Cell, Vol. 102, 109–126, July 7, 2000, Copyright ©2000 by Cell Press

Functional Discovery via a Compendium of Expression Profiles

Timothy R. Hughes,*# Matthew J. Marton,*# Allan R. Jones,* Christopher J. Roberts,* Roland Stoughton,* Christopher D. Armour,* Holly A. Bennett,* Ernest Coffey,* Hongyue Dai,* Yudong D. He,* Matthew J. Kidd,* Amy M. King,* Michael R. Meyer,* David Slade,* Pek Y. Lum,* Sergey B. Stepaniants,* Daniel D. Shoemaker,* Daniel Gachotte,† Kalpana Chakraburtty,‡ Julian Simon,§ Martin Bard,† and Stephen H. Friend*‖

*Rosetta Inpharmatics, Inc.
12040 115th Avenue N.E.
Kirkland, Washington 98034

†Biology Department
Indiana University—Purdue University Indianapolis
723 W. Michigan Street
Indianapolis, Indiana 46202

‡Department of Biochemistry
Medical College of Wisconsin
Milwaukee, Wisconsin 53226

§Fred Hutchinson Cancer Research Center
1100 Fairview Avenue N.
Mail Stop D2-100
Seattle, Washington 98109

Summary

Ascertaining the impact of uncharacterized perturbations on the cell is a fundamental problem in biology. Here, we describe how a single assay can be used to monitor hundreds of different cellular functions simultaneously. We constructed a reference database or "compendium" of expression profiles corresponding to 300 diverse mutations and chemical treatments in *S. cerevisiae*, and we show that the cellular pathways affected can be determined by pattern matching, even among very subtle profiles. The utility of this approach is validated by examining profiles caused by deletions of uncharacterized genes: we identify and experimentally confirm that eight uncharacterized open reading frames encode proteins required for sterol metabolism, cell wall function, mitochondrial respiration, or protein synthesis. We also show that the compendium can be used to characterize pharmacological perturbations by identifying a novel target of the commonly used drug dyclonine.

Introduction

Systematic approaches for identifying the biological functions of novel genes are needed to ensure rapid progress from genome sequence to directed experimentation and applications. A similar need exists for methods to characterize biologically active compounds. Means have been devised to survey gene functions en masse either computationally (Marcotte et al., 1999) or experimentally; among these, highly parallel assays of thousands of mapped mutants (Shoemaker et al., 1996; Ross-Macdonald et al. 1999; Spradling et al., 1999) seem most likely to succeed comprehensively because they screen for phenotypes. However, just as high-throughput screens of chemical libraries require simple assays, highly parallel mutant analyses depend on easily scored phenotypes such as unusual appearance or sensitivity to certain culture conditions. Such convenient phenotypes do not exist for many important cellular functions.

Whole-genome expression profiling, facilitated by the development of DNA microarrays (Schena et al., 1995; Lockhart et al., 1996), represents a major advance in genome-wide functional analysis. In a single assay, the transcriptional response of each gene to a change in cellular state can be measured, whether it is disease, a process such as cell division, or a response to a chemical or genetic perturbation (DeRisi et al., 1997; Heller et al., 1997; Cho et al., 1998; Holstege et al., 1998; Spellman et al., 1998; Galitski et al., 1999). Because the relative abundance of transcripts is often tailored to specific cellular needs, most expression profiling studies to date have focused on the genes that respond to conditions or treatments of interest. For example, Chu et al. (1998) showed that several previously uncharacterized genes induced upon yeast sporulation are required for completion of the sporulation program. By the same token, the cellular state can be inferred from the expression profile, and the idea that the global transcriptional response itself constitutes a detailed molecular phenotype has recently begun to receive attention (DeRisi et al., 1997; Eisen et al., 1998; Gray et al., 1998; Holstege et al., 1998; Marton et al., 1998; Roberts et al., 2000). For example, tumors can be classified by expression profile (Perou et al., 1999; Golub et al., 1999; Alizadeh et al., 2000).

Expression profiling thus holds great promise for rapid genome functional analysis. It is plausible that the expression profile could serve as a universal phenotype: provided that the cellular transcriptional response to disruption of different steps in the same pathway is similar, and that there are sufficiently unique transcriptional responses to the perturbation of most cellular pathways, systematic characterization of novel mutants could be carried out with a single genome-wide expression measurement. Using a comprehensive database of reference profiles, the pathway(s) perturbed by an uncharacterized mutation would be ascertained by simply asking which expression patterns in the database its profile most strongly resembles, in a manner analogous to fingerprinting. An appealing aspect of this approach is that it should be equally effective at determining the consequences of pharmaceutical treatments and disease states. A sufficiently large and diverse set of profiles obtained from different mutants, treatments, and conditions would also result in a relatively comprehensive identification of coregulated transcript groups, allowing additional hypotheses to be drawn regarding

‖To whom correspondence should be addressed (e-mail: sfriend@rii.com).
#These authors contributed equally to this work.

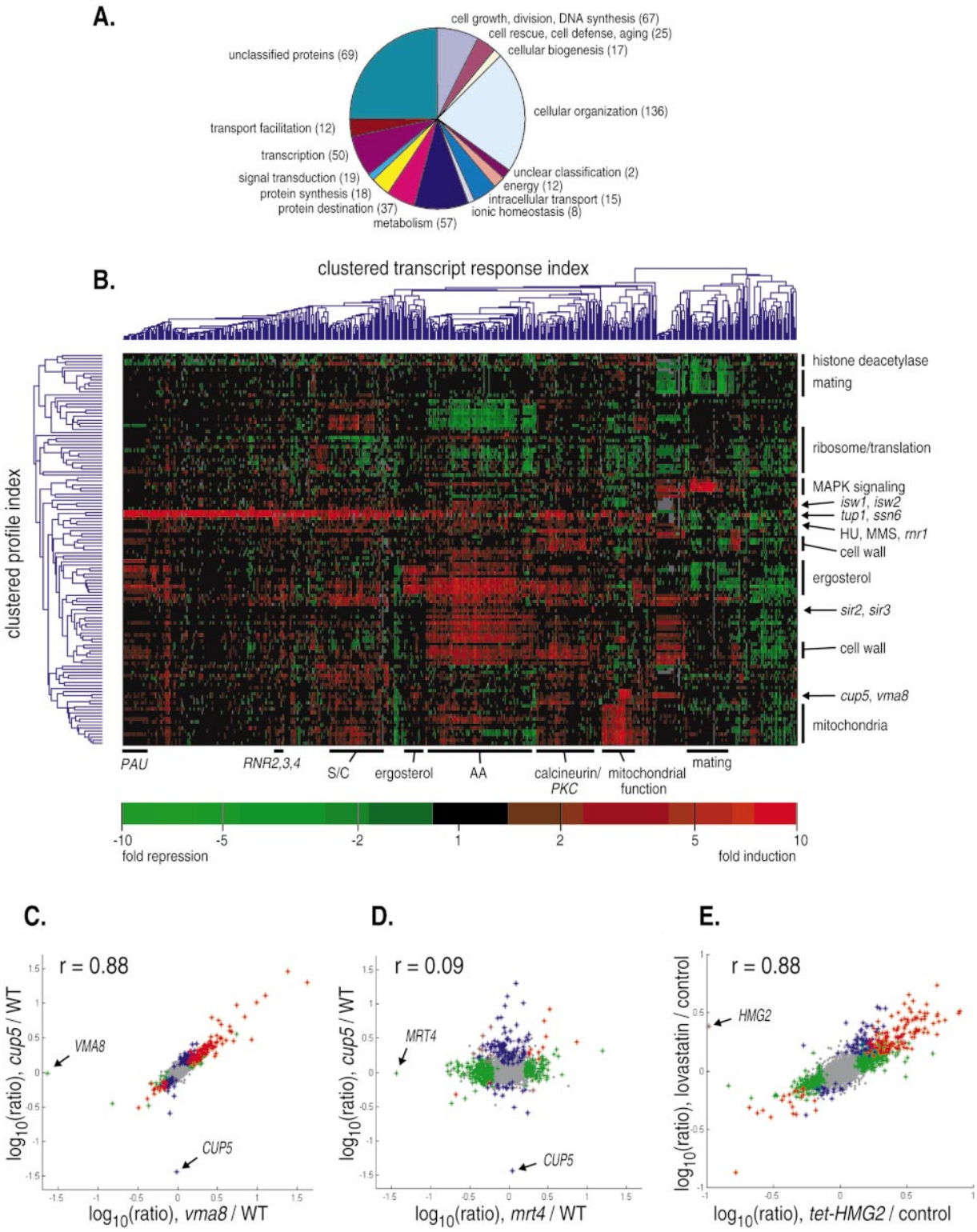

Figure 1. Full-Genome Expression Profiles from Mutations in Over 4% of All Yeast Genes

(A) Distribution of mutants profiled among 15 major classes in the Munich Information Center for Protein Sequence (MIPS) database (http://www.mips.biochem.mpg.de/proj/yeast/catalogues/funcat/index.html) (Mewes et al., 1997). Parentheses indicate the number of mutants profiled in each MIPS class; pie chart reflects total number of genes in each class. Many genes are represented in multiple classes.

(B) Two-dimensional agglomerative hierarchical clustering of 127 experiments and 568 genes, selected to include only experiments with 2 or more genes up- or downregulated greater than 3-fold, and significant at $P \leq 0.01$; and only genes that are up- or downregulated at greater than 3-fold, and at $P \leq 0.01$, in 2 or more experiments. *PAU, PAU* gene family; S/C, stress and carbohydrate metabolism; AA, amino acid biosynthesis; *PKC*, responsive to protein kinase C. See Supplemental Data for a version with all gene and experiment names.

Table 1. Number and Proportion of Significant Changes in the Transcription Profile of Different Classes of Experiments

# Experiments (percentage)	≥1 Gene with ≥2-Fold Induction	Error Model Accounting for Measurement Error Only			Error Model Accounting for Gene-Specific Fluctuations		
		≥5 Genes ($P \leq 0.01$)	≥20 Genes ($P \leq 0.01$)	≥100 Genes ($P \leq 0.01$)	≥5 Genes ($P \leq 0.01$)	≥20 Genes ($P \leq 0.01$)	≥100 Genes ($P \leq 0.01$)
300 compendium experiments	288 (96%)	219 (73%)	170 (57%)	94 (31%)	172 (57%)	122 (41%)	60 (20%)
198 barcoded deletion mutants	189 (95%)	136 (69%)	100 (51%)	51 (26%)	97 (49%)	68 (34%)	28 (14%)
Growth < 90% WT (50 mutants)	50 (100%)	50 (100%)	45 (90%)	32 (64%)	45 (90%)	35 (70%)	22 (44%)
Growth 90–95% WT (26 mutants)	26 (100%)	24 (92%)	20 (77%)	6 (23%)	17 (65%)	11 (42%)	2 (8%)
Growth > 95% WT (122 mutants)	113 (93%)	62 (51%)	35 (29%)	13 (11%)	35 (29%)	22 (18%)	4 (3%)
Named (72)	69 (96%)	48 (67%)	29 (40%)	12 (17%)	31 (43%)	20 (28%)	4 (6%)
Unnamed (50)	44 (88%)	14 (28%)	6 (12%)	1 (2%)	4 (8%)	2 (4%)	0 (0%)
63 control experiments	55 (87%)	11 (17%)	2 (3%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

A quantitative parallel growth assay was used to determine the relative growth rates among 198 of the deletion mutants carrying molecular barcodes (Experimental Procedures) (Shoemaker et al., 1996). A list of the 198 barcoded strains and their relative growth rates is found in the Supplemental Data. "Unnamed" genes are those that are uncharacterized and are thus designated only by an ORF systematic name (e.g., *YER019w*). Error models are described in detail in the Supplemental Data.

the functions of genes based on the regulatory characteristics of their own transcripts (Eisen et al., 1998).

To demonstrate the utility of a database of reference expression profiles as a comprehensive tool for functional classification and discovery, we created a reference database or "compendium" of three hundred full-genome expression profiles in *S. cerevisiae* corresponding to mutations in both characterized genes and uncharacterized open reading frames (ORFs), as well as treatments with compounds with known molecular targets. All experiments were conducted under a single growth condition, allowing direct comparison of all genes over all profiles. As expected, the large, internally consistent expression data set enabled the identification of many coregulated sets of genes, allowing dissection of transcriptional responses and isolation of candidate genes for many cellular processes. Moreover, we show that in the context of a compendium of reference profiles, the expression profile displayed by a mutant serves as a molecular phenotype that predicts phenotypes in conventional assays, as demonstrated by a series of experimentally verified examples in which we use mutant profile similarities to assign uncharacterized ORFs to cellular pathways. Gene functions inferred by profile comparisons can be derived from very subtle transcriptional responses, and there is no requirement for prior knowledge about the functions of the responsive transcripts. We also provide an example in which the compendium is used to identify an unknown target of a commonly used drug.

Results

Expression Profiles from Hundreds of Yeast Deletion Mutants

Using a two-color cDNA microarray hybridization assay (Schena et al., 1995), 300 expression profiles were generated in *S. cerevisiae* in which transcript levels of a mutant or compound-treated culture were compared to that of a wild-type or mock-treated culture (see Experimental Procedures and Supplemental Data, below, for details). Two hundred seventy-six deletion mutants, 11 tetracycline-regulatable alleles of essential genes (Gari et al., 1997), and 13 well-characterized compounds were profiled. One hundred fifty-one of the mutants were profiled in duplicate (i.e., regrown from frozen stocks and completely reanalyzed) to establish the reproducibility of profiles (see Supplemental Data). Deletion mutants were selected such that a variety of functional classifications were represented (Figure 1A). Sixty-nine of the 276 deletions were of uncharacterized open reading frames (ORFs) (Figure 1A). To allow direct comparison of the behavior of all genes in response to all mutations and treatments, experiments were performed under a single condition: cells were grown at 30°C in liquid synthetic complete (SC) medium plus 2% glucose to mid-log phase, with the final optical densities of the experimental and control cultures closely matched (see Experimental Procedures). Because any one growth condition is unlikely to elicit a phenotype from every mutant, one possible outcome of this approach was that many mutants would not display transcriptional alterations. In actuality, nearly all of the experiments resulted in a 2-fold or greater alteration in the abundance of at least one transcript, not including the deleted gene (Table 1, first column).

A Gene-Specific Error Model Compensates for Differences in Variation of Transcript Abundance among Different Yeast Genes

To ensure that observed transcriptional alterations were caused by the mutations or treatments, and not by random fluctuations or systematic biases, we investigated whether the abundance of some transcripts might inherently fluctuate more than others under our culture conditions (Wittes and Friedman, 1999). In parallel with the 300 experiment data set, a series of 63 negative control

(C) Comparison of the transcript profile of a homozygous *cup5* disruption strain to that of a homozygous *vma8* disruption strain. Log_{10} of the expression ratio is plotted for each gene. Genes that changed significantly from wild-type at $P \leq 0.01$ in both experiments are indicated in red; genes changing significantly from wild-type at $P \leq 0.01$ in only *cup5*Δ or *vma8*Δ are indicated in blue or green, respectively; brown, anticorrelated genes at $P \leq 0.01$ in both experiments; gray, genes at $P > 0.01$ in both experiments.
(D) Comparison of the transcript profile of the homozygous *cup5* disruption strain to that of a homozygous *mrt4* disruption strain.
(E) Comparison of the transcript profile of lovastatin treatment to reduction in *HMG2* transcript.

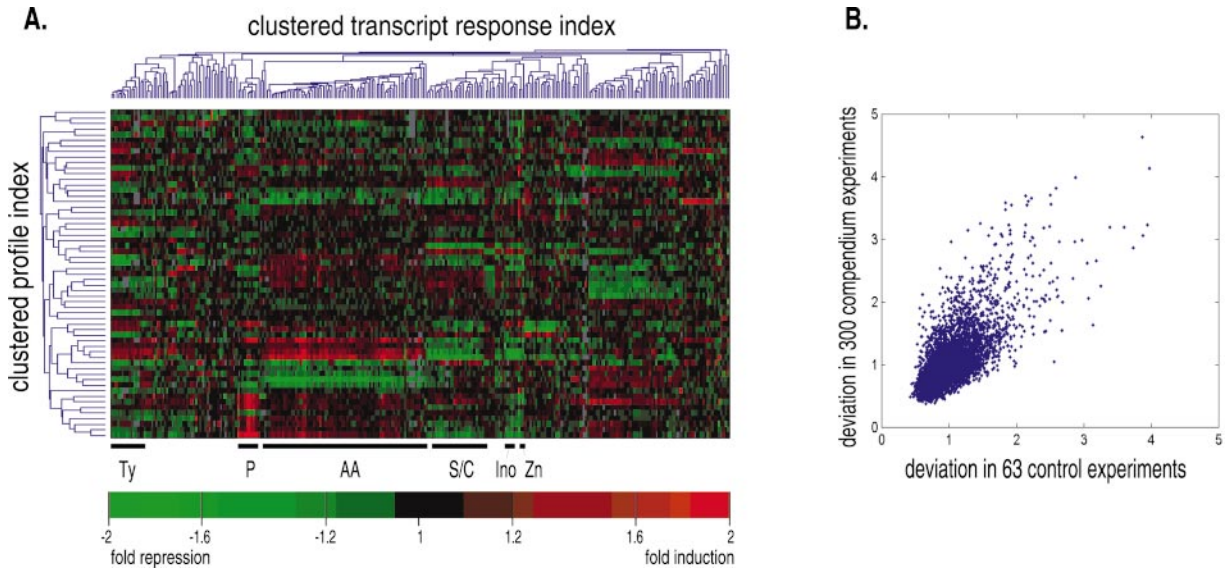

Figure 2. Commonly Occurring Transcriptional Fluctuations Unrelated to Intended Perturbation

(A) Two-dimensional hierarchical clustering of 59 control experiments and 278 genes, selected to include only experiments with two or more genes up- or downregulated at a significance of $P \leq 0.15$; and only genes that are significant at $P \leq 0.15$ in two or more experiments, using a significance model which accounts for measurement error. Because the polarity of transcriptional responses is arbitrary in these experiments, the absolute value of the correlation was used in the dissimilarity measure. The color scale has been adjusted to allow visualization of low-amplitude patterns. Selected gene response clusters are labeled based on the functions of the genes whose transcripts predominate the clusters. Ty, retrotransposons; P, phosphate metabolism; AA, amino acid biosynthesis; S/C, stress and carbohydrate metabolism; Ino, inositol metabolism; Zn, zinc metabolism. See Supplemental Data for a version with all gene and experiment names.

(B) Scatter plot comparing the scale factor Λ derived from 63 control experiments versus Λ derived from 300 compendium experiments for each of 5835 yeast genes. The scale factor for each gene is defined as the standard deviation of $\log_{10}$(ratio)/[error of $\log_{10}$(ratio)] over all experiments (see "Error Model" in Supplemental Data).

experiments were conducted in which simultaneously grown untreated isogenic wild-type cultures were compared to each other. The vast majority of profiles from these negative control experiments also included at least one gene with greater than 2-fold induction or repression (Table 1, first column). Two-dimensional (2-D) clustering analysis (Hartigan, 1975; Eisen et al., 1998) of the control experiments revealed several sets of genes, many known to be regulated by nutrition or stress, that displayed small-magnitude but coordinate differences in transcript abundance between two seemingly identical cultures (Figure 2A; note that the color scale is different from other figures).

We reasoned that these fluctuations represent a form of biological noise. Indeed, the genes with highest variance in the 63 control experiments are among those that fluctuate most in the 300 compendium experiments (Figure 2B), suggesting that the same fluctuations are present in all the experiments. Application of an error model that reduces the significance of the transcriptional regulations of genes in proportion not only to the accuracy to which the measurement was made, but also to their fluctuation in the negative control experiments (see "Error Model" in Supplemental Data) demonstrates that, in most of the mutants with no growth defect, such fluctuations account for virtually all of the transcriptional changes (Table 1, fifth column). Only 29% of deletion mutants with growth $>95\%$ that of wild-type displayed profiles with more than five genes significant at $P \leq 0.01$ in the gene-specific error model, a stringent definition of profile significance as it is fulfilled by none of the 63 negative control experiments. Among the 29%, deletion of a previously characterized (i.e., named) gene had a much higher likelihood of resulting in such a significant profile (43%) than deletion of an uncharacterized ORF (8%). This surprising difference was not accounted for by biases in ORF length, basal transcription, growth rate, or sequence redundancy (data not shown); remaining explanations include functional redundancy, or that these uncharacterized ORFs are required only for very specific processes or to survive particular conditions. 90% of deletion mutants with a growth rate less than 90% that of wild-type displayed profiles with more than five genes significant at $P \leq 0.01$, using the gene-specific error model (Table 1, fifth column). Overall, more than half (172/300) of the experiments resulted in a profile with more than five genes significant at $P \leq 0.01$ (Table 1, fifth column), and $\sim$3/4 of all transcripts (4553) were significantly up- or downregulated at $P \leq 0.01$ in at least one profile (Supplemental Data). Thus, a single growth protocol was sufficient to generate functional data for roughly half of the mutants, and to evoke responses from a large majority of genes.

Transcriptional Landmarks: Profiles from Characterized Mutants Induce Expected Sets of Genes and Form Clusters Corresponding to Phenotypes

We initiated analysis of the biology of the 300-experiment compendium data set by asking whether groups of known or expected coregulated genes were easily

detectable, and whether mutations or treatments known to impact similar cellular processes displayed similar expression profiles. Two-dimensional hierarchical clustering of the most prominent gene behaviors among the experiments with the largest profiles illustrates the gross transcriptional features of the 300 expression profiles (Figure 1B), identifying groups of coregulated transcripts (horizontal axis) as well as groups of experiments with similar profiles (vertical axis). Several large classes of coregulated genes are apparent. Prominent groups whose activation or repression is restricted to specific classes of mutants correspond to proteins involved in mating (Roberts et al., 2000), ergosterol biosynthesis, and mitochondrial respiration (discussed below); others, such as the PKC/calcineurin activated gene cluster (Marton et al., 1998; Roberts et al., 2000), are induced in several types of experiments. The known DNA damage/S phase arrest inducible *RNR2*, *RNR3*, and *RNR4* transcripts are closely coregulated, and are induced almost exclusively by HU, MMS, *rnr1*Δ, *tup1*Δ, and *ssn6*Δ, as expected from previous studies (Zhou and Elledge, 1992). A large group of transcripts including *HIS5*, *ARG4*, *LEU4*, and many others associated with amino acid biosynthesis appear in the largest number of profiles, but because these transcripts also tend to fluctuate in negative control experiments, the error model reduces their significance relative to other genes.

In general, different mutants that affect the same cellular process display related transcript profiles, whether the mutation affects a protein involved directly in transcription or some other cellular process. In most cases, the global profile similarity is sufficient to cause association in the clustering analysis, allowing manual identification of profile clusters that correspond to mutants known to share other phenotypes. For example, as expected, deletion of either component of the Tup1-Ssn6 corepressor (Williams et al., 1991; Keleher et al., 1992; DeRisi et al., 1997) results in a very large and very similar profile (the horizontal red stripe denoted *tup1, ssn6* in Figure 1B). Deletion mutants in *CUP5* or *VMA8*, both of which encode components of the vacuolar H^+-ATPase complex (Eide et al., 1993; Nelson et al., 1995), also share a virtually identical transcript profile (Figures 1B and 1C; for comparison, an example of uncorrelated responses is shown in Figure 1D). Other clusters of profiles that are dominated by deletion mutants known to have similar functional consequences include the groups of mitochondrial respiration-, mating-, and sterol pathway-related experiments, which are easily visualized in Figure 1B because they induce large-magnitude changes in specific groups of transcripts. A number of less-visible groups of lower-magnitude profiles likewise correspond to established functional classes: discrete clusters are formed by mutations in genes encoding silencing factors Sir2p and Sir3p, chromatin-remodeling proteins Isw1p and Isw2p (Tsukiyama et al., 1999), histone deacetylase components, ribosomal proteins, and proteins involved in cell wall function (Figure 1B). Similarly, treatment of cells with an inhibitory compound mimics loss of function of its target in many cases. For example, treatment with lovastatin, which inhibits HMG-CoA reductase (Alberts et al., 1980), results in a transcript profile that correlates highly with the profile produced by reducing expression of *HMG2* (Figure 1E). Additional examples in which profiles caused by inhibitory compounds resemble those of mutants in the affected pathway include itraconazole/*erg11* (Daum et al., 1998), cycloheximide/*yef3*, HU/*rnr1*, tunicamycin/glucosamine/2-deoxy-D-glucose/*gas1*, and FR901,228/*rpd3* (Nakajima et al., 1998) (Figure 1B).

To ask whether the experiment clusters described above (which were identified by the fact that they are composed of mutants in functionally related genes) correspond to those with highest statistical significance, we used a bootstrap method to obtain P values for each of the branch points in the experiment cluster tree shown in Figure 1B (see "Clustering and Correlations" in Supplemental Data). The significance assigned to a branch point is dependent on the number of elements in the branches; thus, the P values of large clusters tend to be more significant than those of smaller clusters, and P values of branches of different sizes may not be directly comparable. Nonetheless, among the fifteen clusters of greater than four experiments and $P \leq 0.01$, 11 of 15 correspond to the groups of mutants discussed above (see Supplemental Data and Discussion). The remaining four clusters are composed largely of profiles from functionally unrelated mutants in which the predominant transcriptional changes are the same as those that drive clustering in the control experiments (labeled "AA" and "S/C" in Figure 1B and Figure 2A, for "Amino Acid biosynthesis" and "Stress/Carbohydrate metabolism," respectively). The presence of these four clusters is consistent with the fact that the gene-specific error model does not eliminate these frequently occurring biases, but rather only reduces their statistical significance. Such potentially misleading experiment clusters can be identified by the groups of genes induced or repressed, or by the fact that their composition makes little biological sense.

Functional Genomics via a Compendium of Expression Profiles: Identification of the Cellular Functions of Uncharacterized ORFs

The juxtaposition of functionally related mutants on the profile index of the clustering analysis in Figure 1B supports the idea that a compendium of profiles could serve as a systematic tool for identification of gene functions, because mutants that display similar expression profiles are likely to share cellular functions. The fact that treatment with pharmacological compounds elicits a response mimicking that of mutation of the target indicates that pathways affected by uncharacterized compounds could be determined as well. We next asked whether the cellular functions of uncharacterized *S. cerevisiae* ORFs could be predicted by comparing the expression profile of the corresponding deletion mutant to profiles of known mutants in the compendium, and whether an unknown drug target could be identified in the same way. We focused on ergosterol biosynthesis, cell wall maintenance, mitochondrial respiration, and protein synthesis, all of which are well-defined pathways and established targets of antifungal and/or antimicrobial compounds.

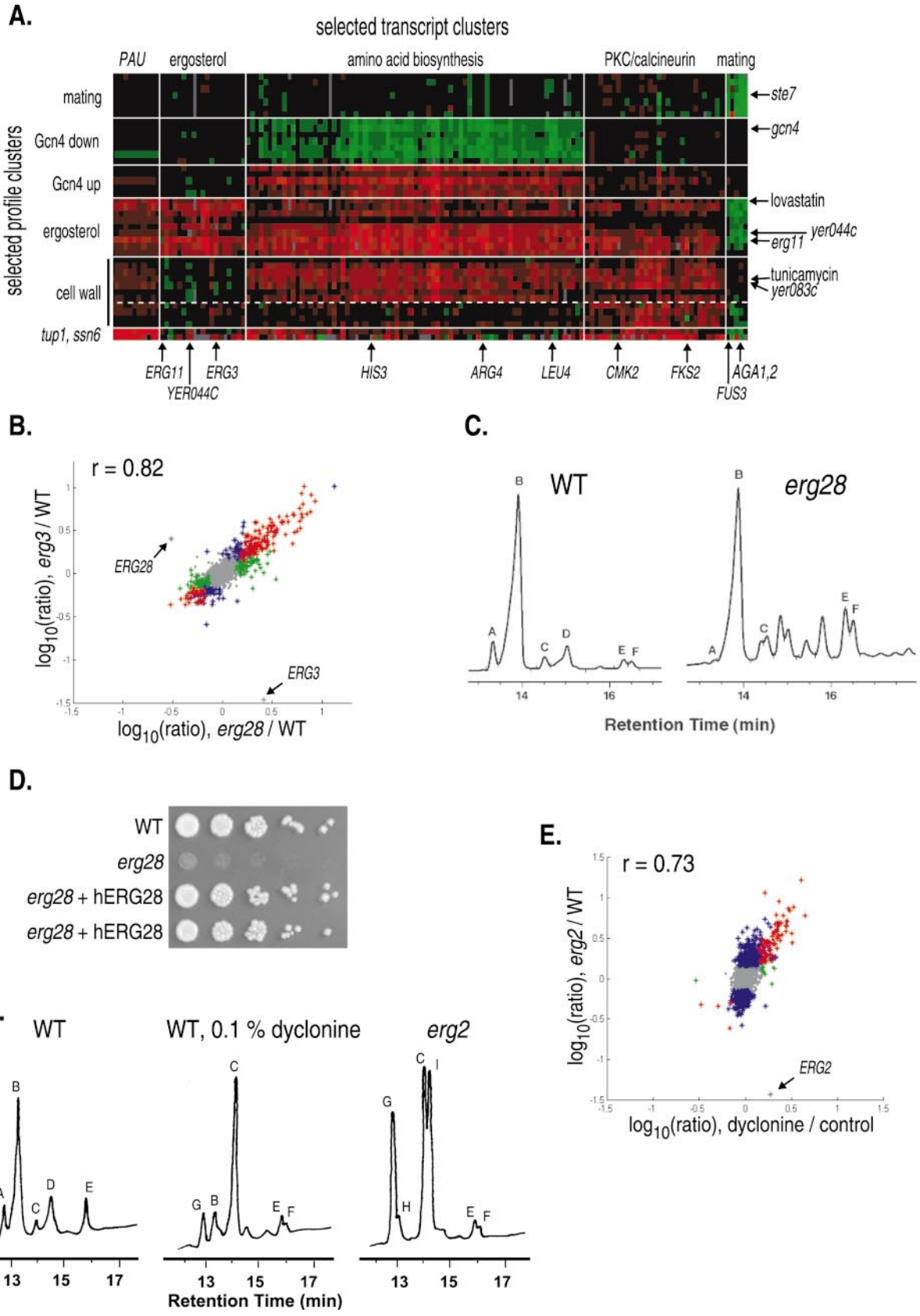

Figure 3. Profile Similarity Identifies Sterol-Pathway Disturbance Resulting from Deletion of Uncharacterized ORF *YER044c* (*ERG28*) and from Dyclonine Treatment

(A) Prominent gene clusters responding to interference with ergosterol biosynthesis, assembled from Figure 1B. Color scale is identical to Figure 1B. See Supplemental Data for a version with all gene and experiment names.

YER044c (*ERG28*) Encodes a Protein Involved in Ergosterol Biosynthesis

The yeast ergosterol biosynthesis pathway is of particular interest as it is the target of numerous antifungal compounds, and shares many features with human cholesterol biosynthesis (Daum et al., 1998). Because inhibition of the pathway results in transcriptional induction of many of the genes encoding pathway enzymes (Daum et al., 1998), study of sterol biosynthesis in yeast should represent an ideal opportunity to identify novel factors by expression profiling. Indeed, mutants and treatments affecting sterol biosynthesis display a characteristic transcript profile, causing these profiles to cluster together. This profile includes several hundred expression changes that can be dissected into at least five major transcript clusters (Figure 3A) and appears to reflect the fact that disrupting sterol homeostasis compromises membrane function, resulting in impaired tryptophan uptake, sensitivity to cations, and decreased mating frequency (Parks et al., 1995). Only one of these transcript clusters (the group of transcripts labeled "ergosterol" in Figure 3A) is induced specifically by ergosterol-related experiments, and corresponds primarily to the ergosterol biosynthetic genes: 7 of 19 encode known components of the ergosterol biosynthetic pathway (see Figure 3A in Supplemental Data).

Clustered among the profiles from sterol pathway mutants *erg2*Δ, *erg3*Δ, and tet-*ERG11* is the profile caused by deletion of the uncharacterized ORF *YER044c* (Figure 3A; Figure 3B shows an example profile similarity; based on results described below, *YER044c* is hereafter referred to as *ERG28*). Consistent with the hypothesis that *ERG28* is involved in sterol biosynthesis, the *ERG28* transcript is itself a member of the ergosterol-specific transcript cluster (Figure 3A). To assess the biological significance of these observations, an *erg28*Δ strain was analyzed in detail. Although *ERG28* is not essential, *erg28*Δ cells grew slowly (~70% of the wild-type rate) (Smith et al., 1996; Winzeler et al., 1999; Figure 3D). Gas chromatography (GC) analysis revealed that *erg28*Δ cells have an unusual sterol content in that additional sterols accumulate that are not seen in the wild-type strain, implicating the involvement of Erg28p in ergosterol biosynthesis (Figure 3C and M. B., unpublished data). The *erg28*Δ cells accumulate significantly less ergosterol (~50%) than wild-type cells, but still in sufficient quantity to support growth (Figure 3C); this explains why *erg28*Δ cells are not resistant to nystatin (data not shown), a compound that binds ergosterol and has been used extensively to screen for mutants in the ergosterol biosynthesis pathway (Bard, 1972; Molzahn and Woods, 1972). Therefore, *ERG28* encodes a novel protein involved in sterol biosynthesis that could not have been identified by the typical primary screen for ergosterol-related mutants.

hERG28: *A Human Gene that Complements the Yeast* erg28Δ *Mutation*

During the course of these experiments, a human homolog of *ERG28* (here designated hERG28) was identified (Veitia et al., 1999) but the function of the gene was not determined. We tested the hypothesis that hERG28 functions in sterol biosynthesis by asking whether hERG28 could complement the yeast *erg28*Δ deletion mutant. The hERG28 open reading frame was PCR amplified and cloned in front of the strong yeast *HOR7* promoter on a 2μ plasmid (see Experimental Procedures). hERG28 restored wild-type growth to an *erg28*Δ mutant (Figure 3D), showing that the gene is functionally conserved, and therefore is potentially a novel component of the human cholesterol biosynthetic pathway.

Drug Target Identification: Dyclonine Inhibits Erg2p, the Yeast Homolog of the Sigma Receptor

In an effort to identify unknown drug targets using the compendium, a number of expression profiles were generated by treating yeast with drugs without known targets (data not shown). Among these, treatment with the commonly used topical anesthetic dyclonine resulted in an expression profile that most closely resembled profiles resulting from perturbation of the ergosterol pathway (correlation with the *erg2*Δ profile is shown in Figure 3E; the dose of dyclonine used is lower than the IC_{50} and therefore the transcriptional response is lower magnitude than the *ERG* pathway deletion mutants). GC analysis confirmed that the sterol content of dyclonine-treated cells was abnormal (Figure 3F, center), and featured a buildup of fecosterol, indicating inhibition of Erg2p, the sterol C-8 isomerase (Figure 3F, right; the differences between *erg2*Δ and dyclonine treatment are due to the fact that the *erg2*Δ mutant has accumulated aberrant products of fecosterol). Consistent with Erg2p being the target of dyclonine, an *erg2*Δ/*ERG2* strain is hypersensitive to dyclonine, whereas heterozygous mutants in other steps of the sterol biosynthetic pathway are not (data not shown). Furthermore, *ERG2* overexpression confers increased resistance to dyclonine (data not shown).

The human gene with the greatest sequence similarity to the yeast Erg2 protein is not the human sterol isomerase, but rather the sigma receptor (Hanner et al., 1996; Kekuda et al., 1996). The sigma receptor is a neurosteroid-interacting protein that positively regulates potassium conductance (Nguyen et al., 1998; Wilke et al., 1999) and binds a number of neuroactive drugs, including haloperidol. Several inhibitory compounds target both the yeast Erg2 protein and the mammalian

(B) Comparison of the transcript profile of an *erg28*Δ strain to that of an *erg3*Δ strain.
(C) Sterol content of wild-type (left) and *erg28*Δ (right) strains.
(D) Complementation of the *erg28*Δ growth defect by plasmids expressing hERG28, the human homolog of *ERG28* (see Experimental Procedures for plasmid details). Cultures of log-phase haploid strains, carrying the *erg28*Δ deletion and hERG28 plasmids as indicated, were equalized for cell density, serially diluted by a factor of six, spotted onto plates containing yeast extract, peptone, and dextrose (YPD), and incubated 2 days at 30°C.
(E) Comparison of the transcript profile resulting from a 6 hr treatment with 0.063 % dyclonine to that of an *erg2*Δ deletion strain.
(F) Sterol content of wild-type (left), dyclonine-treated (middle), and *erg2*Δ (right) strains. Dyclonine treatment was 16 hr.
A, zymosterol; B, ergosterol; C, fecosterol; D, episterol; E, lanosterol; F, 4,4-dimethylzymosterol; G, ergosta-5,8,22-trien-3β-ol; H, ergosta-8,22-trien-3β-ol; I, ergosta-8-en-3β-ol. See Strain Table in Supplemental Data for strain details.

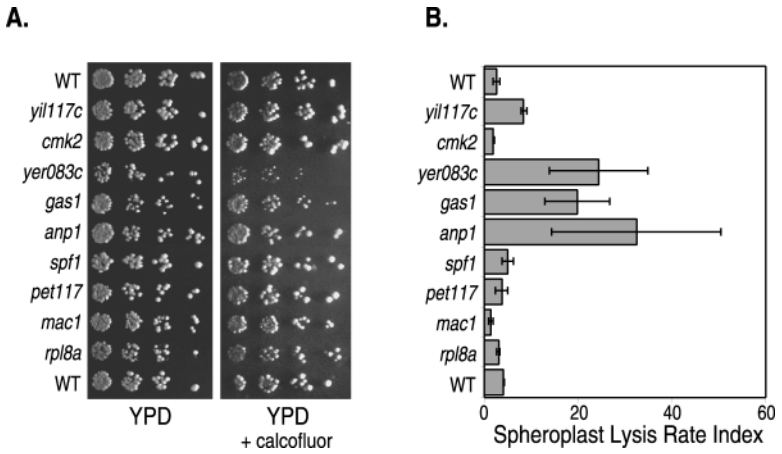

Figure 4. Association of *yer083c*Δ with Phenotypes Indicating Cell Wall Defects

(A) Hypersensitivity of *yer083c*Δ to calcofluor white. Cultures of log-phase wild-type or independently generated haploid deletion strains were equalized for cell density, serially diluted by a factor of four, and spotted onto YPD plates with or without 0.8% calcofluor white (Sigma, Fluorescent Brightener 28), and incubated 2.5 days at 30°C. Most of the calcofluor precipitates when plates are poured, leaving a much lower concentration at the surface of the plate. *gas1*Δ is only slightly sensitive at this concentration. Higher concentrations of calcofluor white inhibited growth of many of the control mutants not related to cell wall function.

(B) Rate Index of spheroplast lysis assays conducted in parallel (see Experimental Procedures). *YER083c*, which encodes a 307 amino acid polypeptide, partially overlaps *YER084w*, a 128 amino acid uncharacterized ORF. However, *YER084w* does not appear to be expressed (Velculescu et al., 1997).

See Strain Table in Supplemental Data for strain details.

sigma receptor: for example, the known Erg2 inhibitor fenpropimorph binds the sigma receptor (Moebius et al., 1997), and haloperidol binds and inhibits the yeast Erg2 enzyme (Moebius et al., 1996). Thus, a potential mechanism for the anesthetic property of dyclonine is that it binds the sigma receptor and inhibits nerve conductance by reducing potassium current.

Functional Discovery without Pathway-Specific Reporters: Involvement of the *YER083c* Gene Product in Yeast Cell Wall Function

Unlike the expression profiles resulting from perturbation of ergosterol biosynthesis, which feature a group of ergosterol-specific transcript inductions, no transcripts could be identified that were induced only by mutants involved in cell wall function (Figures 1B, 3A, and data not shown). Nonetheless, the pattern of combined inductions and repressions was sufficiently unique to cause profiles from cell wall–related mutants *gas1*Δ, *spf1*Δ, and *anp1*Δ to form a discrete cluster together with tunicamycin, glucosamine, and 2-deoxy-D-glucose treatments, all of which impact cell wall function (this profile cluster is labeled "cell wall" in Figure 3A). Grouped together with these cell wall–related profiles is that of *yer083c*Δ, suggesting that *YER083c* is required for normal cell wall function. Indeed, *yer083c*Δ has additional properties characteristic of cell wall mutants: *yer083c*Δ mutants grow slowly (~80% of wild-type growth rate) (Smith et al., 1996; Winzeler et al., 1999), and are hypersensitive to calcofluor white (Figure 4A), a compound frequently used to identify cell wall–related mutants because it binds chitin and interferes with cell wall function (Roncero et al., 1988; Ram et al., 1994; Lussier et al., 1997). Furthermore, *yer083c*Δ cells suffer an increased spheroplast lysis rate (Figure 4B), also indicating alterations in the cell wall (Lipke et al., 1976; Ovalle et al., 1998). Thus, *YER083c* is required for proper cell wall function. The fact that the function of this gene was identified based solely on the composite expression pattern of many genes shows that discovery of novel gene functions via the compendium is independent of pathway-specific reporters. In addition, because the regulation of the *YER083c* transcript itself does not suggest involvement in cell wall function (Supplemental Data), this example shows that identifying a gene's function using the compendium does not require any information on the transcriptional regulation of the gene in question.

Identification of Novel Proteins Required for Mitochondrial Respiration: Subclassification of Gross Phenotypes and Coregulation of the Mitochondrial Ribosome

We next asked whether the transcript profile could yield more specific information regarding the molecular consequences of a perturbation. Defects in mitochondrial respiration are known to be associated with at least two types of nuclear mutations in yeast: those that directly compromise mitochondrial function, and those that are primarily involved in iron regulation, which is in turn required for mitochondrial function (Raguzzi et al., 1988; Eide et al., 1993). We found that the established physiological difference between these two classes of respiratory mutations is mirrored by the expression profile. Two distinct clusters arise from the profiles of respiratory-deficient mutants: a larger group corresponding to mutations in mitochondrial components, and a smaller group composed of deletions mutants in *MAC1*, *VMA8* and *CUP5*, all required for iron regulation (Figures 1B and 5A) (Raguzzi et al., 1988; Eide et al., 1993; Jungmann et al., 1993; Szczypka et al., 1997). The two branches are distinct at $P \leq 0.001$ (Supplemental Data). The major difference between the two groups of profiles is that while all respiratory mutants specifically induce a category of transcripts including iron-homeostatic regulators and a set of major facilitator superfamily genes, only the larger cluster of profiles (corresponding to mutations in mitochondrial components) features upregulation of a set of transcripts that includes several genes encoding citric acid cycle enzymes (Figure 5A). This shows that gross phenotypes can be subclassified by the expression profile into groups that reflect the source of the defect.

To further confirm this subclassification capability, we

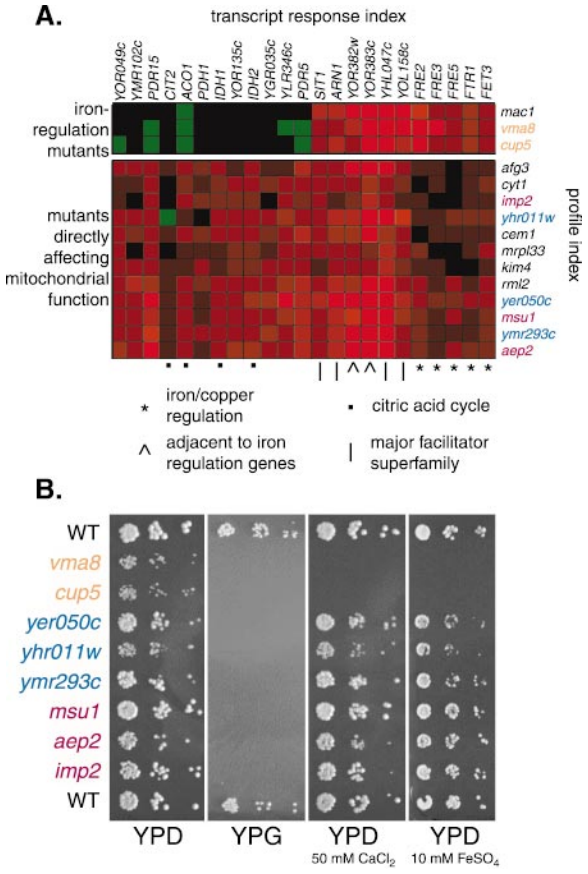

Figure 5. Subclassification of Mitochondrial Dysfunction Mutants by Expression Profile

(A) Two gene/experiment clusters enlarged from Figure 1B. The color scale is identical to that in Figure 1B.

(B) Cultures of log-phase strains were equalized for cell density, serially diluted by a factor of ten, and spotted onto plates containing yeast extract, peptone, and either dextrose (YPD), glycerol (YPG), dextrose plus 50 mM $CaCl_2$ (YPD + 50 mM $CaCl_2$), or dextrose plus 10 mM $FeSO_4$ (YPD + 10 mM $FeSO_4$). The YPD and YPD + 50 mM $CaCl_2$ plates were incubated for 2 days at 30°C; the YPD + 10 mM $FeSO_4$ plate was incubated for 3 days at 30°C; and the YPG plate was incubated for 3.5 days at 30°C.

Different colors of mutant names are used for convenience of locating different mutants in (A) and (B). See Strain Table in Supplemental Data for strain details.

asked whether other laboratory phenotypes of novel mutants followed classifications assigned by the profile clustering tree. Expression profiles displayed by uncharacterized ORF deletions *yhr011w*Δ, *yer050c*Δ, and *ymr293c*Δ are interleaved in the profile cluster tree with those of deletions of mitochondrial components (Figures 1B and 5A), suggesting that these mutants should be unable to grow on glycerol, a nonfermentable carbon source. The specific pattern of transcript inductions, including the set encompassing citric acid–cycle genes, furthermore leads to the prediction that they are not primarily involved in iron regulation and thus will not display the calcium- (Ohya et al., 1991) or iron-sensitivity phenotypes exhibited by the *vma8*Δ and *cup5*Δ mutants. Both of these predictions were confirmed (Figure 5B), verifying the predictive ability of subclassifications on the profile clustering tree. Further consistent with a role in mitochondrial function, these genes all have close homologs in bacteria: Yer050cp is similar to bacterial ribosomal subunit S18, Yhr011wp is similar to bacterial seryl-tRNA synthases, and Ymr293cp is similar to bacterial glutamyl-tRNA amidotransferases.

Subtle Coordinate Regulation of Mitochondrial Ribosome Subunits: Functional Discovery by Transcript Coregulations

Although none of the transcripts of these three new respiratory-related genes (*YHR011w*, *YER050c*, *YMR293c*) is strongly regulated in our data set (none of them is up- or downregulated more than 2-fold in any of the 300 experiments), a search for transcripts most closely regulated with each of them revealed that they are all coregulated with numerous components of the mitochondrial ribosome (data not shown). In fact, at least 37 components of the mitochondrial ribosome are members of a group of at least 114 genes that are coregulated over all 300 experiments (see Figure S1 in Supplemental Data). Because the overwhelming majority of these regulations are less than 2-fold, this coregulated group of genes would have been difficult or impossible to detect without the benefit of a substantial collection of profiles.

Nearly two-thirds of the genes in this group (67/114) have known mitochondrial functions. The majority of the others (41/47) are uncharacterized ORFs. We reasoned that these might also be required for mitochondrial function, and that this would provide an opportunity to test the efficacy of determining gene functions by transcript coregulations alone. Deletion mutants in eight of the mitochondrial ribosome components in this gene cluster all resulted in respiratory deficiency, based on their inability to grow on glycerol (Table 2). Among the uncharacterized ORFs were five genes with similarity to prokaryotic and/or chloroplast ribosomal proteins; deletion mutants in four of these ORFs were tested and were likewise respiratory deficient, suggesting that they are previously unidentified components of the mitochondrial ribosome (Table 2). Among five randomly selected uncharacterized ORFs in this group, however, deletion mutants in only three were respiratory deficient (Table 2). Thus, coordinated transcriptional regulations can be used to enrich for novel genes with a given phenotype, but there is potential for false positives.

Low-Magnitude Profiles Identify Novel Proteins Involved in Protein Synthesis

The unexpected identification of very low-magnitude transcriptional regulation of the mitochondrial ribosome suggests that such low-amplitude but meaningful regulatory patterns might be common, and that they might be sufficient to cause correlations among related experiments. We next asked whether similarity between experiments driven by such low-magnitude changes can be shown to be biologically significant. Mutations in genes involved in protein synthesis provide a test case: although profiles from a number of them form a discrete cluster, only a handful of genes ($<$10) are significantly up- or downregulated more than 3-fold in any of these profiles (Figure 1B and Supplemental Data). Indeed, the profile associations do not rely on these outliers, because the association of all these experiments in the

Table 2. Deletion Phenotypes of Genes Transcriptionally Coregulated with Components of the Mitochondrial Ribosome

Deleted Gene	Growth on Glycerol	Sequence Homologs
MRPS5	−	prokaryotic ribosomal protein S5
MRPL6	−	prokaryotic ribosomal protein L6
MRPL13	−	prokaryotic ribosomal protein L13
MRPL16	−	prokaryotic ribosomal protein L16
MRPL24	−	prokaryotic ribosomal protein L24
MRPL31	−	prokaryotic ribosomal protein L31
MRPL37	−	prokaryotic ribosomal protein L37
MRPL38	−	prokaryotic ribosomal protein L38
YDR115W	−	prokaryotic ribosomal protein L34
YDR116C	+/− (sick)	prokaryotic ribosomal protein L1
YMR188C	−	prokaryotic ribosomal protein S17
YNL177C	−	*T. maritima* ribosomal protein L22
YGR165W	−	−
YHR116W	−	−
YDR175C	−	−
YML030W	+	−
YMR157C	+	−

Deletions mutants were barcoded start-to-stop codon strains obtained from the Saccharomyces Genome Deletions Consortium (Winzeler et al., 1999). Strains were inoculated from glycerol stocks onto YPD plates, then a dilution series was made and replicated to both YPD and YPG plates. All strains grew on YPD; strains not showing healthy colonies on YPG after four days were scored as unable to grow on glycerol. Two independently generated strains were tested in each case. See Strain Table in Supplemental Data for strain details.

profile clustering tree is retained even when all regulations greater than 1.5-fold are masked from the 300-experiment data structure (Table 3, column D). Furthermore, when such small regulations are used as the basis of clustering analysis, this profile cluster gains additional translation-related experiments that are otherwise excluded, such as tet-*YEF3* and cycloheximide treatment (which have the lowest magnitude profiles in this group, probably due to the fact that they are acute treatments) (Figure 6A). This verifies that low-magnitude expression changes can be used to group profiles from related mutants.

Like the ergosterol, cell wall, and mitochondrial respiration profile clusters, the translation-related cluster includes mutants with deletions in uncharacterized ORFs (*yor078w*Δ, *ymr269w*Δ, *yhr034c*Δ) together with known ribosome subunit deletions *rps24a*Δ, *rpl27a*Δ, *rpl8a*Δ, and *rpl12a*Δ (in yeast, many ribosomal proteins are duplicated; deletion of one copy generally results in slow growth). This experiment profile cluster also includes a partial deletion of translation elongation factor 5A (encoded by *YEL034w*) produced by deletion of the overlapping ORF *YEL033w*. Although the comprehensive biochemical analyses to which ribosomes have been subjected (Link et al., 1999) make it seem unlikely that the products of these uncharacterized ORFs are ribosomal proteins, each of these new mutants has both a reduced growth rate and a reduced protein synthesis rate, similar to ribosome subunit deletion strains; other slow-growth mutants tested do not have reduced protein synthesis rates (Figure 6B). *YOR078w*, *YMR269w*, and *YHR034c* could conceivably encode previously unidentified ribosome assembly or stability factors, translational regulatory factors, or nucleolar proteins. This example shows that low-magnitude expression patterns can be used to discover the cellular functions of novel genes.

Discussion

The Compendium Approach: A New Method for Functional Discovery

We present a conceptually simple and straightforward method for determining the cellular pathway(s) affected by uncharacterized perturbations. In this method, which we call the compendium approach, the expression profile caused by an uncharacterized perturbation is compared to a large and diverse set of reference profiles, and the pathway(s) perturbed is determined by matching the profile caused by the uncharacterized perturbation to profiles that correspond to disturbance of a known cellular pathway. Success of the method depends on the existence of distinct expression profiles that identify different pathways. We show that many classes of mutants display recognizable and distinct expression profiles (Figure 2, Table 3). Although we illustrate the basic principle of the method primarily using deletion mutants, we also show that a compendium can be effective in understanding the activities of pharmacological compounds.

Importantly, we find that meaningful expression patterns can involve groups of transcripts whose relative abundance changes at levels considerably less than 2-fold. Clusters of profiles involving not only protein synthesis but ergosterol biosynthesis, mitochondrial respiration, mating, cell wall biosynthesis, histone deacetylation, the vacuolar H^+-ATPase, and others are all largely intact in clustering analyses even after all regulations over 1.5-fold are masked in the data structure (Table 3, column D). This shows that there are layers of regulation that are ignored when array data are interpreted outside the context of a compendium, because ratio measurements less than 2-fold are often considered to be unreliable in an isolated microarray experiment. It also underscores the importance of generating high-quality internally consistent data, and the necessity of developing means to compensate for noise or other biases, if the compendium method is to be used most effectively. Furthermore, it shows that the compendium method is totally unlike the use of highly induced pathway-specific reporters to detect disturbance of particular cellular processes, and relies instead on the composite pattern of many transcriptional changes to "fingerprint" perturbation of different pathways.

On the whole, the specific clusters of mutant profiles are resiliant to a variety of alterations in the clustering parameters (Table 3). However, no one specific set of parameters is capable of correctly grouping all known functionally related mutants, suggesting that identification of novel gene functions can be enhanced by examining multiple analyses. It is interesting that many biologically relevant profile associations in the clustering tree are intact even when all transcripts encoding proteins with known functions or close sequence homologs are

Table 3. Composition and Statistical Significance of Experiment Clusters Resulting from Varied Clustering Parameters

	A	B	C	D	E	F	G	H
clustering parameters:	≥2 genes over 3-fold and P ≤ 0.01 (Fig. 1B)	≥2 genes over 2-fold and P ≤ 0.01	≥2 genes over 1.58-fold and P ≤ 0.1; genes over 2-fold masked	≥2 genes over 1.26-fold and P ≤ 0.1; genes over 1.5-fold masked	≥2 genes at P ≤ 0.01; genes under 2-fold masked	≥2 genes at P ≤ 0.01; genes under 1.5-fold masked	only ORFs with no known functions or homologs	All ORFs, all genes
# experiments:	127	179	168	177	179	179	300	300
# genes:	568	1435	3350	4072	1485	1495	1091	5962
mitochondrial function	(≤0.001) afg3 imp2 **yhr011w** cem1 mrpl33 kim4 rml2 **yer050c** msu1 **ymr293c** aep2 rip1 qcr2 cyt1 *bub3*	(≤0.001) afg3 imp2 **yhr011w** cem1 mrpl33 kim4 rml2 **yer050c** msu1 **ymr293c** aep2 rip1 qcr2 cyt1 pet111 cbp2	(0.021) afg3 imp2 **yhr011w** cem1 mrpl33 kim4 rml2 **yer050c** msu1 **ymr293c** aep2 rip1 qcr2 cyt1 pet111 cbp2 pet117	(≤0.001) afg3 imp2 **yhr011w** cem1 mrpl33 kim4 rml2 **yer050c** msu1 **ymr293c** aep2 rip1 qcr2 cyt1 pet111 cbp2 pet117 *ard1*	(≤0.001) afg3 **yhr011w** cem1 mrpl33 kim4 rml2 **yer050c** msu1 **ymr293c** aep2 rip1 cyt1	(≤0.001) afg3 imp2 **yhr011w** cem1 mrpl33 kim4 rml2 **yer050c** msu1 **ymr293c** aep2 cyt1 pet111 cbp2	(≤0.001) afg3 imp2 **yhr011w** cem1 mrpl33 kim4 rml2 **yer050c** msu1 **ymr293c** aep2 cbp2 pet117 *ard1* doxycycline	(≤0.001) afg3 imp2 **yhr011w** cem1 mrpl33 kim4 rml2 **yer050c** msu1 **ymr293c** aep2 rip1 qcr2 cyt1 pet111 cbp2 pet117 *ard1* doxycycline
cell wall	(0.01) tet-KAR2 tet-CDC42 tet-FKS1 fks1	(≤0.001) tet-KAR2 tet-CDC42 tet-FKS1 fks1 glucosamine 2-deoxy-D-glucose erd1 spf1 anp1 cla4	(≤0.001) tet-KAR2 tet-CDC42 tet-FKS1 fks1 glucosamine 2-deoxy-D-glucose erd1 tet-RHO1 hog1	(≤0.001) tet-KAR2 tet-CDC42 tet-FKS1 fks1 glucosamine 2-deoxy-D-glucose erd1 cla4 tet-RHO1 hog1 kre1 gyp1 *ymr030w*	(0.0026) tet-CDC42 tet-FKS1 fks1	(0.002) tet-KAR2 tet-CDC42 tet-FKS1 fks1 cla4	(≤0.001) tet-KAR2 tet-CDC42 tet-FKS1 fks1 glucosamine 2-deoxy-D-glucose spf1 anp1 tet-RHO1	(0.008) tet-KAR2 tet-CDC42 tet-FKS1 fks1 glucosamine 2-deoxy-D-glucose erd1 spf1 anp1 cla4

(*continued*)

Table 3. Continued

clustering parameters:	A ≥2 genes over 3-fold and P ≤ 0.01 (Fig. 1B)	B ≥2 genes over 2-fold and P ≤ 0.01	C ≥2 genes over 1.58-fold and P ≤ 0.1; genes over 2-fold masked	D ≥2 genes over 1.26-fold and P ≤ 0.1; genes over 1.5-fold masked	E ≥2 genes at P ≤ 0.01; genes under 2-fold masked	F ≥2 genes at P ≤ 0.01; genes under 1.5-fold masked	G only ORFs with no known functions or homologs	H All ORFs, all genes
cell wall (continued)	(≤0.001) tunicamycin **yer083c** gas1 spf1 anp1 glucosamine 2-deoxy-D-glucose	(0.006) tunicamycin **yer083c** gas1	(0.006) tunicamycin **yer083c** gas1	(0.01) tunicamycin **yer083c** gas1	(0.024) tunicamycin **yer083c** gas1 2-deoxy-D-glucose	(0.002) tunicamycin **yer083c** gas1 2-deoxy-D-glucose	(0.006) tunicamycin **yer083c** gas1	(0.006) tunicamycin **yer083c** gas1
protein synthesis	(≤0.001) rps24a yel033w *ymr014w* **yhr034c** rpl27a **ymr269w** mrt4	(≤0.001) rps24a yel033w *ymr014w* **yhr034c** rpl27a **ymr269w** mrt4 **yor078w** rpl8a rpl12a *arg5,6*	(≤0.001) rps24a yel033w *ymr014w* **yhr034c** rpl27a **ymr269w** mrt4 **yor078w** rpl8a rpl12a *arg5,6* top1 rps27b	(0.04) rps24a yel033w *ymr014w* **yhr034c** rpl27a **ymr269w** mrt4 **yor078w** rpl8a rpl12a *arg5,6* top1 rps27b *yor006c*	(≤0.001) rps24a yel033w *ymr014w* **yhr034c** rpl27a **ymr269w** mrt4 **yor078w** rpl8a rpl12a rps27b *rad27*	(≤0.001) rps24a yel033w *ymr014w* **yhr034c** rpl27a **ymr269w** mrt4 **yor078w** rpl8a rpl12a *arg5,6*	(≤0.001) rps24a yel033w *ymr014w* **yhr034c** rpl27a **ymr269w** mrt4 **yor078w** rpl8a rpl12a *arg5,6* *yor006c* *rad27*	(≤0.001) rps24a yel033w *ymr014w* **yhr034c** rpl27a **ymr269w** mrt4 **yor078w** rpl8a rpl12a *arg5,6* top1 *yor006c*
ergosterol biosynthesis	(0.003) lovastatin tet-HMG2 terbinafine hmg1	(≤0.001) lovastatin tet-HMG2 terbinafine hmg1 *top3*	(0.005) lovastatin tet-HMG2	(0.1) lovastatin tet-HMG2	(0.04) lovastatin tet-HMG2 *top3*	(≤0.001) lovastatin tet-HMG2 terbinafine hmg1 *top3*	(0.04) lovastatin tet-HMG2 terbinafine	(≤0.001) lovastatin tet-HMG2
	(0.003) erg3 **erg28** tet-ERG11 itraconazole erg2	(0.006) erg3 **erg28** tet-ERG11 itraconazole erg2	(0.006) erg3 **erg28** tet-ERG11 itraconazole erg2	(0.01) erg3 **erg28** itraconazole erg2	(≤0.001) erg3 **erg28** tet-ERG11 itraconazole erg2	(0.002) erg3 **erg28** tet-ERG11 itraconazole erg2	(0.006) erg3 **erg28** tet-ERG11 itraconazole erg2	(0.006) erg3 **erg28** tet-ERG11 itraconazole erg2

(*continued*)

Table 3. Continued

clustering parameters:	A ≥2 genes over 3-fold and P ≤ 0.01 (Fig. 1B)	B ≥2 genes over 2-fold and P ≤ 0.01	C ≥2 genes over 1.58-fold and P ≤ 0.1; genes over 2-fold masked	D ≥2 genes over 1.26-fold and P ≤ 0.1; genes over 1.5-fold masked	E ≥2 genes at P ≤ 0.01; genes under 2-fold masked	F ≥2 genes at P ≤ 0.01; genes under 1.5-fold masked	G only ORFs with no known functions or homologs	H All ORFs, all genes
mating	(0.006) ste4 ste11 ste7 ste18 ste12 fus3, kss1 ste5	(0.006) ste4 ste11 ste7 ste18 ste12 fus3, kss1 ste5	(0.018) ste4 ste11 ste7 ste18 ste12 fus3, kss1 ste5	(0.128) ste4 ste11 ste7 ste18 ste12 fus3, kss1	(0.006) ste4 ste11 ste7 ste18 ste12 fus3, kss1 ste5	(0.006) ste4 ste11 ste7 ste18 ste12 fus3, kss1 ste5	(≤0.001) ste4 ste11 ste7 ste18 ste12 fus3, kss1 ste5 *yjl107c*	(0.027) ste4 ste11 ste7 ste18 ste12 fus3, kss1 ste5
MAPK activation	(≤0.001) dig1/dig2 fus3 sst2 dig1 hog1 ras2	(0.006) dig1/dig2 fus3 sst2	(≤0.001) dig1/dig2 fus3	(≤0.001) dig1/dig2 fus3	(≤0.001) dig1/dig2 fus3 sst2 hog1 ras2	(0.006) dig1/dig2 fus3 sst2 hog1 ras2	(≤0.001) dig1/dig2 fus3 sst2 dig1 *kin3*	(≤0.001) dig1/dig2 fus3 sst2
rnr1, HU	(0.024) rnr1 HU MMS rad6	(≤0.001) rnr1 HU MMS rad6 swi6	(0.011) rnr1 HU swi6	(≤0.001) rnr1 HU swi6	(≤0.001) rnr1 HU MMS rrm3	(0.01) rnr1 HU MMS rad6 swi6	(0.057) rnr1 HU MMS rad6 swi6	(≤0.001) rnr1 HU MMS rad6 swi6
histone deacetylase	(≤0.001) rpd3 sap30 FR901,228 sin3	(≤0.001) rpd3 sap30 FR901,228 sin3	(≤0.001) rpd3 sap30 FR901,228 rad6	(≤0.001) rpd3 sap30 sin3 rad6	(≤0.001) rpd3 sap30 sin3 stb4	(≤0.001) rpd3 sap30 FR901,228 sin3	(≤0.001) rpd3 sap30 FR901,228 sin3	(≤0.001) rpd3 sap30 FR901,228 sin3

(*continued*)

Table 3. Continued

clustering parameters:	A ≥2 genes over 3-fold and P ≤ 0.01 (Fig. 1B)	B ≥2 genes over 2-fold and P ≤ 0.01	C ≥2 genes over 1.58-fold and P ≤ 0.1; genes over 2-fold masked	D ≥2 genes over 1.26-fold and P ≤ 0.1; genes over 1.5-fold masked	E ≥2 genes at P ≤ 0.01; genes under 2-fold masked	F ≥2 genes at P ≤ 0.01; genes under 1.5-fold masked	G only ORFs with no known functions or homologs	H All ORFs, all genes
isw	(≤0.001) isw1 isw2 isw1/2 *hst3*	(0.185) isw1 isw2 isw1/2	(≤0.001) isw1 isw2 isw1/2	(0.004) isw1 isw2 isw1/2	(0.04) isw1 isw2 isw1/2	(≤0.001) isw1 isw2 isw1/2 *rrm3*	(≤0.001) isw1 isw2 isw1/2	(≤0.001) isw1 isw2 isw1/2
vacuolar ATPase/ iron regulation	(≤0.001) cup5 vma8 mac1	(≤0.001) cup5 vma8 mac1	(0.1) cup5 vma8	(0.1) cup5 vma8	(≤0.001) cup5 vma8 mac1	(0.04) cup5 vma8 mac1	(0.04) cup5 vma8	(0.024) cup5 vma8 mac1
sir	(≤0.001) sir2 sir3	(0.167) sir2 sir3	(0.332) sir2 sir3		(≤0.001) sir2 sir3 sir4	(0.1) sir2 sir3 sir4	(0.1) sir2 sir3	(0.01) sir2 sir3 sir4
tup1, ssn6	(0.024) tup1 ssn6	(≤0.001) tup1 ssn6	(0.157) tup1 ssn6		(0.040) tup1 ssn6	(0.01) tup1 ssn6	(0.006) tup1 ssn6	(≤0.001) tup1 ssn6

See Supplemental Data for full cluster trees. The clusters shown were selected because they clearly correspond to established functional categories; other clusters either could not be ascribed to specific functional categories or were formed due to frequently occurring biases. Columns A and B were generated using the standard clustering procedure described in "Clustering and Correlations" in the Supplemental Data. Columns C, D, E, and F were generated by removing the indicated regulations from the data set, then running the standard clustering procedure. 1.26-fold is $10^{0.1}$; 1.58-fold is $10^{0.2}$. Column G was generated by restricting clustering to the list of 1161 genes in the text file "unknown ORFs" in the Supplemental Data, 1091 of which are in the data set. No cuts were made for column G, although only genes measured in more than 50% of the experiments are included. P values are indicated in parentheses and are assigned using the bootstrap approach described in "Clustering and Correlations" document in the Supplemental Data. Italics indicate genes not previously known to belong in the indicated functional groups, and may represent either false positives or novel assignments of genes to functional classes. Boldface indicates novel genes described in this study.

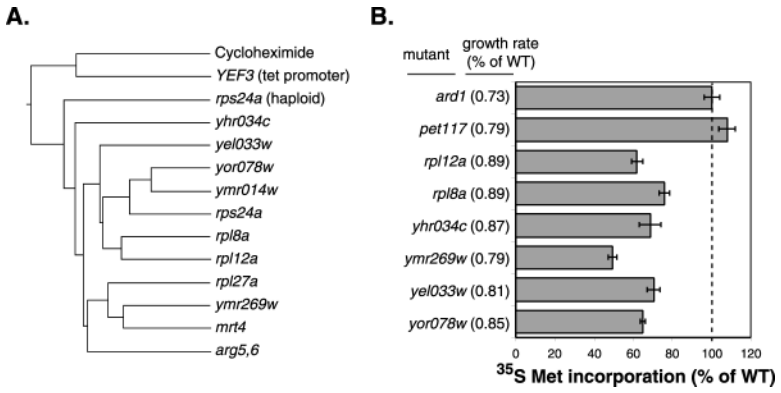

Figure 6. Subtle Transcriptional Consequences of Novel Mutations Affecting Protein Synthesis

(A) Selected experiment cluster resulting from relaxing clustering parameters to accept experiments with 2 or more genes up- or down-regulated by 2-fold or more and significant at $P \leq 0.01$, and genes significant at $P \leq 0.01$ in 2 or more experiments. All characterized mutants in this cluster are previously known to impact ribosome function or translation, with the exception of *arg5,6*. The significance of the branch shown is $P \leq 0.001$; the significance of the division of cycloheximide and tet-*YEF3* from the other mutants is also $P \leq 0.001$.

(B) Bar graph indicates the ^{35}S-methionine incorporation rate relative to a wild-type strain grown in parallel with the mutant; incorporation was normalized to cell OD to compensate for slow-growing mutants. Error bars represent the standard error of the mean of at least five measurements made in duplicate. *ard1*Δ and *pet117*Δ mutants are included as negative controls. See Strain Table in Supplemental Data for strain details.

masked in the analysis (Table 3, column G). This further illustrates that the compendium approach relies entirely on the patterns of the profiles, and not on any knowledge of the specific transcripts that compose the profiles. It also shows that a compendium need not be composed of full-genome profiles.

Discovery of Eight Gene Functions and a Previously Unknown Drug Target

To verify that the compendium approach to functional discovery works, we assign eight uncharacterized genes to four different cellular pathways and determine an unknown target of a common drug. Among the eight new genes, only those involved in mitochondrial function have significant sequence features suggesting specific biochemical roles. The new genes we identify in sterol biosynthesis and protein synthesis would have been very difficult to discover by any conventional primary screening methods. Although discovery of the function of any new gene is significant, the *ERG28* gene is of particular interest because its human homolog has a related or conserved cellular function, suggesting that it is a novel factor in cholesterol biosynthesis. The *YHR034c* gene product, involved in protein synthesis, also has a sequence homolog in the human genome, suggesting that it too may be a functionally conserved protein. The analysis and discoveries we present are by no means comprehensive; only a handful of the uncharacterized mutants were examined in detail, and analysis techniques other than gross profile comparison and 2-D clustering will probably be required to completely circumvent frequent biases and improve resolution of small but significant components in large profiles. We anticipate that these data will be a resource for continued functional discovery by us and other researchers.

Extrapolating to the Entire Genome: Obtaining Significant Expression Profiles from All Yeast Mutants

From our data, it is difficult to predict the total number of unique transcriptional responses yeast can produce. However, the 300 experiments presented here did not approach saturation, as many of the mutant profiles could not be placed into a recognizable category with other mutants. Based on the recurrence of known patterns as a function of number of mutant profiles in our data set, a crude prediction can be made that 300–700 different full-genome transcription patterns would be obtained from a full set of 5000 individual yeast deletion mutants profiled under the single condition used in this study (data not shown). Because the likelihood of detecting recurring transcriptional patterns increases with the number of profiles, as does the ability to estimate the significance of any given measurement or cluster, it is possible that the cellular functions of many more of the uncharacterized ORFs we profiled could be identified using a larger set of reference profiles. An expansion of the current study thus holds considerable potential for determining functions of many of the currently uncharacterized yeast genes.

A major limiting requirement in the use of the expression profile as a completely universal functional assay appears to be obtaining profiles from all of the mutants: under the single condition employed here, the transcript profiles of only about half of the mutants were significantly changed from wild-type. However, significant transcriptional phenotypes were obtained from virtually every mutant affecting growth. This suggests that significant profiles might be obtained from the remaining half of the mutants under conditions that impair growth of the mutant in comparison to wild-type. In fact, we have previously shown that mutants in the calcineurin signaling pathway, which are hypersensitive to cations, display characteristic transcriptional alterations from wild-type when both wild-type and mutant are grown in rich medium containing calcium (Marton et al., 1998), although the same mutants (*cna1 cna2*, for example) display neither a growth defect nor significant transcriptional alterations under the growth conditions used in this study. One could imagine developing a panel of conditions under which growth of a large majority of deletion mutants is compromised at least once in comparison to wild-type, and subsequently generating several parallel compendia encompassing mutation of all genes in the genome. At least two tools exist for accomplishing the task of rapidly identifying the appropriate conditions for each mutant: pooled barcoded deletions, developed by Shoemaker et al. (1996), and

"phenotypic macroarrays," described by Ross-Macdonald et al. (1999).

Advantages and Applications for a Compendium of Expression Profiles

The fundamental advantage of the compendium approach over conventional assays is that it substitutes a single genome-wide expression profile in the place of many conventional (often tedious) assays that measure only a single cellular parameter. Because the compendium approach to determining gene function does not rely at all on the regulatory characteristics of the gene of interest, it has a significant advantage over the widely accepted idea of assessing gene function based on coregulation (Eisen et al., 1998). Furthermore, the same compendium used to characterize mutants can also be used to characterize other perturbations, including treatments with pharmaceutical compounds, and potentially disease states as well. Thus, many practical applications for a compendium of expression profiles can be envisioned. The compendium approach represents a highly parallel alternative to high-throughput screening of chemical libraries: rather than using a simple assay to examine many compounds for a single activity, each compound can be examined for many possible activities in a single assay. The fact that multiple molecular consequences of a given perturbation can be discerned simultaneously suggests that a compendium might be used to discover unanticipated activities of drugs (Marton et al., 1998), or to ensure that only desired treatment effects are occurring in patients. Although this study employed mutants in the yeast *S. cerevisiae*, there is no fundamental barrier to creating a similarly useful compendium of profiles in more complex organisms such as mammals, although separate data sets might be required for different tissues or developmental stages. Current efforts aimed at establishing large collections of mapped insertion mutants in *C. elegans*, *D. melanogaster*, and *A. thaliana* (Pennisi, 1998; Somerville and Somerville, 1999; Spradling et al., 1999) will facilitate similar studies in these metazoan species.

Experimental Procedures

Supplemental Data

Complete data sets for the 300 compendium experiments, the 63 control experiments, and dyclonine treatment, as well as an experiment list, strain table, versions of figures with gene and experiment labels, descriptions of arraying and quantitation methods, a summary of reproducibility, formulas for correlations and error models, and other related tables and documents can be downloaded from http://www.rii.com/tech/pubs/cell_hughes.htm. For the convenience of readers, the Supplemental Data include ratios-only versions of the data sets that can be loaded directly into the public-domain clustering program described by Eisen et al. (1998) (downloadable from http://rana.stanford.edu/software/, but not used in this study). Similarity searches among genes and profiles in this study can be performed at http://www.rii.com/tech/data/cell_search.

A subset of the authors' Supplemental Data—including enlarged versions of Figures 1B, 2A, and 3A—is also available on the *Cell* website (http://www.cell.com/cgi/content/full/101/7/109/DC1).

Yeast Strains and Plasmids

The Saccharomyces Genome Deletions Consortium strain background (Winzeler et al., 1999) was employed in all experiments. In order to minimize the potential impact of unlinked recessive mutations associated with strain construction, homozygous diploid deletion mutants were profiled when possible. All deletion mutants are start-to-stop codon. For experiments involving tet-regulatable genes, the natural promoter on the chromosome was replaced with a heptamerized tet operator fused to a kanamycin-resistance cassette enabling direct integration, and the "tet activator" (tTA*, which dissociates in the presence of doxycyline) was supplied either on a CEN plasmid (Gari et al., 1997) or integrated into the genome. Plasmids expressing human ERG28 (hERG28) were constructed by PCR: the hERG28 ORF was amplified from a mixture of human cDNA libraries and cloned by in vivo recombination in front of the yeast *HOR7* promoter on pDW394 (Acacia Biosciences), a high copy URA3 vector. Inserts were confirmed by sequencing. Plasmids used in Figure 3D are pRP382 and pRP383.

Yeast Culture and cDNA Microarray Expression Analysis

Experimental (mutant or chemical treated) cultures were grown, harvested, and processed in parallel with corresponding wild-type or control cultures. (See "Experiment List" in Supplemental Data.) Several colonies of similar size were picked from freshly streaked YAPD agar plates into liquid Synthetic Complete medium (SC) with 2% glucose, grown overnight at 30°C to mid-log phase, diluted to 0.4–1.0 $\times$ 10^6 cells/ml, and grown an additional 5–7 hr until reaching 0.4–1.0 $\times$ 10^7 cells per ml, at which point they were pelleted by centrifugation for 2 min at room temperature and frozen in liquid nitrogen. The final optical densities of experimental and control cultures were matched as closely as possible. For experiments involving treatments with chemical compounds (including doxycycline), compounds were added at the beginning of the 5–7 hr final growth phase, with equal amounts of solvent or doxycycline added to control cultures as appropriate. Total RNA was prepared by phenol:chloroform extraction followed by ethanol precipitation as described previously (Marton et al., 1998) except that vortexing with glass beads was replaced by a 10 min incubation at 65°C followed by 1 min of vortexing. Poly-A^+ RNA purification, cDNA labeling, microarray production, and microarray hybridization and washing were as described previously (Marton et al., 1998) with measurements taken in fluor-reversed pairs (i.e., each time a mutant was analyzed, it was hybridized to two arrays). Arrays were scanned, images were quantitated and physical artifacts (dust and salt residue) edited as described previously (Marton et al., 1998). Resulting data files were evaluated by a series of quality-control criteria relating first to the image itself and second to known biological artifacts. Experiments flagged as containing biological artifacts were noted but not excluded, for purposes of illustrating the impact of biases. Aneuploid strains are indicated in the data tables and figures in the Supplemental Data by two asterisks followed by the number of the duplicated or missing chromosome (for details see Hughes et al., 2000). The data for each of the mutants profiled in duplicate is a statistical composite of the two replicates.

Clustering and Error Model

Clustering analysis was composed of two steps: first, profiles and transcripts were selected from a data matrix, and second, experiments and responsive genes were grouped by agglomerative hierarchical clustering (Hartigan, 1975) where the similarity measure is the error-weighted correlation coefficient. Genes and experiments (columns and rows, respectively) were reordered according to the resulting clustering similarity trees. Significance (P value) for gene regulations takes the gene measurement error and biological variation in control experiments into account. See Supplemental Data for a complete description of error-weighted correlations, a more detailed description of the clustering procedure, and a summary of the bootstrap method used to calculate branch P values.

Parallel Growth Assay

Among the 276 deletion mutants profiled, 198 corresponded to barcoded homozygous disruption strains (Shoemaker et al., 1996). These were grown as part of a pool of barcoded homozygous deletion strains obtained from the Saccharomyces Genome Deletions Consortium (Winzeler et al., 1999). Seven time points were taken over 20 population doublings of the pool in SC medium containing

2% glucose. The relative abundances of the different tags over 20 population doublings were determined by a two-color hybridization assay using custom-made 24,000-element oligonucleotide arrays (A. Blanchard and S. H. F., unpublished data). Growth rates for each of the strains in the pool were determined by plotting the changes in relative abundance (log ratios) versus population doublings using a linear fit model.

Sterol Profiles

Sterols were identified using pure standards and on the basis of relative retention times. Nonsaponifiables were isolated from yeast as previously described (Molzahn and Woods, 1972). Gas chromatography (GC) analyses of the nonsaponifiable fraction were analyzed on an HP5890 series II GC equipped with the HP chemstation software. The capillary column (DB-5) was 15 m $\times$ 0.25 mm i.d., 0.2 μm film thickness and was programmed from 195°C to 300°C (195°C for 3 min, 5.5°C/min to 300°C then held for 10 min). The linear velocity was 30 cm/s using nitrogen as the carrier gas and all injections were run in the splitless mode.

Spheroplast Lysis Rate Assay

Spheroplast Lysis Rate was determined following methods described previously (Ovalle et al., 1998). Cells were grown to nearly identical densities in mid-log phase, washed 3 times with TE, and resuspended in TE at $OD_{600} = 0.6$. Zymolyase 100T (dissolved in 50% glycerol) was added to 5 μg/ml and OD_{600} was measured every 3–7 min for 1 hr. Rate Index is determined as Lag Time/Maximum Lysis Rate, where Lag Time is the time in which the OD_{600} decreased by 0.05, and the Maximum Lysis Rate is the absolute value of the slope of the least-squares fit line for the portion of the lysis curve (six or more points) with the steepest log-linear decline, as defined previously (Ovalle et al., 1998).

^{35}S-Methionine Incorporation Rate Assay

Yeast strains were grown in synthetic complete medium lacking methionine and cysteine. Cells were grown in 40 ml liquid culture at 30°C to an A_{600} of 0.5–0.7. At zero time, methionine and cysteine were added at 50 mM concentrations along with 1 μCi/ml of ^{35}S-labeled pro-mix (Amersham Pharmacia biotech). Duplicate aliquots of 1 ml were removed from each culture at 20 min intervals for up to 2 hr. Cell densities were monitored at 1 hr intervals. Samples were precipitated with 200 μl of 50% ice-cold TCA (trichloroacetic acid) for 10 min on ice, heated at 90°C for 15 min, chilled on ice for 10 min, filtered through GF/C filters, washed with 15 ml of 2.5% ice-cold TCA followed by 10 ml of cold 95% ethanol, dried and counted in a scintillation counter with 3.5 ml of the scintillant (Bio-safe II from RPI). Labeled methionine/cysteine incorporation into proteins was calculated based on the total amount of TCA precipitable counts.

Acknowledgments

We thank D. Dimster-Denk, D. Gottschling, H. Rao, J. Tong, T. Tsukiyama, T. Ward, S. Whelen, and A. Wolfe for yeast strains; D. Wong for plasmid pDW394; M. Ashby, J. Phillips, P. Cundiffe, and J. King for many helpful discussions, J. Koch, A. Leonardson, and E. Callo for assistance with scanning and image analysis, and L. Hartwell and A. Murray for comments on the manuscript. This work was supported by Rosetta Inpharmatics, Inc., a grant from the National Institutes of Health (AI38598) to M. B., and a grant from the National Cancer Institute (CA78746) to J. S. We apologize to the many researchers whose work could not be cited due to space considerations, and refer the reader to Yeast Protein Database (http://customer.proteome.com) and the Saccharomyces Genome Database (http://genome-www.stanford.edu/Saccharomyces/) for information and references regarding specific gene products.

Received December 17, 1999; revised May 5, 2000.

References

Alberts, A.W., Chen, J., Kuron, G., Hunt, V., Huff, J., Hoffman, C., Rothrock, J., Lopez, M., Joshua, H., Harris, E., et al. (1980). Mevinolin: a highly potent competitive inhibitor of hydroxymethylglutaryl-coenzyme A reductase and a cholesterol-lowering agent. Proc. Natl. Acad. Sci. USA *77*, 3957–3961.

Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., et al. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature *403*, 503–511.

Bard, M. (1972). Biochemical and genetic aspects of nystatin resistance in *Saccharomyces cerevisiae*. J. Bateriol. *111*, 649–657.

Cho, R.J., Campbell, M.J., Winzeler, E.A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T.G., Gabrielian, A.E., Landsman, D., Lockhart, D.J., and Davis, R.W. (1998). A genome-wide transcriptional analysis of the mitotic cell cycle. Mol. Cell *2*, 65–73.

Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P.O., and Herskowitz, I. (1998). The transcriptional program of sporulation in budding yeast. Science *282*, 699–705.

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.

DeRisi, J.L., Iyer, V.R., and Brown, P.O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. Science *278*, 680–686.

Eide, D.J., Bridgham, J.T., Zhao, Z., and Mattoon, J.R. (1993). The vacuolar H(+)-ATPase of *Saccharomyces cerevisiae* is required for efficient copper detoxification, mitochondrial function, and iron metabolism. Mol Gen. Genet. *241*, 447–456.

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.

Galitski, T., Saldanha, A.J., Styles, C.A., Lander, E.S., and Fink, G.R. (1999). Ploidy regulation of gene expression. Science *285*, 251–254.

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.

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., et al. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science *286*, 531–537.

Gray, N.S., Wodicka, L., Thunnissen, A.M., Norman, T.C., Kwon, S., Espinoza, F.H., Morgan, D.O., Barnes, G., LeClerc, S., Meijer, L., et al. (1998). Exploiting chemical libraries, structure, and genomics in the search for kinase inhibitors. Science *281*, 533–538.

Hanner, M., Moebius, F.F., Flandorfer, A., Knaus, H.G., Striessnig, J., Kempner, E., Glossmann, H. (1996). Purification, molecular cloning, and expression of the mammalian sigma1-binding site. Proc. Natl. Acad. Sci. USA *93*, 8072–8077.

Hartigan, J.A. (1975). Clustering Algorithms (New York: John Wiley & Sons).

Heller, R.A., Schena, M., Chai, A., Shalon, D., Bedilion, T., Gilmore, J., Woolley, D.E., and Davis, R.W. (1997). Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. Proc. Natl. Acad. Sci. USA *94*, 2150–2155.

Holstege, F.C., Jennings, E.G., Wyrick, J.J., Lee, T.I., Hengartner, C.J., Green, M.R., Golub, T.R., Lander, E.S., and Young, R.A. (1998). Dissecting the regulatory circuitry of a eukaryotic genome. Cell *95*, 717–728.

Hughes, T.R., Roberts, C.J., Dai. H., Jones, A.R., Meyer, M.R., Slade, D., Burchard, J., Dow, S., Ward, T.R., Kidd, M.J., Friend, S.H. and Marton, M.J. (2000). Widespread aneuploidy revealed by DNA microarray expression profiling. Nat. Genet., in press.

Jungmann, J., Reins, H.A., Lee, J., Romeo, A., Hassett, R., Kosman, D., and Jentsch, S. (1993). *MAC1*, a nuclear regulatory protein related to Cu-dependent transcription factors is involved in Cu/Fe utilization and stress resistance in yeast. EMBO J. *12*, 5051–5056.

Kekuda, R., Prasad, P.D., Fei, Y.J., Leibach, F.H., Ganapathy, V. (1996). Cloning and functional expression of the human type 1 sigma receptor (hSigmaR1). Biochem. Biophys. Res. Commun. *229*, 553–558.

Keleher, C.A., Redd, M.J., Schultz, J., Carlson, M., and Johnson, A.D. (1992). Ssn6-Tup1 is a general repressor of transcription in yeast. Cell *68*, 709–719.

Link, A.J., Eng, J., Schieltz, D.M., Carmack, E., Mize, G.J., Morris, D.R., Garvik, B.M., and Yates, J.R., III (1999). Direct analysis of protein complexes using mass spectrometry. Nat. Biotechnol. *17*, 676–682.

Lipke, P.N., Taylor, A., and Ballou, C.E. (1976). Morphogenic effects of alpha-factor on *Saccharomyces cerevisiae* a cells. J. Bacteriol. *127*, 610–618.

Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., and Brown, E.L. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat. Biotechnol. *14*, 1675–1680.

Lussier, M., White, A.M., Sheraton, J., di Paolo, T., Treadwell, J., Southard, S.B., Horenstein, C.I., Chen-Weiner, J., Ram, A.F., Kapteyn, J.C., et al. (1997). Large scale identification of genes involved in cell surface biosynthesis and architecture in *Saccharomyces cerevisiae*. Genetics *147*, 435–450.

Marcotte, E.M., Pellegrini, M., Thompson, M.J., Yeates, T.O., and Eisenberg, D. (1999). A combined algorithm for genome-wide prediction of protein function. Nature *402*, 83–86.

Marton, M.J., DeRisi, J.L., Bennett, H.A., Iyer, V.R., Meyer, M.R., Roberts, C.J., Stoughton, R., Burchard, J., Slade, D., Dai, H., et al. (1998). Drug target validation and identification of secondary drug target effects using DNA microarrays. Nat. Med. *4*, 1293–1301.

Mewes, H.W., Albermann, K., Heumann, K., Liebl, S., and Pfeiffer, F. (1997) MIPS: a database for protein sequences, homology data and yeast genome information. Nucleic Acids Res. *25*, 28–30.

Moebius, F.F., Bermoser, K., Reiter, R.J., Hanner, M., and Glossmann, H. (1996). Yeast sterol C8-C7 isomerase: identification and characterization of a high-affinity binding site for enzyme inhibitors. Biochemistry *35*, 16871–16878.

Moebius, F.F., Reiter, R.J., Hanner, M., and Glossmann, H. (1997). High affinity of sigma 1-binding sites for sterol isomerization inhibitors: evidence for a pharmacological relationship with the yeast sterol C8-C7 isomerase. Br. J. Pharmacol. *121*, 1–6.

Molzahn, S.W., and Woods, R.A. (1972). Polyene resistance and the isolation of sterol mutants in *Saccharomyces cerevisiae*. J. Gen. Microbiol. *72*, 339–348.

Nakajima, H., Kim, Y.B., Terano, H., Yoshida, M., and Horinouchi, S. (1998). FR901228, a potent antitumor antibiotic, is a novel histone deacetylase inhibitor. Exp. Cell. Res. *241*, 126–133.

Nelson, H., Mandiyan, S., and Nelson, N. (1995). A bovine cDNA and a yeast gene (*VMA8*) encoding the subunit D of the vacuolar H(+)-ATPase. Proc. Natl. Acad. Sci. USA *92*, 497–501.

Nguyen, V.H., Ingram, S.L., Kassiou, M., and Christie, M.J. (1998). Sigma-binding site ligands inhibit K^+ currents in rat locus coeruleus neurons in vitro. Eur. J. Pharmacol. *361*, 157–163.

Ohya, Y., Umemoto, N., Tanida, I., Ohta, A., Iida, H., and Anraku, Y. (1991). Calcium-sensitive *cls* mutants of *Saccharomyces cerevisiae* showing a Pet^- phenotype are ascribable to defects of vacuolar membrane H(+)-ATPase activity. J. Biol. Chem. *266*, 13971–13977.

Ovalle, R., Lim, S.T., Holder, B., Jue, C.K., Moore, C.W., and Lipke, P.N. (1998). A spheroplast rate assay for determination of cell wall integrity in yeast. Yeast *14*, 1159–1166.

Parks, L.W., Smith, S.J., and Crowley, J.H. (1995). Biochemical and physiological effects of sterol alterations in yeast—a review. Lipids *30*, 227–230.

Pennisi, E. (1998). Worming secrets from the *C. elegans* genome. Science *282*, 1972–1974.

Perou, C.M., Jeffrey, S.S., van de Rijn, M., Rees, C.A., Eisen, M.B., Ross, D.T., Pergamenschikov, A., Williams, C.F., Zhu, S.X., Lee, J.C., et al. (1999). Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. Proc. Natl. Acad. Sci. USA *96*, 9212–9217.

Raguzzi, F., Lesuisse, E., and Crichton, R.R. (1988). Iron storage in *Saccharomyces cerevisiae*. FEBS Lett. *231*, 253–258.

Ram, A.F., Wolters, A., Ten Hoopen, R., and Klis, F.M. (1994). A new approach for isolating cell wall mutants in *Saccharomyces cerevisiae* by screening for hypersensitivity to calcofluor white. Yeast *10*, 1019–1030.

Roberts, C.J., Nelson, B., Marton, M.J., Stoughton, R.S., Meyer, M.R., Bennett, H.A., He, Y.D., Dai, H., Walker, W.L., Hughes, T.R., Tyers, M., Boone, C. and Friend, S.H. (2000). Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. Science *287*, 873–880.

Roncero, C., Valdivieso, M.H., Ribas, J.C., and Duran, A. (1988). Isolation and characterization of *Saccharomyces cerevisiae* mutants resistant to Calcofluor white. J. Bacteriol. *170*, 1950–1954.

Ross-Macdonald, P., Coelho, P.S., Roemer, T., Agarwal, S., Kumar, A., Jansen, R., Cheung, K.H., Sheehan, A., Symoniatis, D., Umansky, L., et al. (1999). Large-scale analysis of the yeast genome by transposon tagging and gene disruption. Nature *402*, 413–418.

Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science *270*, 467–470.

Shoemaker, D.D., Lashkari, D.A., Morris, D., Mittmann, M., and Davis, R.W. (1996). Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. Nat. Genet. *14*, 450–456.

Smith, V., Chou, K.N., Lashkari, D., Botstein, D., and Brown, P.O. (1996). Functional analysis of the genes of yeast chromosome V by genetic footprinting. Science *274*, 2069–2074.

Somerville, C., and Somerville, S. (1999). Plant functional genomics. Science *285*, 380–383.

Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Botstein, D., and Futcher, B. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. Mol. Biol. Cell *9*, 3273–3297.

Spradling, A.C., Stern, D., Beaton, A., Rhem, E.J., Laverty, T., Mozden, N., Misra, S., and Rubin, G.M. (1999). The Berkeley *Drosophila* Genome Project gene disruption project: single P-element insertions mutating 25% of vital *Drosophila* genes. Genetics *153*, 135–177.

Szczypka, M.S., Zhu, Z., Silar, P., and Thiele, D.J. (1997). *Saccharomyces cerevisiae* mutants altered in vacuole function are defective in copper detoxification and iron-responsive gene transcription. Yeast *13*, 1423–1435.

Tsukiyama, T., Palmer, J., Landel, C.C., Shiloach, J., and Wu, C. (1999). Characterization of the imitation switch subfamily of ATP-dependent chromatin-remodeling factors in *Saccharomyces cerevisiae*. Genes Dev. *13*, 686–697.

Veitia, R.A., Ottolenghi, C., Bissery, M.C., and Fellous, A. (1999). A novel human gene, encoding a potential membrane protein conserved from yeast to man, is strongly expressed in testis and cancer cell lines. Cytogenet. Cell Genet. *85*, 217–220.

Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett, D.E., Jr., Hieter, P., Vogelstein, B., and Kinzler, K.W. (1997). Characterization of the yeast transcriptome. Cell *88*, 243–251.

Wilke, R.A., Lupardus, P.J., Grandy, D.K., Rubinstein, M., Low, M.J., Jackson, M.B. (1999). K^+ channel modulation in rodent neurohypophysial nerve terminals by sigma receptors and not by dopamine receptors. J. Physiol. (Lond.) *517*, 391–406.

Williams, F.E., Varanasi, U., and Trumbly, R.J. (1991). The CYC8 and TUP1 proteins involved in glucose repression in *Saccharomyces cerevisiae* are associated in a protein complex. Mol. Cell. Biol. *11*, 3307–3316.

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). Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. Science *285*, 901–906.

Wittes, J., and Friedman, H.P. (1999). Searching for evidence of altered gene expression: a comment on statistical analysis of microarray data. J. Natl. Cancer Inst. *91*, 400–401.

Zhou, Z., and Elledge, S.J. (1992). Isolation of *crt* mutants constitutive for transcription of the DNA damage inducible gene *RNR3* in *Saccharomyces cerevisiae*. Genetics *131*, 851–866.