LETTERS

Stochastic protein expression in individual cells at the single molecule level

Long Cai[1]*, Nir Friedman[1]* & X. Sunney Xie[1]

In a living cell, gene expression—the transcription of DNA to messenger RNA followed by translation to protein—occurs stochastically, as a consequence of the low copy number of DNA and mRNA molecules involved[1-6]. These stochastic events of protein production are difficult to observe directly with measurements on large ensembles of cells owing to lack of synchronization among cells. Measurements so far on single cells lack the sensitivity to resolve individual events of protein production. Here we demonstrate a microfluidic-based assay that allows real-time observation of the expression of β-galactosidase in living *Escherichia coli* cells with single molecule sensitivity. We observe that protein production occurs in bursts, with the number of molecules per burst following an exponential distribution. We show that the two key parameters of protein expression—the burst size and frequency—can be either determined directly from real-time monitoring of protein production or extracted from a measurement of the steady-state copy number distribution in a population of cells. Application of this assay to probe gene expression in individual budding yeast and mouse embryonic stem cells demonstrates its generality. Many important proteins are expressed at low levels[7,8], and are thus inaccessible by current genomic and proteomic techniques. This microfluidic single cell assay opens up possibilities for system-wide characterization of the expression of these low copy number proteins.

β-galactosidase (β-gal) has been the standard reporter for gene expression in both prokaryotic and eukaryotic cells[9-12]. Only active as a tetramer, a single molecule of β-gal can produce a large number of fluorescent product molecules by hydrolysing a synthetic fluorogenic substrate (Fig. 1b), leading to amplification of the fluorescent signal, as was first demonstrated in 1961 (ref. 13). However, given its potential as a high-sensitivity cellular reporter, β-gal has an Achilles' heel: its fluorescent products are not retained in the cell[10], because efflux pumps on the cell membrane actively and efficiently expel
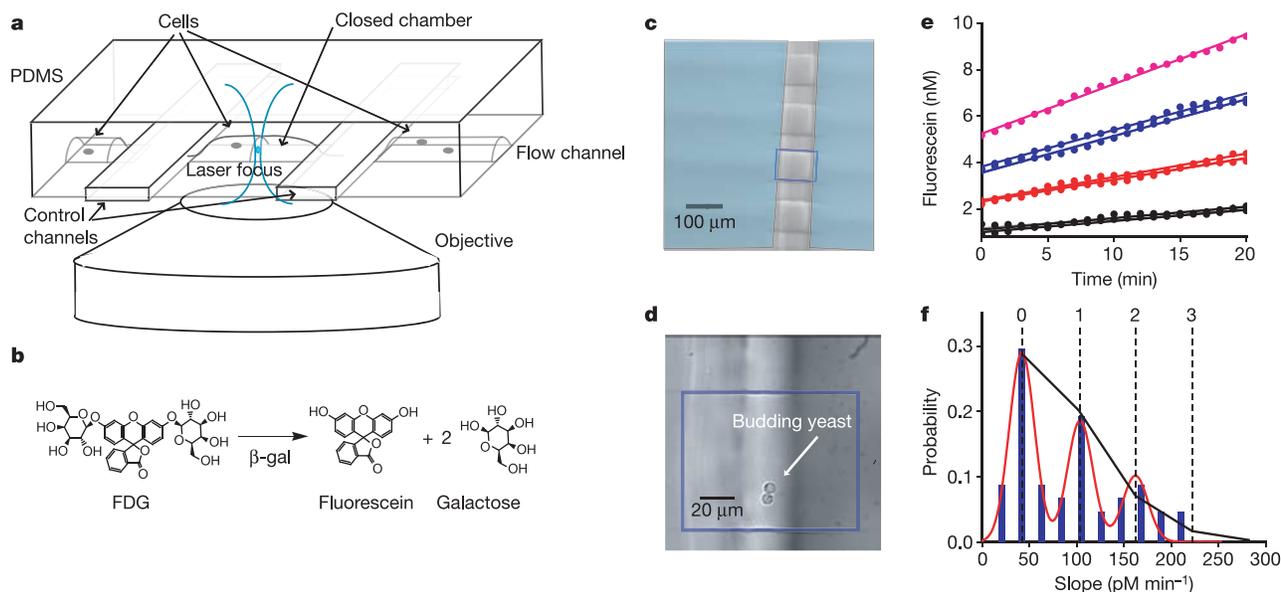
Figure 1 | Single reporter molecule sensitivity in a microfluidic device. a, Schematic diagram of the microfluidic chamber used for the enzymatic assay. Cells are trapped inside a volume of 100 pl, formed by compression of a flow channel by two control channels. **b,** Enzymatic amplification: hydrolysis of the synthetic substrate FDG by the reporter enzyme β-gal yields a fluorescent product, fluorescein. **c,** Differential interference contrast (DIC) image of a microfluidic chip after actuation of the control channels. The boundaries of one closed chamber are boxed in blue. **d,** DIC image of budding yeast cells trapped in a chamber. **e, f,** FDG hydrolysis by purified β-gal. **e,** Fluorescein concentration increases with time in chambers. The discrete slopes are due to, in decreasing order, 3, 2, 1, 0 (autohydrolysis) β-gal molecules. **f,** A histogram of hydrolysis rates (48 chambers). The red curve is a fit to the data with three gaussians of equal widths. The peaks are attributed to 0, 1 and 2 enzyme molecules per chamber. The distribution is well-fitted by a Poisson distribution (black line), with an average of 0.7 β-gal molecules per chamber.

[1]Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA.
*These authors contributed equally to this work.

358

foreign organic molecules from the cytoplasm[14] (see Supplementary Information). As the fluorescent product molecules are pumped to the surrounding medium and rapidly diffuse away, the advantage of enzymatic amplification is lost.

To circumvent the efflux problem, we trap cells in closed micro-fluidic chambers, such that the fluorescent product expelled from the cells can accumulate in the small volume of the chambers, recovering the fluorescence signal due to enzymatic amplification. The fast efflux rate and short mixing time of the fluorescent molecules in the miniature chambers guarantee that the fluorescence signal outside the cells accurately reflects the enzymatic activity inside. The microfluidic device is made of a soft polymer, polydimethyl-siloxane (PDMS), and consists of a flow layer that contains the cells and a top control layer (Fig. 1a)[15,16]. Actuation of two adjacent valves in the control layer forms an enclosure of dimensions $100 \times 100 \times 10\,\mu m^3$ (100 pl) in which cells can be trapped and cultured[17,18] (Fig. 1c, d; see also Supplementary Fig. S1). The micro-fluidic chip is mounted on an inverted fluorescence microscope and translated by a motorized stage, allowing multiplexing of data acquisition by repeatedly scanning the chambers. Typically, 100 chambers can be scanned within less than 2 min. Fluorescence is excited with a tightly focused laser beam (Fig. 1a) that does not directly illuminate the cell, avoiding cellular autofluorescence and photo-damage to the cell.

We first show the ability to detect single enzyme molecules using this technique by injecting a diluted solution of purified β-gal enzyme and 300 μM of the fluorogenic substrate fluorescein-di-β-D-galactopyranoside (FDG) into the chambers[19]. Fluorescence signals from different chambers increase with time, and the slopes give the rates of hydrolysis (Fig. 1e). The distribution of hydrolysis rates measured in the different chambers shows quantized and evenly spaced peaks (Fig. 1f). We attribute these discrete peaks to integer numbers of β-gal molecules. The spacing between the peaks is 60 pM min^{-1}, which gives a calibration for the rate of increase in fluorescein concentration corresponding to one enzyme molecule in a chamber.

Another challenge for using β-gal to monitor gene expression in live cells is that the cell wall acts as a barrier for FDG influx. We quantify this effect in *E. coli* by measuring the hydrolysis rate for live cells compared to cells treated with chloroform, which completely permeabilizes cell membrane (Supplementary Fig. S4a). The ratio of hydrolysis rates between these two cases is defined as the permeability ratio, and is measured to be $R = 13$ at 300 μm FDG. To increase FDG influx, we transformed *E. coli* cells with a plasmid conferring ampicillin resistance and grew the cells in media with β-lactam antibiotics (see Methods). Under these conditions, cell wall synthesis is partially inhibited, making the cells more permeable to FDG, as evident by a lower value of $R = 2 \pm 0.3$ (Supplementary Fig. S4b). In determining the number of enzyme molecules in live cells below, $R = 2$ is used as a correction factor to the *in vitro* calibration (Fig. 1f).

We then monitored gene expression in live *E. coli* cells in real time. β-gal is expressed from the *lacZ* gene on the chromosomal DNA, which is under the control of the *lac* promoter. Cells are grown in glucose-containing medium without inducer to exponential phase; hence the expression level is highly repressed[20]. We observed abrupt changes in hydrolysis rates in chambers with dividing cells, as shown in Fig. 2a, b. These step-wise increases in the rates indicate the stochastic burst-like expression of new β-gal molecules. We attribute the bursts to stochastic and transient dissociation events of the Lac repressor from the promoter, followed by transcription of mRNA, which is then translated into a few copies of the reporter protein before the mRNA is degraded.

The expression of proteins from a given gene can be characterized by two key parameters: the average frequency of expression bursts per cell cycle, *a*; and the average number of protein molecules per burst, *b*. Under conditions of exponential growth in minimal medium, the burst frequency for protein expression from the repressed *E. coli lacZ* gene is measured to be $a = 0.11 \pm 0.03$ bursts per cell cycle. The average burst size is measured to be $b = 5 \pm 2$ enzymes, or 20 ± 8 monomers per burst, which is consistent with biochemical estimates of 25–30 β-gal monomers per mRNA[21,22].

This real-time assay also allows us to measure the distribution of the number of enzymes produced per burst (Fig. 2c). It can be well fitted with an exponential distribution, $P(n) = C\exp(-n/b)$, where *n* is the number of β-gal molecules per burst and *C* is a normalization constant. We attribute this distribution to the fact that the cellular lifetime of the mRNA is exponentially distributed[21,23]. Previously only theoretically predicted[1,24], the exponential $P(n)$ can be accounted for by the competition between mRNA degradation by
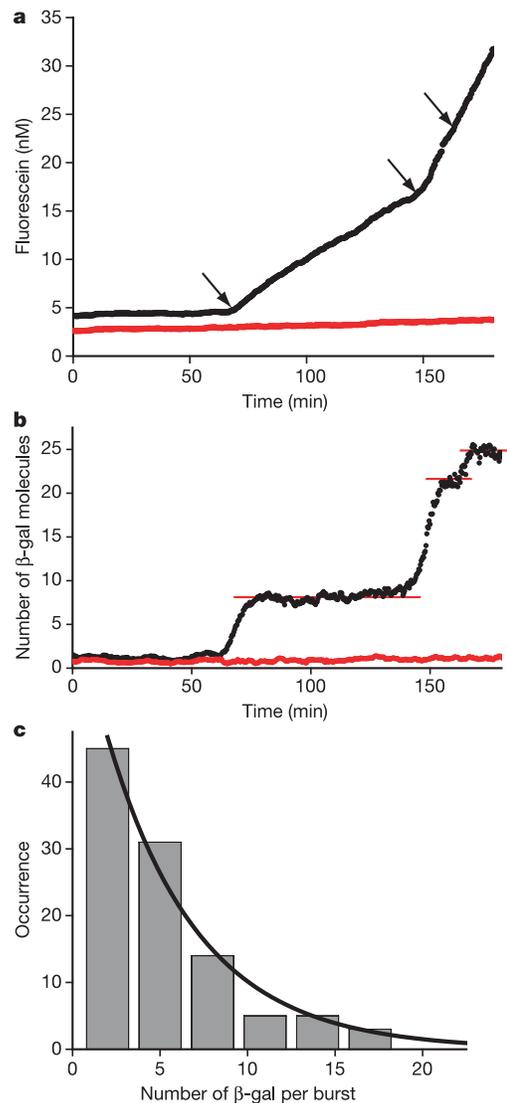
Figure 2 | Quantitative real-time measurement of individual protein expression events in live *E. coli* cells. β-gal is under the control of a repressed *lac* promoter. **a**, Trace of a chamber containing dividing cells shows abrupt changes in hydrolysis rates (arrows on black curve). An empty chamber shows a constant background (red curve). **b**, Discrete jumps in β-gal number are due to burst-like production of proteins. The number of β-gal molecules is calculated by taking the time derivative of the traces in **a** and compensating for fluorescein photobleaching (Supplementary Information). **c**, Histogram of copy number of β-gal molecules per burst. The distribution is well-fitted with an exponential function (black line), with an average of five proteins per burst, and is a consequence of exponential cellular lifetime of the mRNA.
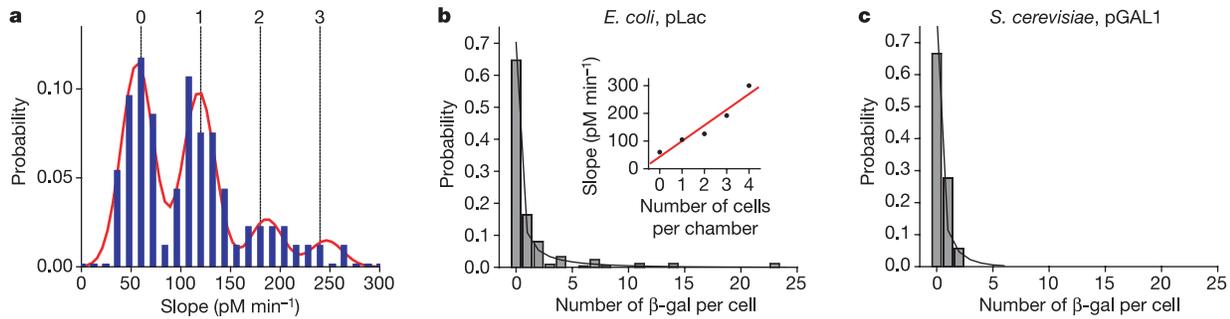
Figure 3 | **Steady-state protein copy number distributions in a population of cells. a**, Histogram of FDG hydrolysis rates from 48 chambers containing chloroform-treated *E. coli* cells shows discrete, evenly spaced peaks, with spacing equal to that measured with purified enzyme (Fig. 1f). **b**, Copy number distribution of β-gal molecules in chambers containing a single *E. coli* cell, under the repressed condition (120 cells). The solid line shows a gamma distribution (equation (1)) with $a = 0.16$ bursts per cell cycle and $b = 7.8$ β-gal molecules per burst. Inset: average hydrolysis rate grows linearly with the number of cells per chamber with a slope of 0.9 enzymes per cell. **c**, Histogram of the number of β-gal molecules in chloroform-treated yeast cells (116 cells) expressing β-gal from the repressed *GAL1* promoter carried on a low copy number plasmid. The solid line is a gamma distribution with $a = 0.2$ and $b = 1.7$.

RNaseE and translation by ribosome. Our single-molecule experiment demonstrates the ability to provide quantitative real-time information on gene expression in a live cell.

Stochasticity of gene expression is exhibited not only in the temporal evolution of protein production in a single cell, but also in the distribution of protein copy number in a population of cells at a particular time. Such distributions have been measured using flow cytometry[3,4] or fluorescence microscopy[2,5,6], but have been limited to only high expression levels. Our method allows measurements of the reporter's distribution in a cell population with single copy sensitivity.

To do this, we treated *E. coli* cells with chloroform, which completely permeabilizes cell membrane and stops protein production, while keeping existing β-gal molecules active inside the cell (Fig. 3b, inset). The treated cells are immediately injected into the microfluidic device together with FDG and the hydrolysis rate is measured. We first demonstrate the detection of single enzyme molecules inside those chloroform-treated cells. The histogram of hydrolysis rates of all chambers shows quantized peaks (Fig. 3a and Supplementary Fig. S12). The spacing between peaks is in agreement with that measured *in vitro* (Fig. 1f).

We compiled a histogram of the number of β-gal molecules in chambers containing only a single cell (Fig. 3b) for cells cultured under the same repressed condition as in the real-time experiment. The mean of this distribution is $m = 1.2 \pm 0.4$ copies of β-gal molecules per cell, and the standard deviation is $\sigma = 3.3$. The mean agrees with the ensemble measurement of 1 ± 0.5 β-gal per cell, as well as the observation that the average number of β-gal molecules per chamber increases with the number of cells trapped in a chamber with a slope of 0.9 enzymes per cell (Fig. 3b inset).

We created a model to relate the measured steady-state protein copy number distributions to the parameters a and b that are defined above. We used a continuous form of the master equation[25] to analytically describe that distribution given the burst dynamics of protein production. We assume that: (1) distribution of protein molecules produced per expression event is exponential; (2) expression events are temporally uncorrelated; and (3) protein molecule copy numbers are halved at cell division. At steady-state, protein production is balanced by protein dilution due to cell growth and division. We found that the steady-state probability distribution of protein number per cell, $p(x)$, is a gamma distribution, which is uniquely determined by the parameters a and b (Supplementary Information):

$$p(x) = \frac{x^{a-1}e^{-x/b}}{b^a\Gamma(a)} \quad (1)$$

where Γ denotes the gamma function. The parameters a and b are related to the mean, m, and standard deviation, σ, by $a = m^2/\sigma^2$ and $b = \sigma^2/m$. Figure 3b shows that the data fits well with equation (1), with $b = 7.8 \pm 2.6$ reporters per burst and $a = 0.16 \pm 0.05$ bursts per cell cycle, which are consistent with the real-time data above within error bars. This demonstrates that the information about the underlying dynamic process, namely the values of a and b, is preserved in both the temporal fluctuations in a single cell and the instantaneous variability within a population.

Notably, the profile of $p(x)$ with $a < 1$ that we measured for the repressed *lac* operon is distinct from $p(x)$ with $a > 1$ that was measured previously for the induced *lac* operon[2]. These correspond to two different regimes of equation (1), schematically illustrated in Fig. 4. For $a < 1$, $p(x)$ peaks at zero and a large fraction of cells contains no β-gal regardless of the burst size b (Fig. 4a); however, for
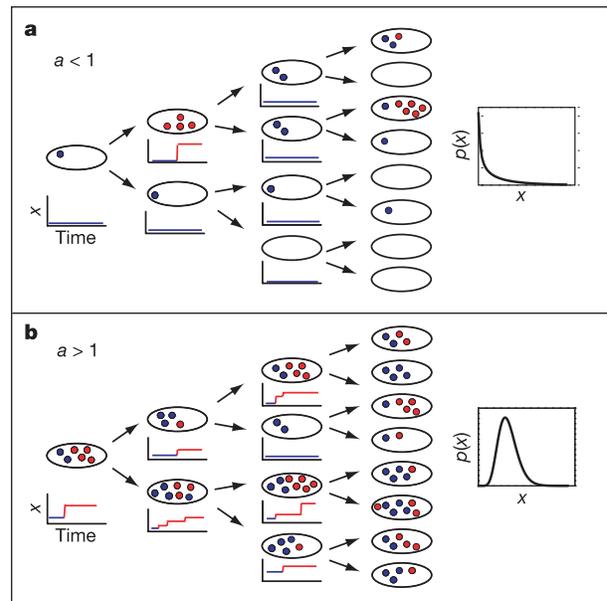
Figure 4 | **Two regimes of stochasticity in protein expression.** Protein molecules are produced in bursts (red), in addition to existing molecules (blue), and diluted upon cell division, leading to a steady-state distribution, $p(x)$, of protein copy number, x, in the cell population. **a**, With $a < 1$, $p(x)$ peaks at $x = 0$ and a large fraction of cells do not contain a single copy of the protein regardless of the burst size, b. The repressed *lac* operon is well described by this regime. **b**, With $a > 1$, most cells contain the protein at some level. Both regimes can be described by the gamma distribution (equation (1)).

$a > 1$, $p(x)$ peaks at a non-zero x and most cells contain β-gal (Fig. 4b).

The profile of $p(x)$ for $a < 1$ may be significant for the response of a population of *E. coli* cells to an increase in lactose concentration, which is transported into the cell by Lac permease. Because β-gal and Lac permease are expressed from the same operon, the large sub-population of cells containing no β-gal would generally have no Lac permease. These cells cannot respond to a rise of lactose concentration in the environment, and thus are distinct from cells containing permease, which can respond and induce further expression of the *lac* operon by way of positive feedback. Thus, $a < 1$ provides a mechanism for 'all-or-none' induction responses that cells exhibit under suboptimal inducer concentrations[11,26].

This microfluidic assay can be applied to other cell types expressing β-gal as a reporter. We demonstrate applicability by measuring low-level expression from repressed promoters in yeast and mammalian cells. Figure 3c shows the distribution of reporter number in a population of chloroform-treated *Saccharomyces cerevisiae* (yKT32) cells expressing β-gal from the repressed *GAL1* promoter. Here, the promoter–reporter fusion is carried on a low copy number plasmid. Cells were grown in the presence of glucose, such that the promoter is expected to be highly repressed. The average expression level is measured to be 0.4 ± 0.07 reporters per cell, with $\sigma = 0.6$. Using the above model (equation (1)) we can extract a and b from the measured mean and σ: $a = 0.2 \pm 0.08$ bursts per cell cycle and $b = 1.7 \pm 0.4$ reporters per burst. This is consistent with our real-time experiments with permeable mutant yeast strain ($pdr1^- \, pdr3^-$) in which we do not observe abrupt large jumps as seen with *E. coli*. The low burst size in this case may reflect the non-optimal codon usage of the bacterial reporter gene in yeast. Finally, we measured β-gal expression also from the ROSA26 promoter in embryonic mouse stem cells (ES 54A). The average expression level was 425 ± 80 enzymes per cell with $\sigma = 375$ (Supplementary Information).

The single cell version of the β-gal assay[9] presented here exhibits single molecule sensitivity for single prokaryotic and eukaryotic cells. This method, together with other newly developed single cell and single molecule techniques at the mRNA[27,28] and protein[29] levels, opens up possibilities for studying low-level gene expression in single cells of diverse cell types (for example, transcription factors involved in gene regulation, or the leakiness of supposedly silenced promoters in differentiated cells). The ease of scaling up the microfluidic assay will enable characterization of gene expression on a whole-genome scale, especially in the unexplored regime of low expression levels.

METHODS

Cell strains and growth conditions. *E. coli* K12 (MG1655) cells are transformed with a medium copy number plasmid, pBR293.3, which confers ampicillin resistance. Cells were cultured in M9 media supplemented with amino acids and vitamins. Experiments were performed at 30 °C. Under these conditions cells grow with a cell cycle time of about 145 min (Supplementary Fig. S1). *S. cerevisiae* cells (yKT0032) express β-gal from the *GAL1* promoter, carried on a low-copy 2 µm plasmid (pYES2-lacZ). Cells were cultured in SD glucose –ura medium (BD Bioscience) at 30 °C. Mouse embryonic stem cells ES54A (Smo + /−, Rosa26 lacZ⁺) and ES17 (lacZ⁻) from a 129SVJ background were cultured in a defined medium to prevent differentiation. Single cell suspension is derived by adding trypsin to the cell culture and resuspending.

Microfluidic chip fabrication. The master mould for the bottom channels is made by spin-coating positive photoresist (Shipley SPR 220-7, MicroChem) onto a silicon wafer (University Wafer) to a height of 7 µm. After patterning with high-resolution transparency mask (PageWorks), the features are rounded by heating[15,16]. The control layer master is patterned with negative photoresist at 40 µm height. PDMS (10:1 Dow Corning Sylgard 184 A:B, Ellsworth Adhesives) is spun onto the bottom master at 2,000 r.p.m. for 1 min and partially cured. Control layer PDMS mould is also partially cured and baked at the thickness of 5 mm. The two PDMS pieces are aligned and cured overnight. The two-layer device is cut out, plasma oxidized for 3 min and bonded permanently to a coverslip.

Fluorescence detection. Fluorescence was excited by a tightly focused beam of a 488-nm Ar-ion laser (Melles Griot), with a power of 27 µW. A high NA objective

lens (Nikon 60 × Plan Apo NA = 1.4) was used in an inverted fluorescence microscope (Nikon) for excitation and collection of the fluorescence signal. A CCD camera (Andor Technology) with on-chip amplification and a large dynamic range was used for detection. Filters used were: excitation, 488 nm narrow bandpass; emission, 535 nm/50 bandpass; dichroic, 495 nm high pass (Chroma Corp). Measurement of signals from multiple chambers was multiplexed by scanning the microfluidic chip using a motorized stage, ASI MS2000 XY (ASI).

Data analysis. Analysis of the fluorescent signal was performed using Matlab (Mathworks). Signal was averaged over the 100 brightest pixels from each image, and background was subtracted. The measured fluorescence time traces $F(t)$ were corrected for photobleaching and time derivative was taken to obtain the number of enzymes in a chamber as a function of time. A median filter of length 11 was used to smooth $F(t)$. With a sampling rate of once every 20 s per chamber, the effective time resolution is about 4 min. The error in calculating the parameters a and b from the copy number distribution was estimated with bootstrap analysis[30]. The original data set was re-sampled 1,000 times. The a and b parameters were calculated for each of those re-sampled data sets, from which the error was estimated. Additional details are provided in the Supplementary Information.

1. McAdams, H. H. & Arkin, A. Stochastic mechanisms in gene expression. *Proc. Natl Acad. Sci. USA* **94**, 814–819 (1997).
2. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
3. 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).
4. Blake, W. J., Kærn, M., Cantor, C. R. & Collins, J. J. Noise in eukaryotic gene expression. *Nature* **422**, 633–637 (2003).
5. Raser, J. M. & O'Shea, E. K. Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811–1814 (2004).
6. 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).
7. Guptasarma, P. Does replication-induced transcription regulate synthesis of the myriad low copy number proteins of *Escherichia coli*? *Bioessays* **17**, 987–997 (1995).
8. Ghaemmaghami, S. *et al.* Global analysis of protein expression in yeast. *Nature* **425**, 737–741 (2003).
9. Miller, J. H. *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, New York, 1972).
10. Russo-Marie, F., Roederer, M., Sager, B., Herzenberg, L. A. & Kaiser, D. β-galactosidase activity in single differentiating bacterial cells. *Proc. Natl Acad. Sci. USA* **90**, 8194–8198 (1993).
11. Novick, A. & Weiner, M. Enzyme induction as an all-or-none phenomenon. *Proc. Natl Acad. Sci. USA* **43**, 553–566 (1957).
12. Nolan, G. P., Fiering, S., Nicolas, J. F. & Herzenberg, L. A. Fluorescence-activated cell analysis and sorting of viable mammalian cells based on β-D-galactosidase activity after transduction of *Escherichia coli* lacZ. *Proc. Natl Acad. Sci. USA* **85**, 2603–2607 (1988).
13. Rotman, B. Measurement of activity of single molecules of β-D-galactosidase. *Proc. Natl Acad. Sci. USA* **47**, 1981–1991 (1961).
14. Nikaido, H. Prevention of drug access to bacterial targets: permeability barriers and active efflux. *Science* **264**, 382–388 (1994).
15. Unger, M. *et al.* Monolithic microfabricated valves and pumps by multilayer soft lithography. *Science* **288**, 113–116 (2000).
16. Wu, H., Wheeler, A. & Zare, R. N. Chemical cytometry on a picoliter-scale integrated microfluidic chip. *Proc. Natl Acad. Sci. USA* **101**, 12809–12813 (2004).
17. McDonald, J. C. *et al.* Fabrication of microfluidic systems in poly(dimethylsiloxane). *Electrophoresis* **21**, 27–40 (2000).
18. Balagadde, F. K., You, L., Hansen, C. L., Arnold, F. H. & Quake, S. R. Long-term monitoring of bacteria undergoing programmed population control in a microchemostat. *Science* **309**, 137–140 (2005).
19. Rondelez, Y. *et al.* Microfabricated arrays of femtoliter chambers allow single molecule enzymology. *Nature Biotechnol.* **23**, 361–365 (2005).
20. Monod, J., Pappenheimer, A. M. Jr & Cohen-Bazire, G. The kinetics of the biosynthesis of β-galactosidase in *Escherichia coli* as a function of growth. *Biochim. Biophys. Acta* **9**, 648–660 (1952).
21. Kennell, D. & Riezman, H. Transcription and translation initiation frequencies of the *Escherichia coli* lac operon. *J. Mol. Biol.* **114**, 1–21 (1977).
22. Sorensen, M. & Pedersen, S. Absolute *in vivo* translation rates of individual codons in *Escherichia coli*. The two glutamic acid codons GAA and GAG are translated with a threefold difference in rate. *J. Mol. Biol.* **222**, 265–280 (1991).
23. Berg, O. G. A model for the statistical fluctuations of protein number in a microbial population. *J. Theor. Biol.* **71**, 587–603 (1975).
24. Yarchuk, O., Jacques, N., Guillerez, J. & Dreyfus, M. Interdependence of translation, transcription and mRNA degradation in the *lacZ* gene. *J. Mol. Biol.* **226**, 581–596 (1992).

25. Paulsson, J., Berg, O. G. & Ehrenberg, M. Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation. *Proc. Natl Acad. Sci. USA* **97**, 7148–7153 (2000).
26. Ozbudak, E. M., Thattai, M., Lim, H. N., Shraiman, B. I. & Van Oudenaarden, A. Multistability in the lactose utilization network of *Escherichia coli*. *Nature* **427**, 737–740 (2004).
27. Levsky, J. M., Shenoy, S. M., Pezo, R. C. & Singer, R. H. Single-cell gene expression profiling. *Science* **297**, 836–840 (2002).
28. Golding, I., Paulsson, J., Zawilski, S. M. & Cox, E. C. Real-time kinetics of gene activity in individual bacteria. *Cell* **123**, 1025–1036 (2005).
29. Yu, J. *et al.* Probing gene expression in live cells, one protein molecule at a time. *Science* **311**, 1600–1603 (2006).
30. Efron, B. & Tibshirani, R. J. *An Introduction to the Bootstrap* (Chapman & Hall, New York, 1993).