

# Gene Regulation at the Single-Cell Level

Nitzan Rosenfeld,<sup>1\*</sup> Jonathan W. Young,<sup>3</sup> Uri Alon,<sup>1</sup>  
Peter S. Swain,<sup>2\*</sup> Michael B. Elowitz<sup>3†</sup>

The quantitative relation between transcription factor concentrations and the rate of protein production from downstream genes is central to the function of genetic networks. Here we show that this relation, which we call the gene regulation function (GRF), fluctuates dynamically in individual living cells, thereby limiting the accuracy with which transcriptional genetic circuits can transfer signals. Using fluorescent reporter genes and fusion proteins, we characterized the bacteriophage lambda promoter  $P_R$  in *Escherichia coli*. A novel technique based on binomial errors in protein partitioning enabled calibration of in vivo biochemical parameters in molecular units. We found that protein production rates fluctuate over a time scale of about one cell cycle, while intrinsic noise decays rapidly. Thus, biochemical parameters, noise, and slowly varying cellular states together determine the effective single-cell GRF. These results can form a basis for quantitative modeling of natural gene circuits and for design of synthetic ones.

The operation of transcriptional genetic circuits (1–5) is based on the control of promoters by transcription factors. The GRF is the relation between the concentration of active transcription factors in a cell and the

rate at which their downstream gene products are produced (expressed) through transcription and translation. The GRF is typically represented as a continuous graph, with the active transcription factor concentration on the  $x$  axis and the rate of production of its target gene on the  $y$  axis (Fig. 1A). The shape of this function, e.g., the characteristic level of repressor that induces a given response, and the sharpness, or nonlinearity, of this response (1) determine key features of cellular behavior such as lysogeny switching (2), developmental cell-fate decisions (6), and oscillation (7). Its properties are also crucial for the design of synthetic genetic networks (7–11). Current models estimate GRFs from in vitro

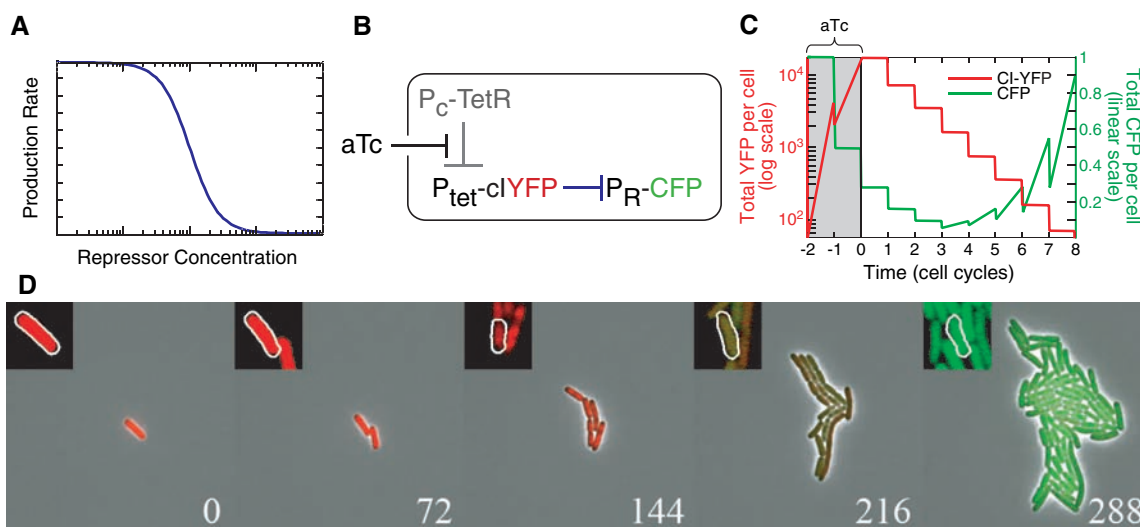
data (12, 13). However, biochemical parameters are generally unknown in vivo and could depend on the environment (12) or cell history (14, 15). Moreover, gene regulation may vary from cell to cell or over time. Three fundamental aspects of the GRF specify the behavior of transcriptional circuits at the single-cell level: its mean shape (averaged over many cells), the typical deviation from this mean, and the time scale over which such fluctuations persist. Although fast fluctuations should average out quickly, slow ones may introduce errors in the operation of genetic circuits and may pose a fundamental limit on their accuracy. In order to address all three aspects, it is necessary to observe gene regulation in individual cells over time.

Therefore, we built “ $\lambda$ -cascade” strains of *Escherichia coli*, containing the  $\lambda$  repressor and a downstream gene, such that both the amount of the repressor protein and the rate of expression of its target gene could be monitored simultaneously in individual cells (Fig. 1B). These strains incorporate a yellow fluorescent repressor fusion protein (*cl-yfp*) and a chromosomally integrated target promoter ( $P_R$ ) controlling cyan fluorescent protein (*cfp*). In order to systematically vary repressor concentration over its functional range (in logarithmic steps), we devised a “regulator dilution” method. Repressor production is switched off in a growing cell, so that its concentration subsequently decreases by dilution as the cell divides and grows into a microcolony (Fig. 1C). We used fluorescence time-lapse microscopy (Fig. 1D; fig. S1 and movies S1 and S2) and computational image analysis to reconstruct the lineage tree (family tree) of descent and sibling relations among the cells in each microcolony (fig.

<sup>1</sup>Departments of Molecular Cell Biology and Physics of Complex Systems, Weizmann Institute of Science, Rehovot, 76100, Israel. <sup>2</sup>Centre for Non-linear Dynamics, Department of Physiology, McGill University, 3655 Promenade Sir William Osler, Montréal, Québec, Canada, H3G 1Y6. <sup>3</sup>Division of Biology and Department of Applied Physics, Caltech, Pasadena, CA 91125, USA.

\*These authors contributed equally to this work  
†To whom correspondence should be addressed.  
E-mail: melowitz@caltech.edu

**Fig. 1.** Measuring a gene regulation function (GRF) in individual *E. coli* cell lineages. (A) The GRF is the dependence of the production rate of a target promoter ( $y$  axis) on the concentration of one (or more) transcription factors ( $x$  axis). (B) In the  $\lambda$ -cascade strains (16) of *E. coli*, CI-YFP is expressed from a tetracycline promoter in a TetR+ background and can be induced by anhydrotetracycline (aTc). CI-YFP represses production of CFP from the  $P_R$  promoter. (C) The regulator dilution experiment (schematic): Cells are transiently induced to express CI-YFP and then observed in time-lapse microscopy as repressor dilutes out during cell growth (red line). When CI-YFP levels decrease sufficiently, expression of the *cfp* target gene begins (green line). (D) Snapshots of a typical regulator dilution



experiment using the  $O_R2^*$ - $\lambda$ -cascade strain (see fig. S3) (16). CI-YFP protein is shown in red and CFP is shown in green. Times, in minutes, are indicated on snapshots. (Insets) Selected cell lineage (outlined in white). Greater time resolution is provided in fig. S1.

S2). For each cell lineage, we quantified over time the level of repressor ( $x$  axis of the GRF) and the total amount of CFP protein (Fig. 2A). From the change in CFP over time, we calculated its rate of production ( $y$  axis of the GRF) (16).

Regulator dilution also provides a natural in vivo calibration of individual protein fluorescence. Using the lineage tree and fluorescence data, we analyzed sister cell pairs just after division (Fig. 2B). The partitioning of CI-YFP fluorescence to daughter cells obeyed a binomial distribution, consistent with an equal probability of having each fluorescent protein molecule go to either daughter (16). Consequently, the root-mean-square error in CI-YFP partitioning between daughters increases as the square root of their total CI-YFP fluorescence. Using a one-parameter fit, we estimated the fluorescence signal of individual CI-YFP molecules (Fig. 2B and supporting online material). Thus, despite cellular autofluorescence that prohibits detection of individual CI-YFP molecules, observation of partitioning errors still permits calibration in terms of apparent numbers of molecules per cell.

The mean GRFs obtained by these techniques are shown in Fig. 3A for the  $P_R$  promoter and a point mutant variant (fig. S3). These are the mean functions, obtained by averaging individual data points (Fig. 3B) in bins of similar repressor concentration, indicating the average protein production rate at a given repressor concentration. Their cooperative nature would have been “smeared out” by population averages (6, 17, 18).

These mean GRF data provide in vivo values of the biochemical parameters underlying transcriptional regulation. Hill functions of the form  $f(R) = \beta/[1 + (R/k_d)^n]$  are often used to represent unknown regulation

functions (1, 6–10). Here,  $k_d$  is the concentration of repressor yielding half-maximal expression,  $n$  indicates the degree of effective cooperativity in repression, and  $\beta$  is the maximal production rate. Hill functions indeed fit the data well (Fig. 3A and Table 1). The measured in vivo  $k_d$  is comparable to previous estimates (2, 12, 13, 19) (see supporting online text). The significant cooperativity observed ( $n > 1$ ) may result from dimerization of repressor molecules and cooperative interactions between repressors bound at neighboring sites (2, 12, 13, 19, 20). A point mutation in the  $O_{R2}$  operator,  $O_{R2}^*$  (20) (fig. S3), significantly reduced  $n$  and increased  $k_d$  (Fig. 3A and Table 1). Note that with similar methods it is even possible to measure effective cooperativity ( $n$ ) for native repressors without fluorescent protein fusions (16).

We next addressed deviations from the mean GRF. At a given repressor concentration, the standard deviation of production rates is  $\sim 55\%$  of the mean GRF value. Such variation may arise from microenvironmental differences (21), cell cycle-dependent changes in gene copy number, and various sources of noise in gene expression and other cellular processes (22). We compared microcolonies in which induction occurs at different cell densities (16). The results suggested that the measured GRF is robust to possible differences among the growth environments in our experiments (fig. S6). We analyzed the effect of gene copy number, which varies twofold over the cell cycle as DNA replicates. The CFP production rate correlated strongly with cell-cycle phase; cells about to divide produced on average twice as much protein per unit of time as newly divided cells (16). Thus, gene dosage is not compensated. Nevertheless, after normalizing production rates to the average cell-cycle phase

(16), substantial variation still remains in the production rates, and their standard deviation is  $\sim 40\%$  of the mean GRF (Fig. 3). The deviations from the mean GRF show a log-normal distribution (see supporting online text and fig. S5).

These remaining fluctuations may arise from processes intrinsic or extrinsic to gene expression. Intrinsic noise results from stochasticity in the biochemical reactions at an individual gene and would cause identical copies of a gene to express at different levels. It can be measured by comparing expression of two identically regulated fluorescent proteins (22). Extrinsic noise is the additional variation originating from fluctuations in cellular components such as metabolites, ribosomes, and polymerases and has a global effect (22, 23). Extrinsic noise is often the dominant source of variation in *E. coli* and *Saccharomyces cerevisiae* (22, 24).

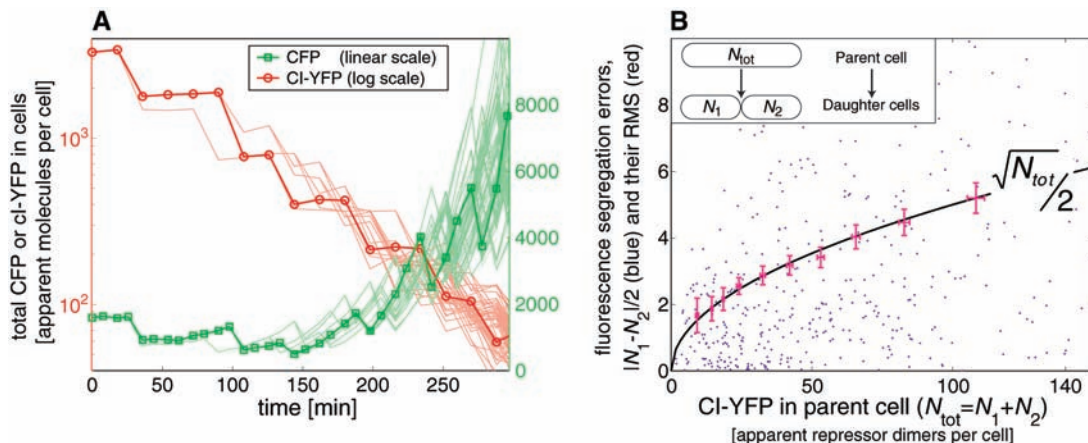
To test whether fluctuations were of intrinsic or extrinsic origin, we used a “symmetric branch” strain (16) that produced CFP and YFP from an identical pair of  $P_R$  promoters (Fig. 4D, movie S3). The difference between CFP and YFP production rates in these cells indicates  $\sim 20\%$  intrinsic noise in protein production [averaged over 8- to 9-min intervals (16)], suggesting that the extrinsic component of noise is dominant and contributes a variation in protein production rates of  $\sim 35\%$ .

Our measurements provide more detailed analysis of extrinsic noise in two ways. First, in previous work (22), extrinsic noise included fluctuations in upstream cellular components, including both gene-specific and global factors. Here, we quantify the extrinsic noise at known repressor concentration, and so extrinsic noise encompasses fluctuations in global cellular components such as polymerases or

**Fig. 2.** Data and calibration.

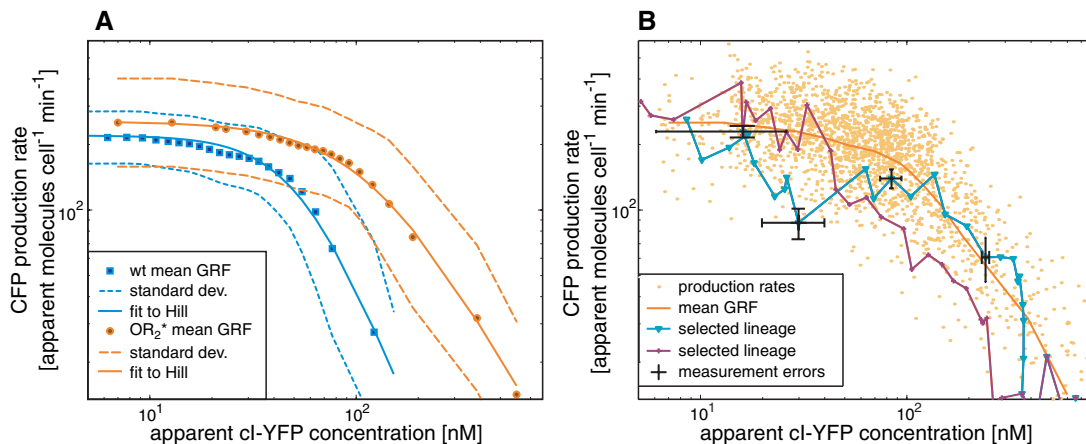
(A) Fluorescence intensities of individual cells are plotted over time for the experiment of Fig. 1D. Red indicates CI-YFP, which is plotted on a logarithmic  $y$  axis to highlight its exponential dilution: As CI-YFP is not produced, each division event causes a reduction of about twofold in total CI-YFP fluorescence. Green indicates CFP, which is plotted on a linear  $y$  axis to emphasize its increasing slope, showing that CFP production rate increases as the CI-YFP levels decrease. A selected cell lineage is highlighted (also outlined in Fig. 1D).

(B) Analysis of binomial errors in protein partitioning to find  $v_y$ , the apparent fluorescence intensity of one independently segregating fluorescent particle (16). Cells containing  $N_{tot}$  copies of a fluorescent particle (total fluorescence  $Y_{tot} = v_y \cdot N_{tot}$ ) undergo division (inset). If each particle segregates independently,  $N_1$  and  $N_2$ , the number of copies received by the two daughter cells, are distributed binomially, and satisfy

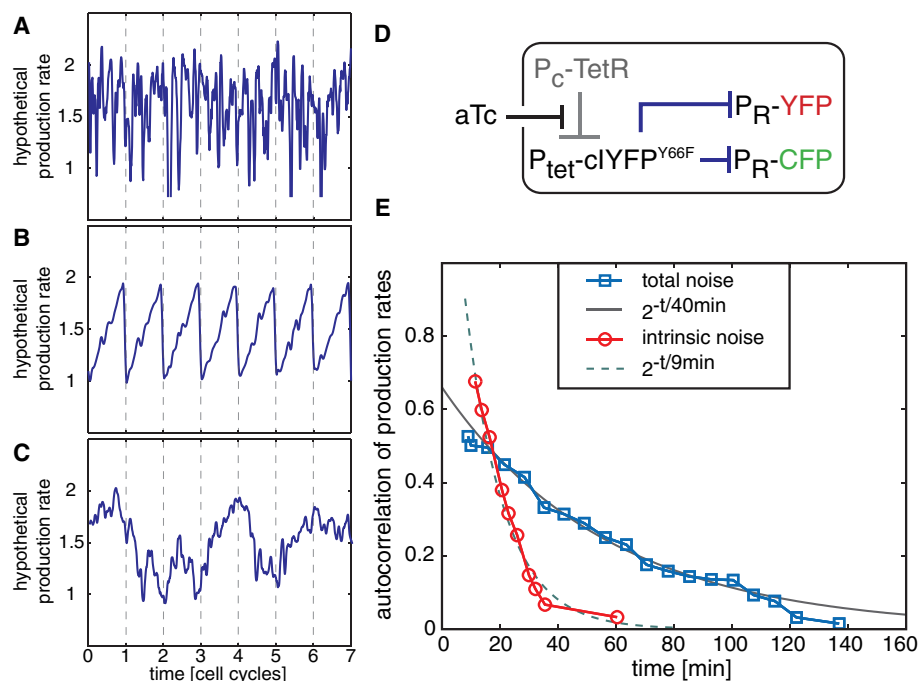


$\sqrt{\langle (N_1 - N_2)^2 \rangle} = \sqrt{N_{tot}}/2$ . A single-parameter fit thus determines the value of  $v_y$ . Here we plot  $|N_1 - N_2|/2$  (in numbers of apparent molecule dimers) versus  $N_{tot} = N_1 + N_2$ . Blue dots show the scatter of individual division events. Crosses (red) show the root-mean-square (RMS) error in protein partitioning and its standard error. The expected binomial standard deviation is shown in black.

**Fig. 3.** The GRF and its fluctuations. (A) The mean regulation function of the wild-type  $\lambda$ -phage  $P_R$  promoter (blue squares) and its  $O_R2^*$ -mutated variant ( $O_R2^*$ , orange circles) are plotted with their respective standard deviations (dashed/dotted lines). Hill function approximations (using parameters from Table 1) are shown (solid lines). (B) Variation in the  $O_R2^*$  GRF. Individual points indicate the instantaneous production rate of CFP, as a function of the amount of CI-YFP in the same cell, for all cells in a microcolony of the  $O_R2^*$ - $\lambda$ -cascade strain. The time courses of selected lineages in this microcolony are drawn on top of the data, showing slow fluctuations around the mean GRF. CI-YFP concentration decreases with time, and



consecutive data points along a trajectory are at 9-min intervals. Typical measurement errors (black crosses) are shown for a few points. Data are compensated for cell cycle-related effects (16).



**Fig. 4.** Fluctuations in gene regulation. (Left) Three types of variability observed here. (A) Fast fluctuations in CFP production, similar to those produced by intrinsic noise. (B) Periodic, cell cycle-dependent oscillations in CFP production, which can result from DNA replication. (C) Slow aperiodic fluctuations, such as extrinsic fluctuations in gene expression. (D) Intrinsic and extrinsic noise can be discriminated using a symmetric-branch strain (16) of *E. coli*, containing identical, chromosomally integrated  $\lambda$ -phage  $P_R$  promoters controlling *cfp* and *yfp* genes. The strain also expresses nonfluorescent CI-YFP from a Tet-regulated promoter. (E) The autocorrelation function of the relative production rates in the  $\lambda$ -cascade strains (blue squares) shows that the time scale for fluctuations in protein production is  $\tau_{\text{corr}} \sim 40$  min (blue). The difference between production rates of YFP and CFP in the symmetric branch strain has a correlation time of  $\tau_{\text{intrinsic}} < 10$  min (red). The data and correlations presented are corrected for cell cycle-related effects (16).

ribosomes but not in the concentration of the repressor, CI. Second, dynamic observations permit us to measure extrinsic noise in the rate of protein expression rather than in the amount of accumulated protein. The present breakdown should be more useful for modeling and design of genetic networks.

In cells, fast and slow fluctuations can affect the operation of genetic networks in dif-

ferent ways. Previous experiments (22, 24–26) used static “snapshots” to quantify noise at steady state and were thus unable to access the temporal dynamics of gene expression. However, a similar steady-state distribution of expression levels can be reached by fluctuations on very different time scales (Fig. 4). Fluctuations can be characterized by their autocorrelation time,  $\tau_{\text{corr}}$  (16). The

**Table 1.** In vivo values of effective biochemical parameters. Molecular units are estimated using binomial errors in protein partitioning (16) (Fig. 2B), which may have systematic errors up to a factor  $\sim 2$ . Concentrations are calculated from apparent molecule numbers divided by cell volumes estimated from cell images (16), with an average volume of  $1.5 \pm 0.5 \mu\text{m}^3$  (for which  $1 \text{ nM} = 0.9 \text{ molecule/cell}$ ).

Parameter	$P_R$	$P_R (O_R2^*)$
$n$ (degree of cooperativity in repression)	$2.4 \pm 0.3$	$1.7 \pm 0.3$
$k_d$ [concentration of repressor yielding half-maximal expression (nM)]	$55 \pm 10$	$120 \pm 25$
$\beta$ [unrepressed production rate (molecules $\cdot$ cell $^{-1}$ $\cdot$ min $^{-1}$ )]	$220 \pm 15$	$255 \pm 40$

magnitude of  $\tau_{\text{corr}}$  compared with the cell-cycle period is crucial: Fluctuations longer than the cell cycle accumulate to produce significant effects, whereas more rapid fluctuations may “average out” as cellular circuits operate (27, 28). In these data, three types of dynamics are observed (Fig. 4, A to C): Fast fluctuations, periodic cell-cycle oscillations due to DNA replication, and aperiodic fluctuations with a time scale of about one cell cycle.

We found that the trajectories of single-cell lineages departed substantially from the mean GRF over relatively long periods (Fig. 3B), with  $\tau_{\text{corr}} = 40 \pm 10$  min (Fig. 4E). This value is close to the cell cycle period,  $\tau_{\text{cc}} = 45 \pm 10$  min, indicating that, overall, fluctuations typically persist for one cell cycle. Therefore, if a cell produces CFP at a faster rate than the mean GRF, this overexpression will likely continue for roughly one cell cycle, and CFP levels will accumulate to higher concentrations than the mean GRF would predict.

In contrast, the autocorrelation of the intrinsic noise (*16*) decays rapidly:  $\tau_{\text{intrinsic}} < 10 \text{ min} \ll \tau_{\text{corr}}$  (Fig. 4E). Thus, the observed slow fluctuations do not result from intrinsic noise; they represent noise extrinsic to CFP expression (see supporting online text). The concentration of a stable cellular factor would be expected to fluctuate with a time scale of the cell cycle period (*7, 10*). For instance, even though intrinsic fluctuations in production rates are fast, the difference between the total amounts of YFP and CFP in the symmetric branch experiments has an autocorrelation time of  $\tau_{\text{total}} = 45 \pm 5 \text{ min}$  (*16*). A similar time scale may well apply to other stable cellular components such as ribosomes, metabolic apparatus, and sigma factors. As such components affect their own expression as well as that of our test genes, extrinsic noise may be self-perpetuating.

These data indicate that the single-cell GRF cannot be represented by a single-valued function. Slow extrinsic fluctuations give the cell and the genetic circuits it comprises a memory, or individuality (*29*), lasting roughly one cell cycle. These fluctuations are substantial in amplitude and slow in time scale. They present difficulty for modeling genetic circuits and, potentially, for the cell itself: In order to accurately process an intracellular signal, a cell would have to average its response for well over a cell cycle—a long time in many biological situations. This problem is not due to intrinsic noise in the output, noise that fluctuates rapidly, but rather to the aggregate effect of fluctuations in other cellular components. There is thus a fundamental tradeoff between accuracy and speed in purely transcriptional responses. Accurate cellular responses on faster time scales are likely to require feedback from their output (*1, 4, 6, 10, 30*). These data provide an integrated, quantitative characterization of a genetic element at the single-cell level: its biochemical parameters, together with the amplitude and time scale of its fluctuations. Such systems-level specifications are necessary both for modeling natural genetic circuits and for building synthetic ones. The methods introduced here can be generalized to more complex genetic networks, as well as to eukaryotic organisms (*18*).

#### References and Notes

- M. A. Savageau, *Biochemical Systems Analysis* (Addison-Wesley, Reading, MA, 1976).
- M. Ptashne, *A Genetic Switch: Phage Lambda and Higher Organisms* (Cell Press and Blackwell Science, Cambridge, MA, ed. 2, 1992).
- H. H. McAdams, L. Shapiro, *Science* **269**, 650 (1995).
- E. H. Davidson *et al.*, *Science* **295**, 1669 (2002).
- S. S. Shen-Orr, R. Milo, S. Mangan, U. Alon, *Nature Genet.* **31**, 64 (2002).
- J. E. Ferrell Jr., E. M. Machleder, *Science* **280**, 895 (1998).
- M. B. Elowitz, S. Leibler, *Nature* **403**, 335 (2000).
- T. S. Gardner, C. R. Cantor, J. J. Collins, *Nature* **403**, 339 (2000).
- A. Becskei, B. Seraphin, L. Serrano, *EMBO J.* **20**, 2528 (2001).
- N. Rosenfeld, M. B. Elowitz, U. Alon, *J. Mol. Biol.* **323**, 785 (2002).
- F. J. Isaacs, J. Hasty, C. R. Cantor, J. J. Collins, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 7714 (2003).
- K. S. Koblan, G. K. Ackers, *Biochemistry* **31**, 57 (1992).
- P. J. Darling, J. M. Holt, G. K. Ackers, *J. Mol. Biol.* **302**, 625 (2000).
- R. J. Ellis, *Trends Biochem. Sci.* **26**, 597 (2001).
- M. Mirasoli, J. Feliciano, E. Michelini, S. Daunert, A. Roda, *Anal. Chem.* **74**, 5948 (2002).
- Materials and methods are available as supporting material on Science Online.
- P. Cluzel, M. Surette, S. Leibler, *Science* **287**, 1652 (2000).
- G. Lahav *et al.*, *Nature Genet.* **36**, 147 (2004).
- I. B. Dodd *et al.*, *Genes Dev.* **18**, 344 (2004).
- B. J. Meyer, R. Maurer, M. Ptashne, *J. Mol. Biol.* **139**, 163 (1980).
- J. A. Shapiro, *Annu. Rev. Microbiol.* **52**, 81 (1998).
- M. B. Elowitz, A. J. Levine, E. D. Siggia, P. S. Swain, *Science* **297**, 1183 (2002).
- P. S. Swain, M. B. Elowitz, E. D. Siggia, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12795 (2002).
- J. M. Raser, E. K. O'Shea, *Science* **304**, 1811 (2004).
- E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, A. van Oudenaarden, *Nature Genet.* **31**, 69 (2002).
- W. J. Blake, M. Kærn, C. R. Cantor, J. J. Collins, *Nature* **422**, 633 (2003).
- H. H. McAdams, A. Arkin, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 814 (1997).
- J. Paulsson, *Nature* **427**, 415 (2004).
- J. L. Spudich, D. E. Koshland Jr., *Nature* **262**, 467 (1976).
- P. S. Swain, *J. Mol. Biol.* **344**, 965 (2004).
- We thank Z. Ben-Haim, R. Clifford, S. Itzkovitz, Z. Kam, R. Kishony, A. J. Levine, A. Mayo, R. Milo, R. Phillips, M. Ptashne, J. Shapiro, B. Shraiman, E. Siggia, and M. G. Surette for helpful discussions. M.B.E. is supported by a CASI award from the Burroughs Wellcome Fund, the Searle Scholars Program, and the Seaver Institute. U.A. and M.B.E. are supported by the Human Frontiers Science Program. P.S.S. acknowledges support from a Tier II Canada Research Chair and the Natural Sciences and Engineering Research Council of Canada. N.R. dedicates this work to the memory of his father, Yasha (Yaakov) Rosenfeld.

#### Supporting Online Material

www.sciencemag.org/cgi/content/full/307/5717/1962/DC1

Materials and Methods

SOM Text

Figs. S1 to S6

References and Notes

Movies S1 to S3

29 October 2004; accepted 4 February 2005

10.1126/science.1106914

## Noise Propagation in Gene Networks

Juan M. Pedraza and Alexander van Oudenaarden\*

Accurately predicting noise propagation in gene networks is crucial for understanding signal fidelity in natural networks and designing noise-tolerant gene circuits. To quantify how noise propagates through gene networks, we measured expression correlations between genes in single cells. We found that noise in a gene was determined by its intrinsic fluctuations, transmitted noise from upstream genes, and global noise affecting all genes. A model was developed that explains the complex behavior exhibited by the correlations and reveals the dominant noise sources. The model successfully predicts the correlations as the network is systematically perturbed. This approach provides a step toward understanding and manipulating noise propagation in more complex gene networks.

The genetic program of a living cell is determined by a complex web of gene networks. The proper execution of this program relies on faithful signal propagation from one gene to the next. This process may be hindered by stochastic fluctuations arising from gene expression, because some of the components in these circuits are present at low numbers, which makes fluctuations in concentrations unavoidable (*1*). Additionally, reaction rates can fluctuate because of stochastic variation in the global pool of housekeeping genes or because of fluctuations in environmental conditions that affect all genes. For example, fluctuations in the number of available polymerases or in any factor that alters the cell growth rate will change the reaction rates for all genes. Recent

experimental studies (*2–5*) have made substantial progress identifying the factors that determine the fluctuations in the expression of a single gene. However, how expression fluctuations propagate from one gene to the next is largely unknown. To address this issue, we designed a gene network (Fig. 1A) in which the interactions between adjacent genes could be externally controlled and quantified at the single-cell level.

This synthetic network (*6*) consisted of four genes, of which three were monitored in single *Escherichia coli* cells by cyan, yellow, and red fluorescent proteins (CFP, YFP, and RFP). The first gene, *lacI*, is constitutively transcribed and codes for the lactose repressor, which down-regulates the transcription of the second gene, *tetR*, that is bicistronically transcribed with *cfp*. The gene product of *tetR*, the tetracycline repressor, in turn down-regulates the transcription of the third gene, reported by YFP. The fourth gene, *rfp*, is under

Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

\*To whom correspondence should be addressed: E-mail: avano@mit.edu