Bayesian Analysis of Gene Expression Levels: Statistical Quantification of Relative mRNA Level across Multiple Strains or Treatments

Citation

Townsend, Jeffrey P. and Daniel L. Hartl. 2002. Bayesian analysis of gene expression levels: statistical quantification of relative mRNA level across multiple strains or treatments. Genome Biology 3(12): research0071.1-research0071.16.

Published Version

http://genomebiology.com/2002/3/12/research/0071

Permanent link

http://nrs.harvard.edu/urn-3:HUL.InstRepos:4457588

Terms of Use

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

Research

Bayesian analysis of gene expression levels: statistical quantification of relative mRNA level across multiple strains or treatments

Jeffrey P Townsend* and Daniel L Hartl

Address: Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA.
*Current address: Miller Institute for Basic Research in Science, University of California Berkeley, 2536 Channing Way, Berkeley, CA 94720-5190, USA.

Correspondence: Jeffrey P Townsend. E-mail: townsend@nature.berkeley.edu

Abstract

Background: Methods of microarray analysis that suit experimentalists using the technology are vital. Many methodologies discard the quantitative results inherent in cDNA microarray comparisons or cannot be flexibly applied to multifactorial experimental design. Here we present a flexible, quantitative Bayesian framework. This framework can be used to analyze normalized microarray data acquired by any replicated experimental design in which any number of treatments, genotypes, or developmental states are studied using a continuous chain of comparisons.

Results: We apply this method to *Saccharomyces cerevisiae* microarray datasets on the transcriptional response to ethanol shock, to *SNF2* and *SWI1* deletion in rich and minimal media, and to wild-type and *zap1* expression in media with high, medium, and low levels of zinc. The method is highly robust to missing data, and yields estimates of the magnitude of expression differences and experimental error variances on a per-gene basis. It reveals genes of interest that are differentially expressed at below the twofold level, genes with high 'fold-change' that are not statistically significantly different, and genes differentially regulated in quantitatively unanticipated ways.

Conclusions: Anyone with replicated normalized cDNA microarray ratio datasets can use the freely available MacOS and Windows software, which yields increased biological insight by taking advantage of replication to discern important changes in expression level both above and below a twofold threshold. Not only does the method have utility at the moment, but also, within the Bayesian framework, there will be considerable opportunity for future development.

Background

Methods for analysis of cDNA microarray data include those that cluster hierarchically [1] by principles of self-organization [2] or by k-means [3]. These methods yield enormous amounts of information about similarities of cell state and coordination of gene regulation, and are useful for grouping genes or transcriptional profiles by similarity. They have the limitation that although experimental replication enhances the significance of groupings observed, the groupings do not inherently quantify signal and noise. A fold-value cutoff originally was used for this purpose [4], and held double duty as a signifier of true signal and a boundary beyond

which observed fold-measures were considered to be reflective of actual fold-change. Other approaches use likelihood-based methods [5,6] to obtain P-values for gene expression differences in replicated comparisons. These methods make the assumptions and have the power of model-based statistics, but as yet are not formulated to handle more than two genotypes, environments, or developmental states within a single, cohesive framework.

One method for analyzing experiments that involve numerous treatments is the use of analysis of variance on microarray data. Methods have been developed that can yield a profusion of information about the sources of experimental variation [7,8] or, at a biological level, about the proportion of variation in expression profile attributable to biological factors such as sex or genotype [9]. These methods can estimate the magnitude of effects as well as significance, but also impose considerable constraints on experimental design [10], and they are not robust to missing or excluded data.

Volcano plots [8] have highlighted well the important distinction between biological and statistical significance. There are effects that may be biologically important that may not be statistically significant, and vice versa. Because many microarray experiments can have a complex and unbalanced design, owing to the technical failure of certain hybridizations and the iterative nature of the work itself, we have developed an approach for assessing statistical significance that could potentially use all the available observations in any transitively connected design. Our goal is to identify effects of biologically significant magnitude to statistically significant precision.

To that end, we introduce a Bayesian analysis of gene expression level (BAGEL) model for statistical inference of gene expression and demonstrate its utility by re-examining cDNA microarray data on the response of yeast to ethanol shock [11], on transcriptional regulation by *SNF2* and *SWI1* [12], and on zinc regulation [13].

Results and discussion

Our model estimates gene-expression levels, μ_i, and error variances, σ_i for each gene by Markov chain Monte Carlo integration of the likelihood function of observed gene-expression ratios, and incorporates a prior distribution for the parameters. With an uninformative prior, statistical analysis within this model is possible as long as there are as many comparisons as there are parameters to be estimated. Unfortunately, many cDNA microarray studies have been carried out with minimal replication, and most use a reference-sample design (for example [4,14-16]) that yields weak statistical information ([7,8]; see also below).

Figure 1 diagrams the experimental design of three recent cDNA microarray studies [11-13] that have incorporated
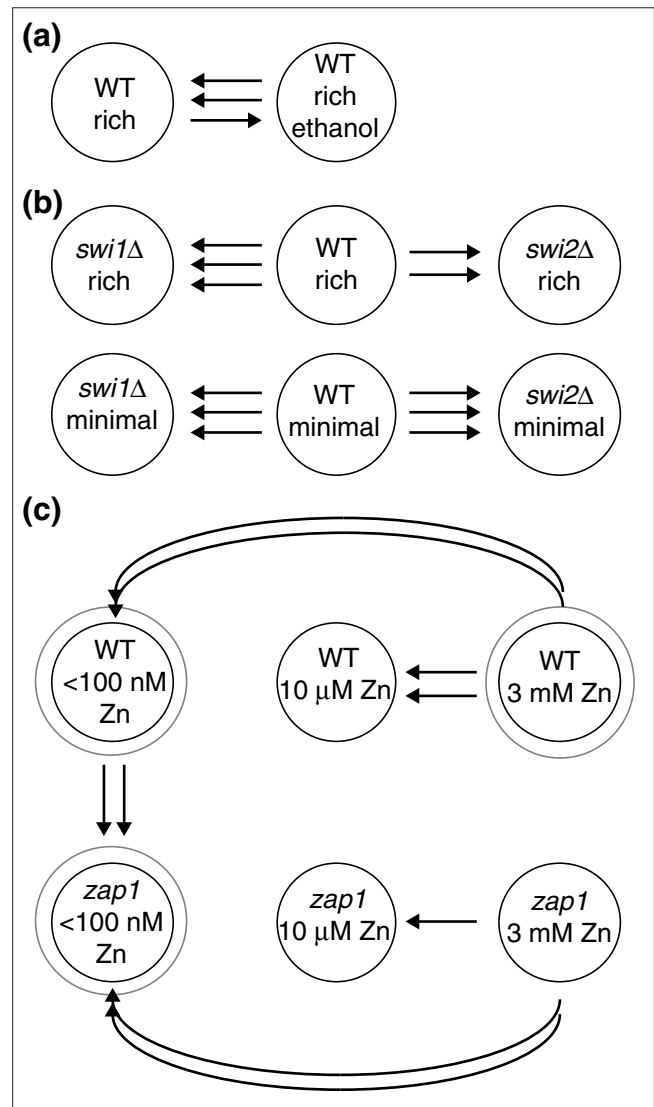

Figure 1

Experimental designs for three studies analyzed using the BAGEL framework. Each circle represents an expression 'node', typically characterized by a particular genotype, environment and developmental state. Here, each circle is inscribed with a genotype and environment. Arrows represent individual two-color cDNA microarray experiments, with the arrowhead pointing toward the sample labeled with the Cy3 fluorophore. **(a)** A comparison of global gene expression of wild-type (WT) yeast in rich medium at log-growth to that of yeast at log-growth after 30 min exposure to 7% v/v ethanol [11]. **(b)** A comparison of *swi1Δ* and *snf2Δ* mutants to wild type in rich and minimal media [12]. **(c)** A comparison of wild-type yeast and *zap1* mutants in synthetic complete medium with three different concentrations of zinc [13], in a long, unbroken, replicated chain of comparisons. Circles surrounded by a double line highlight the samples that turned out to be most important to their study.

some replication. Each expression node is diagrammed as a circle, with genotype and environmental state inscribed. For some studies (Figure 1a), $n = 2$: expression during normal log growth, and expression after 30 minutes of ethanol

shock. For others, n is larger. For instance, the examination of $snf2\Delta$, $swi1\Delta$, and wild-type genotypes in rich and minimal medium, yields an $n = 3$ in each of two conditions, and the study of $zap1$ and wild-type strains in high, medium, and low zinc yields $n = 6$. These studies, in an exemplary fashion [17], have incorporated replication into their experimental design. As the experiments were not originally designed for analysis by this method, analysis of these datasets demonstrates some of the flexibility and utility of the BAGEL statistical framework. Furthermore, data from experiments following any of the replicated experimental designs described in Yang and Speed [18] may be directly and easily imported and analyzed by the BAGEL software.

Ethanol shock

Alexandre *et al.* [11] examined the effect of 30 minutes of ethanol shock on a culture of yeast exposed during log-phase growth, by comparing the global gene expression of the ethanol-exposed cells to the global gene expression of cells in the mid-log phase of growth (Figure 1a). If we arbitrarily assign the node with the lower estimated expression to a level of unity and assume an equal error variance in both treatment and control, there are only two parameters to estimate: the mean expression level μ of the higher-expressed node, and their common error variance σ^2. Figure 2a shows contour plots of the likelihood functions $\prod_k f(z_k)$ (see Materials and methods) for genes for three yeast transcriptional regulators (*MSN4*, *HSF1*
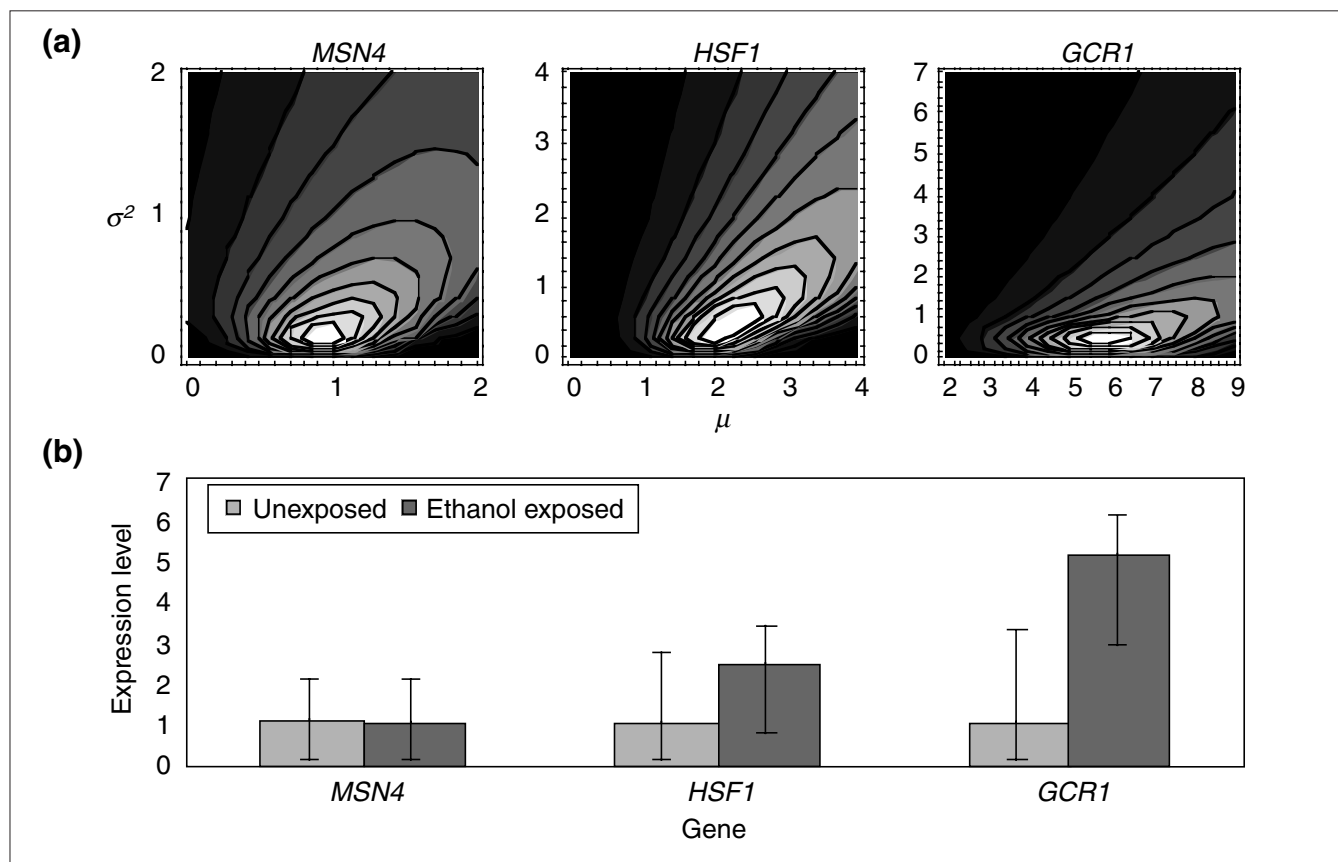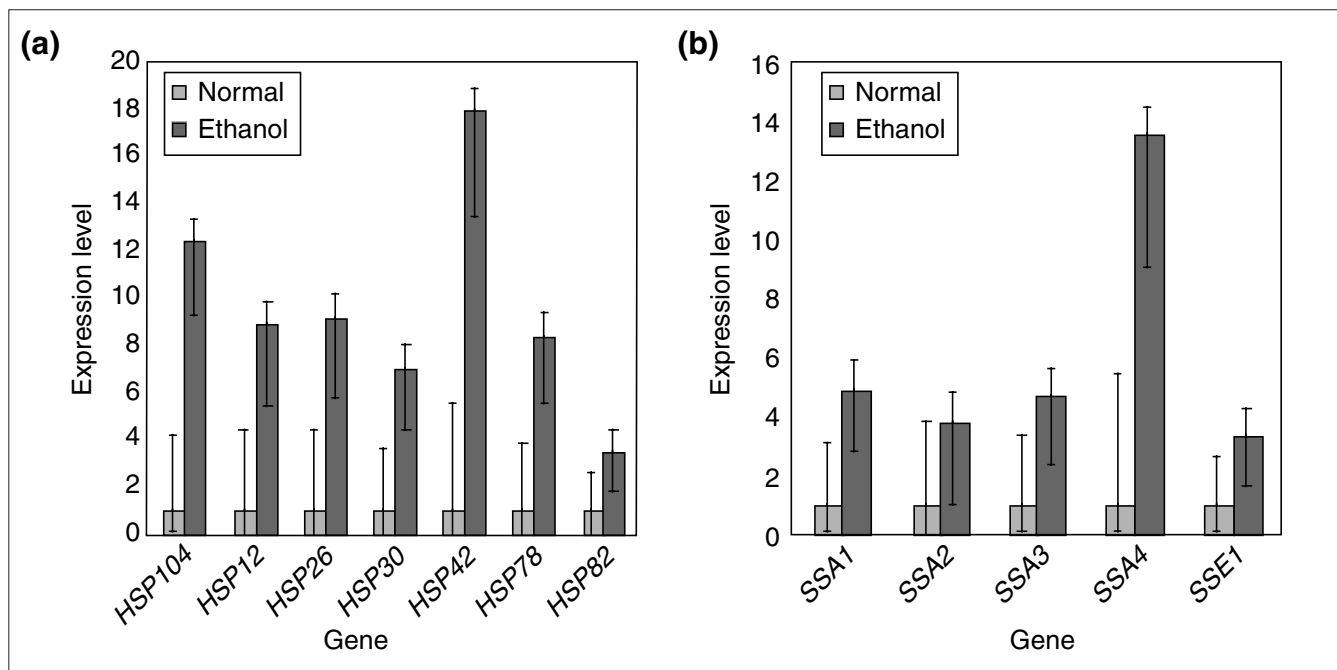

Figure 2
Estimates of gene expression levels for three yeast transcription factors after 30 minutes of exposure to 7% vol/vol ethanol during peak growth phase. **(a)** Likelihood contour plots for the relative gene-expression level (μ) and experimental error variance (σ^2). For each gene, surfaces have a single peak. **(b)** Credible intervals for gene-expression level for three yeast transcription factors that are candidates for the ethanol shock response, taken from cultures at log-phase growth either unexposed or exposed to 30 min of 7% v/v ethanol. Because the slope of the likelihood is much steeper to the right than to the left of this ridge represented in (a), credible intervals with two-sample data such as this are highly asymmetric. More balanced experimental designs have more symmetric credible intervals. Using a conservative gene-by-gene criterion of nonoverlapping 95% credible intervals, none of these transcription factors shows statistically significantly differential gene expression, although *GCR1* is close. *MSN4* is the gene for a zinc-finger transcriptional activator for genes regulated through Snf1p. A general regulator for stress response, it does not seem to be responsible for upregulation of stress genes in response to ethanol. *HSF1* is the gene for a heat-shock transcription factor that binds to the heat-shock DNA element at both normal and elevated temperatures. Although the difference in estimates observed here is suggestive of regulatory action of the heat-shock response, these differences are not statistically significant. *GCR1* encodes a positive regulator of glycolytic genes, including *GLK1*. Although the difference in estimates observed here is suggestive of regulatory action on transcription of transposable elements and the trehalose pathway, the differences are not statistically significant by a conservative criterion of nonoverlapping 95% credible intervals.

Figure 3
Stress-induced chaperonins are abundant in log-phase yeast culture shocked by 30 min exposure to 7% ethanol. **(a)** BAGEL estimates and confidence intervals for seven heat-shock genes. The first four were mentioned by [11], and are significantly abundantly expressed after ethanol shock. *HSP30* and *HSP42* are also significantly abundant, whereas *HSP82* is not. **(b)** Except for *SSA4*, these genes of the *HSP70* family are not statistically significantly differentially expressed after ethanol shock. Note, however, that estimates for their expression level are consistently higher.

and *GCR1*) that could potentially be responsible for observed downstream changes discussed later in this text. These three genes demonstrate typical two-dimensional likelihood surfaces in which the most likely expression levels are equivalent, different, and very different, respectively. Msn4p is a zinc-finger transcriptional activator for many genes regulated by Snf1p. When *MSN4* has a null allele, stress-response genes including *HSP12*, *HSP26*, *HSP42*, *HSP78* and *HSP104* are upregulated [19]. Hsf1p is a heat-shock transcription factor that binds to the heat-shock DNA element at both normal and raised temperatures. Abundant transcripts from *MSN4* and *HSF1* could have a role in the upregulation of heat shock genes after ethanol shock (see below). Abundant transcripts of *MSN4* and *GCR1* could have a role in the concomitant upregulation of the trehalose pathway (see below).

The observed *MSN4* ratios in two replicate microarrays were inconsistent (0.6- and 2-fold). This, in fact, gives very weak support to the hypothesis that the nodes are expressed at nearly the same level. This is in part because, conditional on higher error variances, larger differences in expression become increasingly likely. This effect is seen more dramatically in the contour plot for *HSF1*, for which the highly dispersed, somewhat inconsistent ratios of 0.9-fold, 3-fold and 5-fold were observed. Consistent data, even when relatively dispersed (ratios 3-fold, 5-fold and 10-fold), as for *GCR1*, shows this effect but with a greater slope in the likelihood

surface. Although Figure 2a depicts two-dimensional surfaces from an ordinary treatment-reference experimental design with a common variance term, it shows typical simple topologies of the likelihood surface. Increasing the amount of data in larger datasets accelerates reliable convergence upon the stationary distribution of the Markov chain, which is required for inference of the posterior distributions of the parameters of interest. Furthermore, in larger datasets, these posterior unconditional distributions of the parameters are unimodal. By inference, the multidimensional likelihood surfaces are expected to be fairly simple.

These likelihood plots clearly convey the most probable expression levels. To determine statistical significance, we examine the posterior distributions of the parameters, as determined by Markov chain Monte Carlo simulation. In fact, the credible intervals for the expression level for all three of these genes overlap (Figure 2b). There is not enough information in the replicated comparisons of each gene's ratio of expression to constrain the variance parameter to a small enough value so that expression levels could be inferred to lie within a small range.

In contrast to the lack of statistical significance of the expression differences observed in these three upstream transcription factors, many of the downstream conclusions of the original study are not only consistent with BAGEL estimates, but are
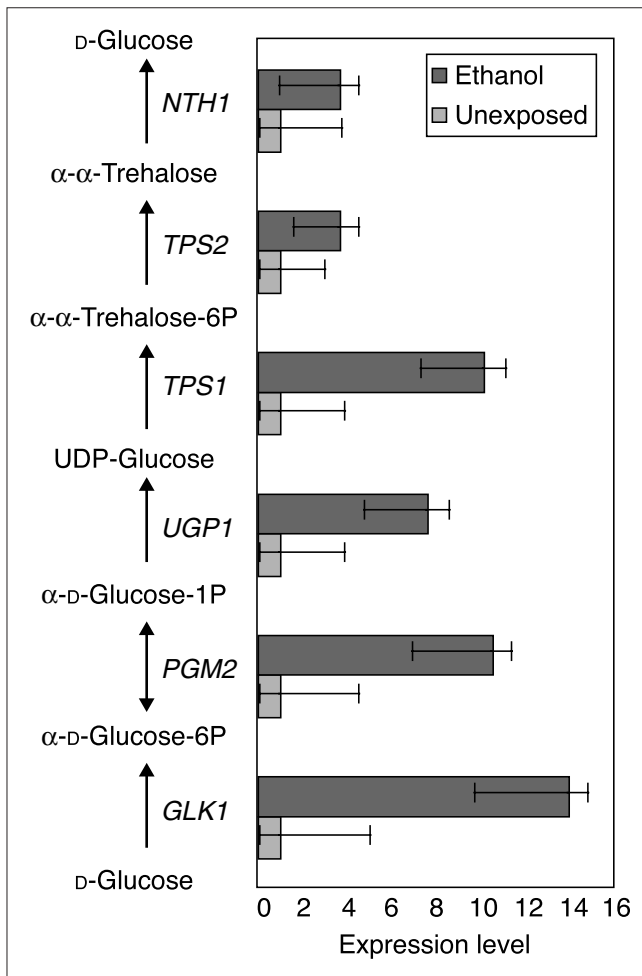
Figure 4
Trehalose pathway transcripts are significantly abundant along the pathway from D-glucose to trehalose 6-phosphate in yeast cultures exposed to 7% ethanol for 30 min at log-phase growth. Trehalose has a protective role after ethanol exposure. Although estimates of trehalase expression are consistent with an increase in transcripts responsible for degradation of trehalose (*NTH1*, shown; *ATH1*, not shown but very similar), neither gene has nonoverlapping 95% credible intervals in the two conditions.

Figure 5
Transposable-element transcripts are abundant after 30 min ethanol shock to a yeast culture in log-phase growth. The high sequence homology of these ORFs means that the effects charted above cannot be considered independent as observed on a cDNA microarray. The high level of general transposable-element transcript abundance, however, remains interesting.

statistically significant by the conservative gene-by-gene criterion of nonoverlapping 95% credible intervals. For instance, the authors concluded that the stress-induced chaperonins ('heat-shock' genes) were upregulated by 30 minutes of exposure to 7% v/v ethanol [11]. Many of these genes are clearly significantly abundant in the stressed state (Figure 3a), although there is not enough data to conclude that all of the set of *HSP70*-family genes is significantly abundant in the stressed state (Figure 3b). On the basis of the large estimated expression differences (nearly always twofold or greater), an effect on the *HSP70* family may yet be large and biologically important.

In addition, the trehalose pathway, whose product aids the cell in dealing with excess ethanol [20], is clearly upregulated
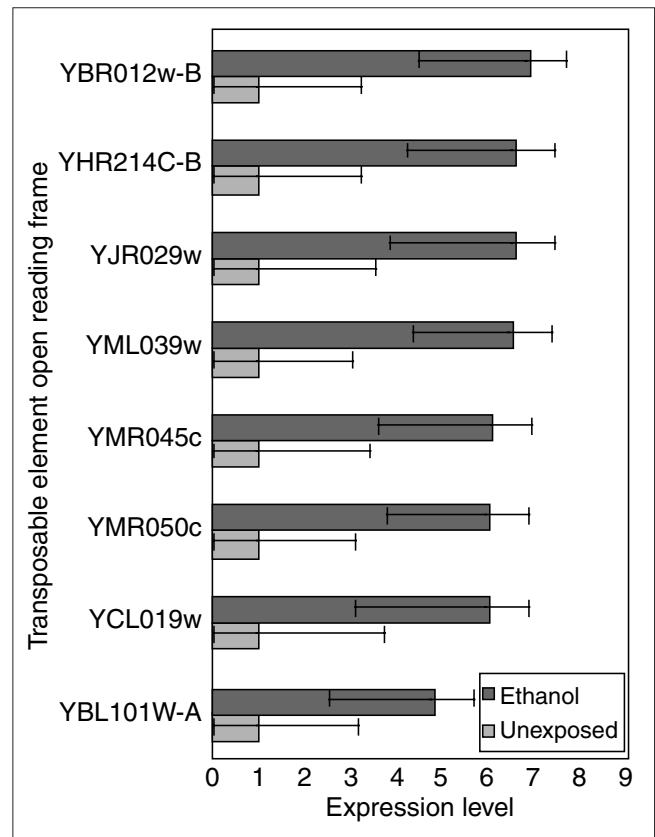
(Figure 4). Interestingly, the six genes *HXK2*, *PGM1*, YHL0012w, *TSL1*, *TPS2* and *ATH1*, form a completely parallel pathway from α-D-glucose to α,α-trehalose to D-glucose. Of those genes, none is significantly upregulated. Teasing apart the biochemical modes of action of these isoforms ('redundant' genes) will lead to a better understanding of the modularity of genomes and of the varied methods of response to environmental challenges of organisms. Alexandre *et al.* propose a futile trehalose cycle after observing high ratios of expression of *NTH1* in their ethanol-shocked sample [11]. However, the wide credible intervals around the expression levels of the genes for the neutral trehalase NTH1 (Figure 4) and the acid trehalase ATH1, the observed accumulation of trehalose in fermentations [21], and the potential for compartmentalization [22] indicate that more data are needed before one can infer the extent of such a cycle.

In addition to the results previously noted, it is interesting to observe significant abundant expression of numerous open
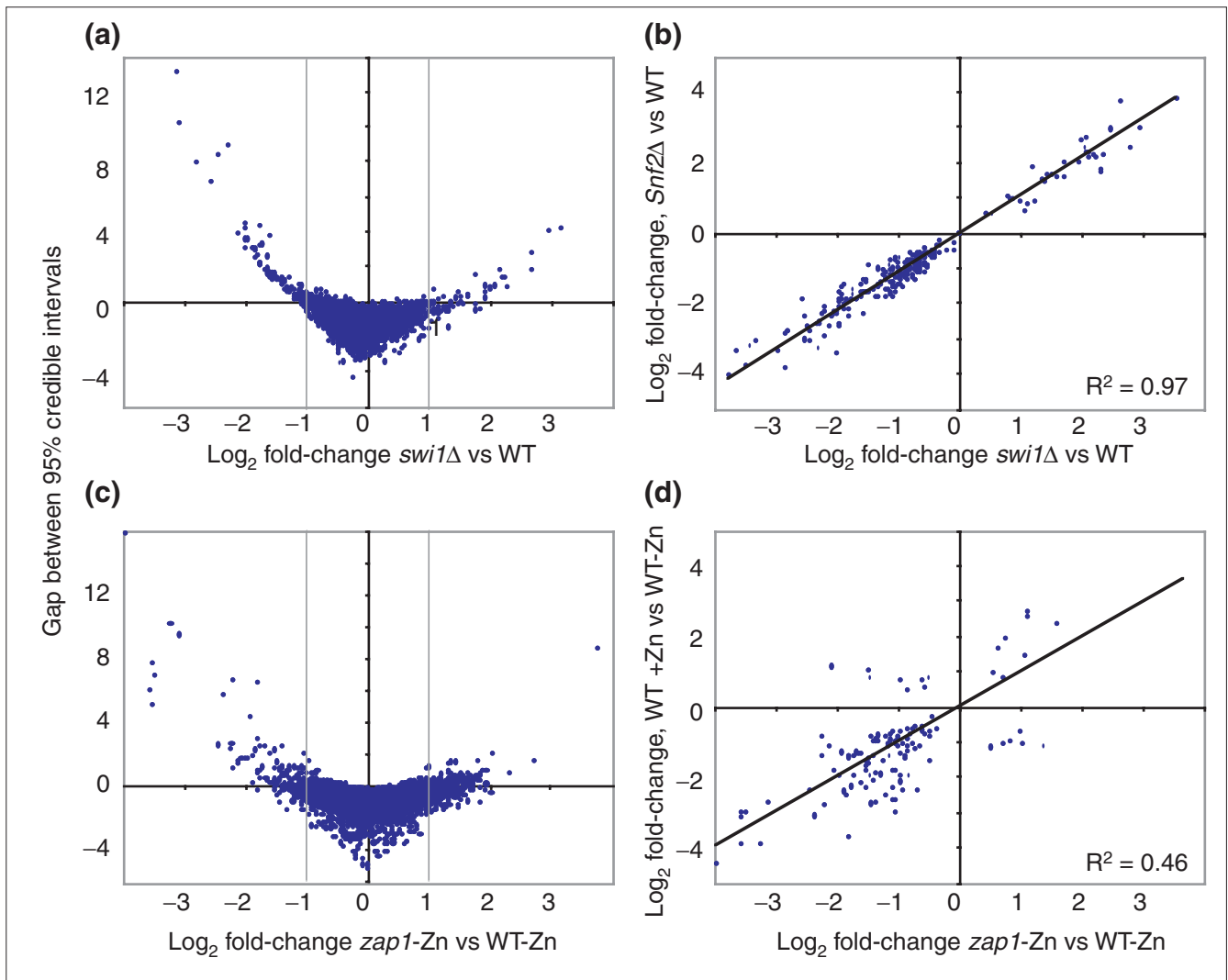
Figure 6
Statistical and biological significance of inferred gene expression levels in *swi1*Δ, *snf2*Δ, and *zap1* mutants compared to wild type (WT). **(a)** Scatterplot of the log$_2$ BAGEL estimate of fold-change in the *swi1*Δ mutant compared to wild type for all genes (on the *x*-axis) versus the negative overlap of BAGEL credible intervals for the *swi1*Δ mutant compared to wild type. Points above zero on the *y*-axis have nonoverlapping 95% credible intervals. Dashed lines indicate a twofold difference in estimate of gene-expression level. Points outside the region defined by these dashed lines and below zero on the *y*-axis have low statistical significance despite their high estimated fold-change. Points within the region defined by these dashed lines and above zero on the *y*-axis have high statistical significance and a low fold-change. Note that these points are very densely packed and appear close to zero on the *y*-axis because the difference between confidence intervals cannot exceed the difference between estimates. **(b)** Scatterplot of log$_2$ fold-change of gene-expression levels in the *swi1*Δ mutant compared to wild type against gene-expression levels of the *snf2*Δ mutant. Genes shown have nonoverlapping 95% credible intervals with wild type. The regression passes through zero without forcing. **(c)** Scatterplot of the log$_2$ BAGEL estimate of fold-change in the *zap1* mutant in zinc-deficient medium compared to wild type in zinc-deficient medium for all genes (on the *x*-axis) versus the negative overlap of BAGEL credible intervals for the *zap1* mutant compared to wild type in zinc-deficient medium. **(d)** Scatterplot of log$_2$ fold-change of gene-expression levels in *zap1* mutant compared to wild type in zinc-deficient medium versus wild type in zinc-replete medium compared to wild type in zinc-deficient medium. The regression is forced through zero; its *y*-intercept and slope are slightly lower when not forced. Genes shown have nonoverlapping 95% credible intervals with wild type.

reading frames (ORFs) containing transposable-element sequence under ethanol shock (Figure 5). Although robust inferences are compromised by the high sequence homology of these ORFs, the fact that six were significantly abundant and none was meagerly expressed suggests an interesting biological phenomenon. One possible effector is the transcriptional regulator Gcr1p, whose BAGEL estimate in ethanol shock is fivefold greater than in normal log growth and whose credible intervals in the two conditions barely overlap (Figure 2). Gcr1p elevates transcription of transposable elements Ty2-917 and Ty1-912 [23] as well as of *GLK1* of the trehalose pathway ([24], and Figure 4).

Table I

Pairwise comparisons of gene-expression level category by mean log$_2$ ratio and by Bayesian estimation with statistical significance

Strain	Environment	Category*	Mean log$_2$ ratio	Significant†	Not significant	Total
S288C	Ethanol shocked	> 2-fold	2,235	87	1,523	1,610
S288C	versus mid-log growth	< 2-fold	3,504	0	2,767	2,767
		Excluded	399		1,761	6,138
*swi1*Δ and *snf2*Δ	Rich medium	> 2-fold	114	170	46	216
versus S288C	Rich medium	< 2-fold	5,987	81	5,704	5,785
		Excluded	41		141	6,142
*swi1*Δ and *snf2*Δ	Minimal medium	> 2-fold	218	243	14	257
versus S288C	Minimal medium	< 2-fold	5,883	361	4,890	5,251
		Excluded	41		634	6,142
DY1457 versus	Zinc-limited	> 2-fold	1,034	419	471	890
DY1457	versus zinc-replete	< 2-fold	5,118	309	4,953	5,262
		Excluded	0		0	6,152
DY1457 versus	Zinc-limited	> 2-fold	678	198	310	508
DY1457 *zap1*	Zinc-limited	< 2-fold	5,474	122	5,522	5,644
		Excluded	0		0	6,152

*Genes were excluded by necessity because of poor quality or insufficient replication of measurement. †Numbers are not corrected for multiple tests.

There are many reasons why data points in microarray experiments are wisely excluded from analysis. Whole sectors occasionally fail due to premature drying during hybridization, spots are occasionally malformed, and experimental signal is frequently so low compared to background that a spot is best excluded from the data. Only 2,041 genes of 6,138 in these data passed the spot background-foreground acceptance criteria (see Materials and methods) in all three experiments. Of these genes, 65 were significantly different by a gene-by-gene criterion, demanding nonoverlapping 95% confidence intervals. A further 2,337 genes had one observation missing; their credible intervals were appropriately wider after BAGEL analysis of the data and only 22 were statistically significant. Genes with one acceptable observation were not analyzed. Use of an informative prior distribution on the variance, however, would allow such analysis.

Table 1 gives an overview of informative pairwise comparisons from the experiments analyzed here. In the ethanol-shock experiments, only the 87 genes listed in Table 1 are statistically significantly different out of the, on-average, 1,851 genes measured as greater than twofold differentially expressed in these single microarray experiments. BAGEL analysis allows unambiguous inference as to the expression levels and confidence intervals of these genes of interest. In this dataset with only three replicates, we detect no significant differences at below a twofold level. Datasets with greater direct replication, or with less dispersed ratios, or with transitively informative comparisons (see below and J.P.T., unpublished data) promise to do so.

Deletion of *SWI1* and *SNF2*

Sudarsanam *et al.* undertook a study to ascertain which genes were controlled by the Snf2p/Swi1p chromatin-remodeling and transcriptional-activator complex subunits, whether the genes had any distinct regulatory roles, and how their regulatory roles compare in rich and minimal media ([12], and Figure 1b). In rich medium, the credible intervals among mutant and wild-type expression levels do not overlap for 251 genes (more than 4% of the genome). Importantly, 46 genes with estimated ratios greater than twofold have overlapping credible intervals (Figure 6a, lower left and lower right quadrants). Moreover, almost one quarter of the 251 significantly different genes have estimated ratios less than twofold (densely packed between the dashed lines because the difference between estimates cannot exceed the distance between credible intervals, Figure 6a). In minimal medium, similar results were acquired; a greater number of significant results were obtained, due in part to greater replication (Figure 1b, Table 1). The credible intervals of mutant and wild-type expression levels do not overlap for 604 genes, almost 10% of the genome. Interestingly, 14 genes with estimated ratios greater than twofold have overlapping credible intervals, and 361 (more than 50%) of the 604 significantly differentially expressed genes are differentially expressed below the twofold level.

Whether these genes represent all of the genes controlled by the Snf/Swi complex depends, of course, on what level of difference is deemed biologically significant, which in many cases depends upon the gene of interest. For many transcription factors, it may be that small changes in expression level
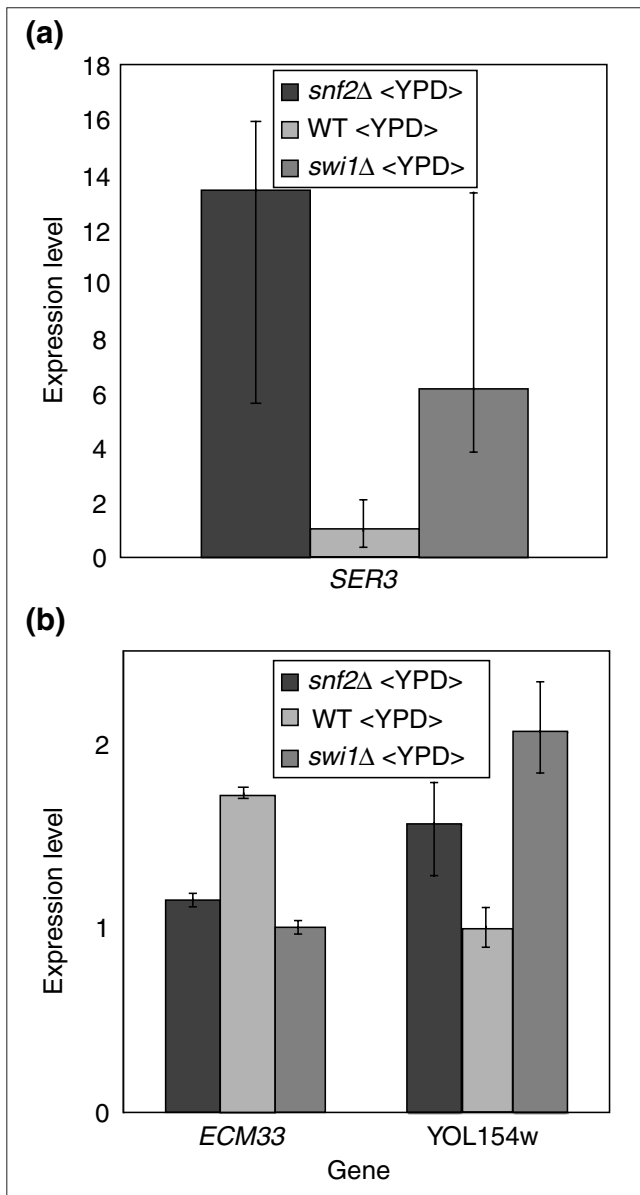
Figure 7
Candidates for genes differentially expressed in mutants *snf2*Δ and *swi1*Δ.
(a) *SER3* is an example of a gene that should not be inferred to be
differentially expressed on the basis of this microarray data despite a
greater than twofold difference in estimated expression level. Credible
intervals for gene expression of *SER3* in *swi1*Δ and *snf2*Δ mutants do not
overlap with the credible interval for wild-type gene expression, but have
extensive overlap with each other. **(b)** Two genes detected by BAGEL as
differentially expressed in the two mutants.

cause very large changes in biological state [25-27], and for
many metabolic genes, very large changes in expression may
change pathway throughput very little [28]. Note also that
these criteria do not distinguish between *cis* and *trans*
effects on expression level. Presumably, many *trans* effects
percolate into much or all of the genome in very small ways;
some of these effects may be overwhelmed by systematic

error induced by the technology, but most should be
detectable with sufficient replication.

Credible intervals determined using BAGEL make it clear
that even *SER3*, which Sudarsanam *et al.* examine as a can-
didate gene differentially affected in the *snf2*Δ and *swi1*Δ
strains on the basis of cDNA microarray and northern data
[12], should not be inferred to be differentially affected
solely on the basis of their cDNA microarray data. Although
estimates of the abundance of *SER3* transcript differ by
more than twofold, the credible intervals for these two
mutants have extensive overlap (Figure 7a). Examining the
rich-media cDNA microarray dataset for other candidates
for differential expression, we observe that there are 27
genes in which only one of the two mutants has nonoverlap-
ping credible intervals with the wild type.

More important, there are just two genes in this global
dataset for which the credible intervals of expression levels
in *snf2* and *swi1* mutants do not overlap. This answers a
question posed by Sudarsanam *et al.* as to whether the dif-
ferences in gene expression between these mutants that they
observed were due to variation in microarray measurements
or to real differences between the mutants [12]. Nearly all
differences they observed in the transcriptional profiles of
these two mutants are potentially due to variation in
microarray experiments. A scatterplot of the $\log_2$ estimates
of gene-expression levels for genes significantly different
from wild type, in the two deletion mutants compared to
wild type, yields a linear correlation of 0.97 (Figure 6b). The
large number of genes meagerly expressed compared to wild
type in both deletion mutants is consistent with the roles of
Snf2p and Swi1p as transcriptional activators. In addition,
globally, not a single gene shows a contrasting change in
expression in comparison to wild type. This affirms the con-
clusion [12] that the two genes work almost entirely in
concert in the media conditions tested. The expression levels
of the two genes detected by BAGEL as having significantly
different expression in the two mutants, *ECM33* and
YOL154w, are shown in Figure 7b. The small magnitude of
the difference observed in *ECM33* between the two strains
may indicate that this highly statistically significant differ-
ence is biologically irrelevant, but certainly further investiga-
tion of the significantly different gene-expression levels
observed in YOL154w is warranted.

Quantitation of short transcript sequence tags by serial
analysis of gene expression (SAGE) [29] yields absolute
counts of expressed sequences in the yeast genome [30].
These quantitations may be used to put relative expression
values obtained from BAGEL on an absolute scale. There are
issues with sampling error and non-uniqueness of sequence
tags in SAGE assays [31] as well as with the linearity of
cDNA microarray ratio measures [17], and with differences
between the 'wild-type' strain used in SAGE and in cDNA
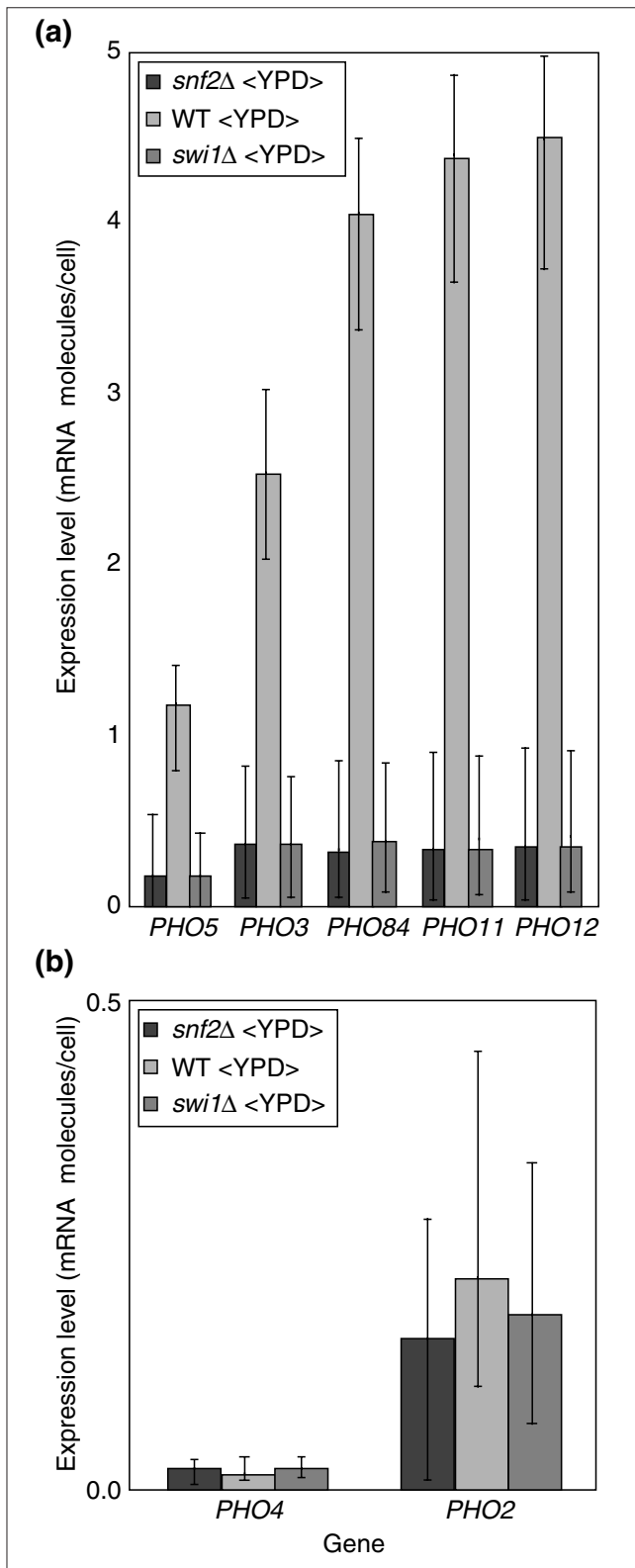microarray experiments. These sources of error are many,

(a)

(b)

Figure 8
Estimates of absolute gene-expression level, in molecules per cell, assuming that the SAGE counts for each gene are typical of wild-type strains. The 95% credible intervals indicated are those for relative gene expression among samples, and do not incorporate sampling variance in the SAGE assay. **(a)** Acid phosphatase genes are similarly meagerly expressed in *swi1*Δ and *snf2*Δ mutants. **(b)** Acid phosphatase regulators *PHO4* and *PHO2* are not significantly differentially expressed in *swi1*Δ and *snf2*Δ mutants, so that Snf2p/Swi1p control may be direct. cDNA microarray measurements of *PHO4* gene-expression levels show that it is not significantly abundant or meager in *swi1*Δ or *snf2*Δ mutants, even though a large SAGE assay could detect no transcript sequence tags for this gene. *PHO4* expression level shown here was arbitrarily set to 0.1 molecules per cell, an order of magnitude below the experiment's detection threshold [30].

normalized so that the wild-type strain on rich medium is assumed to have the same expression levels as indicated by SAGE assay [30]. Figure 8a shows that, as observed by Sudarsanam *et al.* [12], the acid phosphatase genes are meagerly expressed in *snf2*Δ and *swi1*Δ mutants.

Note that although there is esthetic appeal to having the measurements reported on an absolute basis of transcripts per cell, little additional information is obtained by this transformation for most experimental questions, as the result is simply a rescaling of the *y*-axis for each gene. Also, whereas the statistical significance of BAGEL results is robust to considerable deviation from general linearity of cDNA microarray results, the quantitative result using SAGE counts as a basis depends critically upon the linearity of cDNA microarray measurements. Lastly, it is unclear how to scale meagerly expressed genes that had a count of zero in a SAGE analysis. *PHO4* is such a gene; in Figure 8b, we have arbitrarily charted a frequency that would be appropriate if it were present at approximately one order of magnitude below the SAGE experiment's detection threshold of 0.1 molecules per cell. Neither *PHO2* nor *PHO4* shows significant evidence of differential expression in *snf2*Δ/*swi1*Δ mutants. Consideration of the raw ratio data led Sudarsanam *et al.* [12] to thus conclude that *PHO5* gene expression is directly controlled by Snf2p/Swi1p. *PHO3*, a 'constitutive' acid phosphatase, shows the same relative abundance across *snf2*Δ/*swi1*Δ mutants as does *PHO5* (data not shown). The 87% identity of the nucleotide sequence of these genes means that cross-hybridization may confound inference about their regulation. However, unlike *PHO5*, there is currently no evidence that *PHO3* is regulated by *PHO2* or *PHO4*. *PHO5* may share a direct Snf2p/Swi1p regulatory mechanism with *PHO3*.

Analysis presented here supports the result that *SWI1* and *SNF2* work almost entirely in concert within the cell in rich medium [12]. As Sudarsanam *et al.* noted, however, the proteins may not always be produced together in all conditions. Moreover, it should be noted that their experimental

but small and uncorrelated. With these caveats in mind, Figure 8a assumes that SAGE counts are exact in order to provide absolute expression levels inferred using BAGEL,

design has the least power to detect these specific differences. This is because all comparisons of expression level between the mutant strains are transitive comparisons, which inherit the variance associated with the intermediary wild-type expression level as well as the variance associated with the expression levels of the two mutant strains. The result is that on a log scale the credible intervals of gene-expression level are broader for the two mutants than for the wild type. This outcome is a recurrent problem with a repeated reference-sample experimental design: one learns the most information about the reference sample, which is frequently arbitrary and not necessarily of interest. An ideal addition to the experimental design in Figure 1b would be several direct comparisons of the mutant strains. Then, both transitive and direct information would contribute to the statistical power of a BAGEL analysis for any comparison. It is also clear that an increase in number of comparisons would yield more power to detect differentially expressed genes, and would indeed find more of them. Every single gene that was detected as significantly differentially expressed in only one of the two mutants compared to wild type in the rich medium was significantly different in expression level only in the *swi1*Δ mutant. The *swi1*Δ mutant comparison to wild type had one more replicate than did the comparison of the *snf2*Δ mutant (Figure 1b).

Zinc regulation

Lyons *et al*. examined wild-type and *zap1* mutant strains of yeast growing in cultures containing three different concentrations of zinc ([13], and Figure 1c). Zap1p is a transcriptional activator that appears to regulate transcription of the zinc-uptake system genes in response to zinc [32]. Using the nonoverlapping 95% credible interval criterion, a BAGEL analysis on this data reveals 469 genes significantly more abundantly expressed in cells grown in zinc-deficient medium compared to cells grown in zinc-supplemented medium, and reveals 261 genes significantly less abundant in the same comparison. This is a total of about 10% of the genome, and is two thirds of the number found by use of an averaged twofold criterion. A considerable number of the genes viewed as abundant by a twofold criterion, then, are not significantly different by the credible interval criterion (Table 1, Figure 6c, lower left and lower right quadrants). Moreover, 42 of the genes significantly different by the credible-interval criterion are significantly different at a ratio of below twofold (Figure 6c, densely packed between the dashed lines delimiting a twofold change). As one might generally expect, making *ZAP1* nonfunctional in a zinc-deficient environment creates a similar relative effect on most significantly differentially expressed genes to that created by providing zinc to a zinc-deficient wild-type strain (Figure 6d).

A common technique to discover candidate genes of interest in microarray studies is to pick out overlapping sets of genes expressed at a ratio greater than some cutoff from replicate and/or different microarray experiments (for example, Figure 1 of [13], Figure 1 of [33]). BAGEL supplants any need to examine overlapping gene sets in replicate experiments. Moreover, it provides a statistically rigorous method for comparison of multiple different experiments. The equivalent of an overlap in two lists of highly expressed genes is that neither of the credible intervals for gene-expression level in a given gene in two experimental conditions overlaps the reference condition.

For instance, BAGEL analysis of the zinc-regulation data revealed 96 genes whose levels of gene expression in a wild-type genotype and zinc-deficient medium were significantly greater than both zinc-supplemented wild-type and *zap1* in zinc-deficient medium. This is 31 more genes than were detected using the twofold criterion. Figure 9 shows BAGEL results for two genes, *PHM7* and YGL121c, discussed by [13], as well as results on four other genes. A zinc-responsive element (ZRE) consensus DNA sequence, ACCYTNARGGT (in the single-letter amino-acid code, where N is any nucleotide, Y is C or T, and R is A or G) (compare Figure 2 of [13]), is located in close proximity to *CTT1* (chromosome VII, location 65270-80), *MNT2* (VII, 20659-69 and 20774-84), YOL083w (XI, 442975-85) and YNL253w (XIV, 169669). BAGEL results shown in Figure 9 convey not just the significance but also the extent of differential regulation. *CTT1* and YOL083w are like the two genes *PHM7* and YGL121c discussed by [13], in that having the *zap1*::*TRP1* genotype does not entirely eliminate transcriptional response to zinc deficiency. These intermediate levels of expression indicate the action of auxiliary zinc-regulatory mechanisms.

Figure 10 charts several transcription factors that are significantly differentially expressed below the twofold level. For every gene shown, the credible intervals between wild type in 3 mM zinc and wild type in < 100 nM zinc do not overlap. For *VPS20*, in contrast to the other four genes, expression is greater when in zinc-deficient medium. Expression levels in Figure 10 are graphed cosequentially with the experimental design structure (Figure 1c). As a result, the increasing size of credible intervals as one scans from the middle expression nodes to the expression nodes at the ends of the chain of experiments is clear. The largest credible interval is for the *zap1* strain grown in 10 μM zinc, on the far right of each cluster of columns. This is as expected, given that it is at the end of the line of comparisons and is informed by only one cDNA microarray comparison. In a balanced linear set of comparisons such as this, it turns out that for any gene on a log scale, the credible intervals around the gene-expression values of the ends are greater than the credible intervals around those in the middle of the chain (data not shown).

For an analysis such as this one, the statistical power could be improved if expression nodes were compared to more
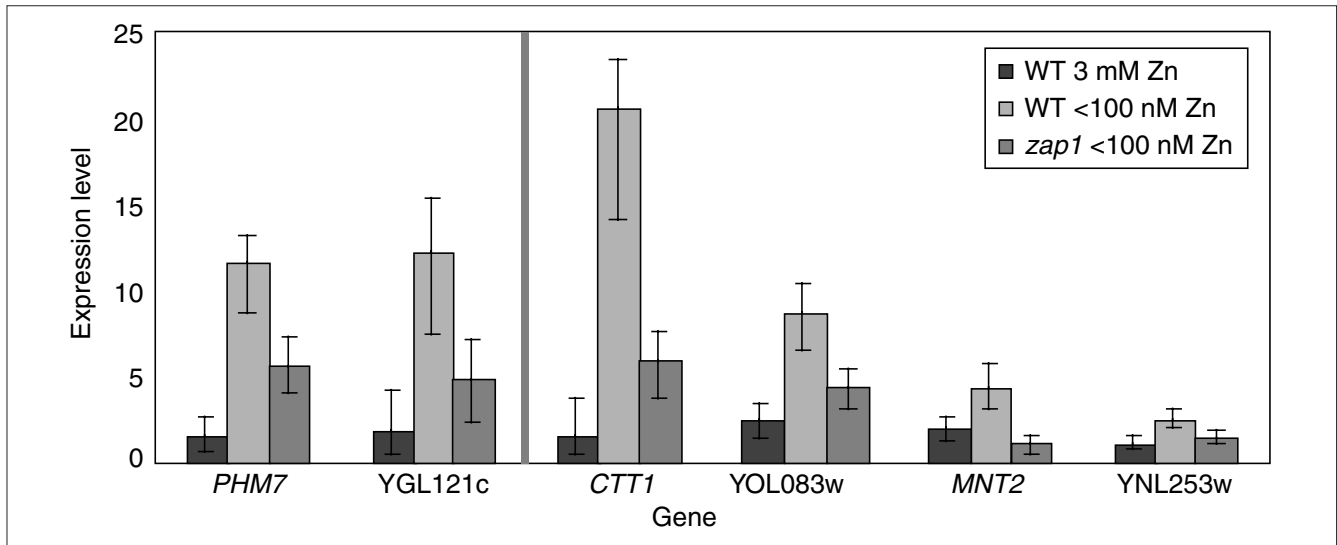
Figure 9
A sample of genes differentially expressed between (wild type, 3 mM zinc) and (wild type, < 100 nM zinc) expression nodes. Six genes located in proximity to zinc-responsive element consensus sequence ACCYTNARGGT that are meagerly expressed in both wild-type genotype grown with 3 mM zinc and *zap1* genotype grown with < 100 nM zinc. *PHM7* and YGL121c were identified and remarked upon in [13] for the difference in expression level between (wild type, < 100 nM zinc) and (*zap1*, < 100 nM zinc), which reflects the action of zinc-regulatory mechanisms auxiliary to Zap1p.
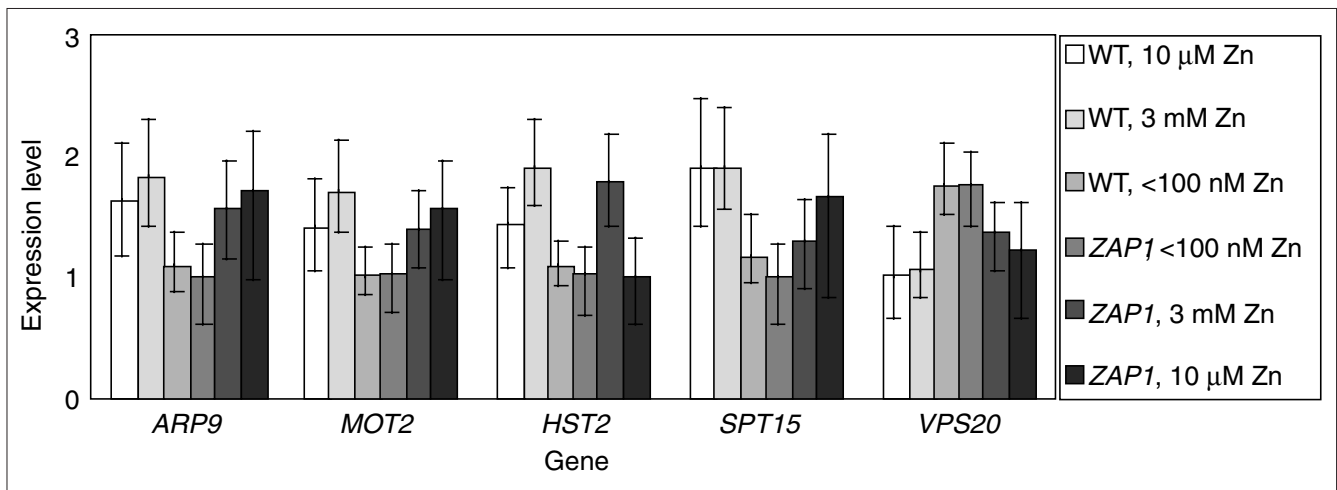

Figure 10
Six transcription factors that are significantly differentially expressed at below the twofold level between wild type, 3 mM zinc and wild type, < 100 nM zinc. Note that each column cluster above is arranged in the linear order of the experimental comparison chain, from wild type, 10 μM zinc to *zap1*, 10 μM zinc; credible intervals for expression levels obtained for expression nodes at the ends of the chain are wider than those obtained for expression nodes in the middle of the chain.

than two neighbors. Specifically, direct comparisons between wild type in 3 mM Zn and *zap1* in 3 mM Zn would complete a circle of comparisons, increasing transitive information on all expression nodes. Cross-circle comparisons would also contribute considerable power. Generally, eliminating 'ends' of chains of comparisons should be a goal of any cDNA microarray experimental design.

Conclusions

Small datasets such as the ethanol dataset are very good candidates for use of a more informative prior distribution to keep variances within a reasonable range and yield better results. Analysis of larger datasets indicates that true variances for microarray data within this model are not larger than one. The impact of using less-vague priors, especially

for the variance, when there are few comparisons, is under investigation by a number of researchers. A hierarchical Bayesian model [34] has been used to analyze ratio data and provide 95% confidence intervals for the log ratio of gene expression from reference to control. This method assumes normal distributions of the log ratios rather than ratios of normal distributions. It has a hierarchical structure that allocates error variance among microarrays and experiments. The authors suggest use of calibration data, or, alternatively, empirical evaluation of the distribution of variances across all other genes in order to construct a prior. A subset of genes of nearly equal intensity can be used to form a prior for variances [35]. This prior was used by the authors to input reliable variances in *t*-tests of significance. Promising advances have been made on Bayesian methods for correcting misleading fold-change measurements made from low-intensity spots [36], using a gamma rather than normal model of ratio results.

All these methods have considerable potential to be incorporated as priors into a framework such as that presented here, so that the prior may be applied to multiple samples from different genotypes, environments or developmental states. Priors such as those above should result in smaller credible intervals and detection of increasingly significant differences because they curtail the exploration of unrealistically high variances that small datasets have too few observations to rule out 'on their own'. Continued work in this area, using an increased amount of non-ratio data provided from scanned microarrays, should be very fruitful [10]. Furthermore, posterior distributions from such analyses of gene-expression level have subsequent use in Bayesian methodologies for clustering [37] and tumor identification [38].

In summary, the model-based approach we have implemented can accommodate complex and unbalanced experimental designs. Some research will continue to be carried out comparing just two samples multiple times. However, complex designs will increase in popularity as investigators explore multiple genotypes, environments and developmental states within a single research project [18]. The utility of this approach in determining levels of gene expression may be maximized if these designs incorporate certain features.

First, compare samples of direct interest directly. When interested in the differences between two samples, compare them to each other rather than to an arbitrary reference sample [7,8]. Whenever possible, study a few expression nodes thoroughly, rather than many superficially.

Second, replicate each comparison at least once [39]. Whether this is done directly by incorporation of dye during reverse transcription, or, preferably, by labeling incorporated amino-allyl-dUTP, reverse the dye labeling to ameliorate any dye effect thereof [9,10,18].

Third, eliminate 'ends' of comparison chains by carrying out hybridizations comparing one end to another. This allows reconciliation of transitive data around a circle of comparisons. The more circles created, the more reconciliation occurs. The smaller the circumference of the circle created, the stronger the transitive power.

Fourth, connect nodes otherwise distantly located on a chain of comparisons with extra cross-comparisons. The number of 'extra' comparisons to make depends on what size of effect is of interest. The observation of a small but significant effect on key regulatory genes may be of greater biological interest than the same observation on a metabolic enzyme. The appropriate weighing of the cost of additional comparisons against the greater precision of measurement depends critically upon the question being asked.

For time-course experiments or any other experiment with an explicit ordered *x*-axis these guidelines may still be followed, as long as replicate comparisons are made among nodes. Inferred estimates at each node are assessed independently of location along the *x*-axis, so that regressions across them are valid. Ultimately, experimental design may be subject to limitation owing to lack of resources or experimental failure. Fortunately, within a framework such as that developed here, missing spots or missing comparisons do not require any special consideration or any change in methodology. Credible intervals acquired for less well determined genes or less well determined expression nodes are correspondingly larger. This quantitative information on gene-expression levels tendered by a thorough analysis of microarray results should be carefully considered in assessments of the biological effects of genetic or environmental differences upon cellular state.

Materials and methods
Normalization
For the ethanol-shock dataset [11], raw data from GenePix files was processed as follows. Any spot was excluded from analysis if both the Cy3 and Cy5 fluorescence signals were within two standard deviations of the distribution of intensities of the background pixels for that spot. These low-intensity spots are those most aberrant in fold-change and are those for which the magnitude is adjusted most by the model of Newton *et al.* [36]. Expression values were normalized by linear scaling of the Cy5 values so that the mean Cy5 and Cy3 values of nonexcluded spots were equal. Two of three experiments thereby achieved a linear log-log intensity plot for included spots, with slope approximately 1. The third was linearized by exponentiation of the Cy3 channel to 0.8, before normalization of the means.

For other experiments, ratio and spot pass criteria were used as reported in the papers. In the dataset released by Sudarsanam *et al.* [12], one of the three reported microarray

hybridizations between wild-type and *snf2* nodes was excluded from analysis because it had an anomalous global mean log ratio of -0.26, whereas for all others that value was very nearly zero. For the zinc-regulation dataset [13], no pass criterion had been used to ensure each spot on each microarray carried considerably more signal than noise. The data appeared to be of high quality. However, in a few cases, one or two misleading data points from low-intensity spots may have led to especially high gene-by-gene error variance estimates and thus concealed otherwise significant differences.

The best normalization method and spot pass criteria are highly dependent on cDNA microarray protocol, methodology, experimental experience and analytical resources. As long as normalization method and spot pass criteria are applied uniformly within a dataset, the resulting ratios should be appropriate for analysis by the model described here.

Model

In microarray experiments, the original idea was that, with current technology, spots on a cDNA microarray had a number of confounding pseudolinear terms - whose variation from experiment to experiment could be minimized but not eliminated - which contributed to the intensity measurements observed when a hybridization was scanned. Model parameters under this scheme differ from those used for high-density oligonucleotide microarrays [40]. These terms included the density and size of the cDNA deposition, the correspondingly larger or smaller amount of labeled mRNA hybridized to the microarray, the hybridization conditions and the sequence of the gene [41]. With these assumptions, the post-normalization intensity in one fluorescence channel at a reporter spot may be modeled linearly as

$$\mu \prod_{m=1}^{q} c_m \sum_{l=q+1}^{t} c_l + \epsilon, \qquad (1)$$

where μ is the absolute quantity of mRNA per cell, the c_m are terms for any q multiplicatively confounding factors, the c_l are terms for any q - t linearly confounding factors, and ϵ is an error term accounting for mild pseudorandom departures from linearity and unavoidable small variations in cell-growth conditions. Note that multiplicative error associated with the sample must have been compensated for by an appropriate normalization ([42], reviewed by [43]); if, on the other hand, there are considerable multiplicative factors that are common to a spot, they are accommodated by the c_m above. This model cannot be directly implemented because experimental replication to estimate each confounding term is technically difficult. These difficulties led to a rapid conclusion that the information that can be best obtained from a cDNA microarray experiment is ratio information rather than absolute quantification. This is because the confounding spot terms above will apply equally to two samples competitively hybridized against a single spot on a single microarray, yielding a ratio that correctly represents the ratio of expression levels in the comparison of interest.

When comparing just two samples, ratio measurements are nearly as good as absolute data. However, when more than one genotype or environmental condition or cell developmental stage is examined, ratio measurements rapidly become cumbersome because comparing across numerous states requires a common unit of measurement. Therefore it is of interest to use these ratio measurements within a statistical model to estimate gene-expression levels in a common (if arbitrary) unit, and also to assess the significance of such a difference.

Consistent with the original interpretation of cDNA microarray data that privileged ratio over absolute quantification [41], the observed ratios of intensities, y_{ij}, may be modeled as

$$y_{ij} = \frac{\mu_i \left(\prod_{m=1}^{q} c_m \sum_{l=q+1}^{t} c_l \right) + \epsilon_i}{\mu_j \left(\prod_{m=1}^{q} c_m \sum_{l=q+1}^{t} c_l \right) + \epsilon_j} , \qquad (2)$$

where μ_i and μ_j are the scaled quantities of mRNA in samples i versus j. This expression is true for any terms that are consistent for both mRNA preparations, regardless of linearity, which is fortunate, as there are known nonlinearities in microarray data acquisition [17]. This yields

$$y_{ij} = \frac{\mu_i + \dfrac{\epsilon_i}{\left(\prod_{m=1}^{q} c_m \sum_{l=q+1}^{t} c_l \right)}}{\mu_i + \dfrac{\epsilon_j}{\left(\prod_{m=1}^{q} c_m \sum_{l=q+1}^{t} c_l \right)}} , \qquad (3)$$

Assuming that error terms are composed of many small, unbiased effects, and scaling the error terms σ_i and σ_j by

$$\prod_{m=1}^{q} c_m \sum_{l=q+1}^{t} c_l,$$

a constant for every microarray spot, so that they are distributed with variances specific to each sample, σ_i^2 and σ_j^2, it follows directly that the observed ratio data, z_{ij}, are drawn from

$$\frac{N(\mu_i, \sigma_i^2)}{N(\mu_j, \sigma_j^2)}.$$

This can be true even if the distribution of intensities across spots and arrays is decidedly not Gaussian, because confounding factors which vary linearly or multiplicatively or even interactively across spots or arrays are commonly presumed not to be different between two labeled samples hybridized to a single spot [41]. Note that samples may,

under this formulation, have different variances as well as different means.

The ratio of standard normal distributions is a Cauchy distribution, which has the unpleasant property that it has no moments. The ratio of nonstandard normal distributions is not much better. Fortunately, the infinite tails of the normal distributions that result in this property are not generally observed in real data; in fact, a model that allows negative gene expression levels is not valid. The joint normal distribution may be truncated at a considerable distance from its peak, along an elliptical probability contour within the positive quadrant [44], yielding a ratio z_{ij} distributed approximately as

$$f(z_{ij}|\mu_i,\sigma_i^2,\mu_j^2) = \frac{\sigma_i^2\mu_j + \sigma_j^2\mu_i z}{\sqrt{2\pi}(\sigma_i^2 + \sigma_j^2 z^2)^{\frac{3}{2}}} \; e^{-\frac{(\mu_i - \mu_j z)^2}{2(\sigma_i^2 + \sigma_j^2 z^2)}} \; . \tag{4}$$

This approximation of the density of the ratio of two nonstandard normal distributions is extremely good as long as $\mu_i \gg \sigma_i$ and $\mu_j \gg \sigma_j$, which is certainly true for informative cDNA microarray experiments. It is not symmetrical like a normal distribution, but skewed with a long right-hand tail, like gene expression data. Moreover, as implied above, if microarray data contain no negative expression levels (a natural proposition), it is likely that the distribution of z above is closer to the true distribution of microarray data than is the distribution of the ratio of true normal distributions.

Let us consider how this result may be used. An ideal statistical framework for the analysis of microarray ratio data could be used to analyze microarray data of any experimental design including any number of treatments, genotypes, or developmental states; would be highly robust to missing data; and would yield estimates of the magnitude of expression differences and measures of statistical significance across all treatments, genotypes, and developmental states.

The number of expression nodes, n, is equal to the number of permutations of strain, treatment, and developmental state that are examined. Unless informative prior information about expression levels or error variance is used, the following (minimal) requirements must be met. First, every node of interest must be present in at least one comparison. Second, every node of interest must be connected to every other node of interest by an unbroken chain of comparisons. And third, there must be as many comparisons as there are mean and variance parameters to be estimated.

If separate error variances for each sample are to be estimated as well as means, the last requirement indicates that there must be at least $2n-1$ measurements when each expression node is assumed to have an independent variance, and there must be at least n measurements when each expression node is assumed to have the same error variance.

A few measurements beyond the minimum contribute greatly to the power to detect differences in gene expression and to the ease with which significance of results is ascertained within the Bayesian framework. Figure 1 shows the comparison structure of experiments examined in this paper. The three-dimensional matrix of ratio results from these comparisons, Z, may be constructed, with dimensions i denoting the sample labeled with one fluorophore, j denoting the sample labeled with another, and k denoting the replicate ordinate of that particular comparison. Then, for any continuous structure of comparisons among the nodes of interest, the likelihood density for the parameters μ_l and σ_l^2, $1 \le l \le n$, is, by Bayes' rule,

$$h(\mu_1,\sigma_i^2...\mu_l,\sigma_l^2|Z) = \frac{\left(\prod_{\substack{i=1\\j=1\\k}}^{n} f(z_{ijk}|\mu_i,\sigma_i^2,\mu_j,\sigma_j^2)\right) g(\mu_i,\sigma_i^2,\mu_j,\sigma_j^2)}{\int_{M_l\Sigma_l}\left(\left(\prod_{\substack{i=1\\j=1\\k}}^{n} f(z_{ijk}|\mu_i,\sigma_i^2,\mu_j,\sigma_j^2)\right) g(\mu_i,\sigma_i^2,\mu_j,\sigma_j^2)\right) d\sigma_l^2 \, d\mu_l}, \tag{5}$$

where $g(\mu_i, \sigma_i^2, \mu_j, \sigma_j^2)$ is the prior distribution of the parameters, and where the probability $f(z_{ijk})$ of empty elements in the data matrix Z is evaluated as 1.

Any genes for which $\sum_{i,j,k} I(z_{ijk} \ne 1) < \upsilon$, where I is an indicator variable and $\upsilon = n$ when variances are common among all samples within a gene and $\upsilon = 2n-1$ when variances are not, yield improper posteriors unless an informative prior is used. These few genes are not analyzed here, nor are genes whose matrix Z is reducible. Note that diagonal entries in Z are control, self-self hybridizations, in which the same expression node has been labeled with two different dyes. In this framework, these controls can and should be included in the dataset if performed with all sources of experimental error included in their preparation. They contribute to estimation of the error variance of genes.

Appropriate informative priors for the variance of microarray data are currently under investigation by a number of groups [34-36]. An informative prior must be clearly justified in order to prevent inappropriate conclusions of statistical significance. In this paper, a noninformative prior distribution, uniform across positive real numbers, has been used for both the expression levels and for their variance. In theory, we use a uniform prior for the variance, bounded from 0 to 100. In practice, the upper limit, beyond 20 or so, makes no difference, as such high values are very improbable and are never sampled by the chain in the datasets analyzed here. This uniform prior gives the microarray data itself the greatest impact on the inferred means and variances, and implies that

credible intervals constructed are close to those that would be found by maximum likelihood if analytical integration of the full multidimensional parameter surface were feasible. A frequently used 'noninformative' prior such as σ^1 [45] is in this case not desirable, because, in practice, the most likely variances observed are so small that this prior has a considerable impact on the posterior distribution.

Fortunately, we may use the constant denominator of the Bayes' rule formulation (Equation 5) to assert that

$$h(\mu_1, \sigma_1^2 ... \mu_l, \sigma_l^2 | \mathbf{Z}) \propto \left(\prod_{\substack{i=1 \\ j=1 \\ k}}^{n} f(z_{ijk} | \mu_i, \sigma_i^2, \mu_j, \sigma_j^2) \right) g(\mu_i, \sigma_i^2, \mu_j, \sigma_j^2).$$

(6)

This may be used to construct a Markov chain whose stationary distribution is the posterior distribution of the parameters given the data. A vector $\overrightarrow{\sigma^2}$ and a vector $\overrightarrow{\mu}$ of initial expression levels, $[\mu_1, ..., \mu_n]$ is chosen such that $\frac{1}{n}\sum_1^n\mu_l = 1$ at step $t = 0$, and subsequent values in the chain are determined iteratively by choosing successive proposed values according to an acceptance rule.

Our proposed values are constructed in two separate steps. First, two of the n gene-expression level parameters from $\overrightarrow{\mu}$ are chosen at random. A step size is drawn at random from a triangular distribution centered at zero with range $[-\Delta_\mu, +\Delta_\mu]$. The first of the two chosen parameters is incremented by the chosen step size, and the second is decremented by the same quantity, so that $\frac{1}{n}\sum_1^n\mu_l' = 1$ is maintained, where the apostrophe indicates a proposed parameter value. In the next iteration, the variance parameters, σ_l^2 are incremented by an amount drawn at random from a triangular distribution with range $[-\Delta_{\sigma^2}, +\Delta_{\sigma^2}]$ to form $\sigma_l^{2'}$. Because these operations have probabilities of transitions from the current state to the proposed state equal to the probabilities that the converse transitions would have, this proposal scheme satisfies Hastings [46] and can be implemented in the Metropolis [47] algorithm. Thus the conjecture is accepted for the next state of the Markov chain if

$$RANDOM(0,1) < \frac{\left(\prod_{\substack{i=1 \\ j=1 \\ k}}^{n} f(z_{ijk} | \mu_i', \sigma_i^{2'}, \mu_j', \sigma_j^{2'}) \right) g(\mu_i', \sigma_i^{2'}, \mu_j', \sigma_j^{2'})}{\left(\prod_{\substack{i=1 \\ j=1 \\ k}}^{n} f(z_{ijk} | \mu_i, \sigma_i^2, \mu_j, \sigma_j^2) \right) g(\mu_i, \sigma_i^2, \mu_j, \sigma_j^2)},$$

(7)

Otherwise the original state is retained for the next iteration of the Markov chain.

These steps are repeated over many generations in order to 'burn in' the chain, so that it converges from the initial

parameter settings to a stationary distribution. Subsequently, states are sampled from the chain at regular intervals to build a posterior distribution for each parameter, integrated across the probable states of all other parameters. An easy-to-use stand-alone software program entitled BAGEL, which implements this Bayesian analysis of gene expression levels on MacOS or Windows platforms, is available on the web with an online manual [48]. It accepts tab-delimited text files of ratio data as input.

To decrease the number of parameters that must be estimated, we estimate a single variance parameter for each gene across all expression nodes, which is equivalent to the assumption that σ_i and σ_j from the same distribution for every sample. On larger, more highly replicated datasets, where all parameters could be estimated, constraining the variance has not led to substantially different results (J.P.T., unpublished data and J.M. Ranz, personal communication). All analyses in this paper were performed with 20,000 generations of burn-in, followed by 200,000 generations during which the chain was sampled every 20 generations to construct the posterior distribution. Runs using multiple starting vectors $\overrightarrow{\mu}$ and $\overrightarrow{\sigma^2}$ were performed and always converged to the same, unimodal, posteriors, indicating that this is a well-behaved multidimensional likelihood surface (see Figure 2a). Results reported here were the outcomes of Markov chains started with the elements of $\overrightarrow{\mu}$ all equal to one, and started with the elements of $\overrightarrow{\sigma^2}$ all equal to 0.03. Step sizes Δ_μ and Δ_{σ^2} were tuned for each gene so that acceptance ratios for each parameter update were in the efficient and well-mixed range (0.15, 0.50) [49]. If acceptance ratios for either parameter jump were less than 0.15 or greater than 0.5, the chain was run again with a better-tuned jump size, until acceptable ratios for both parameters were obtained. In this way, there is no alteration of the jump size during any run. There is only the evaluation of pilot Markov chains to optimize jump size.

Output from the BAGEL software is in the form of a tab-delimited text file with one header row. Each row thereafter displays the results for a single gene, including columns with the estimate of expression level for each sample (the median of the posterior distribution); the additions and subtractions to make 95% upper and lower bounds on that estimate; the stationary acceptance rates for the Monte Carlo steps for that gene; and a column that reads 'TRUE' when those rates are acceptable. Further columns contain the posterior probabilities for whether that gene's expression level in each expression node is greater, or lesser, than that gene's expression level in each other expression node.

Additional data files

Estimates and credible intervals for expression levels of all genes assayed in these experiments in all conditions are available with the online version of this manuscript as

tab-delimited text output files, with columns of data as described in the methods section. The files are entitled 'EtOH.txt', 'SwiSnfMin.txt' and 'SwiSnfRich.txt', and 'Zinc.txt'.

Acknowledgements

References
1. 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.
2. Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, Golub TR: **Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation.** *Proc Natl Acad Sci USA* 1999, **96:**2907-2912.
3. Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM: **Systematic determination of genetic network architecture.** *Nat Genet* 1999, **22:**281-285.
4. DeRisi JL, Iyer VR, Brown PO: **Exploring the metabolic and genetic control of gene expression on a genomic scale.** *Science* 1997, **278:**680-686.
5. Chen Y, Dougherty ER, Bittner ML: **Ratio-based decisions and the quantitative analysis of cDNA microarray images.** *J Biomed Optics* 1997, **2:**364-374.
6. Ideker T, Thorsson V, Siegel A, Hood L: **Testing for differentially-expressed genes by maximum-likelihood analysis of microarray data.** *J Comput Biol* 2000, **7:**805-817.
7. Kerr MK, Martin M, Churchill GA: **Analysis of variance for gene expression microarray data.** *J Comput Biol* 2000, **7:**819-837.
8. Wolfinger RD, Gibson G, Wolfinger ED, Bennett L, Hamadeh H, Bushel P, Afshari C, Paules RS: **Assessing gene significance from cDNA microarray expression data via mixed models.** *J Comput Biol* 2001, **8:**625-637.
9. Jin W, Riley RM, Wolfinger RD, White KP, Passador-Gurgel G, Gibson G: **The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*.** *Nat Genet* 2001, **29:**389-395.
10. Kerr MK, Churchill GA: **Statistical design and the analysis of gene expression microarray data.** *Genet Res* 2001, **77:**123-128.
11. Alexandre H, Ansanay-Galeote V, Dequin S, Blondin B: **Global gene expression during short-term ethanol stress in *Saccharomyces cerevisiae*.** *FEBS Lett* 2001, **498:**98-103.
12. Sudarsanam P, Iyer VR, Brown PO, Winston F: **Whole-genome expression analysis of *snf/swi* mutants of *Saccharomyces cerevisiae*.** *Proc Natl Acad Sci USA* 2000, **97:**3364-3369.
13. Lyons TJ, Gasch AP, Gaither LA, Botstein D, Brown P, Eide D: **Genome-wide characterization of the Zap1p zinc-responsive regulon in yeast.** *Proc Natl Acad Sci USA* 2000, **97:**7957-7962.
14. 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.
15. Shamji A, Kuruvilla FG, Schreiber S: **Partitioning the transcriptional program induced by rapamycin among the effectors of the Tor proteins.** *Curr Biol* 2000, **10:**1574-1580.
16. Furlong EE, Andersen EC, Null B, White KP, Scott MP: **Patterns of gene expression during *Drosophila* mesoderm development.** *Science* 2001, **293:**1629-1633.
17. Ramdas L, Coombes KR, Baggerly K, Abruzzo L, Highsmith WE, Krogmann T, Hamilton S, Zhang W: **Sources of nonlinearity in cDNA microarray expression measurements.** *Genome Biol* 2001, **2:**research0047.1-0047.7.
18. Yang YH, Speed T: **Design issues for cDNA microarray experiments.** *Nat Rev Genet* 2002, **3:**579-588.
19. Treger JM, Schmitt AP, Simon JR, McEntee K: **Transcriptional factor mutations reveal regulatory complexities of heat shock and newly identified stress genes in *Saccharomyces cerevisiae*.** *J Biol Chem* 1998, **273:**26875-26879.
20. Mansure JJC, Panek AD, Crowe LM, Crowe JH: **Trehalose inhibits ethanol effects on intact yeast cells and liposomes.** *Biochim Biophys Acta* 1994, **1191:**309-316.
21. Mansure J, Souza R, Panek A: **Trehalose metabolism in *Saccharomyces cerevisiae*.** *Biotechnol Lett* 1997, **19:**1201-1203.
22. Thevelein JM: **Regulation of trehalose mobilization in fungi.** *Microbiol Rev* 1984, **48:**42-59.
23. Türkel S, Liao X-B, Farabaugh PJ: ***GCR1*-dependent transcriptional activation of yeast retrotransposon Ty2-917.** *Yeast* 1997, **13:**917-930.
24. Herrero P, Flores L, de la Cera T, Moreno F: **Functional characterization of transcriptional regulatory elements in the upstream region of the yeast GLK1 gene.** *Biochem J* 1999, **343:**319-325.
25. Gibson G: **Epistasis and pleiotropy as natural properties of transcriptional regulation.** *Theor Popul Biol* 1996, **49:**58-89.
26. Gibson G, van Helden S: **Is function of the *Drosophila* homeotic gene *Ultrabithorax* canalized?** *Genetics* 1997, **147:**1155-1168.
27. Doebley J, Lukens L: **Transcriptional regulators and the evolution of plant form.** *Plant Cell* 1998, **10:**1075-1082.
28. Hartl DL, Dykhuisen DE, Dean AM: **The limits of adaptation: the evolution of selective neutrality.** *Genetics* 1985, **111:**655-674.
29. Velculescu VE, Zhang L, Vogelstein B, Kinzler K: **Serial analysis of gene expression.** *Science* 1995, **270:**484-487.
30. Velculescu VE, Zhang L, Zhou W, Vogelstein J, Basrai M, Bassett DE, Hieter P, Vogelstein B, Kinzler KW: **Characterization of the yeast transcriptome.** *Cell* 1997, **88:**243-251.
31. Stollberg J, Urschitz J, Urban Z, Boyd CD: **A quantitative evaluation of SAGE.** *Genome Res* 2000, **10:**1241-1248.
32. Zhao H, Eide DJ: **Zap1p, a metalloregulatory protein involved in zinc-responsive transcriptional regulation in *Saccharomyces cerevisiae*.** *Mol Cell Biol* 1997, **17:**5044-5052.
33. Bernstein BE, Tong JK, Schreiber SL: **Genomewide studies of histone deacetylase function in yeast.** *Proc Natl Acad Sci USA* 2000, **97:**13708-13713.
34. Tseng GC, Oh M-K, Rohlin L, Liao JC, Wong WH: **Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects.** *Nucleic Acids Res* 2001, **29:**2549-2557.
35. Baldi P, Long A: **A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes.** *Bioinformatics* 2001, **17:**509-519.
36. Newton MA, Kendziorski CM, Richmond CS, Blattner FR, Tsui KW: **On differential variability of expression ratios: improving statistical inference about gene expression changes from microarray data.** *J Comput Biol* 2001, **8:**37-52.
37. Barash Y, Friedman N: **Context-specific Bayesian clustering for gene expression data.** *J Comput Biol* 2002, **9:**169-191.
38. Ibrahim J, Chen M, Gray R: **Bayesian models for gene expression with DNA microarray data.** *J Am Stat Assoc* 2002, **97:**88-89.
39. Lee M-LT, Kuo FC, Whitmore GA, Sklar J: **Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations.** *Proc Natl Acad Sci USA* 2000, **97:**9834-9839.
40. Theilhaber J, Bushnell S, Jackson A, Fuchs R: **Bayesian estimation of fold-changes in the analysis of gene expression: the PFOLD algorithm.** *J Comput Biol* 2001, **8:**585-614.
41. Eisen MB, Brown PO: **DNA arrays for analysis of gene expression.** *Methods Enzymol* 1999, **303:**179-205.
42. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed T: **Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation.** *Nucleic Acids Res* 2002, **30:**e15.
43. Quackenbush J: **Compuatational analysis of microarray data.** *Nat Rev Genet* 2001, **2:**418-427.
44. Fieller EC: **The distribution of the index in a normal bivariate population.** *Biometrika* 1932, **24:**428-440.
45. Gelman A, Carlin JB, Stern HS, Rubin DB: *Bayesian Data Analysis*. Boca Raton: Chapman & Hall/CRC; 1995.
46. Hastings WK: **Monte Carlo sampling methods using Markov chains and their applications.** *Biometrika* 1970, **57:**97-109.
47. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E: **Equation of state calculations by fast computing machines.** *J Chem Phys* 1953, **21:**1087-1092.
48. **Hartl Lab Publications - Software** [http://www.oeb.harvard.edu/hartl/lab/publications/software.html]
49. Gelman A, Roberts GO, Gilks WR: **Efficient Metropolis jumping rules.** In *Bayesian Statistics 5*. Edited by Bernardo JM, Berger JO, Dawid AP, Smith AFM. Oxford: Oxford University Press; 1996: 599-607.