

Research

# Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae*

Justin C Fay<sup>\*§</sup>, Heather L McCullough<sup>\*</sup>, Paul D Sniegowski<sup>†</sup> and Michael B Eisen<sup>\*‡</sup>

Addresses: <sup>\*</sup>Department of Genome Sciences, Life Sciences Division, Lawrence Berkeley National Laboratory, One Cyclotron Rd, Berkeley, CA 94720, USA. <sup>†</sup>Department of Biology, University of Pennsylvania, 324 Leidy Laboratories, Philadelphia, PA 19104, USA. <sup>‡</sup>Center for Integrative Genomics, Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA. <sup>§</sup>Current address: Department of Genetics, Washington University, 4566 Scott Ave, St. Louis, MO 63110, USA.

Correspondence: Justin C Fay. E-mail: [jfay@genetics.wustl.edu](mailto:jfay@genetics.wustl.edu)

Published: 24 March 2004

*Genome Biology* 2004, 5:R26

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2004/5/4/R26>

Received: 28 January 2004

Revised: 25 February 2004

Accepted: 27 February 2004

© 2004 Fay *et al.*; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

## Abstract

**Background:** The relationship between genetic variation in gene expression and phenotypic variation observable in nature is not well understood. Identifying how many phenotypes are associated with differences in gene expression and how many gene-expression differences are associated with a phenotype is important to understanding the molecular basis and evolution of complex traits.

**Results:** We compared levels of gene expression among nine natural isolates of *Saccharomyces cerevisiae* grown either in the presence or absence of copper sulfate. Of the nine strains, two show a reduced growth rate and two others are rust colored in the presence of copper sulfate. We identified 633 genes that show significant differences in expression among strains. Of these genes, 20 were correlated with resistance to copper sulfate and 24 were correlated with rust coloration. The function of these genes in combination with their expression pattern suggests the presence of both correlative and causative expression differences. But the majority of differentially expressed genes were not correlated with either phenotype and showed the same expression pattern both in the presence and absence of copper sulfate. To determine whether these expression differences may contribute to phenotypic variation under other environmental conditions, we examined one phenotype, freeze tolerance, predicted by the differential expression of the aquaporin gene *AQY2*. We found freeze tolerance is associated with the expression of *AQY2*.

**Conclusions:** Gene expression differences provide substantial insight into the molecular basis of naturally occurring traits and can be used to predict environment dependent phenotypic variation.

## Background

An important question concerning the genetic basis and evolution of complex traits is the relative contribution of gene regulation versus protein structure. If gene-expression differences

make a substantial contribution to phenotypic variation found in nature, the genetic basis of complex traits may be more readily understood through the analysis of gene expression [1]. Furthermore, it would imply that most evolutionary

changes occur through changes in either patterns or levels of gene expression [2,3].

Genome expression studies have shown numerous differences in transcript abundance both within and between closely related species [4-12]. In some instances, genetic variation in gene expression has been associated with phenotypic variation [1,5,10,13-16]. However, gene expression differences correlated with a phenotype may or may not contribute to the phenotype. Distinguishing between these possibilities requires locating the genes responsible for the trait [1,14-16].

To further investigate the relationship between genetic variation in gene expression and phenotypic variation, we measured genome-wide mRNA transcript levels in nine strains of *Saccharomyces cerevisiae* which vary in their sensitivity to copper sulfate ( $\text{CuSO}_4$ ), a strong oxidizing agent often used as an antimicrobial agent in vineyards [17,18].

## Results

### Natural isolates of *Saccharomyces cerevisiae* vary in their sensitivity to copper sulfate

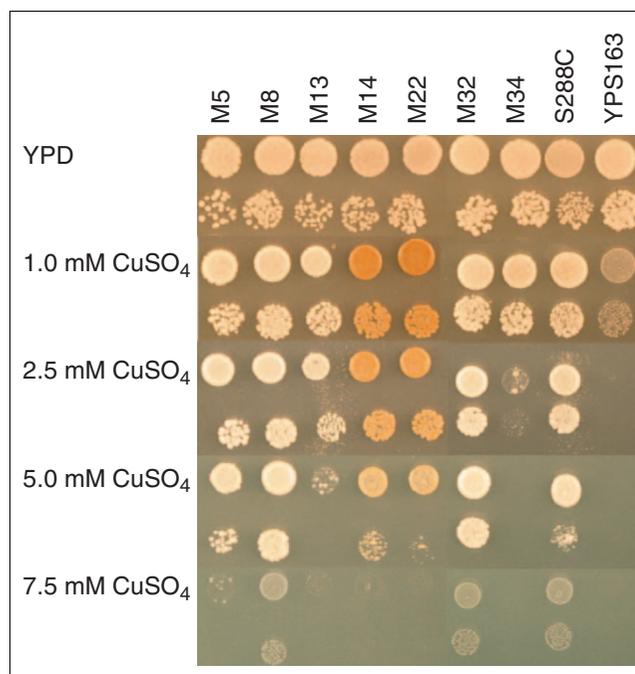
Copper is an oxidizing agent necessary for many single-electron transfer reactions within the cell and is toxic at high concentrations [19]. Natural isolates of *S. cerevisiae* have been reported to vary in their sensitivity to copper sulfate [17,20,21], and resistance to copper sulfate may be a recently acquired adaptation as a result of the application of copper sulfate as a fungicide to treat powdery mildew in vineyards [17,18]. Seven isolates from vineyards in Italy, the sequenced laboratory strain S288C and an isolate from an oak tree in Pennsylvania vary in their sensitivity to copper sulfate (Table 1, Figure 1). Two of the strains produce red/brown or rust-colored colonies in the presence of copper sulfate.

**Table 1**

#### Strains used in this study

Strain*	Location	Year	Reference
M5	Italy	1993/94	[17]
M8	Italy	1993/94	[17]
M13	Italy	1993/94	[17]
M14	Italy	1993/94	[17]
M22	Italy	1993/94	[17]
M32	Italy	1993/94	[17]
M34	Italy	1993/94	[17]
YPS163	PA, USA	1999	[55]
S288C	CA, USA	1938	
YPS125	PA, USA	1999	[55]

\*All strains are diploid and homothallic except S288C, which is *MATa/a*, *GAL2/GAL2*, *Dura3 EcoRV-Stu1/lura3-52 ho*.

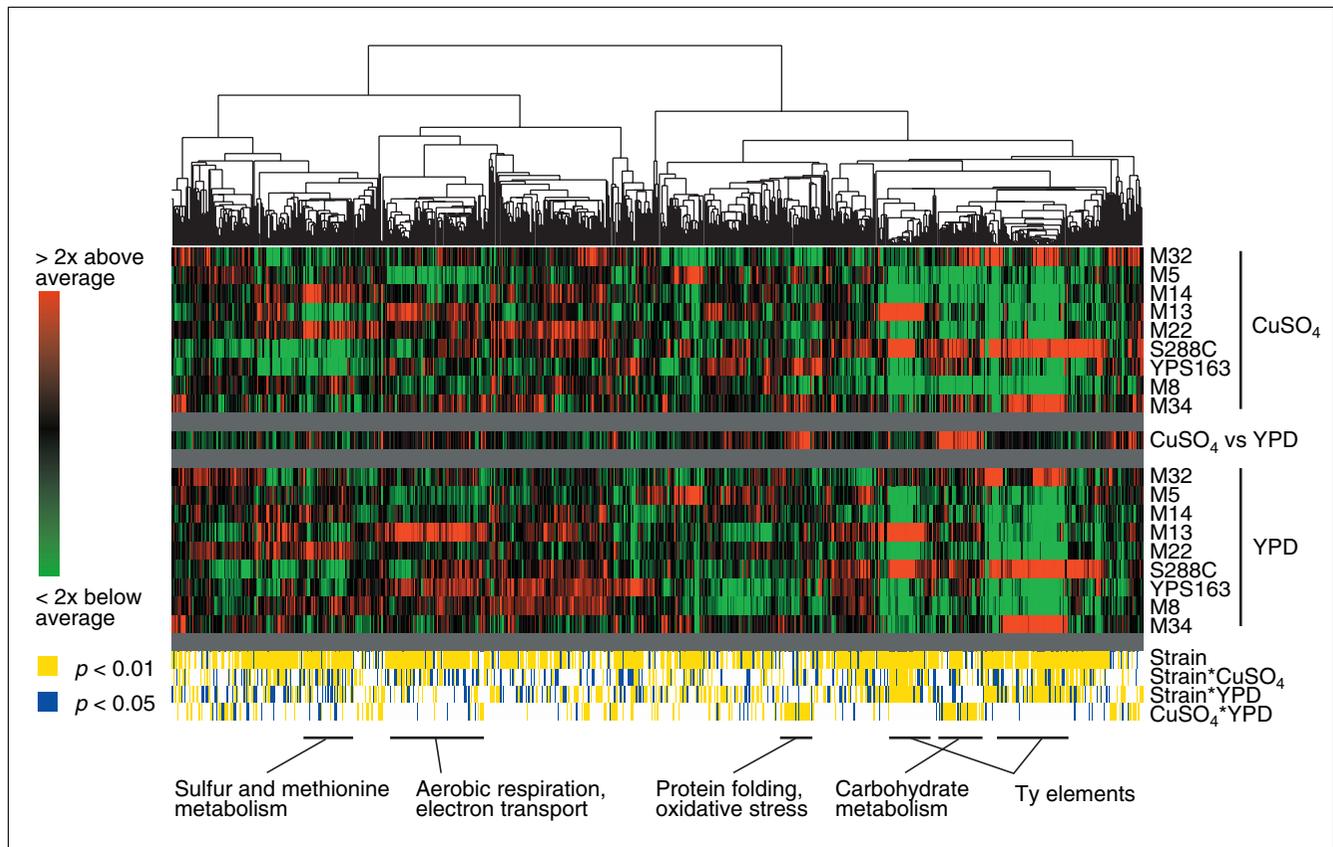


**Figure 1**

Growth of strains on rich medium (YPD) and rich medium supplemented with different concentrations of copper sulfate ( $\text{CuSO}_4$ ). For each condition, a  $10^{-3}$  and a  $10^{-4}$  dilution of cells from an overnight YPD culture are shown.

### Identification of gene expression differences in the presence and absence of copper sulfate

Expression levels were measured using DNA microarrays in the nine strains during exponential growth in rich medium and in rich medium supplemented with copper sulfate (see Materials and methods). The microarrays used in this study are composed of oligonucleotides of 70 base pairs (bp) that are perfect matches to the S288C sequence. Although cDNA prepared from the other eight strains will not always be a perfect match to the sequence on the microarray, we expect fewer than 0.2 differences per 70 bp on average (see Materials and methods), and therefore do not expect the sequence differences to affect our measurements. A reference design was used whereby the RNA of each strain grown in rich medium and rich medium supplemented with copper sulfate was compared to the pooled RNA from all nine strains grown in rich medium and copper sulfate medium, respectively. Using three replicate experiments, four statistical tests were used to identify differentially expressed genes. From an analysis of variance, 194 genes showed significant expression differences among strains grown in copper sulfate medium, 241 genes showed significant expression differences among strains grown in rich medium, and 516 genes showed significant expression differences across both conditions ( $p < 0.01$ ). One hundred and thirty-one genes showed significant differences between the rich medium and copper sulfate medium reference pools ( $t$ -test,  $p < 0.01$ ). Because an analysis of variance



**Figure 2**  
 Hierarchical clustering of differentially expressed genes. Genes with significant expression differences among strains in both media (strain), in copper-sulfate medium (strain\*CuSO<sub>4</sub>), in rich medium (strain\*YPD), and between copper sulfate and rich medium reference pools (YPD vs CuSO<sub>4</sub>) for  $p < 0.05$  (yellow) and  $p < 0.01$  (blue). Groups of functionally related genes are also shown.

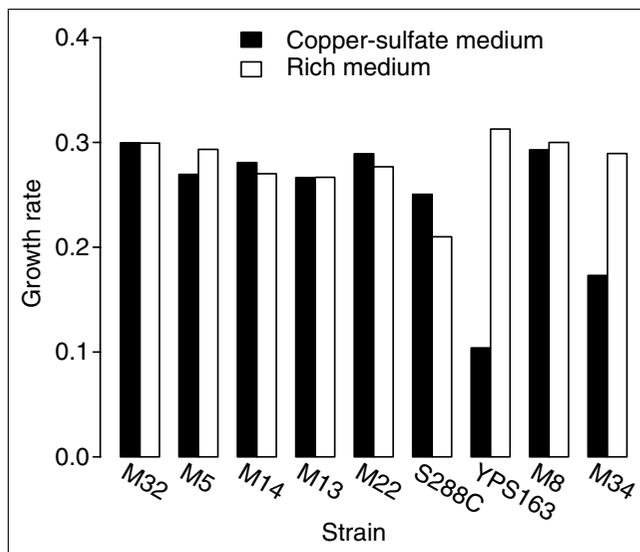
assumes errors are independent and identically distributed, we estimated the rate of false positives using a nonparametric permutation resampling method (see Materials and methods). The estimated number of false positives was 57, 64, 55 and 71, for the test of gene-expression differences among strains in copper sulfate medium, in rich medium, in both media, and between the two reference pools, respectively. We chose a  $p$ -value cutoff of 0.01, as empirically, many significant genes are missed using a  $p$ -value cutoff of 0.001 and numerous false positives are generated using a  $p$ -value cutoff of 0.05 (see Materials and methods).

A total of 731 genes showed significant expression differences by one or more of the four tests. These genes were hierarchically clustered on the basis of the centered correlation coefficient and are presented with their  $p$ -values in Figure 2. Most genes show similar expression patterns in rich medium and copper sulfate medium. Of the 633 genes that were found to be differentially expressed among strains in either one or both treatments, 79 genes and 36 genes were only significant in rich medium and copper sulfate medium, respectively. Manual inspection of these genes revealed that many of the expression patterns significant in one medium showed a

similar, although nonsignificant, expression pattern in the other medium. Through a separate analysis of variance, we found 56 genes specifically differ in their pattern of expression in rich medium compared to copper sulfate medium (see Materials and methods).

**Differentially expressed genes correlated with growth rate in the presence of copper sulfate function in response to oxidative stress**

To identify gene-expression differences correlated with resistance to copper sulfate, we measured the correlation between the differentially expressed genes and sensitivity to copper sulfate. In liquid medium M34 and YPS163 were sensitive to copper sulfate (ANOVA,  $p = 0.00022$ ), whereas no significant differences were measured in rich medium alone (ANOVA,  $p = 0.159$ ; see Materials and methods and Figure 3). Genes correlated with sensitivity to copper sulfate are presented in Figure 4a (see Materials and methods). We used a correlation cutoff of 0.80, which corresponds to a significance of  $p < 0.01$ . Permutation resampling of the expression differences showed that only 13 expression differences are expected to reach a correlation of 0.80 or above (see Materials and methods). Of those genes correlated with sensitivity to copper



**Figure 3**

The average growth rates from three replicates of strains in rich medium and rich medium with 1 mM copper sulfate. Relative growth rates were measured by the slope of the linear regression of cell density on time.

sulfate, eight are expressed at a higher level in the presence of copper sulfate while fewer than one ( $20 \times 131/6,144$ ) is expected (exact test,  $p < 10^{-7}$ ). Thus, there are more genes that are correlated with sensitivity to copper sulfate and that change in response to copper sulfate than expected by chance.

Genes expressed at higher levels in copper-sensitive (M34 and YPS163) compared to resistant strains are known to function in response to oxidative stress. At high concentrations, copper causes oxidative stress resulting in lipid peroxidation, aggregation and fragmentation of proteins and DNA damage [22]. Thioredoxin peroxidase (*TSA1*) and thioredoxin (*TRX2*) function in redox homeostasis and are regulated by the transcription factors Yap1p and Skn7p [23,24]. The heat-shock proteins encoded by *SSA1* and *HSP82* are also regulated by Yap1p and Skn7p and function in protein folding and translocation of misfolded proteins [25]. Sti1p is a member of the Hsp82 protein complex [26]. Kar2p interacts with Ire1p [27] to activate the unfolded protein response, including protein disulfide isomerase, PDI1 [28], which is required for oxidative protein folding in the endoplasmic reticulum [29]. These genes, in addition to functioning in oxidative stress and protein folding, had higher levels of expression in the copper sulfate compared to rich medium reference pool (Figure 4a).

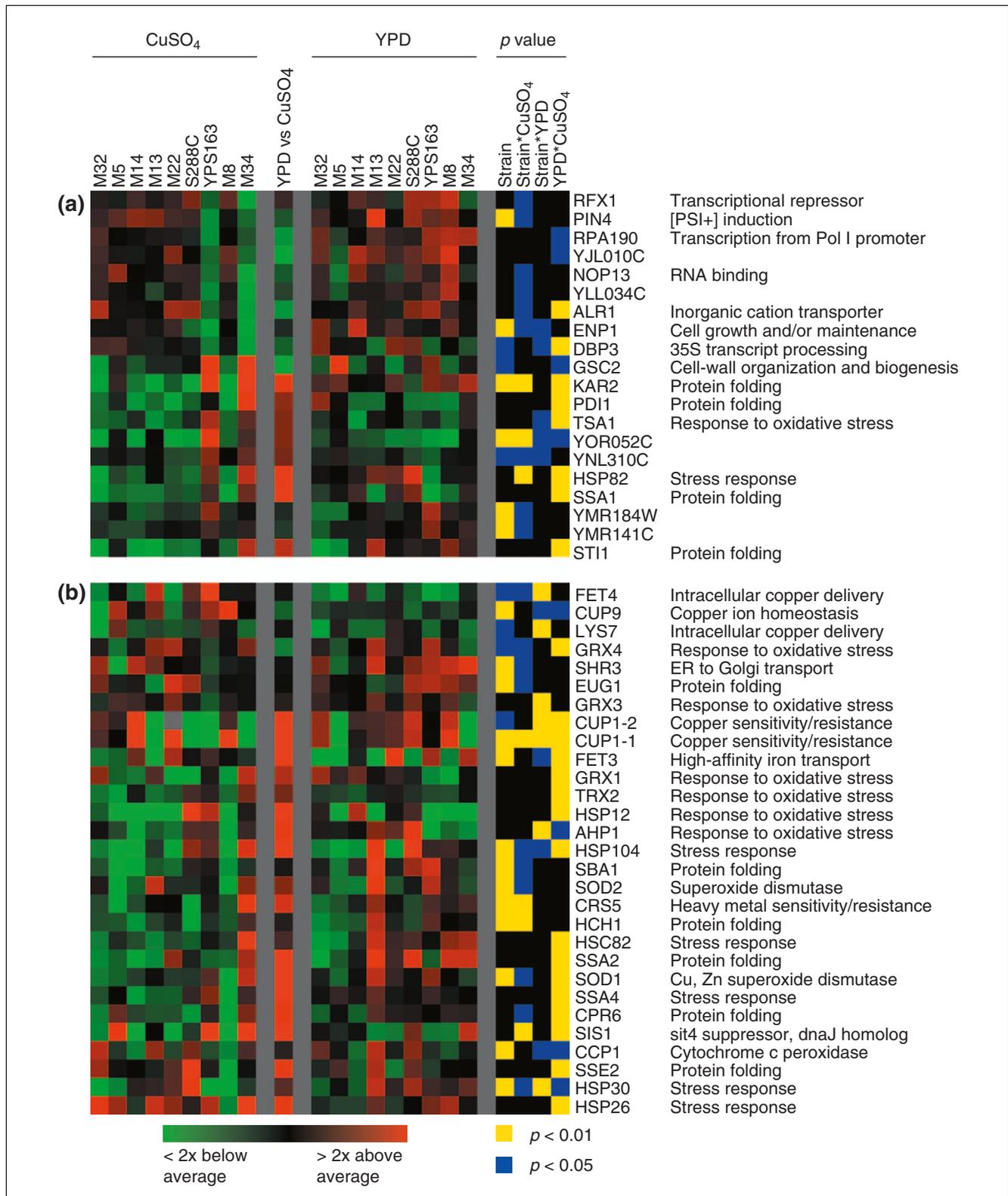
Genes expressed at lower levels in strains sensitive to copper sulfate were expressed at lower levels in the copper sulfate compared to the rich medium reference pool and function in RNA processing. *RFX1* encodes a repressor of RNA polymerase II (Pol II) promoters [30]. *ENP1* encodes a small nucleolar RNA-binding protein involved in rRNA processing [31]. In addition, both YJL010C and YLLO34C show changes in gene

expression similar to other RNA-processing genes [32], which together form a major component of the environmental stress response [33]. The expression of RNA-processing genes may be related to a general stress response and/or the reduced growth rate of copper-sulfate-sensitive strains.

Expression differences weakly correlated with resistance to copper sulfate may also be relevant to understanding the molecular basis of the trait, especially if it is complex. To identify relevant expression differences weakly correlated with resistance to copper sulfate we examined genes annotated as functioning in copper homeostasis, protein folding or oxidative stress (Figure 4b), as well as all genes expressed at higher or lower levels as a result of the presence of copper sulfate (Figure 5). Some genes show a weak correlation with resistance to copper sulfate. For instance, the superoxide dismutase gene *SOD2* was found expressed at higher levels in the copper sulfate reference pool, and at higher levels in M13 and M34, two of the three most copper-sensitive strains (Figure 4b). Also, the copper, zinc superoxide dismutase *SOD1* was found expressed at intermediate levels in M13 and at higher levels in YPS163 and M34 (Figure 4b), in correspondence with the strains' sensitivity to copper sulfate (Figure 1). Superoxide dismutases protect cells against reactive oxygen species and are induced in response to oxidative stress [22].

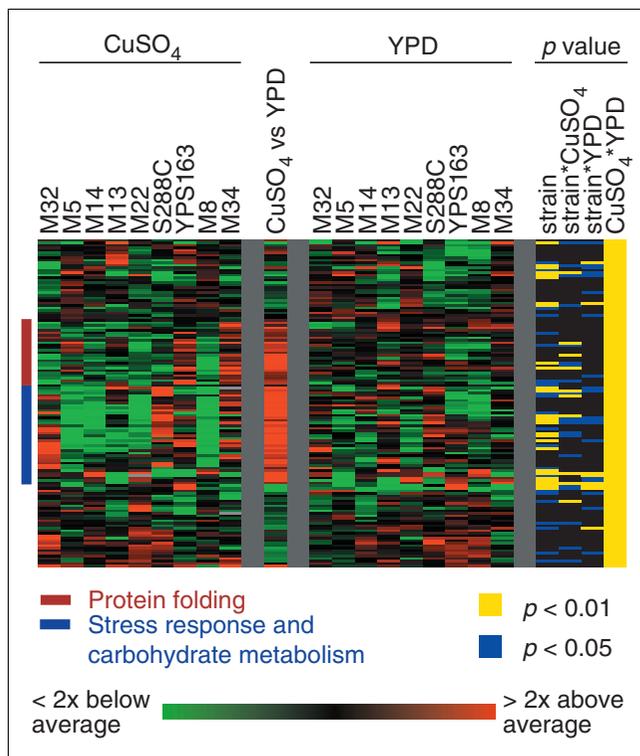
Of those genes found to change in response to copper sulfate (Figure 5), the genes expressed at lower levels in the presence of copper sulfate are not functionally related, and the genes expressed at higher levels in the presence of copper sulfate are significantly enriched in genes known to function in protein folding, stress response and metabolism (see Materials and methods). Of the 131 genes, 24 were expressed at twofold or higher levels in the presence of copper sulfate and one, *ZRT1*, encoding a high-affinity zinc transporter, was expressed at half the level in the presence of copper sulfate. Of these 24 genes, seven are known to function in the stress response (*ALD3*, *DDR2*, *HSP12*, *HSP104*, *TSL1*, *YGP1*, *YRO2*), four in protein folding (*SSA1*, *SSA2*, *SSA4*, *SIS1*), four in metabolism (*ALD4*, *GLK1*, *HXX1*, *PGM2*), five in copper homeostasis (*CUP1-1*, *CUP1-2*, *FET3*, *FTR1*, *SOD1*), two are uncharacterized (YHR087W, YMR315W), one encodes a lipid-binding protein (*TFS1*), and one gene is involved in meiotic sister-chromatid recombination (*MSC1*).

Of those genes expressed at higher levels in the presence of copper sulfate, many are also expressed at higher levels in YPS163 and M34 (Figure 5). However, the response differs among the copper-sulfate-resistant strains. The expression pattern in the copper-resistant strains delineates two major clusters enriched for genes known to function in protein folding (Figure 5, red bars) and stress response and metabolism (Figure 5, blue bars). The group enriched for genes functioning in protein folding tends to be expressed at higher levels in YPS163, M34 and, to some extent, M5. Whereas M5 is resistant to copper in rich medium, it is quite sensitive in SD



**Figure 4**

Genes associated with resistance to copper sulfate. **(a)** Genes correlated with sensitivity to copper sulfate ( $r > 0.8$ ,  $p < 0.01$ ) that are differentially expressed among strains in the presence of copper sulfate or between the rich medium and copper sulfate reference pools. **(b)** Genes differentially expressed and annotated as functioning in copper homeostasis, protein folding or response to oxidative stress.

**Figure 5**

Genes with different expression levels in the copper sulfate compared to the rich medium reference pool. Groups of genes enriched for functions in protein folding (red bar) and stress response and metabolism (blue bar) are shown.

or SC medium (see Additional data file 1). One of the genes expressed at higher levels in M5, YPS163 and M34 is *SIS1*, encoding an HSP40 family chaperone required for the initiation of translation [34], and known to regulate the protein-folding activity of the heat-shock protein Ssa1p [35]. The group enriched for genes functioning in the stress response and carbohydrate metabolism tends to be expressed at higher levels in the two copper-sensitive strains, YPS163 and M34, but also tends to be expressed in S288C and M32, two of the three most resistant strains.

#### Differentially expressed genes correlated with rust coloration function in the sulfur assimilation/methionine pathway

To identify those genes associated with the rust color phenotype, the expression of genes in copper sulfate was correlated with rust coloration in the presence of copper sulfate (Figure 6). Twenty-four genes differentially expressed in the presence of copper sulfate were found tightly correlated with rust coloration ( $r > 0.8$ ,  $p < 0.01$ ). Only 13 genes are expected by changes, as determined by permutation resampling. Genes with higher levels of expression in M14 and M22 often had the same pattern in both the presence and absence of copper sulfate (Figure 6). Of the 24 genes, 10 (*MET1*, *MET3*, *MET10*, *ECM17*, *MET17*, *MET22*, *SAM1*, *SAM2*, *SAM3*, *SAH1*) are

known to function in the sulfur assimilation/methionine metabolism pathway. Many of these genes are known to be regulated by the transcription factor complexes Cbf1p/Met4p/Met28p [36] and Met31p/Met32p [37]. The 14 other genes are not obviously related to each other or to the rust coloration phenotype.

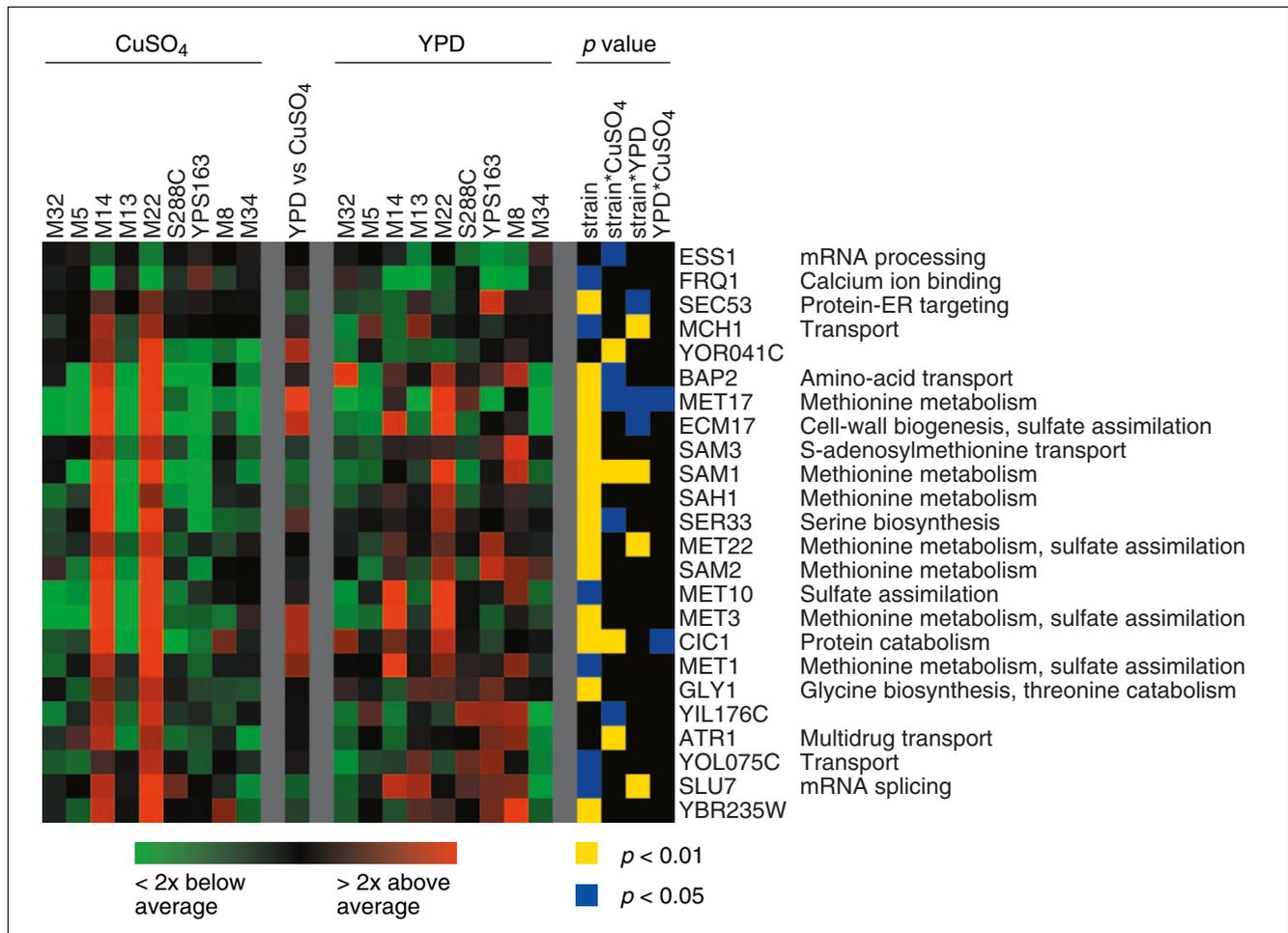
#### A candidate phenotype, freeze tolerance, is associated with the differential expression of the aquaporin gene AQY2

Gene-expression differences not associated with either copper sulfate phenotype may have fitness effects under other environmental conditions. The expression level of the aquaporin gene *AQY2* has been shown to affect freeze tolerance [38]. YPS163 shows a 2.6- and 5.3-fold greater level of expression of *AQY2* compared to the other strains in copper sulfate and rich media, respectively. We hypothesized that YPS163 may show more freeze tolerance as a result of this expression difference. As predicted, the growth of YPS163 is not significantly different following a  $-30^{\circ}\text{C}$  compared to a  $4^{\circ}\text{C}$  treatment, whereas all the other strains showed a significantly reduced growth rate ( $p < 10^{-8}$ , paired *t*-test) following a  $-30^{\circ}\text{C}$  compared to a  $4^{\circ}\text{C}$  treatment (Figure 7).

#### Genes that respond to the presence of copper sulfate show no correlation with sequence divergence between strains

Most expression differences are not associated with either resistance to copper sulfate or rust coloration in the presence of copper sulfate. The differential expression of these genes could be due to a lack of selective constraint on their expression levels or could be due to some form of natural selection. For instance, they may be present due to a balance between mutation and purifying selection or diversifying selection due to environmental heterogeneity. One common method of testing whether a phenotype has been driven by natural selection is to test whether phenotypic differences among species conflict with their known phylogenetic relationship [39-42]. We sequenced three genes to determine the phylogenetic relationship among the strains used in this study (Figure 8). While the three genes show similar levels of divergence among strains, their phylogeny cannot be resolved, as expected for a species with sexual recombination. However, even if multiple genealogies exist across the genome, expression differences are expected to accumulate monotonically as a function of time and mutation rate under an infinite allele model for both single-gene and polygenic characters [43,44]. Thus, we expect neutral differences in gene expression to be correlated with divergence time between strains.

The number of pairwise gene-expression differences found between strains is significantly correlated with the estimated time to coalescence, measured by the number of pairwise sequence differences found in three genes (see Materials and methods and Figure 9a). Because pairwise measures of



**Figure 6**  
Genes correlated ( $r > 0.8$ ,  $p < 0.01$ ) with rust coloration and differentially expressed among strains in the presence of copper sulfate.

divergence are not independent of one another, the correlation may be spurious. A Mantel test is a nonparametric test of association between two dissimilarity matrices that accounts for this nonindependence [45]. Using this test, a significant association was found between divergence in gene expression and DNA sequence divergence ( $p = 0.043$ ). If the expression of genes that respond to the presence of copper sulfate were driven by adaptive evolution, the correlation between divergence in gene expression and DNA sequence divergence may be weaker or even not present. In contrast to overall patterns of gene expression, the expression of genes that respond to the presence of copper sulfate (Figure 6) was not found associated with DNA sequence differences among strains (Figure 9b).

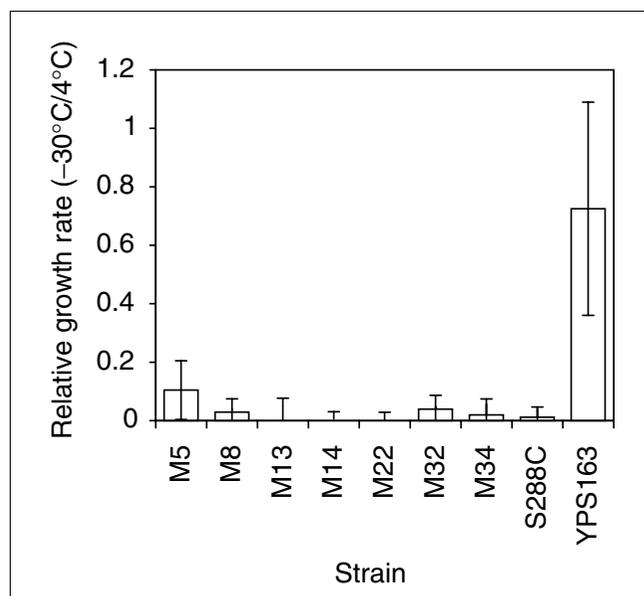
### Discussion

We have examined the association between gene-expression differences and two copper-sulfate-related phenotypes. Whereas the function of these genes implies that they are not

casually associated with the trait, the gene-expression differences may be a response to the phenotype (correlative) or may cause the phenotype (causative). Distinguishing between these possibilities is important to understanding the molecular basis and evolution of complex traits and why transcriptional variation is present in natural populations.

### Resistance to copper sulfate

Resistance to high levels of copper ions is mediated through the copper-binding transcription factor *ACE1*, which induces the metallothionein gene *CUP1* [46], the metallothionein-like gene *CRS5* [47] and the copper, zinc superoxide dismutase gene, *SOD1* [48]. A global analysis of gene expression in response to copper sulfate using DNA microarrays identified *FET3* and *FTR1*, encoding two high-affinity iron transporters and *FIT2*, encoding another iron transporter, as being induced in the presence of copper along with the previously characterized induction of *CUP1*, *SOD1* and *CRS5* [49]. Consistent with these studies, we found that *CUP1*, *SOD1*, *FET3* and *FTR1* were expressed at higher levels in the



**Figure 7**  
Relative rates of growth at 30°C subsequent to a -30°C compared to a 4°C treatment. Growth rates were measured as the change in OD<sub>600</sub> over 4 h. Error bars are one standard deviation.

presence of 1 mM copper sulfate medium compared to rich medium (Figures 4, 5). In addition to these four genes, we found another 127 genes expressed at significantly different levels in the presence of copper sulfate, 20 of which showed a twofold or greater level of expression in the presence of copper sulfate and one, *ZRT1*, encoding a high-affinity zinc transporter, which showed a 50% lower expression level in the presence of copper sulfate (Figure 5). Our study differed from previous studies because we measured expression 180 minutes subsequent to copper treatment in rich medium for three replicate experiments, whereas the other studies measured gene expression 30 minutes subsequent to copper treatment in synthetic complete medium.

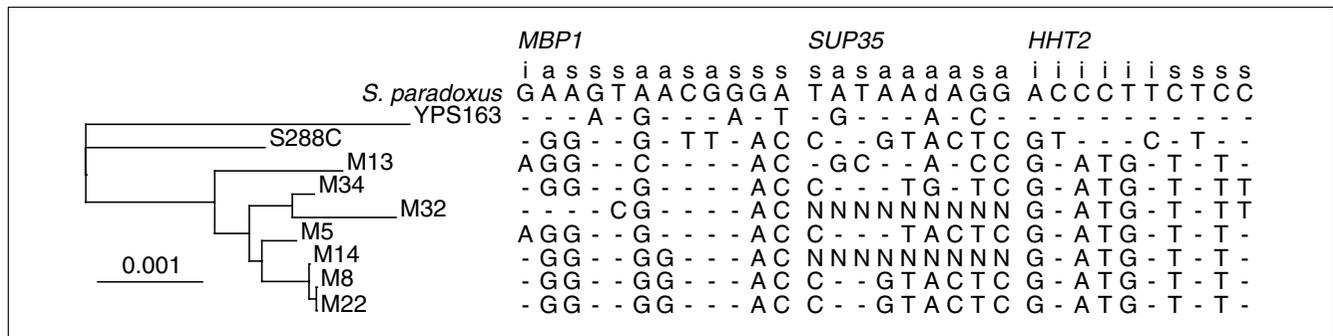
Different levels of copper resistance among strains of *S. cerevisiae* have been attributed to variation in the number of tandem copies of the *CUP1* locus [19,20] and could be due to use of copper sulfate in vineyards as a fungicide against powdery mildew since the 1880s [18]. We have found an incomplete association between *CUP1* expression and resistance to copper sulfate. In the presence of copper sulfate, *CUP1* was expressed at higher levels in strains M14, M22 and M8. These strains are resistant to 5 mM copper sulfate (Figure 1), but so are M5, M32 and S288C. *CUP1* was expressed at the lowest levels in M13, S288C, YPS163 and M34, and while M13, YPS163 and M34 are the most copper-sensitive strains (Figure 1), S288C is one of the most resistant. Because previous studies examined resistance to copper sulfate on synthetic complete (SC) medium, we examined growth on SC

medium with 0.1 mM copper sulfate. Only M8, M13, M32 and M34 grew on synthetic minimal (SD) medium or SC medium supplemented with 0.1 mM copper sulfate (see Additional data file 1). S288C did not grow on either SD or SC medium in the absence of copper sulfate, and M14 and M22 grew weakly in its absence. Thus, YPS163 and M5 are the most sensitive to copper sulfate in SD or SC medium, in contrast to rich medium. Genetic studies will be needed to determine whether resistance to copper sulfate is mediated by loci other than the *CUP1* locus and whether the different transcriptional responses among strains contribute to resistance in the presence of copper sulfate in rich medium or in other growth or environmental conditions.

Genes tightly correlated with sensitivity to copper sulfate (Figure 4a) are likely to be correlated characters and do not contribute to levels of resistance. The oxidative stress response involves numerous genes, many of which were found differentially expressed between strains (Figure 4). However, if genes that respond to oxidative stress were protecting resistant but not sensitive strains, we would expect them to be expressed at higher levels in the resistant rather than the sensitive strains. The opposite is observed. Thus, it appears that many of the genes tightly associated with sensitivity to copper sulfate are likely to be differentially expressed as part of a coordinated response to a toxic cellular environment. Ultimately, the genetic basis of resistance to copper sulfate must be mapped to identify any expression differences that contribute to resistance.

#### Rust coloration

Previous studies of other rust-colored strains using electron microscopy [50] and treatment with potassium cyanide [51] have suggested that the rust color produced in the presence of copper sulfate is due to the formation of copper sulfide (CuS) mineral lattices on cell surfaces. The two rust-colored strains, M14 and M22, often produced a distinct smell of hydrogen sulfide (H<sub>2</sub>S) during fermentation in both the presence and absence of copper sulfate. Hydrogen sulfide production in M14 and M22 may be attributed to the conversion of hydrogen sulfite to hydrogen sulfide by sulfite reductase, Met10p/Ecm17p [52], proteins that are expressed at higher levels in both M14 and M22. The rust coloration may be due to the formation of copper sulfide as a consequence of hydrogen sulfide production. Hydrogen sulfide is often produced during wine fermentation [53], and, because of the resulting undesirable flavors, may be a trait that has been selected against in yeast strains used for wine production. In addition, copper sulfate is often used to remove unwanted sulfides, including hydrogen sulfide, produced during wine production. Segregants from a heterozygous Italian strain were found to co-segregate differential expression of the sulfur-assimilation/methionine metabolism pathway with a filigreed colony morphology produced during starvation [21]. However, neither M14 nor M22 showed the filigreed phenotype at any time during starvation.



**Figure 8**  
DNA sequence differences found in three genes (*SUP35*, *MBP1*, *HHT2*). Intergenic (i), amino-acid-altering (a), and synonymous (s) polymorphic sites are shown in reference to the *S. paradoxus* sequence. d indicates an insertion or deletion and N indicates missing data.

The differential expression of the sulfur-assimilation pathway may be responsible for the rust coloration phenotype as the differential expression of the pathway is not due to the presence of copper sulfate. The production of hydrogen sulfide, the differential expression of sulfur-assimilation genes in the absence of copper sulfate and the absence of a response by the sulfur-assimilation genes to the presence of copper sulfate (Figure 6), suggest that the expression of the sulfur-assimilation pathway is not due to the presence of copper sulfate.

**Gene-by-environment interactions**

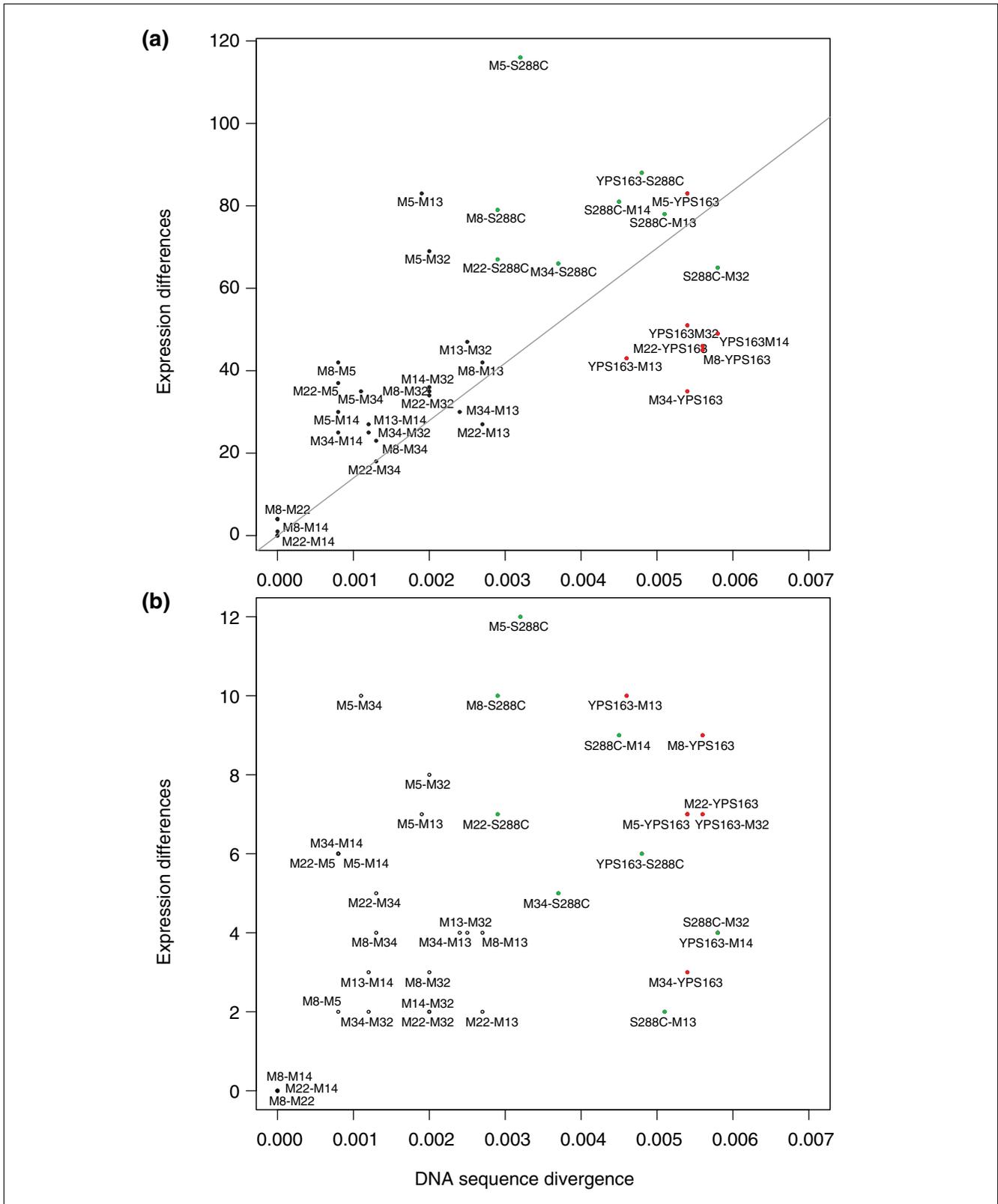
The lack of any obvious phenotype associated with the genes differentially expressed in rich medium suggests that many expression differences may only be associated with phenotypic variation under certain environmental conditions, or may not be associated with any phenotype at all. Because most expression differences persist in the presence and absence of copper sulfate, they may persist under different environmental conditions and may be associated with phenotypic variation under those conditions. This is the case for the sulfur-assimilation/methionine pathway, which is associated with rust coloration only in the presence of copper sulfate. This is also the case for the expression of the aquaporin gene, *AQY2*, which was used to predict phenotype variation among strains subsequent to a freeze-thaw cycle. Our ability to predict phenotype from expression data is not unique. The expression of arsenic-resistance genes was used to correctly predict sensitivity to arsenic among four natural isolates of *S. cerevisiae* [10]. Gene-expression patterns from tumors have been found to predict clinical outcome, for example [54]. Thus, the molecular phenotypes revealed by gene-expression patterns may provide valuable insights into the molecular genetic basis of complex traits, especially those that are environment dependent.

**Rate of divergence in gene expression**

Most expression differences were not associated with either resistance to copper sulfate or rust coloration in the presence of copper sulfate. The differential expression of these genes could be due to a lack of selective constraint on their expres-

sion levels or could be due to some form of natural selection. For instance, they may be the result of a balance between mutation and purifying selection or could be a result of diversifying selection mediated by environmental heterogeneity. We found a significant correlation between divergence in gene expression and DNA sequence divergence for overall patterns of gene expression but not for those that respond to the presence of copper sulfate. While this implies that different explanations are needed for the two groups of genes, it is difficult to ascribe neutral or selective explanations with high levels of confidence. First, gene-expression differences are also expected to accumulate with divergence time if selection is uniform in its pressure across all strains. Second, many factors can influence the variance in the number of expression differences between two strains, so the significance of the association between divergence in gene expression with DNA sequence divergence is difficult to interpret. Regardless, the relationship between rates of protein divergence and divergence in gene expression are useful to understanding biological diversity at the molecular level.

The average rate of change in gene expression was estimated to be 5,448 expression changes across the genome per synonymous substitution per site, or 0.887 (5,448/6,144) expression changes in each gene per synonymous substitution per site (see Materials and methods). The average number of synonymous substitutions per site, amino-acid-altering substitutions per site, and intergenic substitutions per site between strains in the three sequenced regions, was estimated as  $6.87 \times 10^{-3}$ ,  $1.20 \times 10^{-3}$ , and  $2.00 \times 10^{-3}$ , respectively. Therefore, the rate of change in gene expression per synonymous substitution is higher than the rate of amino-acid substitution per synonymous substitution (0.175) or the rate of intergenic substitution per synonymous substitution (0.291). If intergenic sites were neutral, the expected rate of intergenic substitution per synonymous substitution is 1. The ratio of rates of intergenic to synonymous substitution suggests that purifying selection constrains about 70% of intergenic sites found 5' of the *HHT2*, *MBP1* and *SUP35* genes. Because we do not know the effective number of sites in the



**Figure 9** Pairwise differences in gene expression compared to pairwise DNA sequence divergence. **(a)** Genes differentially expressed among strains, and **(b)** genes different between copper-sulfate and rich medium. Distances with S288C (green) and with YPS163 (red) are distinguished.

genome which when mutated alter gene-expression levels in copper sulfate and rich medium, we cannot determine how many differentially expressed genes would be expected in the absence of any selective constraints on changes in gene expression. The mutation variance for gene expression is needed to estimate the amount of selective constraint on gene expression levels.

## Materials and methods

### Strains

Yeast strains were selected from a larger collection of strains surveyed for variation in sensitivity to copper sulfate and those used are listed in Table 1. Of the nine strains, seven were isolated from vineyards in Italy between 1993 and 1994 by R. Mortimer [17]. The diploid, sequenced lab strain, S288C, was obtained from the Botstein lab (DBY8268). The lab strain S288C is mostly derived from EM93, which was isolated from a rotting fig in California in 1938 [55]. The woodland strain, YPS163, and the *S. paradoxus* strain, YPS125, were isolated from oak tree exudates in Lima, Pennsylvania in 1999 [56]. The strains were chosen from a screen of around 100 natural isolates for variation in resistance to copper sulfate.

### Resistance to copper sulfate

Strains were grown in 2 ml overnight rich medium cultures (YPD: 1% yeast extract, 2% peptone, 2% dextrose), diluted by a factor of  $10^3$  and  $10^4$  and plated onto the following media: rich medium (YPD + 2% agar) and rich medium plates supplemented with 1.0, 2.5, 5.0 and 7.5 mM copper sulfate; minimal medium (SD: 0.67% yeast nitrogen base with ammonium sulfate, 2% dextrose, 2% agar); SD supplemented with 0.1 mM copper sulfate; synthetic complete medium (CM: 0.67% yeast nitrogen base with amino acids and ammonium sulfate, 2% dextrose, 2% agar); and CM supplemented with 0.1 mM copper sulfate.

### Expression data

Three complete replicate experiments were done on different days. Each replicate started from a 2-ml overnight rich medium culture (YPD). Strains were grown at 30°C in either rich medium or copper sulfate medium (YPD supplemented with 1 mM  $\text{CuSO}_4$ ) to an  $\text{OD}_{600}$  of 1 (optical density of one at 600 nm is approximately  $1 \times 10^7$  cells/ml), at which point they were diluted to an  $\text{OD}_{600}$  of 0.1 in 10 ml of either rich medium or copper sulfate medium. When the strains had again reached an  $\text{OD}_{600}$  of 0.8-1.0 (about 3 h later) they were spun for 3 min at 1,500g, lysed in 0.5 ml lysis buffer (10 mM Tris-Cl pH 7.4, 10 mM EDTA, 0.5% SDS) and frozen in liquid nitrogen. RNA was extracted using hot phenol and chloroform. At 3 h the strains were in exponential growth and any residual expression differences from the previous culture were not likely to be present. Total RNA was reverse transcribed using aminoallyl-dUTP then coupled to either a Cy3 or Cy5 fluorescent dye (Amersham Pharmacia) and hybridized overnight to microarrays on which 6,144, 70-bp

oligonucleotides (Qiagen Operon) had been spotted as described at [57].

A reference design was used whereby the RNA from each strain grown in rich or copper sulfate medium was compared to a pool of the RNA from all strains grown in rich or copper sulfate medium, respectively. The reference pool was constructed using equal samples of RNA from each strain. While a loop design provides more statistical power from a given number of microarrays [58], in a loop design, a biased slide may bias estimates of all other treatment effects in the loop, while in a reference design, a biased slide only biases the estimate of a single treatment effect. We chose to use a separate rich medium and copper sulfate reference pool so as to maximize our ability to detect strain differences. For example, an expression difference between rich medium and copper sulfate of 1 to 1,000 units in strain A and 2 to 1,000 units in strain B may not be distinguishable unless strain A and B are compared in rich and copper sulfate medium separately. To identify genes expressed at different levels in rich medium compared to copper sulfate, the two reference pools were directly compared.

Arrays were scanned using a GenePix 4000A scanner and GenePix 4.0 software (Axon). An average of 762 spots per slide were manually flagged as unusable. The raw expression data are available in the GEO database under the ID GSE1073. Each array was print-tip normalized (mean normalized as a function of spot intensity) using the SMA package of the R statistics software with a span parameter of 0.7 [59]. A span parameter of 1.0 is equal to no intensity-dependent normalization and a span parameter of 0.7 normalizes by intensity as a function of 70% of the data. Three replicate experiments were carried out, resulting in 54 arrays (9 strains, 3 replicates, 2 conditions) to measure differences among strains and six arrays to measure differences between the two reference pools. For one of the replicates a dye-swap was performed, where Cy3 instead of Cy5 was used to label the reference sample. Of the six comparisons between reference pools, two were dye-swaps. Significant differences in gene expression among strains were obtained by applying an analysis of variance (ANOVA) to each gene individually using the model:  $y_i = u + V_i + e_i$  where  $y_i$  is the ratio of transcripts in strain  $i$  compared to the reference pool,  $u$  is the average ratio across all strains,  $V_i$  is the effect of strain  $i$  on the transcript ratio, and  $e_i$  is the error. An analysis of variance on transcript levels rather than on ratios of transcripts was also done and produced similar results. A  $t$ -test was used to identify genes differentially expressed between the rich medium and copper sulfate medium reference pools.

Permutation resampling was used to estimate the number of false positives generated for different  $p$ -value cutoffs. For each gene, the expression data was randomized with respect to strain 100 times. Each resampling produced very similar rates of false positives. Using a  $p$ -value cutoff of  $p < 0.05$ , 922

genes showed significant expression differences across both conditions and the number of false positives was estimated to be 255 from permutation resampling. Using a cutoff of  $p < 0.01$ , 516 genes showed expression differences across both conditions with only 57 estimated false positives. Using a cutoff of  $p < 0.001$ , 277 genes showed expression differences across both conditions with only two estimated false positives. While the less stringent cutoff produces 667 (922 - 255) compared to 459 (516 - 57) significant expression differences, 28% compared to 11% of significant genes are false positives. While the most stringent cutoff produces only two false positives, only 275 compared to 459 (516 - 57) expression differences are detected. We chose a  $p$ -value cutoff of 0.01 to maximize the number of significant genes and hold the rate of false positives to a minimum.

To identify genes with expression patterns that specifically differ between rich medium and copper sulfate treatments, we performed an analysis of variance using the model  $y_i = u + V_i + V_i M_j + e_i$  where  $y_i$  is the ratio of transcripts in strain  $i$  compared to the reference pool,  $u$  is the average ratio across all strains,  $V_i$  is the effect of strain  $i$  on the transcript ratio,  $V_i M_j$  is the interaction between the ratio of transcripts in strain  $i$  and medium treatment  $j$  (either rich medium or copper sulfate), and  $e_i$  is the error. Using this model, 56 genes were found to differ in their rich medium compared to copper sulfate medium among strain expression patterns.

Significant genes were hierarchically clustered using Cluster and visualized using Treeview [60]. Groups of functionally related genes were annotated by hand and are presented in Figure 2. The microarray images, GenePix gpr files and tab-delimited Cluster files are available on the Faylab homepage [61].

### Growth data

Growth rate was measured as the slope of the regression of  $\log_{10}$  of cell density, as measured by  $OD_{600}$ , and  $\log_{10}$  of time, measured in minutes. Although the cell populations followed a logistic growth curve, copper sulfate treatment affected both the growth rate and carrying capacity parameters of the logistic growth model. Thus, a simple linear rather than logistic regression was used.

### DNA sequence data

Three genes, *HHT2*, *MBP1* and *SUP35*, were sequenced using Big Dye (PerkinElmer) termination sequencing of purified PCR products, GenBank accession number AY553984-AY554008. No polymerase chain reaction (PCR) product could be obtained from M14 and M32 for the *SUP35* gene. *HHT2* on chromosome 14 encodes a histone, *SUP35* on chromosome 4 encodes a translation termination factor, and *MBP1* on chromosome 4 encodes a transcription factor functioning in the cell cycle and DNA replication, 455 kb away from *SUP35*. The forward and reverse primers for *HHT2* were 5'ACCACCTTTACTCTACCGG and 5'AAATTCGCTT-

TATATTCATG, respectively; for *MBP1*, 5'TTACCGATAAGGAGGGGTAGAG and 5'CGGGAAATCGCTCTTCAAA, respectively; and for *SUP35*, 5'AAAATCCCAACCCTACGGTA and 5'CCACTGTAGCCGATACTGGCA, respectively. For each strain both DNA strands were sequenced and analyzed using Phred, Phrap and Consed [62] and polymorphic sites were identified manually. The first polymorphic site found at *MBP1* was heterozygous in both the M5 and M13 strains, and only the site different from the consensus was used in the analysis. Replacement, synonymous and intergenic polymorphic sites were identified using DNASP [63]. A total of 31 segregating sites was found in 3,747 bp surveyed in the nine strains. These include seven intergenic sites, 14 synonymous sites, and 10 replacement sites.

### Comparison of expression and phenotype data

Gene-expression differences in the presence of copper sulfate were correlated with sensitivity to copper sulfate and rust coloration. We used a binary vector to represent differences in growth rate and difference in color in the presence of 1 mM copper sulfate. From Figure 1, both M34 and YPS163 have a reduced rate of growth compared to the other seven strains. Thus, growth rate was a vector of 0s for unaffected strains and 1s for affected strains. Rust coloration was represented by a vector of 0s for white strains and 1s for M14 and M22, the two rust-colored strains.

### Comparison of expression and DNA sequence data

Divergence in gene expression was measured as the number of pairwise differences in gene expression among strains. Because duplicate genes can cross-hybridize on the microarrays, each 70-bp probe was tested for its cross-hybridization potential. Probes susceptible to cross-hybridization were identified as those probes with 70% or greater nucleotide identity to coding sequences other than to the gene the probe was designed to detect. This cutoff was chosen because many genes with 70% sequence identity showed little or no correlation in their expression pattern. Of the 731 genes, 66 families of potentially cross-hybridizing probes were found. Of the 66 families, 44 families only contained one probe among the significant genes and the remaining 26 families contained 158 probes among the significant genes. The largest family was of 71 *gag* or *pol* genes present in Ty transposable elements. All but one member of each potentially cross-hybridizing gene family was removed. The remaining 599 genes were tested for pairwise differences in gene expression using a  $t$ -test ( $p < 0.05$ ). Of these, 436 showed at least one or more differences in gene expression among strains in rich medium, copper sulfate medium or both. A Mantel test showed that these expression differences are significantly associated with DNA sequence differences ( $P = 0.043$ ). The slope of the regression of expression differences on synonymous DNA sequence differences was 5,448. Synonymous DNA sequence differences were used to estimate the rate of change in gene expression as amino-acid altering and intergenic rates of divergence are heavily influenced by purifying selection.

DNA sequence differences between strains were unlikely to affect hybridization. We expect about 0.2 mismatches per 70-bp oligonucleotide, given the sequence divergence between strains at synonymous and amino-acid-altering sites and given that about 70% of coding sites are amino-acid altering. Thus, 82% of hybridizations should contain no mismatches. The remaining probes are not likely to affect the results. After the removal of potentially cross-hybridizing probes, we found many expression patterns that were nearly identical, even though low levels of sequence divergence existed. For example many Ty elements contain one or two mismatches out of 70. Thus, low levels of sequence divergence are unlikely to affect the results. In addition, S288C expression data could be removed and we would expect the results to be the same since the reference pool contains only a small portion of S288C cDNA.

Pairwise gene-expression differences were obtained from the 131 genes found to differ between the two media treatments (Figure 6). Of these genes, two clusters were enriched for genes functioning in protein folding, stress response and carbohydrate metabolism ( $p < 10^{-5}$  [32]). Five genes were removed because of potentially cross-hybridizing probes, and of the remainder, 48 genes showed at least one or more pairwise gene-expression difference between strains in either rich medium, copper sulfate medium, or both ( $t$ -test,  $p < 0.05$ ). A Mantel test found no significant association between DNA sequence divergence and the expression differences of these 48 genes between strains ( $p > 0.05$ ).

### Freeze tolerance

Overnight cultures were resuspended in 4 ml of YPD at an OD<sub>600</sub> of approximately 1 and grown at 30°C for 2 h. Cultures were then split and treated for 1 h at 4°C or at -30°C in an ethanol bath. The frozen cultures were then returned to 4°C by shaking in a 20°C water bath. Both 4°C and -30°C treated cultures were then grown in a 30°C shaker. Relative growth rate was measured by the increase in OD<sub>600</sub> from the time of treatment to 4 h later for three replicate experiments.

### Additional data files

A figure (Additional data file 1) showing strains grown on minimal media and synthetic complete media in the presence and absence of copper sulfate is available with the online version of this article.

### Acknowledgements

We thank A. Moses for stimulating discussions, A. Gasch for the suggestion that differential expression of AQY2 may confer freeze tolerance, and members of the Eisen lab, J. Townsend and D. Crawford for comments on the manuscript. This research was supported by a Sloan Postdoctoral Fellowship to J.C.F. M.B.E. is a Pew Scholar in the Biomedical Sciences. This work was conducted under the US Department of Energy contract number ED-AC03-76SF00098.

### References

- Schadt EE, Monks SA, Drake TA, Lusic AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, et al.: **Genetics of gene expression surveyed in maize, mouse and man.** *Nature* 2003, **422**:297-302.
- Wilson AC, Maxson LR, Sarich VM: **Two types of molecular evolution: evidence from studies of interspecific hybridization.** *Proc Natl Acad Sci USA* 1974, **71**:2843-2847.
- Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA: **The evolution of transcriptional regulation in eukaryotes.** *Mol Biol Evol* 2003, **20**:1377-1419.
- Enard W, Khaitovich P, Klose J, Zollner S, Heissig F, Giavalisco P, Nieselt-Struwe K, Muchmore E, Varki A, Ravid R, et al.: **Intra- and interspecific variation in primate gene expression patterns.** *Science* 2002, **296**:340-343.
- Oleksiak MF, Churchill GA, Crawford DL: **Variation in gene expression within and among natural populations.** *Nat Genet* 2002, **32**:261-266.
- Rockman MV, Wray GA: **Abundant raw material for cis-regulatory evolution in humans.** *Mol Biol Evol* 2002, **19**:1991-2004.
- Cheung VG, Conlin LK, Weber TM, Arcaro M, Jen KY, Morley M, Spielman RS: **Natural variation in human gene expression assessed in lymphoblastoid cells.** *Nat Genet* 2003, **33**:422-425.
- Bray NJ, Buckland PR, Owen MJ, O'Donovan MC: **Cis-acting variation in the expression of a high proportion of genes in human brain.** *Hum Genet* 2003, **113**:149-153.
- Rifkin SA, Kim J, White KP: **Evolution of gene expression in the *Drosophila melanogaster* subgroup.** *Nat Genet* 2003, **33**:138-144.
- Townsend JP, Cavaliere D, Hartl DL: **Population genetic variation in genome-wide gene expression.** *Mol Biol Evol* 2003, **20**:955-963.
- Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL: **Sex-dependent gene expression and evolution of the *Drosophila* transcriptome.** *Science* 2003, **300**:1742-1745.
- Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA, Brown PO: **Individuality and variation in gene expression patterns in human blood.** *Proc Natl Acad Sci USA* 2003, **100**:1896-1901.
- Brem RB, Yvert G, Clinton R, Kruglyak L: **Genetic dissection of transcriptional regulation in budding yeast.** *Science* 2002, **296**:752-755.
- Toma DP, White KP, Hirsch J, Greenspan RJ: **Identification of genes involved in *Drosophila melanogaster* geotaxis, a complex behavioral trait.** *Nat Genet* 2002, **31**:349-353.
- Wayne ML, McIntyre LM: **Combining mapping and arraying: an approach to candidate gene identification.** *Proc Natl Acad Sci USA* 2002, **99**:14903-14906.
- Yvert G, Brem RB, Whittle J, Akey JM, Foss E, Smith EN, Mackelprang R, Kruglyak L: **Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors.** *Nat Genet* 2003, **35**:57-64.
- Mortimer RK: **Evolution and variation of the yeast (*Saccharomyces*) genome.** *Genome Res* 2000, **10**:403-409.
- Besnard E, Chenu C, Robert M: **Influence of organic amendments on copper distribution among particle-size and density fractions in Champagne vineyard soils.** *Environ Pollut* 2001, **112**:329-337.
- Hamer DH, Thiele DJ, Lemontt JE: **Function and autoregulation of yeast copperthionein.** *Science* 1985, **228**:685-690.
- Fogel S, Welch JW: **Tandem gene amplification mediates copper resistance in yeast.** *Proc Natl Acad Sci USA* 1982, **79**:5342-5346.
- Cavaliere D, Townsend JP, Hartl DL: **Manifold anomalies in gene expression in a vineyard isolate of *Saccharomyces cerevisiae* revealed by DNA microarray analysis.** *Proc Natl Acad Sci USA* 2000, **97**:12369-12374.
- Toledano MB, Delaunay A, Bitau B, Spector D, Azevedo D: **Oxidative stress response in yeast.** *Yeast Stress Resp* 2003, **1**:241-303.
- Morgan BA, Banks GR, Toone WM, Raitt D, Kuge S, Johnston LH: **The *Skn7* response regulator controls gene expression in the oxidative stress response of the budding yeast *Saccharomyces cerevisiae*.** *EMBO J* 1997, **16**:1035-1044.
- Lee J, Godon C, Lagniel G, Spector D, Garin J, Labarre J, Toledano MB: **Yap1 and Skn7 control two specialized oxidative stress response regulons in yeast.** *J Biol Chem* 1999, **274**:16040-16046.
- Mayer MP, Bukau B: **Hsp70 chaperone systems: diversity of cellular functions and mechanism of action.** *Biol Chem* 1998, **379**:261-268.
- Chang HC, Lindquist S: **Conservation of Hsp90 macromolecular**

- complexes in *Saccharomyces cerevisiae*. *J Biol Chem* 1994, **269**:24983-24988.
27. Okamura K, Kimata Y, Higashio H, Tsuru A, Kohno K: **Dissociation of Kar2p/BiP from an ER sensory molecule, Ire1p, triggers the unfolded protein response in yeast.** *Biochem Biophys Res Commun* 2000, **279**:445-450.
  28. Mori K, Ogawa N, Kawahara T, Yanagi H, Yura T: **Palindrome with spacer of one nucleotide is characteristic of the cis-acting unfolded protein response element in *Saccharomyces cerevisiae*.** *J Biol Chem* 1998, **273**:9912-9920.
  29. Frand AR, Kaiser CA: **The ERO1 gene of yeast is required for oxidation of protein dithiols in the endoplasmic reticulum.** *Mol Cell* 1998, **1**:161-170.
  30. Huang M, Zhou Z, Elledge SJ: **The DNA replication and damage checkpoint pathways induce transcription by inhibition of the Crt1 repressor.** *Cell* 1998, **94**:595-605.
  31. Chen W, Bucaria J, Band DA, Sutton A, Sternglanz R: **Enp1, a yeast protein associated with U3 and U14 snoRNAs, is required for pre-rRNA processing and 40S subunit synthesis.** *Nucleic Acids Res* 2003, **31**:690-699.
  32. ***Saccharomyces* Genome Database: expression connection** [<http://genome-www.stanford.edu/Saccharomyces/>]
  33. Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO: **Genomic expression programs in the response of yeast cells to environmental changes.** *Mol Biol Cell* 2000, **11**:4241-4257.
  34. Zhong T, Arndt KT: **The yeast SISI protein, a DnaJ homolog, is required for the initiation of translation.** *Cell* 1993, **73**:1175-1186.
  35. Lu Z, Cyr DM: **Protein folding activity of Hsp70 is modified differentially by the hsp40 co-chaperones Sis1 and Ydj1.** *J Biol Chem* 1998, **273**:27824-27830.
  36. Kuras L, Cherest H, Surdin-Kerjan Y, Thomas D: **A heteromeric complex containing the centromere binding factor I and two basic leucine zipper factors, Met4 and Met28, mediates the transcription activation of yeast sulfur metabolism.** *EMBO J* 1996, **15**:2519-2529.
  37. Blaiseau PL, Isnard AD, Surdin-Kerjan Y, Thomas D: **Met31p and Met32p, two related zinc finger proteins, are involved in transcriptional regulation of yeast sulfur amino acid metabolism.** *Mol Cell Biol* 1997, **17**:3640-3648.
  38. Tanghe A, Van Dijck P, Dumortier F, Teunissen A, Hohmann S, Thevelein JM: **Aquaporin expression correlates with freeze tolerance in baker's yeast, and overexpression improves freeze tolerance in industrial strains.** *Appl Environ Microbiol* 2002, **68**:5981-5989.
  39. Felsenstein J: **Phylogenies and the comparative method.** *Am Nat* 1985, **125**:1-15.
  40. Lynch M: **The rate of morphological evolution in mammals from the standpoint of the neutral expectation.** *Am Nat* 1990, **136**:246-256.
  41. Garland T: **Rate tests for phenotypic evolution using phylogenetically independent contrasts.** *Am Nat* 1992, **140**:509-519.
  42. Harmon LJ, Schulte JA 2nd, Larson A, Losos JB: **Tempo and mode of evolutionary radiation in iguanian lizards.** *Science* 2003, **301**:961-964.
  43. Kimura M, Crow JF: **The number of alleles that can be maintained in a finite population.** *Genetics* 1964, **49**:725-738.
  44. Lynch M, Hill WG: **Phenotype evolution by neutral mutation.** *Evolution* 1986, **40**:915-935.
  45. Sokal RR, Rohlf FJ: *Biometry: The Principles and Practice of Statistics in Biological Research* New York: WH Freeman; 1995:887.
  46. Thiele DJ: **ACE1 regulates expression of the *Saccharomyces cerevisiae* metallothionein gene.** *Mol Cell Biol* 1988, **8**:2745-2752.
  47. Culotta VC, Howard WR, Liu XF: **CRS5 encodes a metallothionein-like protein in *Saccharomyces cerevisiae*.** *J Biol Chem* 1994, **269**:25295-25302.
  48. Gralla EB, Thiele DJ, Silar P, Valentine JS: **ACE1, a copper-dependent transcription factor, activates expression of the yeast copper, zinc superoxide dismutase gene.** *Proc Natl Acad Sci USA* 1991, **88**:8558-8562.
  49. Gross C, Kelleher M, Iyer VR, Brown PO, Winge DR: **Identification of the copper regulon in *Saccharomyces cerevisiae* by DNA microarrays.** *J Biol Chem* 2000, **275**:32310-32316.
  50. Ashida J, Higashi N, Kikuchi T: **An electronmicroscopic study on copper precipitation by copper-resistant yeast cells.** *Protoplasma* 1963, **57**:27-32.
  51. Yu W, Farrell RA, Stillman DJ, Winge DR: **Identification of SLF1 as a new copper homeostasis gene involved in copper sulfide mineralization in *Saccharomyces cerevisiae*.** *Mol Cell Biol* 1996, **16**:2464-2472.
  52. Thomas D, Surdin-Kerjan Y: **Metabolism of sulfur amino acids in *Saccharomyces cerevisiae*.** *Microbiol Mol Biol Rev* 1997, **61**:503-532.
  53. Spiropoulos A, Bisson LF: **MET17 and hydrogen sulfide formation in *Saccharomyces cerevisiae*.** *Appl Environ Microbiol* 2000, **66**:4421-4426.
  54. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, et al.: **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.** *Proc Natl Acad Sci USA* 2001, **98**:10869-10874.
  55. Mortimer RK, Johnston JR: **Genealogy of principal strains of the yeast genetic stock center.** *Genetics* 1986, **113**:35-43.
  56. Sniegowski PD, Dombrowski PG, Fingerman E: ***Saccharomyces cerevisiae* and *Saccharomyces paradoxus* coexist in a natural woodland site in North America and display different levels of reproductive isolation from European conspecifics.** *FEMS Yeast Res* 2002, **1**:299-306.
  57. **Microarrays.org** [<http://derisilab.ucsf.edu/microarray/protocols.html>]
  58. Kerr MK, Churchill GA: **Experimental design for gene expression microarrays.** *Biostatistics* 2001, **2**:183-201.
  59. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP: **Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation.** *Nucleic Acids Res* 2002, **30**:e15.
  60. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns.** *Proc Natl Acad Sci USA* 1998, **95**:14863-14868.
  61. **Justin C. Fay Lab homepage** [<http://www.genetics.wustl.edu/jflab/>]
  62. **Phred, Phrap, Consed** [<http://www.phrap.org/>]
  63. Rozas J, Rozas R: **DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis.** *Bioinformatics* 1999, **15**:174-175.