

# Control of Stochasticity in Eukaryotic Gene Expression

Jonathan M. Raser and Erin K. O'Shea\*

Noise, or random fluctuations, in gene expression may produce variability in cellular behavior. To measure the noise intrinsic to eukaryotic gene expression, we quantified the differences in expression of two alleles in a diploid cell. We found that such noise is gene-specific and not dependent on the regulatory pathway or absolute rate of expression. We propose a model in which the balance between promoter activation and transcription influences the variability in messenger RNA levels. To confirm the predictions of our model, we identified both *cis*- and *trans*-acting mutations that alter the noise of gene expression. These mutations suggest that noise is an evolvable trait that can be optimized to balance fidelity and diversity in eukaryotic gene expression.

The stochastic, or random and probabilistic, nature of chemical reactions may create variation in an identical population of cells (1). The reactions underlying gene expression involve small numbers of molecules (e.g., transcription factors, DNA, and mRNAs) and may therefore exhibit stochastic fluctuations that could generate population variation when phenotypic diversity would be advantageous or could act as a theoretical obstacle when fidelity in cellular behavior is required (2, 3).

To measure the stochasticity of eukaryotic gene expression, we implemented the dual-reporter technique, developed in *Escherichia coli* (4, 5), in the budding yeast *Saccharomyces cerevisiae* (6). We constructed diploid yeast strains that express both cyan and yellow fluo-

rescent proteins (CFP and YFP) from identical promoters, integrated at the same locus on homologous chromosomes (Fig. 1A). Two types of noise are distinguished in our analysis: intrinsic noise attributable to stochastic events during gene expression, and extrinsic noise due to any existing cellular heterogeneity that affects gene expression or to stochastic events in upstream signal transduction (5). For each population of cells, we calculated the variability in terms of two metrics: the noise, defined as the standard deviation divided by the mean, which we present to convey the magnitude of variability as a percentage of the level of gene expression; and the noise strength, or variance divided by the mean, which we use for our analysis because it is independent of population mean for a single stochastic process (supporting online text).

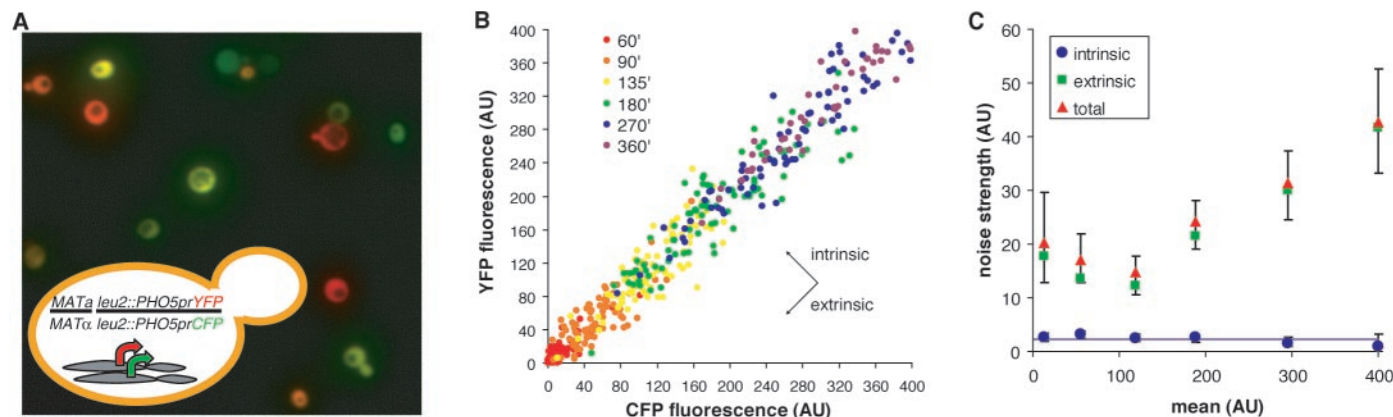
We induced the expression of CFP and YFP from the budding yeast *PHO5* promoter and measured the fluorescence of single cells in random subpopulations at multiple times after induction (Fig. 1B). The total noise of

gene expression from the *PHO5* promoter was dominated by the contribution from extrinsic factors (Fig. 1C); the intrinsic noise strength, although larger than the error of measurement in our system (supporting online text), represented only between 2% and 20% of the total noise strength. For a second promoter, *GALI*, the intrinsic noise strength represented less than 3% of the total noise strength (fig. S1). We conclude that the stochasticity of gene expression is not necessarily reflected by measurements of total noise that employ single-reporter techniques.

Heterogeneity in a number of factors that affect gene expression may underlie extrinsic noise, including heterogeneity in cell size and shape, cell cycle stage, or gene-specific signaling. We tested if extrinsic noise factors could be eliminated by flow cytometry to isolate subpopulations of cells that are homogeneous in size and shape. We found that, although diminished by this process, extrinsic noise predominated relative to intrinsic noise in these subpopulations (fig. S2A). Similarly, neither correction for individual cellular volume nor segregation by cell cycle stage (6) resulted in more than a ~25% decrease in extrinsic noise (fig. S2B). To distinguish extrinsic noise that is gene-specific from a global, nonspecific source, we examined the correlation between CFP and YFP fluorescence in a strain that expresses CFP from one promoter and YFP from a second, distinctly regulated promoter. Expression from the *PHO84* and *GALI* promoters was correlated, with a  $R^2$  value of 0.88 (fig. S3A); additionally, expression from the *PHO84* and *ADHI* promoters was correlated ( $R^2 = 0.93$ ) (fig. S3B). Therefore, the majority of extrinsic noise in these cases is not promoter-specific and will cause gene

Department of Biochemistry and Biophysics, Howard Hughes Medical Institute, University of California—San Francisco (UCSF), 600 16th Street, Room S472D, San Francisco, CA 94143–2240, USA.

\*To whom correspondence should be addressed. E-mail: oshea@biochem.ucsf.edu



**Fig. 1.** Separation of intrinsic and extrinsic noise for the *PHO5* promoter. **(A)** A false-color overlay of YFP (red) and CFP (green) fluorescence micrographs from a diploid yeast strain that expresses YFP and CFP from identical promoters at homologous loci, as diagrammed in the inset. **(B)** Scatter plots showing CFP and YFP values for each cell (solid circles) during a time course of *PHO5* induction by phosphate starvation. Populations from different

time points (in minutes) are indicated with different colors. Extrinsic noise is manifested as scatter along the diagonal and intrinsic noise as scatter perpendicular to the diagonal. AU, arbitrary units of fluorescence. **(C)** Total, extrinsic, and intrinsic noise strength as functions of population mean for **(B)**. The solid line represents expectations for a single stochastic process, and error bars represent bootstrap values (6).

products in a cell to be maintained in constant relative concentration.

To characterize stochasticity in eukaryotic gene expression further, we measured at different rates of gene expression the intrinsic noise strength of the *PHO5* and *PHO84* promoters, which are regulated by the same transcriptional activator, and the *GAL1* promoter (6). The *GAL1* and *PHO84* promoters display a low level of intrinsic noise strength that does not substantially vary with changes in the rate of gene expression (Fig. 2, A and B). In contrast to the other promoters, *PHO5* has a larger intrinsic noise strength that decreases with increasing

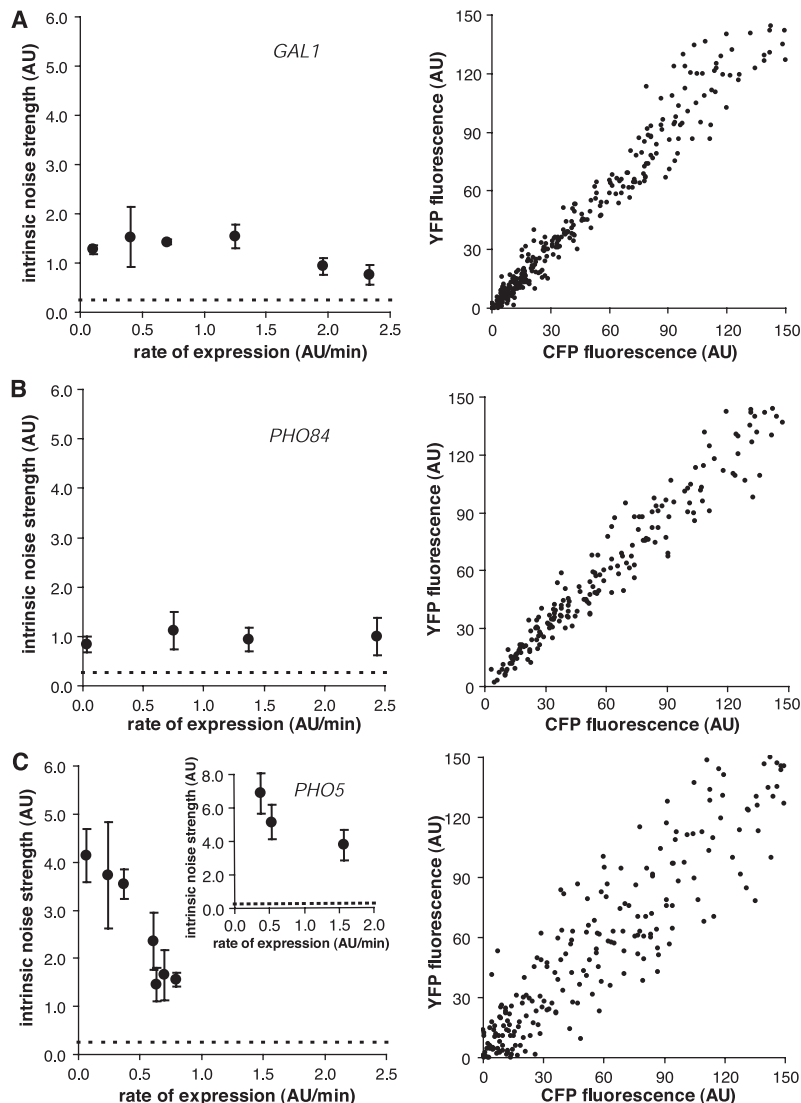
rate of gene expression (Fig. 2C); at maximal expression, *PHO5* displays less than half the intrinsic noise strength that it does at a low expression rate. This decrease in intrinsic noise strength as the rate of gene expression increases is not sensitive to the stimulus used to induce *PHO5* expression (Fig. 2C, inset). We conclude that noise intrinsic to gene expression is promoter-specific and does not depend absolutely on the rate of expression, the induction stimulus, or the identity of the sequence-specific transcriptional activator.

A previous model of noise generation in gene expression predicted that noise

strength would not vary with a change in the rate of mRNA production (7, 8), similar to our observations of *GAL1* and *PHO84*. This model does not predict the noise profile of the *PHO5* promoter, which is known to be regulated by a promoter transition step that is upstream of and independent of transcription (9). We constructed a model of stochastic gene expression, elaborating on previous models (5, 10–13), that incorporates two distinct promoter states: an inactive state not permissive for transcription and an active state that is competent for transcription (Fig. 3A).

We distinguish among three kinetic mechanisms of promoter transcriptional activation (Fig. 3B). In case I, the activation step is infrequent relative to transcription and the active promoter state is stable ( $k_a, \gamma_a \ll k_m$ , where  $k_a$  is the rate of promoter activation,  $\gamma_a$  is the rate of promoter inactivation, and  $k_m$  is the rate of transcription). This could correspond to a promoter that is activated by a slow chromatin-remodeling step in which positioned nucleosomes are removed from the DNA and that is slowly inactivated by replacement of the nucleosomes. In case II, the activation step is infrequent relative to transcription and the active promoter state is unstable ( $k_a \ll \gamma_a, k_m$ ), corresponding to a relatively infrequent but rapidly reversible activation step such as nucleosomal sliding or prokaryotic promoter DNA looping. In case III, the activation step is frequent relative to transcription and the activated promoter is highly unstable ( $k_m \ll k_a, \gamma_a$ ). The third case is equivalent to the previously proposed prokaryotic model (8) and could represent rapid activator binding-dissociation reactions in which transcription occurs only for a fraction of the binding events.

We calculated the solution of the noise strength equation in terms of the stochastic kinetic constants of the model (supporting online text) and also performed stochastic simulations to approximate our experiments in which each intrinsic noise measurement is the average of multiple time points of finite cellular populations (6). We examined how varying the kinetic constants in the model to change the steady-state mean of gene expression affects the intrinsic noise strength for the three different cases (Fig. 3, C to F). The noise strength profile of *PHO5* (Fig. 2C) is similar to the predictions made for case I when the promoter activation rate is changed (Fig. 3C). Therefore, we hypothesized that noise generation at *PHO5* is dependent on the rate of a slow upstream promoter transition. In the inactive state, the *PHO5* promoter displays positioned nucleosomes (14); upon binding of the Pho4



**Fig. 2.** (A) Intrinsic noise strength as a function of rate of expression for *GAL1* (left) and a scatter plot of *GAL1*-expressing cells at maximal induction (right). To produce different rates of expression, cells were induced with different galactose concentrations. (B) Intrinsic noise strength as a function of rate of expression for *PHO84* and a scatter plot of *PHO84*-expressing cells at maximal induction. Cells were induced with different phosphate concentrations. (C) Intrinsic noise strength as a function of rate of expression for *PHO5* and a scatter plot of *PHO5*-expressing cells at the maximal level of induction by phosphate starvation. Cells were induced with various levels of chemical inhibition of an upstream kinase (left) or with various organic phosphate concentrations (inset). The dashed line indicates the intrinsic error of measurement, and error bars represent standard deviations. Scatter plots contain cells from multiple time points.

transcription factor to upstream activating sequences UAS1 and UAS2, chromatin-remodeling complexes are recruited (15, 16) and catalyze removal of nucleosomes from the promoter region (17, 18). Mutation of these UAS sites causes a defect in the disruption of positioned nucleosomes during activation (9). As predicted by the model (case I, decreasing  $k_a$ ), the two *PHO5* UAS mutant promoters had increased intrinsic noise strength compared to the wild-type *PHO5* promoter (Fig. 4A). This observation is consistent with a model in which chromatin remodeling is the upstream stochastic promoter transition for the *PHO5* promoter.

Multiple chromatin-remodeling complexes, including SWI/SNF, INO80, and SAGA, participate in remodeling at the *PHO5* promoter (15, 16, 19, 20). To test further the hypothesis that chromatin remodeling is the stochastic promoter activation step, we examined the noise strength of the maximally induced *PHO5* promoter in yeast strains that lack single components of these three chromatin-remodeling complexes (Fig. 4B). The deletion of components of SWI/SNF (*snf6Δ*), INO80 (*arp8Δ*), or SAGA (*gcn5Δ*) each resulted in increased intrinsic noise strength, consistent with the predictions of the noise model. There were substantial differences in the noise strength among the mutants, which may reflect different roles of these complexes in the promoter transition process.

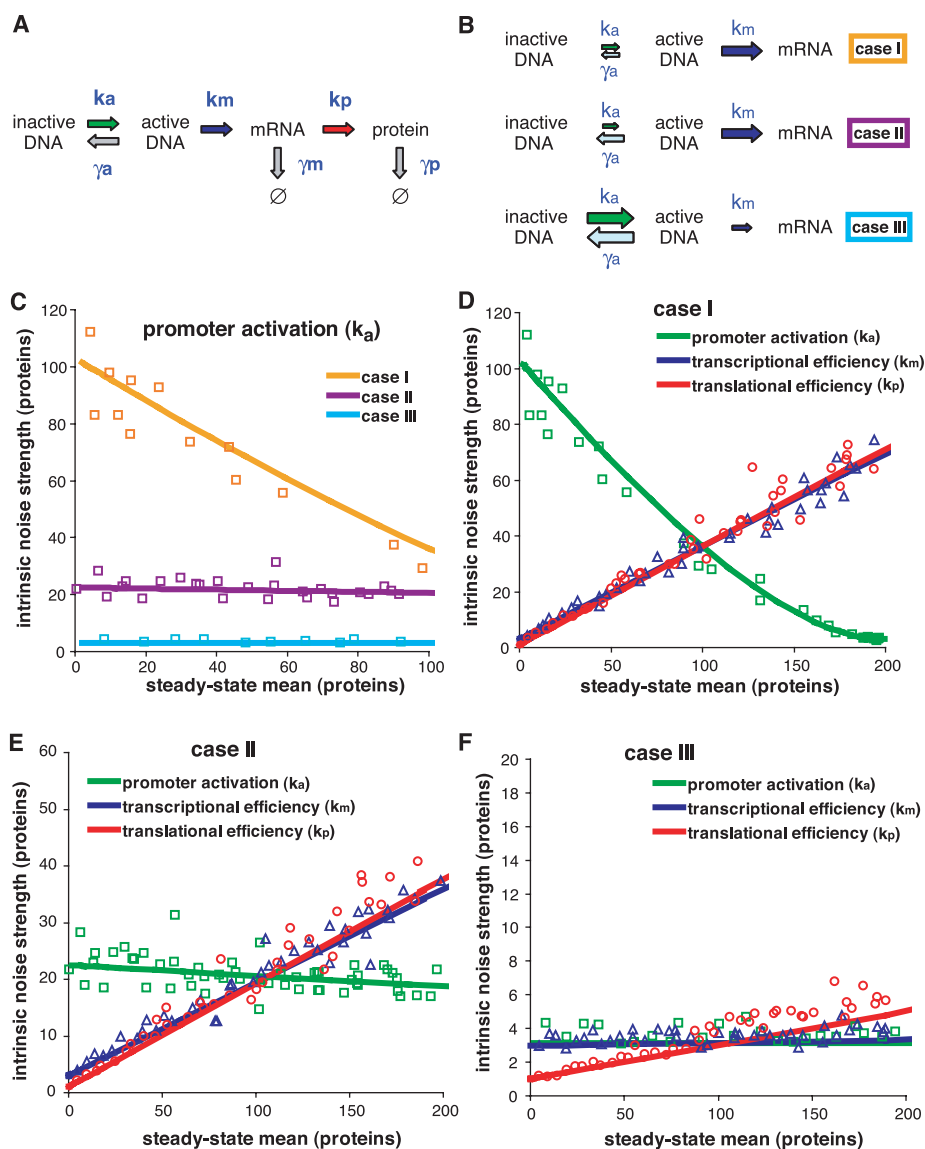
The TATA element of the *PHO5* promoter is required for efficient transcription but dispensable for chromatin remodeling (9). To confirm the prediction that noise strength should scale with the efficiency of a transcription step downstream of promoter activation, we measured, at maximal induction, the noise strength of a series of *PHO5* promoters with various TATA box sequences (Fig. 4C). As predicted, the mutant TATA box promoters displayed decreasing noise strength with a decreasing rate of gene expression. These observations suggest that our model represents a useful framework for the rationalization of noise generation at the *PHO5* promoter.

Our results support a model for noise generation that is applicable to both eukaryotic and prokaryotic promoters, in which the relative rates of activation, deactivation, and transcription determine variability in mRNA levels. Our noise measurements and model contradict the previous assertion that noise generation in gene expression is fundamentally different between prokaryotes and eukaryotes (13). Rather, we assert that the diverse mechanisms of gene regulation in these systems fall into three general categories of noise profiles, dependent upon the relative rates of promoter reactions. Two promoters can

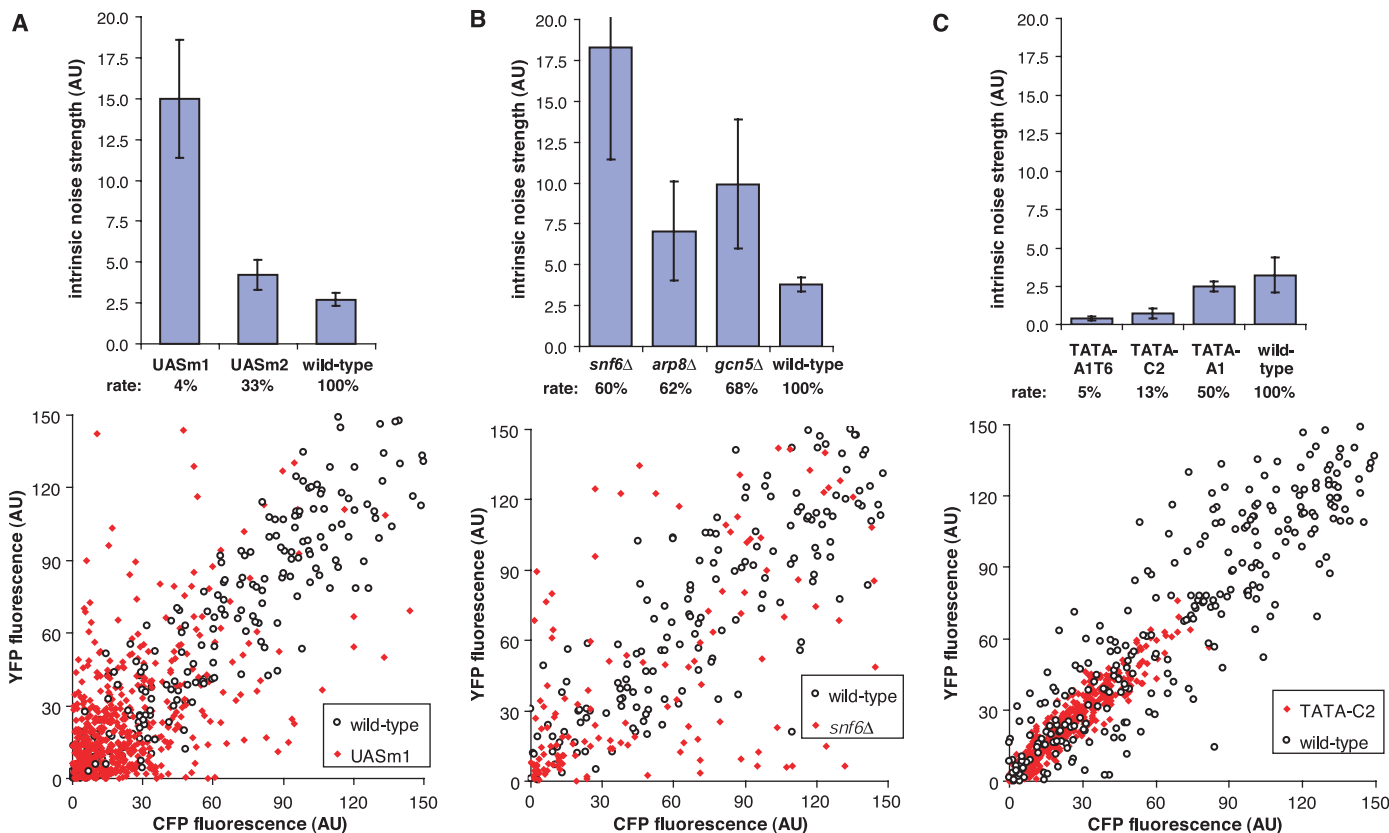
produce the same mean mRNA population with different noise characteristics: a promoter that undergoes frequent activation steps followed by inefficient transcription will produce a cellular population with little variability, whereas a promoter that undergoes infrequent activation steps followed by efficient transcription can display large differences from cell to cell. We have identified simple sequence mutations in the *PHO5* promoter that exemplify these two extremes: the UASm1 and TATA-A1T6 variants have similar rates of gene expres-

sion on a population level but different levels of stochasticity—a more than 30-fold change in intrinsic noise strength. Because the intrinsic noise is altered by small changes in promoter sequence and independent of the absolute rate of gene expression, we speculate that stochasticity is an evolvable characteristic of each eukaryotic gene, determined by both *cis*- and *trans*-acting factors.

Stochasticity intrinsic to gene expression has been invoked as a source of phenotypic variation in the lambda phage lysis-



**Fig. 3.** General stochastic model of gene activation and expression. (A) Schematic of reactions with stochastic rate constants of production ( $k$ ) and degradation ( $\gamma$ ).  $\emptyset$  indicates the null product of degradation. a, active DNA; m, mRNA; p, protein. (B) Three cases, where the relative size of the arrows indicates the relative magnitude of the constants within each case. (C) The effect on intrinsic noise strength of changing the promoter activation rate to change the steady-state mean of expression for case I (orange  $\square$ ), case II (violet  $\square$ ), and case III (cyan  $\square$ ). (D to F) The effect on intrinsic noise strength of changing promoter activation (green  $\square$ ), transcriptional efficiency (red  $\circ$ ), and translational efficiency (purple  $\triangle$ ) to change the steady-state mean of expression for (D) case I, (E) case II, and (F) case III. For (C) to (F), the solid lines display the predicted values, and the open symbols are averages from stochastic simulations (6).



**Fig. 4.** Mutational analysis of the *PHO5* promoter. (A) Intrinsic noise strength and rate of expression of wild-type, UASm1, and UASm2 *PHO5* promoter variants at maximal induction (top) and a scatter plot for the wild type (○) and UASm1 (red ◆) (bottom). The intrinsic noise of the UASm1 promoter is 77% at a mean of 26 AU, compared to 43% intrinsic noise at a mean of 29 AU for the wild-type promoter. (B) Intrinsic noise strength and rate of expression of the *PHO5* promoter at maximal induction in the wild-type background or in strains that

lack *SNF6*, *ARP8*, or *GCN5* (top) and a scatter plot for the wild type (○) and *snf6Δ* (red ◆) (bottom). The intrinsic noise of the *snf6Δ* strain is 73% at a mean of 35 AU. (C) Intrinsic noise strength and rate of expression of wild-type and TATA mutant *PHO5* promoters at maximal induction (top) and a scatter plot for the wild type (○) and TATA-C2 (red ◆) (bottom). The intrinsic noise of the TATA-C2 promoter is 15% at a mean of 25 AU. Error bars represent standard deviations. Scatter plots contain cells from multiple time points.

lysogeny switch (21), mammalian olfactory neuronal receptor choice (22), and tumor formation in response to transcription factor haploinsufficiency (23). The variability of gene expression due to stochasticity in promoter transitions is predicted to be proportional to the inverse square root of the gene copy number, providing a rationale for the preservation of multiple copies of the same gene and the effects of haploinsufficiency when variability is deleterious. If the majority of extrinsic noise is not gene-specific, stochasticity will create variability in the ratio of one gene product to the other within a population; the preservation of the stoichiometric ratio between gene products may be important for components of multisubunit complexes (24). Additionally, stochasticity may cause variation in the ratio of expression of two alleles with distinct functions, allowing a heterozygous population of cells to display multiple phenotypes, including those of each corresponding homozygote. Such phenotypic variability may be beneficial in a variable environment and may contribute to the phenomenon of hybrid vigor. Stochas-

ticity in gene expression is not necessarily an obstacle to invariant cellular behavior but may constitute an evolvable source of advantageous population diversity.

**References and Notes**

1. H. H. McAdams, A. Arkin, *Trends Genet.* **15**, 65 (1999).
2. E. Schroedinger, *What Is Life?* (Cambridge Univ. Press, Cambridge, 1944).
3. C. V. Rao, D. M. Wolf, A. P. Arkin, *Nature* **420**, 231 (2002).
4. M. B. Elowitz, A. J. Levine, E. D. Siggia, P. S. Swain, *Science* **297**, 1183 (2002).
5. P. S. Swain, M. B. Elowitz, E. D. Siggia, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12795 (2002).
6. Materials and methods are available as supporting material on Science Online.
7. E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, A. van Oudenaarden, *Nature Genet.* **31**, 69 (2002).
8. M. Thattai, A. van Oudenaarden, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8614 (2001).
9. K. D. Fascher, J. Schmitz, W. Horz, *J. Mol. Biol.* **231**, 658 (1993).
10. J. Peccoud, B. Ycart, *Theor. Pop. Biol.* **48**, 222 (1995).
11. J. Paulsson, *Nature* **427**, 415 (2004).
12. M. S. Ko, *J. Theor. Biol.* **153**, 181 (1991).
13. W. J. Blake, M. Kaern, C. R. Cantor, J. J. Collins, *Nature* **422**, 633 (2003).
14. A. Almer, W. Horz, *EMBO J.* **5**, 2681 (1986).
15. D. J. Steger, E. S. Haswell, A. L. Miller, S. R. Wentz, E. K. O'Shea, *Science* **299**, 114 (2003).
16. S. Barbaric, H. Reinke, W. Horz, *Mol. Cell. Biol.* **23**, 3468 (2003).

17. K. D. Fascher, J. Schmitz, W. Horz, *EMBO J.* **9**, 2523 (1990).
18. H. Boeger, J. Griesenbeck, J. S. Strattan, R. D. Kornberg, *Mol. Cell* **11**, 1587 (2003).
19. R. Ebbert, A. Birkmann, H. J. Schuller, *Mol. Microbiol.* **32**, 741 (1999).
20. S. Barbaric, J. Walker, A. Schmid, J. Q. Svejstrup, W. Horz, *EMBO J.* **20**, 4944 (2001).
21. A. Arkin, J. Ross, H. H. McAdams, *Genetics* **149**, 1633 (1998).
22. S. Serizawa *et al.*, *Science* **302**, 2088 (2003).
23. J. A. Magee, S. A. Abdulkadir, J. Milbrandt, *Cancer Cell* **3**, 273 (2003).
24. H. B. Fraser, A. E. Hirsh, G. Giaever, J. Kumm, M. B. Eisen, *PLoS Biol.* **2**, e137 (2004); published online 27 April 2004 (10.1371/journal.pbio.0020137).
25. We thank J. Paulsson, D. Wykoff, J. Weissman, H. Li, R. Kishony, D. Schaffer, H. Madhani, W. Marshall, and members of the O'Shea lab for discussion and critical commentary. Supported by the Howard Hughes Medical Institute, NIH grant no. GM51377, and the David and Lucile Packard Foundation (E.K.O.) and by an American Diabetes Association Medical Scholars fellowship and the UCSF Medical Scientist Training Program (J.M.R.).

**Supporting Online Material**

www.sciencemag.org/cgi/content/full/1098641/DC1  
 Materials and Methods  
 SOM Text  
 Figs. S1 to S5  
 References and Notes

1 April 2004; accepted 13 May 2004  
 Published online 27 May 2004;  
 10.1126/science.1098641  
 Include this information when citing this paper.