LETTERS

Origins of extrinsic variability in eukaryotic gene expression

Dmitri Volfson[1,2]*, Jennifer Marciniak[1]*, William J. Blake[3], Natalie Ostroff[1], Lev S. Tsimring[2] & Jeff Hasty[1]

Variable gene expression within a clonal population of cells has been implicated in a number of important processes including mutation and evolution[1,2], determination of cell fates[3,4] and the development of genetic disease[5,6]. Recent studies have demonstrated that a significant component of expression variability arises from extrinsic factors thought to influence multiple genes simultaneously[7–10], yet the biological origins of this extrinsic variability have received little attention. Here we combine computational modelling[11–18] with fluorescence data generated from multiple promoter–gene inserts in *Saccharomyces cerevisiae* to identify two major sources of extrinsic variability. One unavoidable source arising from the coupling of gene expression with population dynamics leads to a ubiquitous lower limit for expression variability. A second source, which is modelled as originating from a common upstream transcription factor, exemplifies how regulatory networks can convert noise in upstream regulator expression into extrinsic noise at the output of a target gene[9]. Our results highlight the importance of the interplay of gene regulatory networks with population heterogeneity for understanding the origins of cellular diversity.

To investigate variability in eukaryotic gene expression, we used the native *GAL1* promoter of the yeast *Saccharomyces cerevisiae* with *yEGFP* (yeast-enhanced green fluorescent protein) as a quantifiable marker. We constructed five yeast strains by varying the number of copies integrated into the *GAL1-10* locus on chromosome II. These multiple-copy constructs can be used to determine whether the variability in gene expression is due to intrinsic or extrinsic sources. Adopting the standard nomenclature, intrinsic noise originates from the small number of regulatory molecules participating in the inherently noisy biochemical reactions leading to expression, whereas extrinsic variability arises from sources such as the variation of the elements of the transcriptional machinery common to all genes or fluctuating environmental variables. This extrinsic variability can have stochastic or deterministic origins, and we reserve the use of the term noise to denote a random, as opposed to a deterministic, origin for the variations.

In the absence of glucose, the *GAL1* promoter is activated in response to galactose, and our cell strains were induced to produce *yEGFP* using varying amounts of galactose ranging from 0.1% to 2%. Single-cell fluorescence data were collected using flow cytometry, and representative distributions are depicted in Fig. 1a for fixed galactose concentration. Information regarding the underlying transcriptional process can be extracted by analysing the scaling properties of the data sets with respect to the copy number. For example, the mean fluorescence increases as a function of galactose concentration and the copy number (Fig. 1c), and when fluorescence is divided by the corresponding copy number, the induction curves for all of the strains 'collapse' to a single curve (Fig. 1c, d). This indicates that each promoter–gene pair transcribes at the same mean rate upon insertion.

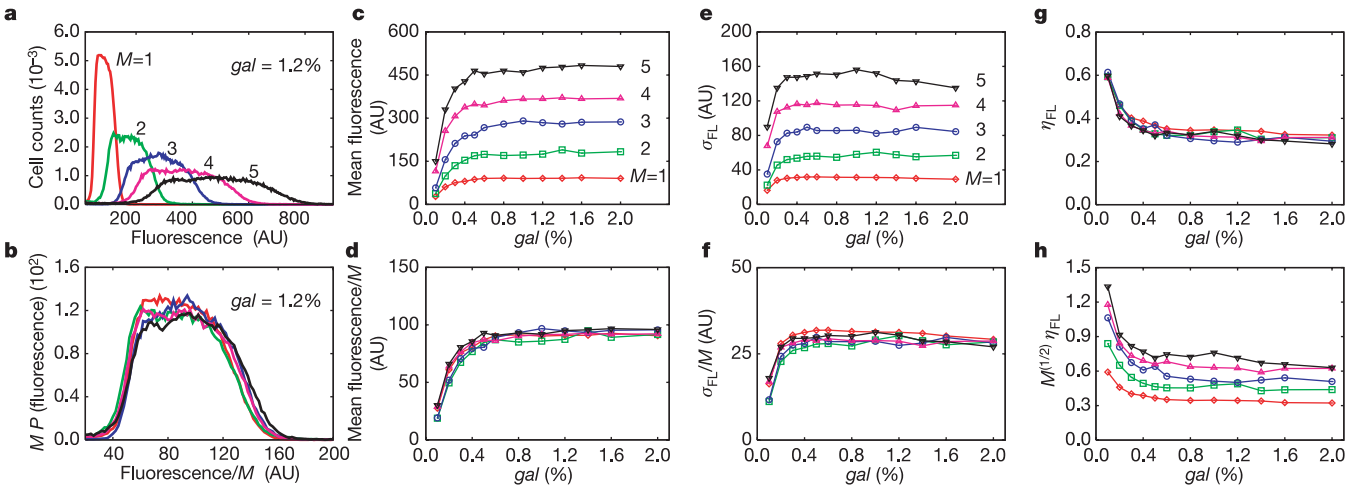


Figure 1 | Experimental results for GFP expression for different copy numbers and galactose concentrations. **a**, Histogram of GFP measurements for copy numbers from $M = 1$ to $M = 5$ above the saturation ($gal = 1.2\%$). **b**, The collapse of GFP distributions under the transformation $F \to F/M$, $P(F) \to MP(F/M)$ implies an extrinsic source of variability. AU, arbitrary units. **c**, Induction curves for copy numbers from $M = 1$ to $M = 5$. **d**, Collapse of the induction curves implies that transcription from each promoter is independent. **e**, Standard deviations of GFP corresponding to induction curves. **f**, The collapse of the standard deviation implies an extrinsic source of variability. **g**, The collapse of the coefficient of variation for different copy number implies an extrinsic source of variability. **h**, Lack of collapse implies that the variability is not of intrinsic origin.

[1]Department of Bioengineering, and [2]Institute for Nonlinear Science, University of California San Diego, La Jolla, California 92093, USA. [3]Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215, USA.
*These authors contributed equally to this work.

We can use scaling in an analogous manner to deduce information about the origins of variability in the system. Depending on the relative magnitude of intrinsic noise versus extrinsic variability, we expect to observe different dependencies (scalings) on the copy number M.

In order to illustrate how scaling can be used to determine whether the observed variability is intrinsic or extrinsic, suppose there are M promoter–gene pairs producing GFP molecules, and let the level of production of the ith copy be the sum of two terms representing the mean and fluctuations about the mean: $g_i = \langle g_i \rangle + \tilde{g}_i$. The time-average of the fluctuating term is zero ($\langle \tilde{g}_i \rangle = 0$) and the variance is given by $v_i = \langle g_i^2 \rangle - \langle g_i \rangle^2$. Then the mean and variance for the total amount of GFP produced by M copies are $\langle G_M \rangle = \sum_{i=1}^{M} \langle g_i \rangle$ and $V_M = \sum_{i=1}^{M} \left[v_i + \sum_{j \neq i} \langle \tilde{g}_i \tilde{g}_j \rangle \right]$, respectively. The first term in the expression for V_M represents the sum of the variances of the individual copies, and the second term is the contribution of cross-correlations between them. For identical copies, $\langle g_i \rangle = \langle g \rangle$, $v_i = v$, $\langle \tilde{g}_i \tilde{g}_j \rangle = c$ and the total mean is proportional to the copy number, $\langle G_M \rangle = M \langle g \rangle$, whereas the total variance is given by $V_M = Mv + M(M-1)c$. When $c = 0$, the fluctuations are completely uncorrelated and the variance is proportional to the copy number. This is the intrinsic noise limit, where expression events from individual copies are completely independent. When $c = v$, the variance is proportional to the square of the copy number. In this limit the fluctuations are completely correlated and often caused by extrinsic factors. Using the results for the variance and mean, the coefficient of variation η (standard deviation divided by the mean) is expected to scale as $\eta \propto M^{-1/2}$ for the case of purely intrinsic noise and be independent of the copy number in the case of extrinsic variability. Thus, if the system is dominated by intrinsic factors, the quantity $M^{1/2}\eta$ should collapse to the same galactose-dependent curve for all copy numbers. On the other hand, if the system is dominated by extrinsic sources, then η itself should collapse for all copy numbers.

In order to ascertain whether the GAL system is dominated by intrinsic noise or extrinsic variability, we plot the unscaled (η) and scaled ($M^{1/2}\eta$) coefficient of variation as a function of galactose (Fig. 1g, h). We observe that even though different strains exhibit small differences, the collapse of the data in the unscaled case indicates that the coefficient of variation is independent of the copy number. An even more striking collapse is presented in Fig. 1b, where we rescale fluorescence distributions $P(F)$ for a representative value of galactose concentration above saturation ($gal = 1.2\%$) and for different strains. In obtaining this collapse, we first normalized the distributions and rescaled by the copy number as follows: $F \rightarrow F/M$, $P(F) \rightarrow MP(F/M)$. This collapse implies that not only is the standard deviation proportional to the copy number, but generally the nth moment of the fluorescence distribution is proportional to M^n. These results corroborate findings by a recent study[8] that used a 'two-colour' approach[7] to demonstrate that extrinsic sources dominate the variations observed in the GAL system. In the Supplementary Information, we demonstrate that our findings are unchanged when the genes are moved to different chromosomes, and we describe how our data analysis can be generalized for intermediate cases where there is mixture of scaling arising from the contribution of both intrinsic and extrinsic sources.

Our analysis of the experimental data provides an important constraint when attempting to deduce the biological origin of the variability, because a non-trivial test of any model description is that it must exhibit correct copy-number scaling. We first construct a fully deterministic model that couples a mesoscopic description of cell growth and division with a microscopic description of gene expression (Box 1, model A). For colony growth, we generalize the classical model introduced by ref. 19, which is known to capture the major experimental observations regarding the cell size distributions. We assume that each cell grows exponentially at a rate α and divides after reaching the same size $x0$. The division is asymmetric so that the size of a mother cell after division is ξ/κ larger than that of a daughter cell ($\xi + \kappa = 1$, $\xi > \kappa$). After a transient, such a colony reaches a stationary distribution of sizes with two characteristic peaks corresponding to subpopulations of daughter and mother cells (see Fig. 2a, where the theoretical distribution is compared with experimental data obtained from the flow cytometry forward scatter data). In addition to the process of growth and division, model A assumes that each cell is producing GFP at a constant rate γ, and GFP is distributed at division in the same proportion as the volume. We show in the Supplementary Information that the joint probability of finding a cell of a given size x, and with GFP content g, has the functional form $\mathscr{P}_{vg}(x,g) = \mathscr{P}_v(x)\mathscr{P}_g(g/a,\kappa)$, where $a = \gamma/\alpha$ is the ratio of the GFP production rate and volume growth rate. Using this result, we then derive an explicit expression for the distribution $\mathscr{P}_g(g/a,\kappa)$ (see Fig. 2b), and from there generate expressions for the mean and standard deviation of the stationary distribution of GFP, $\langle g \rangle = a$ and $\sigma_g = a\eta_g(\kappa)$, where $\eta_g(\kappa) = [1 - \frac{2}{3\kappa\xi}(\kappa^2 \ln(\kappa)^2 + \kappa\xi \ln(\kappa)\ln(\xi) + \xi^2 \ln(\xi)^2]^{1/2}$.

The model results for the mean and standard deviation have non-trivial implications consistent with our experimental findings. Because the rate of GFP production $\gamma = a\alpha$ is proportional to the copy number M, the functional form $\mathscr{P}_g(g/a,\kappa)$ implies that distributions of GFP for different copy numbers M should collapse after the transformation $g \rightarrow g/M$, $\mathscr{P}_g \rightarrow M\mathscr{P}_g$. In particular, because the mean value of the GFP distribution is proportional to the copy number and the standard deviation is proportional to the mean, the coefficient of variation is independent of M as observed in the experiments. In addition, the model yields an explicit dependence of the coefficient of variation on the partition ratio κ (see Supplementary Information), and using a previously reported value for this parameter[20] we obtain a coefficient of variation that is in close quantitative agreement with our experiments for high concentrations of galactose (Fig. 2e, straight line). Importantly, this

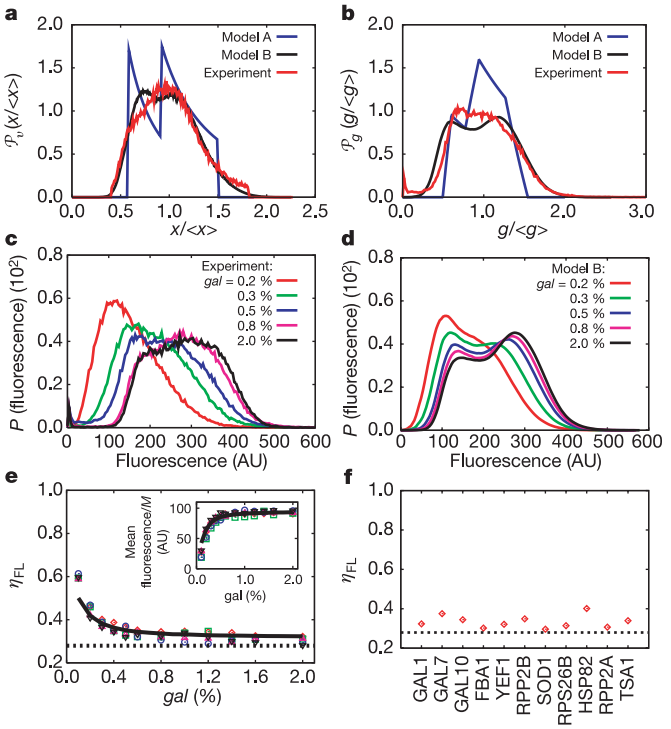


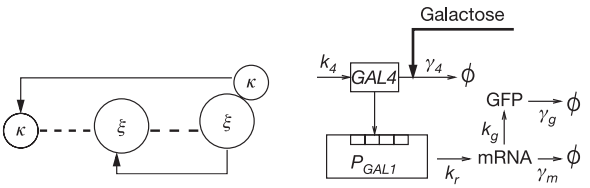
Figure 2 | Comparison of models A and B with experimental findings. **a**, Distributions of cell sizes. **b**, Distributions of GFP at fixed galactose concentration and $M = 1$. **c**, Distributions of GFP for $M = 3$ and for a number of galactose concentrations. **d**, The same as **c** but from model B. **e**, Collapse of the induction curves (symbols) along with the results of simulations of model B (black solid line). **f**, Validation of model predictions. The model noise floor (dashed black line) along with experiments using different fusions (symbols) are shown.

Box 1 | Coupling gene expression to population dynamics

Our two models are constructed by coupling a subsystem for gene expression from the *GAL1* promoter to a subsystem representing the growth and division of individual cells in the population. For the gene expression subsystem, we introduce a simple scheme analogous to the previously reported model[15]. In our model each of the M identical promoters may be in either an inactive state (O_1) or an active one (O_2). The latter represents a preinitiation complex with all the components necessary for transcription in place. The production of mRNA proceeds from state O_2 with rate k_r. After mRNA has been produced, translation ensues with rate k_g. Both mRNA and GFP are allowed to decay at rates γ_m and γ_g, respectively. The effect of the inducer is modelled by taking into account the presence of GAL4p proteins, which are responsible for the activation of the P_{GAL1} promoter. These proteins dimerize and bind to four binding sites of P_{GAL1}, activating the process of transcription. When multiple copies of the promoter–gene pair are inserted, they share the pools of Gal4p, mRNA and GFP. We assume that Gal4p monomers are constitutively produced with rate α_4. Another protein (Gal80p) binds to Gal4p and renders it unable to activate the promoter. We model this by taking the degradation rate of Gal4p (γ_4) to be inversely proportional to the galactose concentration. Combining these assumptions, the differential equations describing the deterministic model for gene expression may be deduced from the chemical reactions. The rate equations for Gal4p, mRNA and GFP, respectively, read.

$$\dot{g}_4 = \alpha_4 - \gamma_4 g_4, \quad \dot{r} = Mk_r H(g_4) - \gamma_m r, \quad \dot{g} = k_g r - \gamma_g g$$

where $H(g_4) = Kg_4^z/(1+Kg_4^z)$ is the Hill function for the activation process, and we use the cooperativity $z = 8$, which stems from the binding of four Gal4p dimers; K is a lumped parameter that describes the net effect of dimerization/dissociation and binding/unbinding; and the dependence on the galactose concentration arises from $\gamma_4 \propto gal^{-1}$ (see Supplementary Information for a description of our modelling methods, along with a derivation of this model from a more general model of the galactose signalling pathway).



For model A, our primary motivation is to elicit the extrinsic variability arising from cell growth and division, so we first develop a deterministic model for the gene expression component. The characteristic timescale here is set by the growth rate of the cells. The stability of GFP implies a negligible degradation rate γ_g, and because the rates of degradation of Gal4p and mRNA are fast compared with the cellular growth rate, we can eliminate the dynamical equations for g_4 and m. Using these considerations, the dynamics of GFP are then reduced to production at a constant rate, $\dot{g} = \Gamma = M[k_g k_r H(\alpha_4/\gamma_4)/\gamma_m]$.

We next couple a deterministic model for cellular growth and division with the deterministic production of GFP. To account for population dynamics we consider an asynchronous population of cells, each growing exponentially with rate α and dividing in an asymmetric fashion[19] after reaching a certain maximum size. We let the size of a mother cell after division be ξ/κ times larger than that of the daughter cell, where $\kappa < 1/2 (\xi + \kappa = 1)$. The process of growth and division repeats indefinitely, producing a growing structured population of cells of multiple generations where each cell is described by a position within the cell cycle and genealogical age. The coupling between this growth model and gene expression arises through the partitioning of GFP content at division. The system of evolution equations describing the size (x) and number of GFP molecules (g) is given by

$$\dot{x} = \alpha x, \quad \dot{g} = \Gamma$$

which should be supplemented by the renewal equations (division rules), according to $x_0 \rightarrow \kappa x_0 + \xi x_0, g(x_0) \rightarrow \kappa g(x_0) + \xi g(x_0)$ where x_0, $g(x_0)$ are the cell size and GFP content at the moment of division, respectively. The additional assumptions behind this model may be summarized as follows. We distinguish only two generations of cells—daughters and mothers—thus neglecting the difference between mothers of second generation and higher. We assume that cells are growing with the same rate regardless of their age, and that both daughters and mothers divide after reaching the same size and in the same proportion. Finally, we assume that GFP is distributed among daughter cells in the same proportion as the size (fast diffusion of GFP molecules). Overall, the statistics of the population is modelled in terms of the equation for the joint probability $\mathscr{P}_{vg}(x, g)$ to find a cell of a given size x and with a certain amount of GFP g, which is solved analytically (see Supplementary Information).

Model B generalizes model A by including several realistic features of both gene expression and population dynamics. For gene expression we use a hybrid approach whereby production and degradation of Gal4p proteins is modelled stochastically using the Gillespie algorithm[24], whereas production of mRNA and GFP is simulated with the set of ordinary differential equations introduced above. Consistent with our experimental observations, our hybrid approach eliminates all intrinsic fluctuations except for those induced by Gal4p monomers. This intrinsic noise is then transferred to the level of mRNA production through fluctuations in the production rate of mRNA, which is proportional to the Hill function $H(g_4)$. Here, the intrinsic noise manifests itself as extrinsic fluctuations common to all M copies. For the population dynamics we introduce genealogical age by using different values of κ depending on the generation[20], and introduce a stochastic component in division by drawing a cell size at division from a narrow gaussian distribution centred at x_0. In addition, upon division the amounts of Gal4p, mRNA and GFP are distributed among the offspring in a binomial fashion.

result was obtained without any additional fitting (in this model, the coefficient of variation is shown to be a function of κ only). As an additional confirmation of the role of the population dynamics in genetic variability we calculated the coefficient of variation in gated subpopulations of cells within narrow windows of sizes and found that gating reduces the coefficient of variation by as much as 50% (see Supplementary Information).

Because the variability in the model arises from the coupling of population effects and deterministic gene expression, an interesting model prediction is that there is a 'noise floor' that provides the lower limit for expression variability. If there are no other sources of variability, this implies that for various highly expressed genes the observed variability should be similar. We tested this prediction by constructing GFP fusions to ten highly expressed *S. cerevisiae* genes. As predicted, we observed a near-constant level of variability for all of the constructed fusions (Fig. 2f). This level is consistent with the lower noise floor observed in the GAL system (Fig. 2e), and can be taken as the lower limit for the variability in a growing population of cells. These results suggest that for high expression levels, population effects dominate and gene expression within a single cell is mostly deterministic.

Although our fully deterministic model A accounts for the majority of the observed variations in the inducible GAL1–GFP system, there are still certain limitations of this simple model. In particular, the theoretical GFP distribution shows a large peak that is not present in the experimental data (Fig. 2b). It occurs because model A ignores the stochastic fluctuations in the cell size at division, as well as the binomial nature of division of GFP among mother and daughter cells, and thus the GFP distribution is narrow. Furthermore, model A does not account for the increase of variability at small galactose values. We next generalize the model to describe the variability at small galactose values (model B; see Box 1). In this regime, the variability increases as the number of GFP molecules decreases. Although this dependence is consistent with an intrinsic noise source,

the scaling of the coefficient of variation with copy number implies an extrinsic source, and thus fluctuations in expression from uncorrelated *GAL1* promoters cannot be the origin. We therefore look upstream of the *GAL1* promoter. The *GAL1* promoter is activated by Gal4p protein dimers, so a common stochastic component of gene expression may originate from fluctuations in the level of the Gal4p activators. We tested the idea of activator-mediated fluctuations with a quantitative model incorporating the galactose-dependent degradation of Gal4p (Fig. 2a–c). The generalized model shows excellent agreement for the distributions and scaling of the coefficient of variation as a function of galactose. In addition, model B predicts that other Gal4p-regulated promoters should exhibit extrinsic scaling at low galactose, and we present experimental validation of this prediction in the Supplementary Information.

We have shown how experimental classification of variability can be used to constrain the space of possible models and lead to the development of a model incorporating variations arising from the coupling of population growth and gene expression. Although this dominant source of variability is likely to exist in all systems, the generalization of our model to include upstream activators of the *GAL1* promoter is system-specific. It would be interesting to explore other inducible eukaryotic systems to see whether the scaling of the variability also implies an extrinsic source for low induction levels. In the context of the reported correlations observed in the transcripts of proximally located genes[21], our results suggest that the biological origin of these correlations could be a common upstream transcriptional regulator.

METHODS

Plasmids, yeast strains and growth conditions. All plasmid backbones were derived from pRS403, pRS404, pRS405 or pRS406 shuttle vectors (Stratagene). The *GAL1-10* promoter region and the *yEGFP* were obtained from pESC1-yG. All yeast strains (see Supplementary Table II) in this study are derived from the *S. cerevisiae* parent strain YPH500 (α, *ura3-52*, *lys2-801*, *ade2-101*, *trp1Δ63*, *his3Δ200*, *leu2Δ1*) (Stratagene). The strains were created by targeted chromosomal integration of shuttle vector constructs at either the *GAL1-10* locus on chromosome II or the *TRP1* locus on chromosome IV. All fusion strains were constructed as described previously[22] (see also Supplementary Information). The yeast strains were grown in synthetic drop-out (SD) medium supplemented for selection of correct integrands containing 2% glucose at 30 °C. The strains were selected for the proper number of promoter–gene pairs by growing the cells on 2% galactose and assaying the cells for average fluorescence. All cloning steps were performed in *Escherichia coli* XL10-Gold (Stratagene).

Gene expression experiments. Exponentially growing yeast cells were diluted from 1:40 to 1:70 into medium containing 2% raffinose and a range of galactose concentrations (0.1–2.0% galactose) as inducer. After 14–18 h, the cells were assayed at mid-log phase at an absorbance at 600 nm of 055 ± 0.15. Expression data were collected using a Becton-Dickinson FACSCalibur flow cytometer with a 488-nm argon excitation laser and a 515- to 545-nm emission filter (FL1) at a low flow rate. Forward scatter values and fluorescence values were collected for 100,000 cells. The standard list-mode files obtained from the flow cytometer were converted to ASCII format with MFI (E. Martz, http://www.umass.edu/microbio/mfi) and analysed using Matlab (The MathWorks, Inc.).

Modelling studies. Model A was solved analytically; details of the calculations are presented in the Supplementary Information. For the comparison with the experiment we used an effective ratio of the volumes of daughter cells after division to mother cells before the division, $\kappa = 0.34$, representing the weighted average of κ among several generations[20] (Fig. 1g). Model B was solved numerically using a hybrid technique[23] whereby the stochastic dynamics of Gal4p protein have been modelled with the Direct Gillespie algorithm[24], the intermediate reactions of dimerization and binding of Gal4p dimers were eliminated using quasi-stationary assumptions for the sake of simplicity, and the dynamics of mRNA and GFP were simulated using the rate equations presented in Box 1. For the population dynamics we used previously reported measurements regarding the structure of the population of *S. cerevisiae*[20]. Namely, for the first generation, we used $\kappa = 0.4$ and for subsequent generations $\kappa = 0.3$; the mean size of the cells at divisions was chosen to be 1.089, 1.179 and 1.268 for genealogical ages from 2 to 4, respectively, and 1.357 for all older generations; the size at division for a particular cell was drawn from a narrow gaussian distribution near these mean values with coefficient of variation 0.15. The parameters of the simulations in units of average growth rate are: $k_r = 80$, $\gamma_m = 4$, $k_g = 4$, $\gamma_g = 0$, $k_4 = 8$, $\gamma_4 = 0.5ga1^{-1}$, $K = 0.05$. The protocol of the simulations closely resembled the experimental procedure. We started from a small collection of cells (typically 10^3), with gaussian distributions of sizes and quantities of GAL4p, mRNA and GFP near the expected mean values. This small population was 'grown' until it reached 100,000 cells. The state of the population was recorded within a narrow time window right after that. We checked that this measurement was consistent with the time-ensemble average over a longer period of time, indicating that the evolution of the population reached the stationary state.

Received 3 June; accepted 29 September 2005.
Published online 21 December 2005.

1. Rao, C., Wolf, D. M. & Arkin, A. P. Control, exploitation and tolerance of intracellular noise. *Nature* **420,** 231–237 (2002).
2. Reanney, D. Genetic noise in evolution? *Nature* **307,** 318–319 (2002).
3. Sternberg, P. W. & Felix, M. Evolution of cell lineage. *Curr. Opin. Genet. Dev.* **7,** 543–550 (1997).
4. Grossman, A. D. Genetic networks controlling the initiation of sporulation and the development of genetic competence in *Bacillus subtilis*. *Annu. Rev. Genet.* **29,** 477–508 (1995).
5. Kemkemer, R., Schrank, S., Vogel, W., Gruler, H. & Kaufmann, D. 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 (2002).
6. Cook, D. L., Gerber, A. N. & Tapscott, S. J. Modeling stochastic gene expression: Implications for haploinsufficiency. *Proc. Natl Acad. Sci. USA* **95,** 15641 (1998).
7. Elowitz, M., Levine, A., Siggia, E. & Swain, P. Stochastic gene expression in a single cell. *Science* **297,** 1183–1186 (2002).
8. Raser, J. M. & O'Shea, E. K. Control of stochasticity in eukaryotic gene expression. *Science* **304,** 1811–1814 (2004).
9. Pedraza, J. M. & van Oudenaarden, A. Noise propagation in gene networks. *Science* **307,** 1965–1969 (2005).
10. Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S. & Elowitz, M. B. Gene regulation at the single-cell level. *Science* **307,** 1962–1965 (2005).
11. Arkin, A., Ross, J. & McAdams, H. H. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected *Escherichia coli* cells. *Genetics* **149,** 1633–1648 (1998).
12. Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nature Genet.* **31,** 69–73 (2002).
13. Kepler, T. & Elston, T. Stochasticity in transcriptional regulation: origins, consequences and mathematical representations. *Biophys. J.* **81,** 3116–3136 (2001).
14. Hasty, J., McMillen, D. & Collins, J. J. Engineered gene circuits. *Nature* **420,** 224–230 (2002).
15. Blake, W. J., Kaern, M., Cantor, C. R. & Collins, J. J. Noise in eukaryotic gene expression. *Nature* **422,** 633–637 (2003).
16. Isaacs, F. J., Hasty, J., Cantor, C. R. & Collins, J. J. Prediction and measurement of an autoregulatory genetic module. *Proc. Natl Acad. Sci. USA* **100,** 7714–7719 (2003).
17. Kaern, M., Elston, T., Blake, W. & Collins, J. Stochasticity in gene expression: from theories to phenotypes. *Nature Rev. Genet.* **6,** 451–464 (2005).
18. Simpson, M. L., Cox, C. D. & Sayler, G. S. Frequency domain analysis of noise in autoregulated gene circuits. *Proc. Natl Acad. Sci. USA* **100,** 4551–4556 (2003).
19. Hartwell, L. H. & Unger, M. Unequal division in *Saccharomyces cerevisiae* and its implications for the control of cell division. *J. Cell Biol.* **75,** 422–435 (1977).
20. Woldringh, C. L., Huls, P. G. & Vischer, N. E. Volume growth of daughter and parent cells during the cell cycle of *Saccharomyces cerevisiae* a/α as determined by image cytometry. *J. Bacteriol.* **175,** 3174–3181 (1993).
21. Cohen, B. A., Mitra, R. D., Hughes, J. D. & Church, G. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nature Genet.* **26,** 183–186 (2000).
22. Sheff, M. A. & Thorn, K. S. Optimized cassettes for fluorescent protein tagging in *Saccharomyces cerevisiae*. *Yeast* **21,** 661–670 (2004).
23. Adalsteinsson, D., McMillen, D. & Elston, T. Biochemical network stochastic simulator (bionets): Software for stochastic modeling of biochemical networks. *BMC Bioinformatics* **5,** 24 (2004).
24. Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81,** 2340–2361 (1977).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the NSF Division of Molecular and Cellular Bioscience and the Alfred P. Sloan Foundation.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.H. (hasty@ucsd.edu).