LETTERS

Stochastic gene expression out-of-steady-state in the cyanobacterial circadian clock

Jeffrey R. Chabot[1]†, Juan M. Pedraza[1], Prashant Luitel[1] & Alexander van Oudenaarden[1]

Recent advances in measuring gene expression at the single-cell level have highlighted the stochastic nature of messenger RNA and protein synthesis[1–3]. Stochastic gene expression creates a source of variability in the abundance of cellular components, even among isogenic cells exposed to an identical environment. Recent integrated experimental and modelling studies[4–13] have shed light on the molecular sources of this variability. However, many of these studies focus on systems that have reached a steady state and therefore do not address a large class of dynamic phenomena including oscillatory gene expression. Here we develop a general protocol for analysing and predicting stochastic gene expression in systems that never reach steady states. We use this framework to analyse experimentally stochastic expression of genes driven by the *Synechococcus elongatus* circadian clock. We find that, although the average expression at two points in the circadian cycle separated by 12 hours is identical, the variability at these two time points can be different. We show that this is a general feature of out-of-steady-state systems. We demonstrate how intrinsic noise sources, owing to random births and deaths of mRNAs and proteins, or extrinsic noise sources, which introduce fluctuations in rate constants, affect the cell-to-cell variability. To distinguish experimentally between these sources, we measured how the correlation between expression fluctuations of two identical genes is modulated during the circadian cycle. This quantitative framework is generally applicable to any out-of-steady-state system and will be necessary for understanding the fidelity of dynamic cellular systems.

So far, the expression reporter used in the cyanobacterium *S. elongatus* PCC7942 has been bacterial luciferase[14,15]. This technique allows population-level measurements but limits studies of expression fluctuations between individuals, because of the very weak bioluminescence signal emitted from single cells[16]. We therefore developed a single-cell fluorescent reporter assay that provides a much stronger signal facilitating quick and accurate measurements of gene expression levels in large numbers of individual cells (Supplementary Fig. 1). We found that an SsrA-tagged[17] yellow-shifted variant of green fluorescent protein, YFP–SsrA(LVA), in which the last three amino acids in the SsrA tag are leucine, valine and alanine, was able to report faithfully the periodic activity of the *S. elongatus kaiBC* promoter (P_{kaiBC}) (Fig. 1, Supplementary Movie 1).

We used this fluorescence assay to measure YFP–SsrA(LVA) expression in individual *Synechococcus* cells using flow cytometry. Using two known neutral loci in the *Synechococcus* genome, defined as NS I and NS II (ref. 18), to insert the P_{kaiBC}–*yfp*–*ssrA(LVA)* construct, we found that the mean YFP–SsrA(LVA) expression (averaged over 10^5 individual cells) exhibits a clear periodic expression pattern with levels that are very similar between the two neutral sites (Fig. 2a). The phase, however, differs from previously reported mRNA oscillations, peaking approximately 4 to 6 h later than the mRNA maximum[14]. From the observed half-life of YFP–SsrA(LVA), we expect a phase delay of 4.3 h, consistent with the observed phase shift (Supplementary Information). Interestingly, the key circadian clock protein KaiC is similarly long-lived, with a half-life of about 10 h (ref. 19). This indicates that this endogenous protein will also display a similar phase shift (up to a maximum of 6 h) with respect to the phase of the corresponding *kaiC* mRNA concentration. Using the measured lifetime of YFP–SsrA(LVA) and an estimate for the half-life of the corresponding mRNA of about 15 min (ref. 19), the mRNA creation rate $k_R(t)$ can be inferred from the YFP–SsrA(LVA)

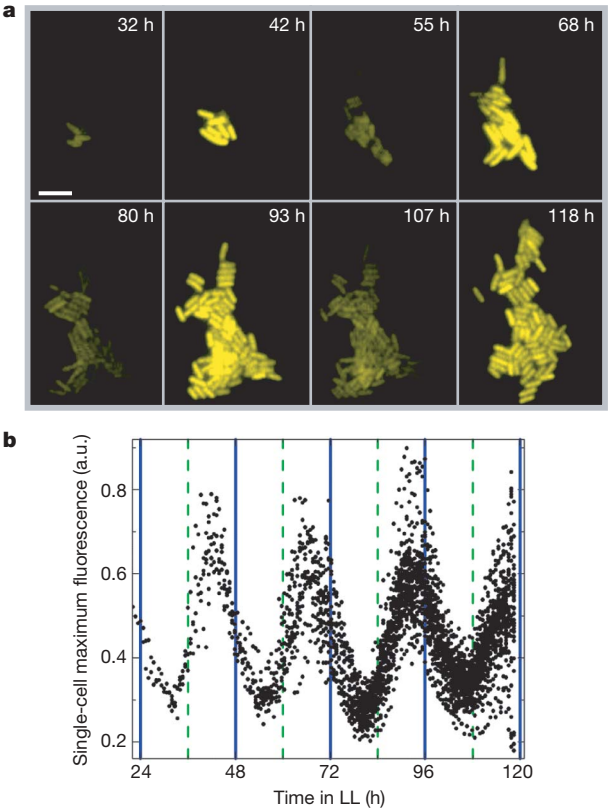

Figure 1 | Monitoring circadian oscillations in single *S. elongatus* PCC7942 cells using fluorescence microscopy. a, Montage of fluorescence microscopy images demonstrating circadian oscillations in single cells. Time is reported in hours (h). The scale bar represents 4 μm. **b**, Analysis of maximum pixel fluorescence within individual cells (measured in arbitrary units, a.u.) as a function of time. LL, constant light.

[1]Department of Physics and George Harrison Spectroscopy Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. †Present address: Systems Biology Group, Pfizer Research Technology Center, Cambridge, Massachusetts 02139, USA.

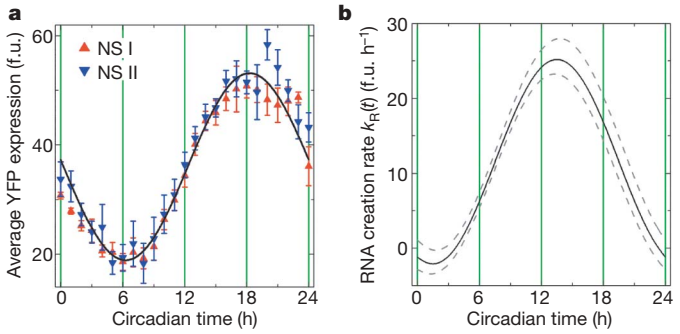

Figure 2 | Circadian oscillations in single cells. a, Average YFP–SsrA(LVA) expression as a function of circadian time for strain JRCS32 (red triangles) and JRCS35 (blue triangles). Fluorescence was obtained by averaging the fluorescence of at least 10^4 single cells. Cells were synchronized by two 12:12 LD cycles before the start of the experiment. The black solid line is a fit to both data sets with a cosine function with a period of 24 h. f.u., fluorescence units. Error bars, 1 s.e.m. **b**, Inferred RNA creation rate $k_R(t)$ as a function of circadian time. Dashed lines indicate the error bars given the experimental uncertainty of the YFP–SsrA(LVA) half-life (5.6 ± 1.0 h).

dynamics (Fig. 2b, Supplementary Information). We find that the *yfp–ssrA(LVA)* mRNA concentration peaks slightly after a circadian time of 12 h and vanishes 6 h before and after this peak consistent with northern blot analyses[14].

This single-cell assay allows us to probe expression variability quantitatively from cell to cell, and to explore the effect of stochastic gene expression in this oscillating system. The expression variability was quantified by the coefficient of variation, which is the standard deviation of the expression distribution normalized by the average expression. Figure 3a displays the squared coefficient of variation as a function of the inverse averaged expression. We chose this representation because in steady-state one would expect a linear relationship between these variables[8,20]. However, we find that during the circadian cycle the experimental data trace out a loop (Fig. 3a). In other words, for a particular average expression, two values of coefficient of variation are possible: a low variability early in the circadian cycle and a high variability later in the cycle. This property cannot be understood by previous theoretical work predicting the stochastic properties of gene expression assuming that protein and mRNA creation and destruction reached a steady state[8,20–22]. This assumption is violated in this circadian clock because the oscillation period is comparable to the protein half-life and therefore the system will never reach even a steady-state in which protein production is balanced with protein destruction. We therefore developed a new theoretical framework that allows us to calculate the cell-to-cell variability in systems that are not in steady-state. Using this model, we indeed find that the loops can be reproduced qualitatively using either an intrinsic noise model (Fig. 3b) or an extrinsic noise model (Fig. 3c). Remarkably, the chirality of the loop depends on how rapidly the extrinsic noise fluctuates.

To characterize the relative contributions of intrinsic and extrinsic sources experimentally, we devised a strategy to measure the correlation between expression fluctuations at NS I and those at NS II. A powerful strategy would be to use two different coloured fluorescent proteins at the two neutral sites and compare the expression values in single cells[5,7]. However, because YFP is the only reliable reporter in *Synechococcus*, we decided to take an alternative approach. We constructed a strain that contains P_{kaiBC}–*yfp–SsrA(LVA)* in both NS I and NS II. As expected, the average YFP expression doubles (Fig. 4a, black symbols). However the variability observed on doubling of the gene dose displays a more complicated behaviour (Fig. 4b).

The total variance observed for one copy is partly caused by a noise source that introduces correlated expression fluctuations between NS I and NS II, and partly caused by another noise source that introduces uncorrelated fluctuations. We will designate the correlated source as global and the uncorrelated source as local. These two sources contribute independently to the total variance, and therefore the total variance for a single copy is given by $\sigma_1^2(t) = \sigma_{\text{local}}^2(t) + \sigma_{\text{global}}^2(t)$. In the case of two copies, the local component of the variance will double and the global component will quadruple[23,24]: $\sigma_2^2(t) = 2\sigma_{\text{local}}^2(t) + 4\sigma_{\text{global}}^2(t)$. Therefore, by using these two equations, the local and global component of the variance can be explicitly found as a function of time (t) (Fig. 4c, d). We find that $\sigma_{\text{local}}^2(t)$ and $\sigma_{\text{global}}^2(t)$ have comparable magnitude but peak at different times of the day. In the next step we use the dynamics of the local and global component of the variance to infer what the

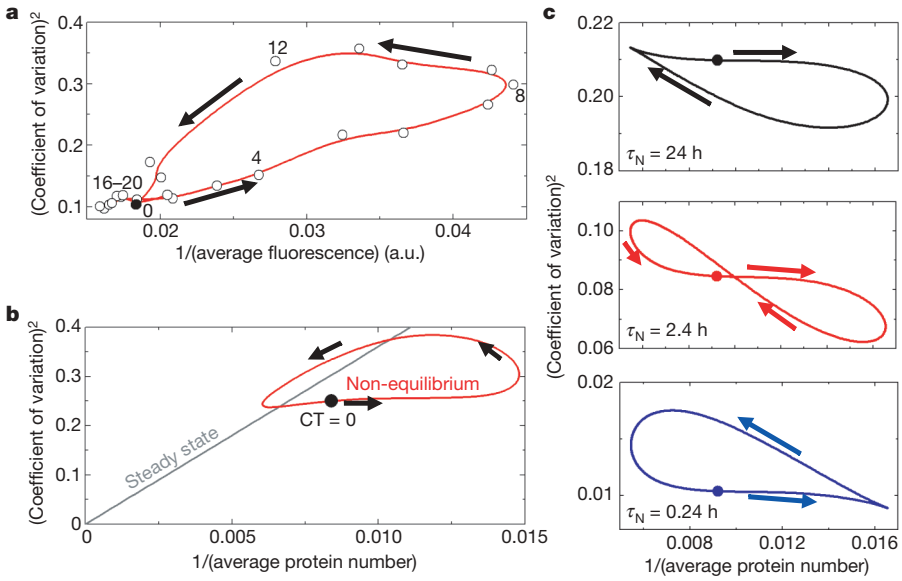

Figure 3 | A dynamic analysis of stochastic gene expression reveals noise loops. a, The squared coefficient of variation as a function of the inverse average fluorescence for strain JRCS32 (open circles). The filled circle denotes the start of the circadian cycle (circadian time = 0) and the numbers denote the circadian time. The red line serves as a guide to the eye. **b**, The squared coefficient of variation as a function of the inverse average protein number using an intrinsic noise model (red solid line). The linear relationship expected in steady state is shown in grey. **c**, The squared coefficient of variation as a function of the inverse average protein number using an extrinsic noise model (solid lines) for different noise decorrelation times ($\tau_N = \frac{\ln 2}{\gamma_N}$).

variability introduced at the transcriptional level had to be to explain the magnitude and dynamics of $\sigma^2_{\mathrm{local}}(t)$ and $\sigma^2_{\mathrm{global}}(t)$.

To determine the magnitude of the fluctuations introduced at the transcriptional level, we consider the following model. The dynamics of the *yfp–SsrA(LVA)* mRNA concentration (x) and YFP–SsrA(LVA) protein concentration (y) can be described generally by the following dynamical system of equations:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = k_{\mathrm{R}}(t) - \gamma_{\mathrm{R}}x + q(t)\xi(t)$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = k_{\mathrm{P}}x - \gamma_{\mathrm{P}}y$$

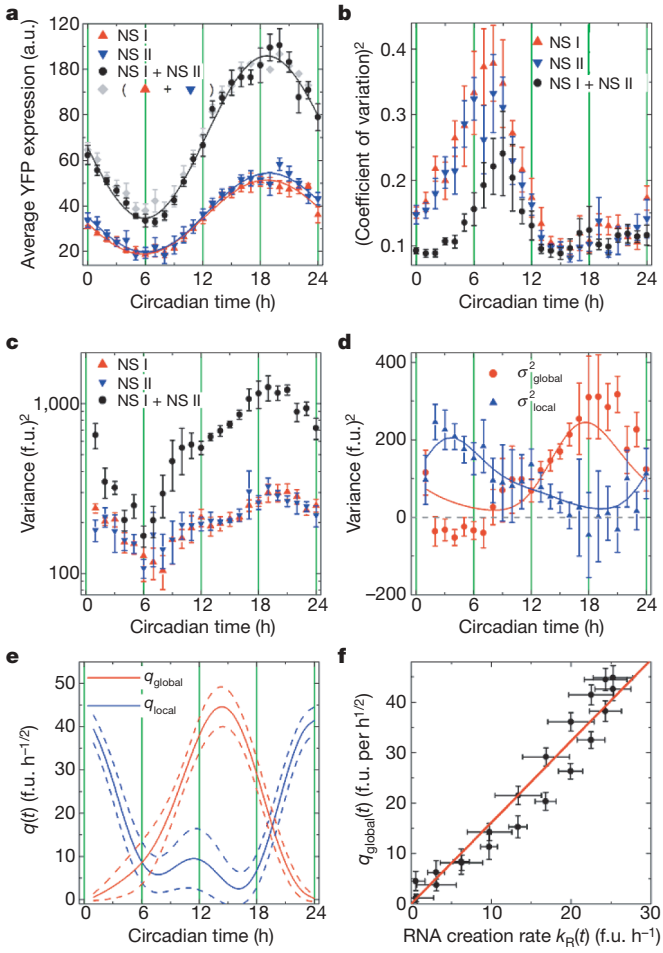

Figure 4 | Cell to cell variability in single- and double-copy-number constructs. a, Average YFP–SsrA(LVA) expression as a function of circadian time for strain JRCS32 (red triangles), JRCS35 (blue triangles) and JRCS70 (black circles). Fluorescence was obtained by averaging the fluorescence of at least 10^4 single cells. Cells were synchronized by two 12:12 LD cycles before the start of the experiment. Solid lines denote fits to a cosine function with a period of 24 h. The grey diamonds note the addition of the two fluorescence values of strain JRCS32 and JRCS35. Error bars denoted 1 s.e.m. obtained from day-to-day reproductions. **b**, The squared coefficient of variation as a function of circadian time. This was computed from at least 10^4 single cells. Error bars represent standard errors computed from day-to-day reproductions. **c**, The variance of the expression distribution as a function of circadian time using the same strains as for **a**. The variances were obtained by averaging the fluorescence of at least 10^4 single cells. Error bars (1 s.e.m) indicate day-to-day variation determined by repeated reproduction. **d**, Component of variance that reflects local (blue triangles) and global (red circles) fluctuations as a function of circadian time. Error bars (1 s.e.m) were obtained by error propagation. **e**, Computed functions $q_{\mathrm{local}}(t)$ (blue) and $q_{\mathrm{global}}(t)$ (red) as a function of circadian time. Dashed lines indicate error bars on the basis of error propagation. **f**, Correlation between $q_{\mathrm{global}}(t)$ and the RNA transcription rate $k_{\mathrm{R}}(t)$. The solid red line represents a linear fit.

where γ_{R} and γ_{P} are the rate constants for RNA degradation and protein degradation, respectively. The estimated mRNA half-life of 15 min (ref. 19) gives $\gamma_{\mathrm{R}} \sim 2.8\,\mathrm{h}^{-1}$. Using the experimentally determined YFP–SsrA(LVA) protein degradation rate (Supplementary Fig. 1) we find $\gamma_{\mathrm{P}} \sim 0.12\,\mathrm{h}^{-1}$. The rate constant for translation, k_{P}, is unknown. The function $k_{\mathrm{R}}(t)$ describes the periodic transcription rate. The absolute transcription rates are unknown; however, we obtained a good estimate of the relative dynamics (Fig. 2b). The noise term $\xi(t)$ reflects gaussian noise with zero mean, $\langle\xi(t)\rangle = 0$, and exponentially decaying correlation, $\langle\xi(t)\xi(\tau)\rangle = \frac{\gamma_{\mathrm{N}}}{2}\mathrm{e}^{-\gamma_{\mathrm{N}}|t-\tau|}$, with decay rate γ_{N}. This type of noise is consistent with recent experimental data[25] that show that γ_{N} is of the order of the cell-doubling rate. We assume that all the stochastic fluctuations are introduced at the level of mRNA production/decay, and that protein production/decay follows deterministic dynamics. Using this model we can determine $q(t)$ from the experimentally determined variances (Fig. 4c, d). The function $q(t)$ is a periodic function that describes the magnitude of the stochastic fluctuations introduced at the transcriptional level. Similarly to the separation of the total variance in terms of $\sigma^2_{\mathrm{local}}(t)$ and $\sigma^2_{\mathrm{global}}(t)$, we can separate $q(t)\xi(t)$ into a local and a global component: $q_{\mathrm{local}}(t)\xi_{\mathrm{l}}(t)$ and $q_{\mathrm{global}}(t)\xi_{\mathrm{g}}(t)$, which can be determined from the experimentally obtained $\sigma^2_{\mathrm{local}}(t)$ and $\sigma^2_{\mathrm{global}}(t)$ (Fig. 4e).

We find that $q_{\mathrm{global}}(t)$ reaches a maximum around a circadian time of 14 h. These dynamics closely resemble the dynamics of $k_{\mathrm{R}}(t)$ (Fig. 2b). Figure 4f demonstrates explicitly the strong correlation between $k_{\mathrm{R}}(t)$ and $q_{\mathrm{global}}(t)$. This correlation is consistent with global fluctuations introduced at the level of RNA creation (Supplementary Information): $q_{\mathrm{global}}(t) = \sqrt{2/\gamma_{\mathrm{N}}}\eta_{\mathrm{G}}k_{\mathrm{R}}(t)$. We use the slope of the $k_{\mathrm{R}}(t) - q_{\mathrm{global}}(t)$ curve (Fig. 4f) to estimate the magnitude η_{G} of these global fluctuations: $\eta_{\mathrm{G}} \sim 0.4$. Here we assume that the noise decay rate γ_{N} is set by the rate constant of protein destruction γ_{P}. This indicates that at any particular circadian time the global transcription rate of the *kaiBC* promoter varies about 40% from cell to cell.

The function $q_{\mathrm{local}}(t)$ reaches a maximum around a circadian time of 0 h. A potential candidate for the local noise source could be intrinsic fluctuations caused by random births and deaths of mRNA molecules. However if $q_{\mathrm{local}}(t)$ would be dominated by intrinsic noise, one would expect $q_{\mathrm{local}}(t) \propto \sqrt{k_{\mathrm{R}}(t)}$, and therefore $q_{\mathrm{local}}(t)$ should peak around a circadian time of 14 h (see ref. 26), which is not observed (Fig. 4e). This rules out not only intrinsic noise but also a wide class of mechanisms, because any noise that is a monotonically increasing function of transcription rate is inconsistent with this observation. One potential candidate is the difference in local cellular environments between the two neutral sites. This hypothesis is consistent with the 'oscilloid' model[27] in which the condensation or supercoiling state of the cyanobacterial chromosome changes with the circadian rhythm. Indeed, recent experimental evidence has demonstrated a periodic cycle of compaction and relaxation of the cyanobacterial chromosome[28].

How fluctuations in clock components affect the fidelity of an oscillator has been modelled in detail[29,30]. Here we explore the stochastic expression of a gene driven by the clock. We show that in oscillatory systems a phase difference arises between signal and noise. This results in noise loops that cannot be explained by current steady-state models. This quantitative framework expands the current stochastic analysis toolbox to include a large class of non-steady-state phenomena such as oscillatory and transient dynamics.

METHODS SUMMARY

Growth media and cell culture. All strains were grown at 30 °C in BG11 media, supplemented with appropriate antibiotics (40 μg ml^{-1} spectinomycin for NS I inserts; 50 μg ml^{-1} kanamycin for NS II inserts). 'Light' environments were maintained at ~2,400 lx (lux) from soft-fluorescent sources (Sylvania); cells in the 'dark' environments experienced less than 50 lx. To synchronize samples, cells were exposed to at least two full 12 h dark:12 h light (12:12 LD) cycles before beginning data collection.

Out-of-steady-state model for stochastic gene expression. It can be shown (Supplementary Information) that the variance in the mRNA concentration $\sigma_x^2(t)$ is given by the following convolution: $\sigma_x^2(t)=\left\langle (q(t)\xi(t)\otimes f(t))^2\right\rangle$, where $f(t)=\exp[-\gamma_R t]$. Similarly, the variance in the protein concentration $\sigma_y^2(t)$ can be calculated: $\sigma_y^2(t)=k_P^2\left\langle (q(t)\xi(t)\otimes g(t))^2\right\rangle$, where $g(t)=\frac{\exp[-\gamma_P t]}{\gamma_R-\gamma_P}+\frac{\exp[-\gamma_R t]}{\gamma_P-\gamma_R}$. This procedure can be applied to $\sigma_{\text{global}}^2(t)$ and $\sigma_{\text{local}}^2(t)$ (Fig. 4d) to determine $q_{\text{global}}(t)$ and $q_{\text{local}}(t)$ (Fig. 4e):

$$\sigma_{\text{global}}^2(t)=\frac{1}{2}\left[\sigma_2^2(t)-2\sigma_1^2(t)\right]=k_P^2\left\langle \left(q_{\text{global}}(t)\xi(t)\otimes g(t)\right)^2\right\rangle$$

$$\sigma_{\text{local}}^2(t)=2\sigma_1^2(t)-\frac{1}{2}\sigma_2^2(t)=k_2P^2\left\langle (q_{\text{local}}(t)\xi(t)\otimes g(t))^2\right\rangle$$

Received 13 August; accepted 16 October 2007.

1. Kaern, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression: from theories to phenotypes. *Nature Rev. Genet.* **6,** 451–464 (2005).
2. Kaufmann, B. B. & van Oudenaarden, A. Stochastic gene expression: from single molecules to the proteome. *Curr. Opin. Genet. Dev.* **17,** 107–112 (2007).
3. Maheshri, N. & O'Shea, E. K. Living with noisy genes: how cells function reliably with inherent variability in gene expression. *Annu. Rev. Biophys. Biomol. Struct.* **36,** 413–434 (2007).
4. 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).
5. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297,** 1183–1186 (2002).
6. Blake, W. J., Kaern, M., Cantor, C. R. & Collins, J. J. Noise in eukaryotic gene expression. *Nature* **422,** 633–637 (2003).
7. Raser, J. M. & O'Shea, E. K. Control of stochasticity in eukaryotic gene expression. *Science* **304,** 1811–1814 (2004).
8. Paulsson, J. Summing up the noise in gene networks. *Nature* **427,** 415–418 (2004).
9. Cai, L., Friedman, N. & Xie, X. S. Stochastic protein expression in individual cells at the single molecule level. *Nature* **440,** 358–362 (2006).
10. Newman, J. R. *et al.* Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* **441,** 840–846 (2006).
11. Bar-Even, A. *et al.* Noise in protein expression scales with natural protein abundance. *Nature Genet.* **38,** 636–643 (2006).
12. Sigal, A. *et al.* Variability and memory of protein levels in human cells. *Nature* **444,** 643–646 (2006).
13. Geva-Zatorsky, N. *et al.* Oscillations and variability in the p53 system. *Mol. Syst. Biol.* **2,** 0033 (2006).
14. Kondo, T. *et al.* Circadian rhythms in prokaryotes: luciferase as a reporter of circadian gene expression in cyanobacteria. *Proc. Natl Acad. Sci. USA* **90,** 5672–5676 (1993).
15. Liu, Y., Golden, S. S., Kondo, T., Ishiura, M. & Johnson, C. H. Bacterial luciferase as a reporter of circadian gene expression in cyanobacteria. *J. Bacteriol.* **177,** 2080–2086 (1995).
16. Mihalcescu, I., Hsing, W. & Leibler, S. Resilient circadian oscillator revealed in individual cyanobacteria. *Nature* **430,** 81–85 (2004).
17. Andersen, J. B. *et al.* New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl. Environ. Microbiol.* **64,** 2240–2246 (1998).
18. Andersson, C. R. *et al.* Application of bioluminescence to the study of circadian rhythms in cyanobacteria. *Methods Enzymol.* **305,** 527–542 (2000).
19. Xu, Y., Mori, T. & Johnson, C. H. Cyanobacterial circadian clockwork: roles of KaiA, KaiB and the *kaiBC* promoter in regulating KaiC. *EMBO J.* **22,** 2117–2126 (2003).
20. Thattai, M. & van Oudenaarden, A. Intrinsic noise in gene regulatory networks. *Proc. Natl Acad. Sci. USA* **98,** 8614–8619 (2001).
21. Swain, P. S., Elowitz, M. B. & Siggia, E. D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl Acad. Sci. USA* **99,** 12795–12800 (2002).
22. Kepler, T. B. & Elston, T. C. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys. J.* **81,** 3116–3136 (2001).
23. Becskei, A., Kaufmann, B. B. & van Oudenaarden, A. Contributions of low molecule number and chromosomal positioning to stochastic gene expression. *Nature Genet.* **37,** 937–944 (2005).
24. Volfson, D. *et al.* Origins of extrinsic variability in eukaryotic gene expression. *Nature* **439,** 861–864 (2006).
25. 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).
26. Pedraza, J. M. & van Oudenaarden, A. Noise propagation in gene networks. *Science* **307,** 1965–1969 (2005).
27. Johnson, C. H. Global orchestration of gene expression by the biological clock of cyanobacteria. *Genome Biol.* **5,** 217 (2004).
28. Smith, R. M. & Williams, S. B. Circadian rhythms in gene transcription imparted by chromosome compaction in the cyanobacterium *Synechococcus elongatus*. *Proc. Natl Acad. Sci. USA* **103,** 8564–8569 (2006).
29. Vilar, J. M., Kueh, H. Y., Barkai, N. & Leibler, S. Mechanisms of noise-resistance in genetic oscillators. *Proc. Natl Acad. Sci. USA* **99,** 5988–5992 (2002).
30. Gonze, D., Halloy, J. & Goldbeter, A. Robustness of circadian rhythms with respect to molecular noise. *Proc. Natl Acad. Sci. USA* **99,** 673–678 (2002).

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

Acknowledgements We thank S. S. Golden and J. L. Ditty for assistance with the initial phase of this work and their gifts of plasmids and strains. We acknowledge I. Lipchin and M. J. T. O'Kelly for assistance with data collection, cloning and bioluminescence measurements. We acknowledge A. Tolonen, S. W. Chisholm, M. Thattai, H. Lim, J. C. Gore and A. Raj for discussions and suggestions. This work was performed in part at the MIT Laser Biomedical Research Center. This work was supported by NSF and NIH grants.

Author Contributions J.R.C. and P.L. performed the experiments. J.M.P. developed the model. J.R.C., .J.M.P. and A.v.O. designed the experiments, interpreted the results and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.v.O. (avano@mit.edu).