

# Positive selection for elevated gene expression noise in yeast

Zihua Zhang, Wenfeng Qian and Jianzhi Zhang\*

Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA

\* Correspondence: Department of Ecology and Evolutionary Biology, University of Michigan, 1075 Natural Science Building, 830 North University Avenue, Ann Arbor, MI 48109, USA. Tel.: +1 734 763 0527; Fax: +1 734 763 0544; E-mail: jianzhi@umich.edu

Received 18.11.08; accepted 9.7.09

**It is well known that the expression noise is lessened by natural selection for genes that are important for cell growth or are sensitive to dosage. In theory, expression noise can also be elevated by natural selection when noisy gene expression is advantageous. Here we analyze yeast genome-wide gene expression noise data and show that plasma-membrane transporters show significantly elevated expression noise after controlling all confounding factors. We propose a model that explains why and under what conditions elevated expression noise may be beneficial and subject to positive selection. Our model predicts and the simulation confirms that, under certain conditions, expression noise also increases the evolvability of gene expression by promoting the fixation of favorable expression level-altering mutations. Indeed, yeast genes with higher noise show greater between-strain and between-species divergences in expression, even when all confounding factors are excluded. Together, our theoretical model and empirical results suggest that, for yeast genes such as plasma-membrane transporters, elevated expression noise is advantageous, is subject to positive selection, and is a facilitator of adaptive gene expression evolution.**

*Molecular Systems Biology* 5: 299; published online 18 August 2009; doi:10.1038/msb.2009.58

*Subject Categories:* functional genomics; simulation and data analysis

*Keywords:* adaptive evolution; expression noise; plasma-membrane transporters; positive selection; yeast

This is an open-access article distributed under the terms of the Creative Commons Attribution Licence, which permits distribution and reproduction in any medium, provided the original author and source are credited. This licence does not permit commercial exploitation or the creation of derivative works without specific permission.

## Introduction

Gene expression, as other biological processes, is subject to noise (Schrodinger, 1944), which is defined as the stochastic variation in the expression level of a gene among isogenic cells under the same condition. Here and elsewhere in the paper, expression level refers to the level of the protein product of the gene, as expression noise is usually measured at the level of protein. Gene expression noise has been measured in prokaryotes (Elowitz *et al.*, 2002; Ozbudak *et al.*, 2002; Rosenfeld *et al.*, 2005), unicellular eukaryotes (Blake *et al.*, 2003; Raser and O'Shea, 2004), and mammalian cells (Ramsey *et al.*, 2006). These and other studies showed that the level of expression noise varies substantially among genes, is determined genetically, and is selectable (Blake *et al.*, 2006; Newman *et al.*, 2006; Maheshri and O'Shea, 2007; Ansel *et al.*, 2008). Expression noise has both intrinsic and extrinsic sources (Orphanides and Reinberg, 2002; Rao *et al.*, 2002; Blake *et al.*, 2003; Kaern *et al.*, 2005; Raser and O'Shea, 2004, 2005; Bar-Even *et al.*, 2006; Newman *et al.*, 2006; Volfson *et al.*, 2006). Stochastic events in gene expression, including those in transcription initiation, mRNA degradation, translation initia-

tion, and protein degradation, generate intrinsic noise (Raser and O'Shea, 2005). Differences between cells, either in local environment or in the concentration or activity of any factor influencing gene expression, generate extrinsic noise (Raser and O'Shea, 2005). We focus on intrinsic noise in this study because only intrinsic noise is an intrinsic property of a gene.

Gene expression noise is often considered a two-edged sword. On one hand, the noise could be deleterious because it ruins cellular homeostasis in metabolism and developmental programs, affects precise controls of biochemical processes in cells, and breaks the stoichiometric balances among members of protein complexes (Fraser *et al.*, 2004; Batada *et al.*, 2006; Lehner, 2008). Increased gene expression noise has been reported to result in disease (Cook *et al.*, 1998; Kemkemer *et al.*, 2002; Bahar *et al.*, 2006). Several studies showed direct and indirect evidence for lessened expression noise of genes that are important to cell growth or sensitive to dosage (Fraser *et al.*, 2004; Newman *et al.*, 2006; Batada and Hurst, 2007; Lehner, 2008). Furthermore, various molecular mechanisms and regulatory network structures (e.g. negative feedbacks) are found to attenuate expression noise (Becskei and Serrano, 2000; Pedraza and van Oudenaarden, 2005). On the other

hand, several benefits of expression noise have been suggested. In particular, it has been argued that stochastic noise is essential in cell-fate determination (Colman-Lerner *et al*, 2005; Kaern *et al*, 2005; Losick and Desplan, 2008) and thus is important in the development of multicellular organisms. In unicellular organisms, it has been shown both theoretically and experimentally that stochastic switching of expression level or high expression noise could be beneficial in the face of fluctuating environments or under acute environmental stresses (Thattai and van Oudenaarden, 2004; Blake *et al*, 2006; Acar *et al*, 2008). It is thus plausible that a certain fraction of genes in a genome have elevated expression noise driven by positive selection. Indeed, a study of 43 yeast genes showed that stress-related genes are noisier than the rest of the genes (Bar-Even *et al*, 2006). A subsequent characterization of expression noise of thousands of yeast genes identified several Gene Ontology (GO) categories with significantly elevated noise, compared with the genomic average (Newman *et al*, 2006). These GO categories include amino-acid biosynthesis, oxidative phosphorylation, heat shock, and stress response. Although it is tempting to suggest that the higher-than-average noise of these genes is a result of positive selection (Newman *et al*, 2006; Lopez-Maury *et al*, 2008; Raj and van Oudenaarden, 2008), one cannot exclude the possibility that 'it is a mere result of lack of constraint on the variability in expression of such genes', as has been previously argued (Bar-Even *et al*, 2006). The genome-wide analysis (Newman *et al*, 2006) also suffered from a lack of statistical correction for multiple testing when many GOs were evaluated. Hence, it is not clear whether there are genuinely noisier-than-average GOs.

In this study, we test the hypothesis of positive selection for elevated gene expression noise by controlling multiple factors potentially associated with the relaxation of purifying selection. We identify plasma-membrane transporters as the only group in yeast that shows significantly greater noise than the neutral expectation. We propose a model explaining why and under what conditions high noise may be beneficial. We further show theoretically and empirically that high noise facilitates adaptive gene expression evolution.

## Results

### Plasma-membrane transporters are significantly noisier than the neutral expectation

Newman *et al* (2006) measured the expression noise for over 2000 genes of the budding yeast *Saccharomyces cerevisiae* in rich (YPD) medium. As they controlled for several extrinsic factors, their noise estimates can be approximately regarded as intrinsic noise (Newman *et al*, 2006). The noise level is commonly measured by the coefficient of variation (CV), which is the s.d. of the expression level divided by the mean. Newman *et al* (2006) found a genome-wide pattern of lower CV for genes with higher mean expression (also see Bar-Even *et al* (2006)). To control the influence of mean expression level on noise and allow among-gene comparison of noise levels, they used a new measure of noise named DM. For a given gene, DM is the difference of its CV from the median CV of those genes that have a similar mean expression as the focal gene (Newman *et al*, 2006).

As there is good evidence that the expression noise is lessened by natural selection for genes important for cell growth (Fraser *et al*, 2004; Batada and Hurst, 2007; Lehner, 2008), we need to control for the 'importance' of a gene when evaluating whether it is noisier than the expectation. The importance of a gene in yeast cell growth can be measured by the reduction in growth rate (i.e. fitness) in YPD upon deletion of the gene from the genome. Fortunately, such data exist for virtually every yeast gene (Giaever *et al*, 2002; Steinmetz *et al*, 2002). We separate all genes with expression noise data into 21 bins of different importance levels, with the fitness of the deletion strains being in the ranges of  $<0.05$ ,  $0.05-0.10$ ,  $0.10-0.15$ , ...,  $0.95-1.00$ , and  $>1.00$ , respectively. The last bin is not empty because the fitness value of a gene-deletion strain was originally measured relative to the mean of all viable gene-deletion strains, rather than to the wild-type strain (Steinmetz *et al*, 2002). To test whether the noise level of genes belonging to a given GO category exceeds the expectation, we randomly draw genes (with replacement) from the genome-wide expression noise data to form a gene set that has the same number of genes in each of the 21 bins as the focal GO has. We repeat this process 20 000 times and calculate the proportion of times when the mean noise level of the GO is lower than that of the randomly constructed gene set. If this probability ( $P$ -value in Table I) is lower than 5%, we regard the GO to be significantly noisier than expected. As we examine numerous GO categories, we further control for multiple testing using a 5% false discovery rate (Storey and Tibshirani, 2003). That is, only GOs with a  $Q$ -value  $<0.05$  are considered as truly significant. To ensure that there is sufficient statistical power to detect elevated noise of a GO, only those GOs with at least 30 genes were examined.

GO categories are organized into three groups: biological process, cellular component, and molecular function (Ashburner *et al*, 2000). The three groups characterize different aspects of a gene's function and are thus examined separately in our analysis. We found that in terms of biological process, 18 GOs related to metabolism and transport show significantly higher-than-expected noise (Table I). In terms of cellular component, five GOs related to organelles (particularly mitochondrion) have high noise (Table I). In terms of molecular function, four GOs related to catalytic activity and transporter activity have high noise (Table I). The high expression noise of proteins localized to the mitochondrion (and other low copy-number organelles) was noted before and was thought to be caused by unequal partitioning of mitochondria (and other organelles) during mitosis (Newman *et al*, 2006). Further evidence for this explanation came from the experiment showing that the same protein expressed from the same promoter and locus is noisier when targeted to low copy-number organelles than when localized to the cytosol (Newman *et al*, 2006). Thus, the high noise of mitochondrial proteins is unlikely the result of positive selection for elevated noise. Further, the high noise of enzymes is probably due to their special insensitivity to dosage, rather than positive selection for high noise, because it is well known that, in a metabolic pathway, even a considerable change in the concentration of an enzyme has a minimal effect on the flux of the pathway (Kacser and Burns, 1981). This phenomenon arises from the kinetic connection through the shared

**Table I** GO categories with significantly greater-than-expected expression noise after the control for gene importance

	GO ID	GO term	# of genes	P-value	Q-value
<i>All genes</i>					
BP	GO0006732	Coenzyme metabolic process	57	$<5.00 \times 10^{-5}$	$<2.54 \times 10^{-3}$
	GO0015980	Energy derivation by oxidation of organic compounds	112	$<5.00 \times 10^{-5}$	$<2.54 \times 10^{-3}$
	GO0044262	Cellular carbohydrate metabolic process	81	$<5.00 \times 10^{-5}$	$<2.54 \times 10^{-3}$
	GO0051186	Co-factor metabolic process	71	$<5.00 \times 10^{-5}$	$<2.54 \times 10^{-3}$
	GO0006811	Ion transport	45	$1.00 \times 10^{-4}$	$2.54 \times 10^{-3}$
	GO0006807	Nitrogen compound metabolic process	116	$3.00 \times 10^{-4}$	$5.08 \times 10^{-3}$
	GO0006812	Cation transport	37	$3.00 \times 10^{-4}$	$5.08 \times 10^{-3}$
	GO0032787	Monocarboxylic acid metabolic process	48	$3.50 \times 10^{-4}$	$5.47 \times 10^{-3}$
	GO0006519	Amino acid and derivative metabolic process	104	$4.00 \times 10^{-4}$	$5.80 \times 10^{-3}$
	GO0009060	Aerobic respiration	32	$5.00 \times 10^{-4}$	$6.77 \times 10^{-3}$
	GO0009117	Nucleotide metabolic process	56	$5.50 \times 10^{-4}$	$6.98 \times 10^{-3}$
	GO0006520	Amino-acid metabolic process	99	$6.50 \times 10^{-4}$	$7.76 \times 10^{-3}$
	GO0009056	Catabolic process	173	$7.50 \times 10^{-4}$	$8.46 \times 10^{-3}$
	GO0016051	Carbohydrate biosynthetic process	32	$8.00 \times 10^{-4}$	$8.55 \times 10^{-3}$
	GO0044248	Cellular catabolic process	168	$8.50 \times 10^{-4}$	$8.63 \times 10^{-3}$
	GO0016310	Phosphorylation	59	$2.15 \times 10^{-3}$	$2.08 \times 10^{-2}$
	GO0008152	Metabolic process	1227	$2.55 \times 10^{-3}$	$2.35 \times 10^{-2}$
	GO0044249	Cellular biosynthetic process	402	$4.30 \times 10^{-3}$	$3.80 \times 10^{-2}$
	GO0044237	Cellular metabolic process	1197	$4.60 \times 10^{-3}$	$3.82 \times 10^{-2}$
	GO0044271	Nitrogen compound biosynthetic process	66	$4.70 \times 10^{-3}$	$3.82 \times 10^{-2}$
CC	GO0005739	Mitochondrion	388	$<5.00 \times 10^{-5}$	$<4.21 \times 10^{-4}$
	GO0005740	Mitochondrial envelope	100	$<5.00 \times 10^{-5}$	$<4.21 \times 10^{-4}$
	GO0005743	Mitochondrial inner membrane	49	$<5.00 \times 10^{-5}$	$<4.21 \times 10^{-4}$
	GO0005759	Mitochondrial matrix	92	$<5.00 \times 10^{-5}$	$<4.21 \times 10^{-4}$
	GO0009277	Chitin- and $\beta$ -glucan-containing cell wall	36	$<5.00 \times 10^{-5}$	$<4.21 \times 10^{-4}$
	GO0019866	Organelle inner membrane	52	$<5.00 \times 10^{-5}$	$<4.21 \times 10^{-4}$
	GO0005618	Cell wall	36	$5.00 \times 10^{-5}$	$4.21 \times 10^{-4}$
MF	GO0003824	Catalytic activity	794	$<5.00 \times 10^{-5}$	$<1.05 \times 10^{-3}$
	GO0015077	Monovalent inorganic cation transporter activity	33	$<5.00 \times 10^{-5}$	$<1.05 \times 10^{-3}$
	GO0015078	Hydrogen ion transporter activity	32	$<5.00 \times 10^{-5}$	$<1.05 \times 10^{-3}$
	GO0005215	Transporter activity	159	$1.00 \times 10^{-4}$	$1.05 \times 10^{-3}$
	GO0015075	Ion transporter activity	68	$5.50 \times 10^{-4}$	$4.62 \times 10^{-3}$
	GO0008324	Cation transporter activity	61	$7.50 \times 10^{-4}$	$5.25 \times 10^{-3}$
	GO0016829	Lyase activity	35	$6.00 \times 10^{-3}$	$3.60 \times 10^{-2}$
<i>After excluding mitochondrial proteins and enzymes</i>					
CC	GO0005886	Plasma membrane	56	$1.00 \times 10^{-4}$	$3.80 \times 10^{-3}$
MF	GO0005215	Transporter activity	73	$7.00 \times 10^{-4}$	$5.60 \times 10^{-3}$

BP, biological process; CC, cellular component; MF, molecular function.

P-values were calculated by the randomization test described in the main text. When none of the 20 000 random samples show greater mean noise than the observed mean noise of a GO category, we consider  $<1$  random sample to have a greater noise than the observed.

substrates/products of adjacent biochemical reactions such that the effect of changing the catalytic activity in one reaction tends to be buffered by the response to this of the other reactions (Kacser and Burns, 1981). Thus, to be conservative, we removed all mitochondrial proteins and all enzymes, and re-tested each GO. This time, we identified plasma membrane as the only cellular component GO category and transporter activity as the only molecular function GO category that show significantly higher-than-expected noise (Table I). No biological process GO category is significantly noisier than expected. Our results are robust to the variation of the number of bins used (11–26) in controlling the effect of gene importance on noise (Supplementary Tables S1 and S2).

Haploinsufficient genes are sensitive to expression noise and should have reduced noise, as has been shown (Cook *et al*, 1998; Batada and Hurst, 2007; Lehner, 2008). To test whether the high noise of plasma-membrane proteins and transporters is simply because they are less likely to be haploinsufficient than other genes in the genome, we further removed all

haploinsufficient genes (Deutschbauer *et al*, 2005) from the genome and re-tested every GO. We found that both ‘plasma membrane’ and ‘transporter activity’ GOs remain significantly noisier than expected ( $Q=0.0038$  and  $0.0056$ , respectively) and that the Q-values are similar to those obtained without the control for haploinsufficient genes (Table I). This is probably due to the paucity of haploinsufficient genes in the yeast genome (Deutschbauer *et al*, 2005). For this reason, haploinsufficient genes are no longer controlled for in subsequent analysis unless otherwise noted. Previous direct and indirect evidence suggested that components of stable protein complexes are also sensitive to dosage and thus have reduced noise (Fraser *et al*, 2004; Lehner, 2008). We found that after the control for gene importance, protein complex members no longer have lower noise than other proteins ( $P=0.76$ , Mann–Whitney  $U$  test). Thus, there is no need to further control for protein complex membership in our analysis. Overexpressions of certain genes are detrimental; these genes could have reduced expression noise as well (Lehner, 2008). However, we

found no significant correlation between gene expression noise (DM) and the fitness of gene overexpression strains (Sopko *et al*, 2006) (Spearman's rank correlation  $\rho=0.03$ ,  $P=0.15$ ). Thus, there is no need to consider the potential selection on noise due to gene overexpression. Taken together, the high expression noise of plasma-membrane genes and transporter genes cannot be explained by relaxation of purifying selection because all known factors that could potentially lead to the relaxation of purifying selection on noise have been excluded; positive selection for elevated noise remains the most plausible explanation of their higher-than-expected noise.

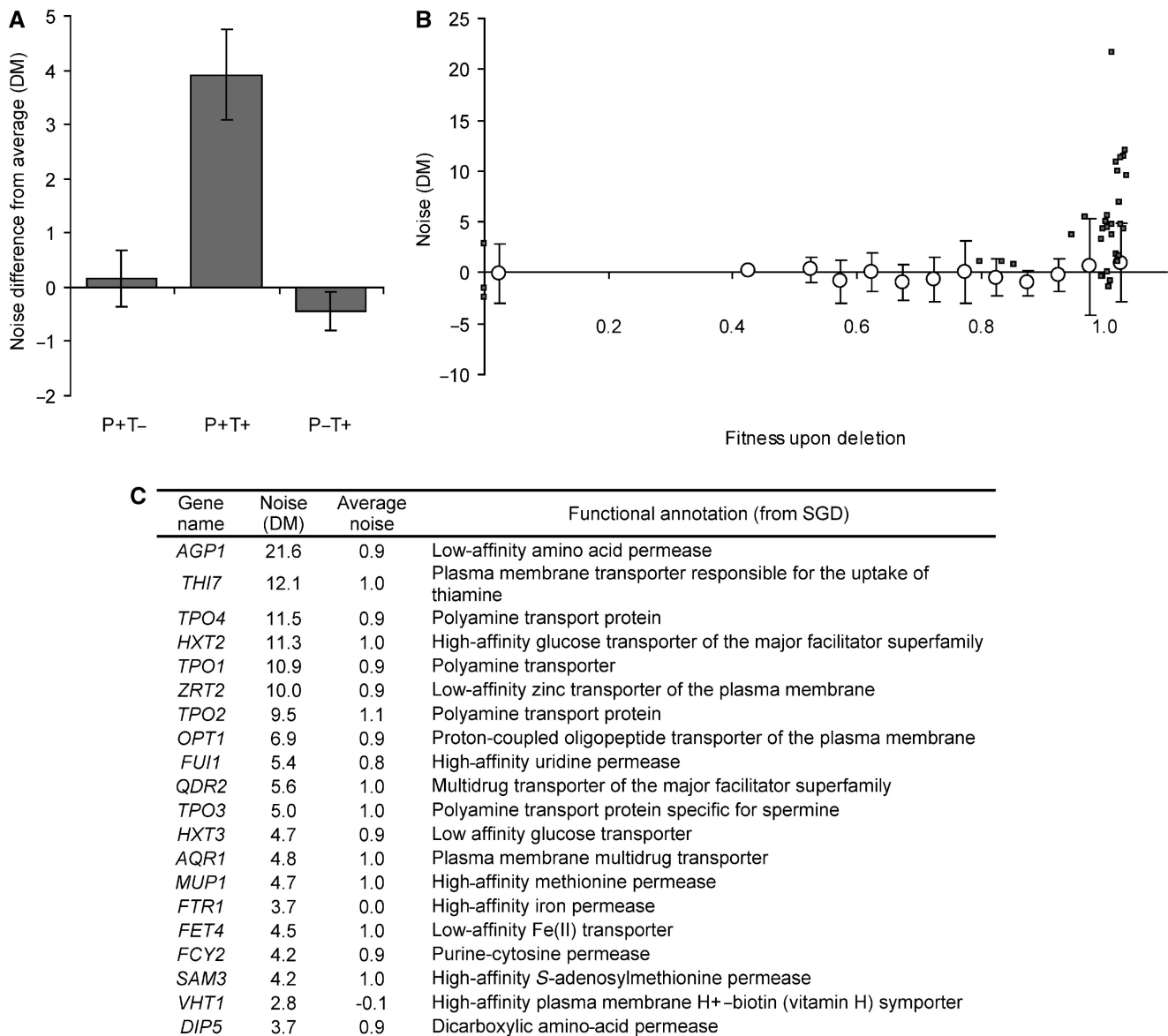
We suspect that the significant results from 'plasma membrane' and 'transporter activity' GOs are because of the high noise of plasma-membrane transporters. Indeed, plasma-membrane transporters are significantly noisier than expected after the control for gene importance and the removal of enzymes and mitochondrial proteins ( $P=3.3 \times 10^{-6}$ ; two-tail Z-test), whereas plasma-membrane proteins that are non-transporters ( $P=0.77$ ) and transporters that are not localized to the plasma membrane ( $P=0.21$ ) are not significantly different from the expectation (Figure 1A). A careful examination shows that the majority of plasma-membrane transporters (79%) belong to the last bin of gene importance (i.e. fitness of the gene-deletion strain  $>1.00$ ) (Figure 1B). For this bin, the genomic average noise level is  $DM=0.87 \pm 0.16$ , only slightly, although significantly, greater than the mean noise ( $-0.10 \pm 0.18$ ) of the first bin (i.e. fitness  $<0.05$ ), suggesting that the effect of negative selection in reducing the expression noise of important genes is overall relatively small (Figure 1B). By contrast, the mean noise of the plasma-membrane transporters in the last bin is  $DM=5.62 \pm 1.00$ , suggesting that the effect of positive selection in elevating expression noise can be substantial (Figure 1B). Again, the above comparison is based on the dataset after the removal of enzymes and mitochondrial proteins. Figure 1C lists the 20 noisiest plasma-membrane transporters. These proteins transport a diverse array of chemicals, such as amino acids, glucose, ions, thiamine, polyamine, oligopeptides, and nucleotides, across the cell membrane. They are involved in the uptake of nutrients and ions, excretion of end products of metabolism and deleterious substances, and communication between cells and the environment. We also examined the yeast expression noise data obtained under the minimal (SD) medium (Newman *et al*, 2006) and confirmed that plasma-membrane transporters is the only group with significantly greater noise than expected after all the controls (i.e. gene importance, enzymes, and mitochondrial proteins) (Supplementary Table S3). We also confirmed that this result is robust to the variation of the number of bins used (11–26) in controlling the effect of gene importance on noise (Supplementary Tables S3–S5).

### High noise can be beneficial when the mean expression level is suboptimal

Why would high noise be beneficial to plasma-membrane transporters? It is likely that the optimal expression level of each transporter depends on environmental factors such as the nutrients available to the cell. The underexpression of a

transporter may limit the nutrient uptake rate and hence limit the cell's Darwinian fitness. However, overexpression of a transporter could also be disadvantageous for two reasons. First, overexpression has a fitness cost due to the waste of energy in transcription and translation (Wagner, 2005; Stoebel *et al*, 2008). Second and more importantly, presence of unwanted transporters could reduce the metabolic efficiency and hence the fitness. For example, imagine that two carbon sources  $C_1$  (e.g. maltose) and  $C_2$  (e.g. lactose) are both present in the medium, but  $C_1$  is energetically more efficient than  $C_2$  for the cell to use. If the total number of carbon source molecules that the cell can catabolize per unit time is limited, it would be better for the cell to use  $C_1$  rather than  $C_2$ . Thus, an overexpression of the transporter for  $C_2$  will reduce the number of carbon source molecules catabolized by the cell per unit time and thus will be deleterious. Certainly, many transporter genes are under transcriptional regulation such that the transporter concentrations differ under different environments. However, changes of expression by gene regulation take time and are energetically costly (Perez-Ortin *et al*, 2007). More importantly, the cell does not have regulatory responses to all possible environmental changes. Thus, high expression noise of transporters allows, at least, some cells to have high fitness in an unpredictable environment. Below we show mathematically that, under certain conditions, genotypes with high expression noise can have greater Darwinian fitness than those with low noise.

Let us consider two genotypes A and B. The only difference between them is that A has a higher level of expression noise than B for gene X. The mean expression level ( $m$ ) of X is identical between the two genotypes. The distribution of the expression noise ( $e$ ) for gene X is described by probability density functions  $g_A(e)$  and  $g_B(e)$  for the two genotypes, respectively. Genome-wide expression noise data showed that  $e$  generally follows a normal distribution (Bar-Even *et al*, 2006; Newman *et al*, 2006). Let us assume that a population, having A and B cells, experiences an environmental change such that the mean expression level of X becomes suboptimal. Let  $f(x)=f(m+e)$  be the fitness of the cell that has an expression level of X equal to  $x$ . So, the fitness of genotype A, or the mean fitness of A cells, equals  $F_A = \int_{-\infty}^{\infty} f(m+e)g_A(e)de$ . Similarly, the fitness of genotype B equals  $F_B = \int_{-\infty}^{\infty} f(m+e)g_B(e)de$ . It can be shown that (i) when  $f(x)$  is a convex function (i.e. the second derivative of  $f(x)$  is positive),  $F_A > F_B$ ; (ii) when  $f(x)$  is a concave function,  $F_A < F_B$ ; and (iii) when  $f(x)$  is linear,  $F_A = F_B$  (Figure 2; Supplementary Figure S1; Supplementary information 1). As  $f(x)$  may not be concave or convex for all possible values of  $x$ , what matters is whether  $f(x)$  is concave or convex for the range of  $x$  realized in the majority (e.g. 95% or 99%) of A and B cells. Note that in our model, the optimal expression level can be either higher or lower than  $m$  (Figure 2). Although the shape of  $f(x)$  is generally unknown, it is reasonable to assume that, at least, for many genes if not most genes, it is bell shaped with the optimal expression level in the center (Kacser and Burns, 1981; Hartl *et al*, 1985; Bedford and Hartl, 2009). In such cases,  $f(x)$  is concave when  $x$  is close to the optimal expression level, but convex when  $x$  is far from the optimal. Thus, big environmental changes tend to generate conditions under which high noise is beneficial. Note that although we compared mean fitness values of cells with two different

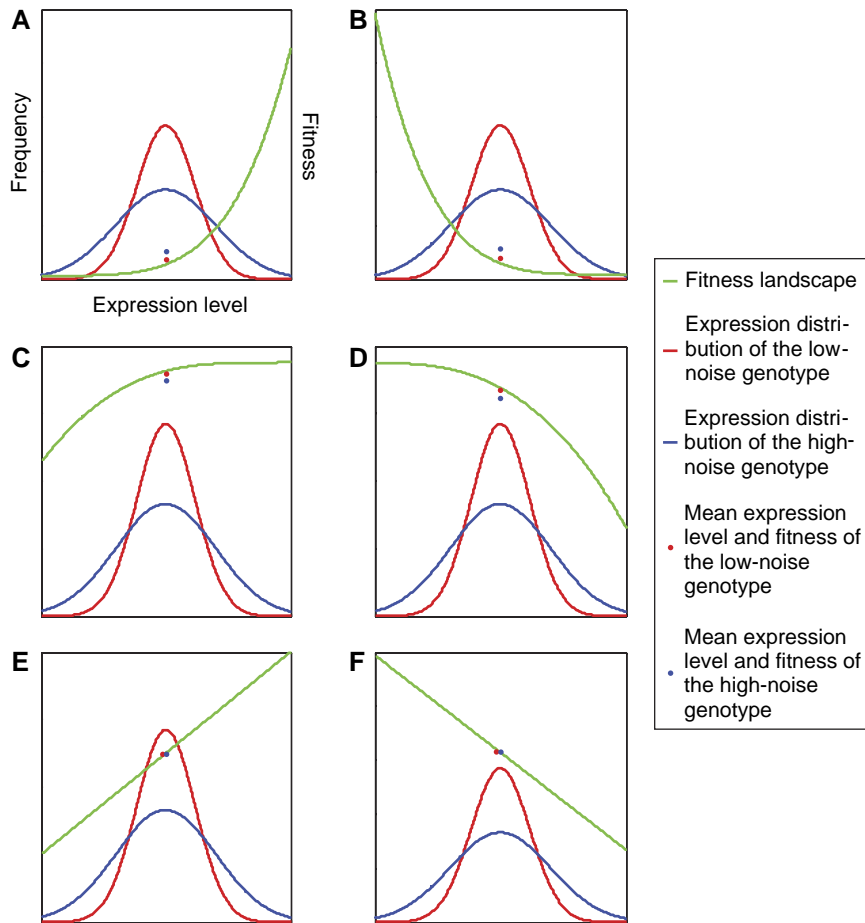


**Figure 1** Higher-than-expected expression noise of plasma-membrane transporters in yeast. **(A)** Plasma-membrane transporters (P + T +) are significantly noisier than the neutral expectation. By contrast, non-transporter plasma-membrane proteins (P + T -) and non-plasma-membrane transporters (P - T +) are not noisier than the expectation. The expectation is computed by the mean DM of all genes in the genome with the same level of gene importance (after the removal of enzymes and proteins localized to mitochondrion). Error bars represent one standard error. **(B)** The noise levels of plasma-membrane transporters, in comparison with those of all genes in the genome (after the removal of enzymes and proteins localized to mitochondrion). Genes are divided into 21 bins based on the fitness of the gene-deletion yeast strains. The mean and s.d. of the noise level for each bin is shown by an open circle and error bars, respectively. No circle is shown if a bin contains no gene, and no error bar is shown if a bin contains only one gene. Plasma-membrane transporters are shown by small squares. **(C)** Twenty noisiest plasma-membrane transporters in yeast. The expected noise level is computed by the mean DM of all genes in the genome with the same level of gene importance (after the removal of enzymes and proteins localized to mitochondrion). Functional annotations of the genes are based on *Saccharomyces* genome database (SGD).

genotypes, there is no involvement of group selection in our model. When  $f(x)$  is convex, in a population fixed with the wild type, a mutant with a higher level of noise is expected to increase its frequency in the population because its fitness is greater than that of the wild type.

To see how large  $F_A - F_B$  is when realistic parameters are used in our model, we examined a few numerical examples. As the effective population size of yeast is of the order of  $10^7$  (Wagner, 2005), a fitness differential greater than  $10^{-7}$  can be detected by natural selection. We found that  $F_A - F_B$  is easily greater

than  $10^{-7}$ . For example, in Figure 2A, we assumed  $f(x) = e^{-(x-\mu)^2/(2\sigma^2)} = \sigma\sqrt{2\pi} \frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}}$ . That is,  $f(x)$  is scaled by  $\sigma\sqrt{2\pi}$  from the probability density function of normal distribution  $N(\mu, \sigma)$ , where  $\mu$  and  $\sigma$  are the mean and s.d. of the normal distribution, respectively. We used  $\mu=6.2$  and  $\sigma=1$ . We further assumed that the expression noise in genotypes A and B follows  $N(0, 1.2)$  and  $N(0, 0.7)$ , respectively, and that the mean expression levels of the two genotypes are both  $m=3$ . Given these parameters, we found that  $F_A - F_B=0.0728-$



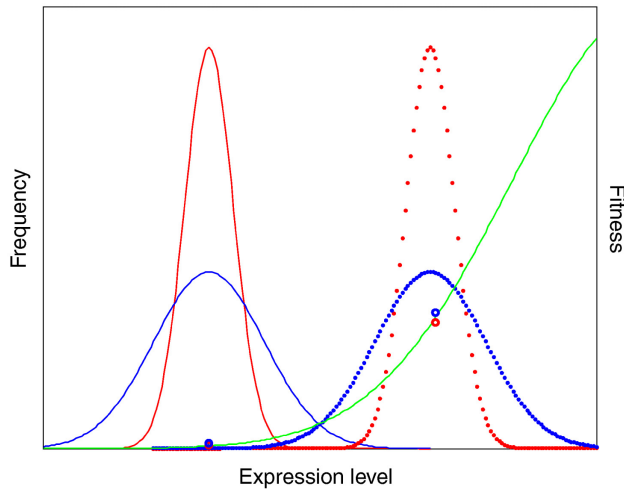
**Figure 2** Fitness landscape affects the relative fitness of high-noise and low-noise genotypes. In each panel, the green curve shows  $f(x)$ , the fitness of the cell with the expression level of gene X equal to  $x$ . The blue and red curves show the frequency distributions of the expression levels ( $x$ ) of the high-noise and low-noise genotypes, respectively. The blue and red dots are the mean fitness of the high-noise and low-noise genotypes, respectively. When  $f(x)$  is convex, the mean fitness of the high-noise genotype is greater than that of the low-noise genotype, no matter whether the optimal expression level is higher (A) or lower (B) than the mean expression levels of the two genotypes. When  $f(x)$  is concave, the fitness of the high-noise genotype is smaller than that of the low-noise genotype, no matter whether the optimal expression level is higher (C) or lower (D) than the mean expression levels of the two genotypes. When  $f(x)$  is linear, the fitness of the high-noise genotype equals that of the low-noise genotype, no matter whether the optimal expression level is higher (E) or lower (F) than the mean expression levels of the two genotypes.

0.0264=0.0464, five orders of magnitude greater than  $10^{-7}$ . Further analysis indicates that a large parameter space allows  $F_A - F_B$  to be substantially greater than  $10^{-7}$ , and this is true even for genes with a tiny fitness effect (e.g. <1%) upon deletion (Supplementary Figure S2). A previous site-directed mutagenesis study showed that a single point mutation in the *GALI* promoter of yeast can more than triple the level of expression noise (measured by the s.d. of the expression level) (Blake *et al*, 2006). So, the assumed noise difference between genotypes A and B here can arise simply by a point mutation. This and other numerical examples, we tried, suggest that conditions under which the benefit of high noise is detectable by natural selection arise easily.

### Our model predicts faster adaptive expression evolution of noisier genes

Under our model described above, it can be shown that, when the mean expression level of a genotype is lower than the

optimal level and the third derivative of  $f(x)$  is positive, or when the mean expression level of a genotype is higher than the optimal level and the third derivative of  $f(x)$  is negative, a given amount of change in mean expression level towards the optimal level will result in a greater fitness increase for the genotype with a higher level of noise (Figure 3; Supplementary Figure S3; Supplementary information 1). This is because, under the above conditions, the same advantageous mutation increases the mean fitness of the noisier genotype more than that of the quieter genotype (Figure 3). Consequently, the strength of positive selection for the same advantageous mutation that changes the same amount of mean expression level is stronger in a noisier genotype than in a quieter genotype. For instance, in the numerical example depicted in Figure 3, we used  $f(x) = e^{-(x-\mu)^2/(2\sigma^2)}$ , where  $\mu=11$  and  $\sigma=2.5$ . The expression noise in genotypes A and B follows  $N(0, 1.2)$  and  $N(0, 0.7)$ , respectively, and the mean expression levels of the two genotypes are both  $m=3.0$ . The advantageous mutation shifts the mean expression of both genotypes to  $n=7.1$ . Under such conditions, the fitness of genotype A



**Figure 3** The same advantageous mutation can generate a greater fitness gain in high-noise than low-noise genotypes under certain conditions. The green curve shows  $f(x)$ , the fitness of the cell with the expression level of gene  $X$  equal to  $x$ . The solid blue and red curves show the frequency distributions of the expression levels ( $x$ ) of the high-noise and low-noise genotypes, respectively. The dotted blue and red curves show the frequency distributions of the expression levels ( $x$ ) of the high-noise and low-noise genotypes with the advantageous mutation that shifts the mean expression level toward right. The two blue circles show the mean fitness value of the high-noise genotype with and without the advantageous mutation, respectively. The two red circles show the mean fitness value of the low-noise genotype with and without the advantageous mutations, respectively.

increases from 0.0178 to 0.4195 because of the mutation, whereas the fitness of genotype B increases from 0.0107 to 0.3978 because of the same mutation. Thus, the fitness gain for genotype A (0.4017) is greater than that (0.3871) for genotype B. We observed this trend in a large parameter space examined (Supplementary Figure S4). Here we assumed that the mutation size ( $n-m=4.1$ ) is  $\sim 3.5$  times the noise level of genotype A and  $\sim 6$  times the noise level of genotype B. This assumption is realistic, because a previous study on the yeast *GAL1* promoter showed that a single point mutation can change the mean expression level by more than 10 times the noise level (Blake *et al*, 2006).

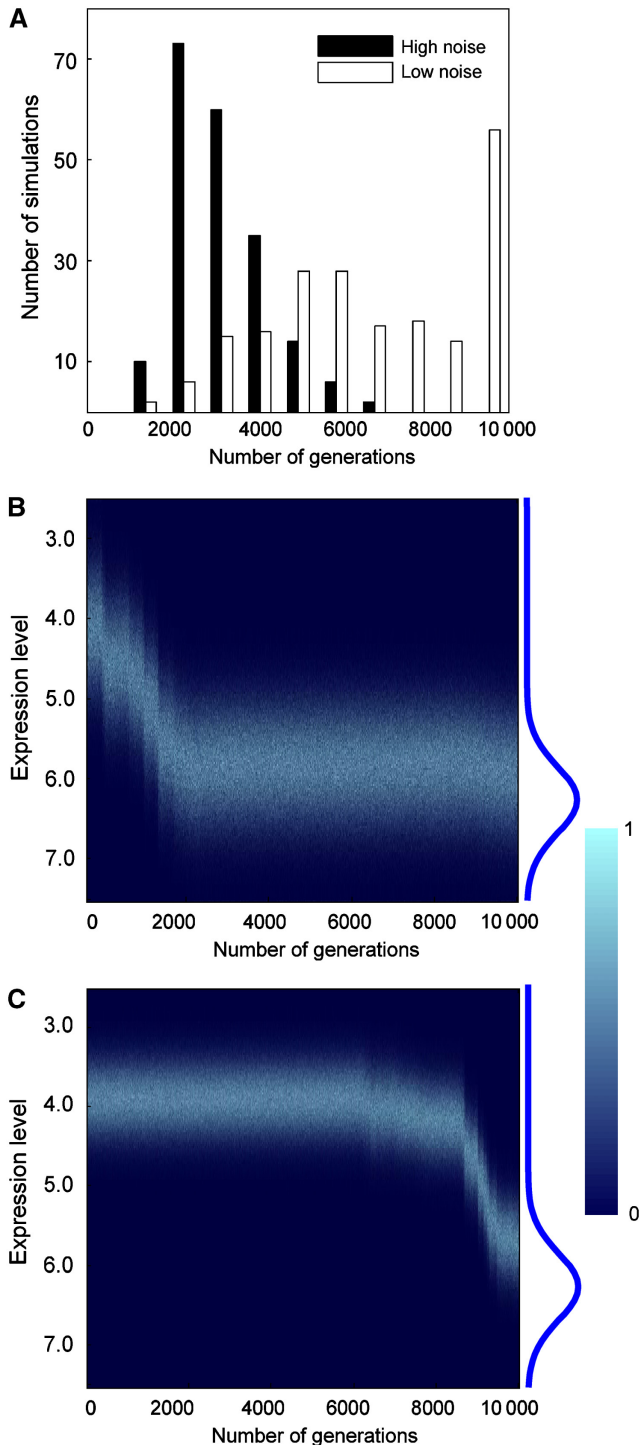
As the same advantageous mutation can enhance the fitness of the noisier genotype more than the quieter genotype, we predict faster adaptive evolutionary changes in mean expression level for noisier genes than for quieter genes. To see to what extent the noise level impacts the rate of adaptation, we conducted a computer simulation. Let us consider a population of yeast cells all with genotype A and another population all with genotype B. The two genotypes have the same mean expression level that is suboptimal. Genotype A has a higher expression noise level than genotype B. The two populations have the same population size, mutation rate, and mutation spectrum. Mutations are randomly generated with a size that follows a normal distribution. Here, mutation size refers to the difference between the mean expression level of the mutant and that of the wild type. We assume that the level of expression noise does not change. As shown in Figure 4A, under the parameters detailed in Methods section, genotype A adapts its expression level to the optimal level significantly faster than genotype B ( $P < 10^{-48}$ ,  $t$ -test), and the difference in speed is on average 2.56-fold. Figure 4 also shows the

adaptation process from one simulation replication, in which the noisier genotype (Figure 4B) adapts to the optimal expression level in about one-fifth the time required for the quieter genotype (Figure 4C). Thus, at least under some conditions, high expression noise leads to a substantially enhanced rate of adaptation of gene expression level because noise can facilitate positive selection for advantageous mutations. Note that although the number of generations required for adaptation seems very large in Figure 4, the actual time required can be much shorter if the mutations are larger or the mutation rate is higher. We found that our simulation result holds in a broad parameter space when we vary the mutation rate and the noise ratio of the high-noise and low-noise genotypes (Supplementary Figure S5).

Above, we considered beneficial mutations. In the case of deleterious mutations, it can be similarly shown that, under our model, when the mean expression level of a genotype is lower than the optimal level and the third derivative of  $f(x)$  is positive, or when the mean expression level of a genotype is higher than the optimal level and the third derivative of  $f(x)$  is negative, a deleterious mutation that renders the mean expression level further away from the optimal level will result in a larger fitness loss for the genotype with a higher level of noise (Supplementary Figure S3; Supplementary information 1). In other words, under such conditions, negative selection against deleterious mutations that affect the mean expression level will be stronger for noisier genotypes.

### Empirical data show greater expression divergence of noisier genes

It will be interesting to empirically verify our prediction of higher rates of adaptive expression evolution in noisy genes than in quiet genes. As the available expression noise data are from one strain of *Saccharomyces cerevisiae* (Newman *et al*, 2006), it is better to estimate the evolutionary rate of gene expression using closely related species or even intraspecific strains such that the noise level may be considered constant in the evolution of gene expression level. We first compared two strains of *S. cerevisiae*, a laboratory strain BY4716 (derived from s288c) and a wild isolated strain RM11-1a, using whole genome microarray gene expression data generated under the same condition (synthetic complete medium at 30°C) (Brem and Kruglyak, 2005). We observed a positive correlation between gene expression noise and gene expression divergence between the two strains (Spearman's rank correlation coefficient  $\rho=0.241$ ,  $P < 10^{-26}$ ) (Figure 5A; Table II). Using microarray gene expression data from five different conditions, we next measured the divergence of mean expression level among four closely related species of the *Saccharomyces sensu stricto* complex (Tirosh *et al*, 2006), and found that the divergence is also positively correlated with expression noise ( $\rho=0.291$ ,  $P < 10^{-32}$ ; Figure 5B; Table II), as was previously observed (Lehner, 2008). However, it is not trivial to show that the correlation reflects the prediction of our model rather than some other mechanism. Below we examine possible alternative mechanisms by calculating partial correlation (Fisher, 1924) and show that although some of them do have a role, they cannot fully explain the observed correlations.



**Figure 4** Computer simulation shows faster adaptive evolution of expression level for a noisier genotype than a quieter genotype. **(A)** The noisier genotype reaches the optimal expression level sooner than the quieter genotype during evolution. **(B)** A typical case of expression evolution of a noisy genotype. The blue curve to the right of the figure is the fitness function  $f(x)$ . Each vertical line in the heat map represents the frequency distribution of  $x$  in the population in a given generation, with different colors representing different frequencies of cells with given  $x$ . **(C)** A typical case of expression evolution of a quiet genotype.

First, it is possible that the number of mutation sites or the effect of each mutation (in terms of changing the mean expression level) varies among genes of different levels of

noise. Indeed, as was previously noted (Landry *et al*, 2007), expression evolution, measured by expression variance among mutation-accumulation lines of yeast, is positively correlated with expression noise ( $\rho=0.24$ ,  $P<10^{-27}$ ). As the effective population size was controlled to be  $\sim 10$  cells in the mutation-accumulation experiment (Landry *et al*, 2007), the majority of non-lethal mutations behave neutrally and thus the rate of expression evolution of these lines reflect the rate and size of expression mutation, hereby collectively referred to as mutational effect. If the difference in mutational effect is the sole reason for the correlation between gene expression noise and expression divergence shown in Figure 5, we should not observe this correlation after the control for the expression variance in mutation-accumulation lines. However, in fact, the partial correlation between gene expression noise and expression divergence remains significantly positive in both between-strain ( $\rho=0.203$ ,  $P<10^{-18}$ ; Table II) and between-species ( $\rho=0.247$ ,  $P<10^{-23}$ ; Table II) comparisons.

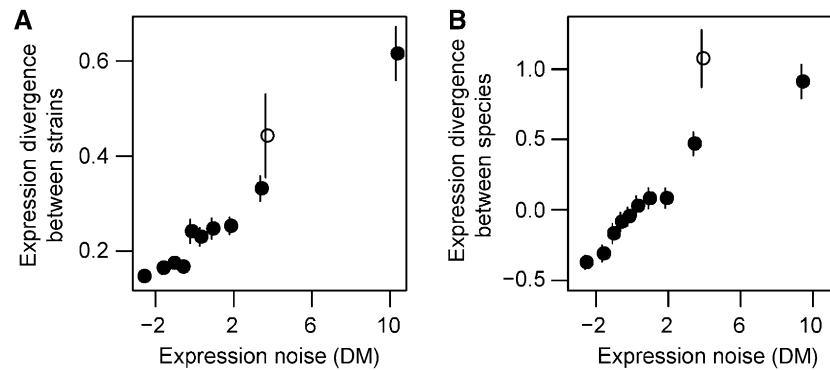
Second, negative selection against expression noise can also generate a positive correlation between the noise level and the rate of expression evolution, because important genes tend to have both low noise (Cook *et al*, 1998; Batada and Hurst, 2007; Lehner, 2008) and low rate of expression evolution (Tirosh and Barkai, 2008). However, after further controlling gene importance, we found the positive correlation between the noise level and the rate of expression evolution to remain significant in both between-strain ( $\rho=0.191$ ,  $P<10^{-10}$ ; Table II) and between-species ( $\rho=0.223$ ,  $P<10^{-11}$ ; Table II) comparisons.

Third, the above control of gene importance does not fully eliminate the among-gene variation in the level of negative selection against noise. Thus, we further removed mitochondrial proteins, enzymes, and haploinsufficient proteins from our dataset. We found the positive correlation between the noise level and expression divergence to remain significant in both between-strain ( $\rho=0.131$ ,  $P<10^{-4}$ ; Table II) and between-species ( $\rho=0.216$ ,  $P<10^{-8}$ ; Table II) comparisons. As shown earlier, after the control for gene importance, membership in stable protein complexes no long correlates with expression noise. We therefore did not further control for complex membership here. Together, the above results provide empirical support to the prediction of our model that high noise can facilitate adaptive gene expression evolution.

## Discussion

By analyzing the yeast genome-wide gene expression noise data, we identified plasma-membrane transporters as the only group that shows significantly greater-than-expected noise after the exclusion of multiple factors related to the relaxation of negative selection against noise. Although this result suggests that the elevation of the expression noise in plasma-membrane transporters is driven by positive selection, an alternative hypothesis is that the high noise is a by-product of selection for something else rather than the direct target of selection. One particularly relevant subject here is the differential use of TATA boxes in the promoters of different groups of genes (Basehoar *et al*, 2004). For example, TATA-containing genes are associated with responses to stress, are highly regulated, and preferentially utilize SAGA rather than





**Figure 5** Expression divergence is positively correlated with expression noise when the divergence is measured (A) between BY4716 and RM11-1a strains of *Saccharomyces cerevisiae* or (B) among four closely related *Saccharomyces sensu stricto* species. All genes with expression noise and expression divergence data are equally divided into 10 consecutive bins according to the noise level. The mean noise level and mean expression divergence ( $\pm$  one standard error) for each bin are indicated by solid circles. The means for plasma-membrane transporters are indicated by open circles.

**Table II** Correlations between the expression noise level and expression divergence of yeast genes

Factors examined <sup>a</sup>	Divergence between strains		Divergence between species	
	$\rho^b$	<i>P</i> -value	$\rho^b$	<i>P</i> -value
Noise, divergence	0.241	$1.77 \times 10^{-27}$	0.291	$8.56 \times 10^{-33}$
Noise, divergence   mutational effect	0.203	$2.24 \times 10^{-19}$	0.247	$4.92 \times 10^{-24}$
Noise, divergence   mutational effect, gene importance	0.191	$7.12 \times 10^{-11}$	0.223	$2.63 \times 10^{-12}$
Noise, divergence   mutational effect, gene importance, mitochondrial proteins, enzymes, haploinsufficient genes	0.131	$5.31 \times 10^{-5}$	0.216	$3.12 \times 10^{-9}$

<sup>a</sup>Factors before | are those being correlated, whereas factors after | are those being controlled for.

<sup>b</sup>Spearman's rank correlation coefficient.

TFIID when compared with TATA-less genes (Basehoar *et al.*, 2004). Interestingly, TATA-containing genes have significantly larger expression noise than TATA-less genes (Newman *et al.*, 2006). Hence, the high noise of plasma-membrane transporters could potentially be a by-product of the use of TATA-containing promoters, if plasma-membrane transporter genes require TATA-containing promoters for gene regulation. After removing enzymes and mitochondrial proteins, our dataset contains 1088 genes that have the information about gene importance, expression noise, and the presence/absence of a TATA box. Although only 13.4% of these 1088 genes contain a TATA box, the corresponding number is 54.5% among plasma-membrane transporters ( $P < 10^{-7}$ ,  $\chi^2$  test). Nevertheless, even among TATA-containing genes, the expression noise is significantly higher for plasma-membrane transporters than for the other genes after the control for gene importance, enzymes, and mitochondrial proteins ( $P < 0.001$ ; two-tail Mann-Whitney *U* test). The same is true among TATA-less genes ( $P < 0.02$ ). Thus, the high noise of plasma-membrane transporters is not fully attributable to the preferential use of TATA-containing promoters, supporting direct positive selection for elevated expression noise of these genes. The result further suggests that multiple molecular mechanisms are used to achieve the high expression noise of plasma-membrane transporters. We note that even if the high noise of plasma-membrane transporters were fully attributable to the preferential use of TATA-containing promoters, the hypothesis of direct selection for high noise could not be rejected, because of

the possibility that the preferential use of TATA-containing promoters is a by-product of the selection for high noise; it would then become necessary to differentiate which is the direct target of selection and which is the by-product.

When controlling for gene importance, we used the data of fitness reduction by gene deletion measured in laboratory rich media, which may not resemble closely the natural environments of yeast. However, the fact that this gene importance index significantly inversely correlates with the expression noise level (Lehner, 2008), also measured under the rich media, suggests that using this gene importance index in analyzing the rich media noise data is meaningful. Our subsequent analysis of the noise data from rich and minimal media showed that despite the large nutritional difference between the two media, the noise levels under the two media are highly correlated ( $\rho = 0.53$ ,  $P < 10^{-15}$ ), suggesting that the noise levels measured in lab conditions may be good proxies for the true values in nature. Moreover, expression noise data from both the rich and minimal media identified plasma-membrane transporters as the only group with higher-than-expected noise. Thus, it is unlikely that our result is an artifact due to the use of various data generated from lab conditions that are different from the natural environments of yeast.

One could argue that plasma-membrane transporters may be regarded as special enzymes and thus their high noise may be related to the dosage-buffering effect that enzymes generally suffer from (Kacser and Burns, 1981). We compared

the noise level of enzymes and plasma-membrane transporters by separating the enzyme genes into 21 bins based on their fitness effects upon deletion and drawing enzyme genes randomly according to the importance levels of the plasma-membrane transporters. Repeating this process 10 000 times, we found that plasma-membrane transporters are on average 2.92 times noisier than enzymes after the control for gene importance ( $P=0.001$ , Mann-Whitney  $U$  test). Thus, the high noise of plasma-membrane transporters cannot be explained by the buffering effect even if the transporters behave as enzymes, further supporting positive selection as the evolutionary force behind their high noise.

We proposed a simple mathematical model and showed that a high-noise genotype will have a greater fitness than a low-noise genotype with the same mean expression, as long as the fitness function is convex. The key question is whether the cellular fitness, as a function of the expression level of a plasma-membrane transporter, has a convex region. To our knowledge, there has been only one study that empirically determined the fitness function (Dekel and Alon, 2005). This study reported the relationship between the fitness of *Escherichia coli* cells and the expression level of Lac proteins; the fitness function seems to be concave in the ranges examined. However, this result does not preclude the existence of a convex region in unexamined expression ranges of Lac proteins, nor does it tell us the fitness functions of other genes. It is likely that the shape of the fitness function varies depending on the specific cellular role each gene has. Owing to the lack of sufficient empirical data on the fitness function, we decide to examine the theoretical possibilities, especially in the context of plasma-membrane transporters. A simple theoretical model shows the existence of convex regions in the fitness function (Supplementary information 2 and Supplementary Figure S6). Although the jury is still out as whether the fitness function indeed contains a convex region, our theoretical modeling supports this possibility. As the natural environment of yeast may change abruptly and frequently and because plasma-membrane transporters are directly involved in the interactions between a cell and its biotic and abiotic environment, conditions under which high expression noise of plasma-membrane transporters is favored over low noise may arise relatively easily. By contrast, non-plasma-membrane transporters (e.g. those localized to the nuclear envelope) and plasma-membrane proteins that are not transporters (e.g. those attaching the cell wall to the plasma-membrane) generally do not face relevant environmental changes that are unpredictable, abrupt, and frequent. It should be noted that because mitochondrial proteins and enzymes were removed in our GO analysis, adaptive elevation of noise could not be tested for genes belonging to these two categories. A previous study identified genes related to stress to be noisier than the genomic average (Newman *et al*, 2006). In our analysis, the biological process GO category of 'response to stress' (GO0006950) was significantly noisier than the genomic average before the control for multiple testing ( $P=0.012$ ), but not after the control ( $Q=0.082$ ) (Supplementary Table S6). Regardless, although our analysis does not preclude the possibility that even some individual non-plasma-membrane-transporter genes have elevated noise driven by positive selection, plasma-membrane transporters are apparently

among those that most frequently face large and unpredictable environment fluctuations. Hence, our results are biologically sensible. The strength of positive selection for high noise depends on how frequently favorable conditions occur and on the fitness functions of a gene under such conditions. It should be noted that high noise is advantageous only under unpredictable environmental changes. Repeated switching among a fixed set of environments may lead to the evolutionary emergence of gene regulation, with which low noise could be beneficial.

Our model of why high noise can be favored over low noise differs significantly from a previous model that is based on the bimodal distribution of gene expression (Thattai and van Oudenaarden, 2004). In the model, the expression level of a gene in a cell switches between two states. Given that empirical data show a normal distribution of noise (Bar-Even *et al*, 2006; Newman *et al*, 2006), our model is more realistic and general. Our model is also more general than another earlier model in which the fitness function  $f(x)$  is assumed to be either 0 or 1 (Blake *et al*, 2006).

Our mathematical model further predicts faster adaptive evolution of gene expression toward the optimum for noisier genotypes than for quieter genotypes under certain conditions. Our model, again, is significantly different from previous models that are based on multiple expression attractors (Kaneko and Furusawa, 2008). Our prediction is supported by our observation of higher expression divergence between yeast strains and between yeast species for genes of higher noise even when all confounding factors are controlled for. We note that our result does not rely on the assumption that all or most gene expression divergence between strains (or species) is adaptive. The fact that, after all controls, expression noise explains only several percent of the among-gene variation in expression divergence, is not inconsistent with the hypothesis that the majority of expression divergence is neutral (Khaitovich *et al*, 2006).

Taken together, high expression noise is not only selected for in certain yeast genes under unpredictable environmental changes, it also facilitates adaptive expression evolution when a directional environmental change occurs. We expect that all unicellular organisms that face unpredictable and frequent environmental changes would show a similar pattern of elevated expression noise in those genes whose expression levels are often suboptimal, and it will be interesting to test this prediction in the future when genome-wide expression noise data become available for additional species. The power and versatility of natural selection in seizing and utilizing even seemingly harmful biological properties such as the stochasticity in gene expression to enhance organismal fitness is a wonderful tribute to the theory of evolution by means of natural selection.

## Materials and methods

### Data analysis

The yeast genome-wide datasets of normalized gene expression noise level (DM) in rich and minimal media were from the study by Newman *et al* (2006). Gene expression data from BY4716 and RM11-1a strains (Brem and Kruglyak, 2005) were retrieved using GEOquery in Bioconductor (Gentleman *et al*, 2004; Sean and Meltzer, 2007).

Expression divergence was estimated by the  $\log_2$  ratio of the relative intensity of hybridization signals in microarray experiments (Brem and Kruglyak, 2005). Gene expression divergence among four yeast species was similarly estimated and was directly taken from the study by Tirosh *et al* (2006). Gene expression divergence among mutation-accumulation lines of yeast was measured by the variance of expression signal across lines and was supplied by Landry *et al* (2007). Gene importance was measured by the reduction in fitness upon gene deletion, and was acquired from earlier studies (Giaever *et al*, 2002; Steinmetz *et al*, 2002). Data on fitness effects of gene overexpression are obtained from Sopko *et al* (2006), in which the growth rates of gene overexpression strains are divided into five levels, from 1 (no growth) to 5 (normal growth). GoMiner (Zeeberg *et al*, 2003) was used to retrieve GO (Ashburner *et al*, 2000) information for yeast genes. The information about the presence and absence of TATA boxes in yeast genes was acquired from the study by Basehoar *et al* (2004).

## Computer simulation

Computer simulation was conducted to examine to what extent the noise level impacts the rate of adaptive expression evolution, in the face of mutation, drift, and selection. We considered a population of yeast cells all with genotype A and another population all with genotype B. The only difference between the two genotypes is that the expression noise for gene X is higher in A than in B. We assumed that the expression level of gene X in individual cells of the two populations follows the normal distribution  $N(\mu, \sigma_1)$  and  $N(\mu, \sigma_2)$ , respectively. The two populations have the same population size  $L$ , mutation rate  $m$ , and mutation size distribution. Here, mutation size refers to the difference between the mean expression level of the mutant and that of the wild type. Mutations were randomly generated with a size that follows the normal distribution  $N(0, \sigma')$ . We assumed that the expression noise level does not change. The fitness function  $f(x)$  used was bell shaped (see Figure 4B and C):  $f(x) = a + \frac{1-a}{\phi(b)} \phi(x)$ , where the normal probability density function  $\phi(x)$  equals  $\frac{e^{-(x-b)^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}}$ , where  $a$  is a fitness scaling factor and  $b$  the optimal expression level. We used the following parameters in our simulation:  $L=1000$ ,  $a=0.1$ ,  $b=6$ ,  $c=0.4$ ,  $m=10^{-4}$ ,  $\mu=4$ ,  $\sigma_1=0.6$ ,  $\sigma_2=0.3$ ,  $\sigma'=0.1$ . We started the simulation by generating a population of  $L$  haploid cells. We then generated mutations in each cell. The relative frequency of each allele in the next generation was determined by its relative fitness in the population as well as genetic drift. When the mean expression level of the population reached within one s.d. from the optimal expression level, we considered the adaptive evolution to be successful and recorded the number of generations used. If after  $10^4$  generations the mean expression level of the population still had not reached the above cutoff, we stopped the simulation and recorded the time used as  $10^4$  generations. We conducted 200 simulation replications for each of the two populations.

## Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website ([www.nature.com/msb](http://www.nature.com/msb)).

## Acknowledgements

We thank Christian Landry for providing the data of expression variance among mutation-accumulation yeast lines and Meg Bakewell, Zhi Wang, and three anonymous reviewers for valuable comments. This study was supported by research grants from the National Institutes of Health and University of Michigan Center for Computational Medicine and Biology to JZ.

## Conflict of interest

The authors declare that they have no conflict of interest.

## References

- Acar M, Mettetal JT, van Oudenaarden A (2008) Stochastic switching as a survival strategy in fluctuating environments. *Nat Genet* **40**: 471–475
- Ansel J, Bottin H, Rodriguez-Beltran C, Damon C, Nagarajan M, Fehrmann S, Francois J, Yvert G (2008) Cell-to-cell stochastic variation in gene expression is a complex genetic trait. *PLoS Genet* **4**: e1000049
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**: 25–29
- Bahar R, Hartmann CH, Rodriguez KA, Denny AD, Busuttill RA, Dolle ME, Calder RB, Chisholm GB, Pollock BH, Klein CA, Vijg J (2006) Increased cell-to-cell variation in gene expression in ageing mouse heart. *Nature* **441**: 1011–1014
- Bar-Even A, Paulsson J, Maheshri N, Carmi M, O'Shea E, Pilpel Y, Barkai N (2006) Noise in protein expression scales with natural protein abundance. *Nat Genet* **38**: 636–643
- Basehoar AD, Zanton SJ, Pugh BF (2004) Identification and distinct regulation of yeast TATA box-containing genes. *Cell* **116**: 699–709
- Batada NN, Reguly T, Breitkreutz A, Boucher L, Breitkreutz BJ, Hurst LD, Tyers M (2006) Stratus not altocumulus: a new view of the yeast protein interaction network. *PLoS Biol* **4**: e317
- Batada NN, Hurst LD (2007) Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nat Genet* **39**: 945–949
- Becskei A, Serrano L (2000) Engineering stability in gene networks by autoregulation. *Nature* **405**: 590–593
- Bedford T, Hartl DL (2009) Optimization of gene expression by natural selection. *Proc Natl Acad Sci USA* **106**: 1133–1138
- Blake WJ, Kærn M, Cantor CR, Collins JJ (2003) Noise in eukaryotic gene expression. *Nature* **422**: 633–637
- Blake WJ, Balazsi G, Kohanski MA, Isaacs FJ, Murphy KF, Kuang Y, Cantor CR, Walt DR, Collins JJ (2006) Phenotypic consequences of promoter-mediated transcriptional noise. *Mol Cell* **24**: 853–865
- Brem RB, Kruglyak L (2005) The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc Natl Acad Sci USA* **102**: 1572–1577
- Colman-Lerner A, Gordon A, Serra E, Chin T, Resnekov O, Endy D, Pesce CG, Brent R (2005) Regulated cell-to-cell variation in a cell-fate decision system. *Nature* **437**: 699–706
- Cook DL, Gerber AN, Tapscott SJ (1998) Modeling stochastic gene expression: implications for haploinsufficiency. *Proc Natl Acad Sci USA* **95**: 15641–15646
- Dekel E, Alon U (2005) Optimality and evolutionary tuning of the expression level of a protein. *Nature* **436**: 588–592
- Deutschbauer AM, Jaramillo DF, Proctor M, Kumm J, Hillenmeyer ME, Davis RW, Nislow C, Giaever G (2005) Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics* **169**: 1915–1925
- Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. *Science* **297**: 1183–1186
- Fisher RA (1924) The distribution of the partial correlation coefficient. *Metron* **3**: 329–332
- Fraser HB, Hirsh AE, Giaever G, Kumm J, Eisen MB (2004) Noise minimization in eukaryotic gene expression. *PLoS Biol* **2**: e137
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**: R80
- Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, Dow S, Lucau-Danila A, Anderson K, Andre B, Arkin AP, Astromoff A, El-Bakkoury M, Bangham R, Benito R, Brachat S, Campanaro S, Curtiss M, Davis K, Deutschbauer A *et al.* (2002) Functional

- profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**: 387–391
- Hartl DL, Dykhuizen DE, Dean AM (1985) Limits of adaptation: the evolution of selective neutrality. *Genetics* **111**: 655–674
- Kacser H, Burns JA (1981) The molecular basis of dominance. *Genetics* **97**: 639–666
- Kaern M, Elston TC, Blake WJ, Collins JJ (2005) Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet* **6**: 451–464
- Kaneko K, Furusawa C (2008) Relevance of phenotypic noise to adaptation and evolution. *IET Syst Biol* **2**: 234–246
- Kemkemmer R, Schrank S, Vogel W, Gruler H, Kaufmann D (2002) Increased noise as an effect of haploinsufficiency of the tumor-suppressor gene neurofibromatosis type 1 *in vitro*. *Proc Natl Acad Sci USA* **99**: 13783–13788
- Khaitovich P, Enard W, Lachmann M, Paabo S (2006) Evolution of primate gene expression. *Nat Rev Genet* **7**: 693–702
- Landry CR, Lemos B, Rifkin SA, Dickinson WJ, Hartl DL (2007) Genetic properties influencing the evolvability of gene expression. *Science* **317**: 118–121
- Lehner B (2008) Selection to minimise noise in living systems and its implications for the evolution of gene expression. *Mol Syst Biol* **4**: 170
- Lopez-Maury L, Marguerat S, Bahler J (2008) Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nat Rev Genet* **9**: 583–593
- Losick R, Desplan C (2008) Stochasticity and cell fate. *Science* **320**: 65–68
- Maheshri N, O'Shea EK (2007) Living with noisy genes: how cells function reliably with inherent variability in gene expression. *Annu Rev Biophys Biomol Struct* **36**: 413–434
- Newman JR, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, Weissman JS (2006) Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* **441**: 840–846
- Orphanides G, Reinberg D (2002) A unified theory of gene expression. *Cell* **108**: 439–451
- Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A (2002) Regulation of noise in the expression of a single gene. *Nat Genet* **31**: 69–73
- Pedraza JM, van Oudenaarden A (2005) Noise propagation in gene networks. *Science* **307**: 1965–1969
- Perez-Ortin JE, Alepuz PM, Moreno J (2007) Genomics and gene transcription kinetics in yeast. *Trends Genet* **23**: 250–257
- Raj A, van Oudenaarden A (2008) Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**: 216–226
- Ramsey S, Ozinsky A, Clark A, Smith KD, de Atauri P, Thorsson V, Orrell D, Bolouri H (2006) Transcriptional noise and cellular heterogeneity in mammalian macrophages. *Philos Trans R Soc Lond B Biol Sci* **361**: 495–506
- Rao CV, Wolf DM, Arkin AP (2002) Control, exploitation and tolerance of intracellular noise. *Nature* **420**: 231–237
- Raser JM, O'Shea EK (2004) Control of stochasticity in eukaryotic gene expression. *Science* **304**: 1811–1814
- Raser JM, O'Shea EK (2005) Noise in gene expression: origins, consequences, and control. *Science* **309**: 2010–2013
- Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB (2005) Gene regulation at the single-cell level. *Science* **307**: 1962–1965
- Schrodinger E (1944) *What is Life?* Cambridge, UK: Cambridge University Press
- Sean D, Meltzer PS (2007) GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **23**: 1846–1847
- Sopko R, Huang D, Preston N, Chua G, Papp B, Kafadar K, Snyder M, Oliver SG, Cyert M, Hughes TR, Boone C, Andrews B (2006) Mapping pathways and phenotypes by systematic gene overexpression. *Mol Cell* **21**: 319–330
- Steinmetz LM, Scharfe C, Deutschbauer AM, Mokranjac D, Herman ZS, Jones T, Chu AM, Giaever G, Prokisch H, Oefner PJ, Davis RW (2002) Systematic screen for human disease genes in yeast. *Nat Genet* **31**: 400–404
- Stoebel DM, Dean AM, Dykhuizen DE (2008) The cost of expression of *Escherichia coli* lac operon proteins is in the process, not in the products. *Genetics* **178**: 1653–1660
- Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* **100**: 9440–9445
- Thattai M, van Oudenaarden A (2004) Stochastic gene expression in fluctuating environments. *Genetics* **167**: 523–530
- Tirosh I, Weinberger A, Carmi M, Barkai N (2006) A genetic signature of interspecies variations in gene expression. *Nat Genet* **38**: 830–834
- Tirosh I, Barkai N (2008) Evolution of gene sequence and gene expression are not correlated in yeast. *Trends Genet* **24**: 109–113
- Volfson D, Marciniak J, Blake WJ, Ostroff N, Tsimring LS, Hasty J (2006) Origins of extrinsic variability in eukaryotic gene expression. *Nature* **439**: 861–864
- Wagner A (2005) Energy constraints on the evolution of gene expression. *Mol Biol Evol* **22**: 1365–1374
- Zeeberg BR, Feng W, Wang G, Wang MD, Fojo AT, Sunshine M, Narasimhan S, Kane DW, Reinhold WC, Lababidi S, Bussey KJ, Riss J, Barrett JC, Weinstein JN (2003) GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol* **4**: R28



*Molecular Systems Biology* is an open-access journal published by *European Molecular Biology Organization* and *Nature Publishing Group*.

This article is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 Licence.