–6405.

Wu, L.F., Hughes, T.R., Davierwala, A.P., Robinson, M.D., Stoughton, R., and Altschuler, S.J. (2002). Large-scale prediction of Saccharomyces cerevisiae gene function using overlapping transcriptional clusters. Nat. Genet. *31*, 255–265.

Xing, F., Hiley, S.L., Hughes, T.R., and Phizicky, E.M. (2004). The specificities of four yeast dihydrouridine synthases for cytoplasmic tRNAs. J. Biol. Chem. *279*, 17850–17860. Published online February 16, 2004. 10.1074/jbc.M401221200.

Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J., and Speed, T.P. (2002). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res. *30*, e15.

Zhang, N., Osborn, M., Gitsham, P., Yen, K., Miller, J.R., and Oliver, S.G. (2003). Using yeast to place human genes in functional categories. Gene *303*, 121–129.

Zhao, Y., Sohn, J.H., and Warner, J.R. (2003). Autoregulation in the biosynthesis of ribosomes. Mol. Cell. Biol. *23*, 699–707.

**Accession Numbers**

Individual microarray experiments have been submitted to the Gene Expression Omnibus with accession number GSE1404.