nature genetics

General properties of transcriptional time series in *Escherichia coli*

Lok-hang So[1,2], Anandamohan Ghosh[3,4], Chenghang Zong[2], Leonardo A Sepúlveda[1,5], Ronen Segev[3,4] & Ido Golding[1,2,5,6]

Gene activity is described by the time series of discrete, stochastic mRNA production events. This transcriptional time series shows intermittent, bursty behavior. One consequence of this temporal intricacy is that gene expression can be tuned by varying different features of the time series. Here we quantify copy-number statistics of mRNA from 20 *Escherichia coli* promoters using single-molecule fluorescence *in situ* hybridization in order to characterize the general properties of these transcriptional time series. We find that the degree of burstiness is correlated with gene expression level but is largely independent of other parameters of gene regulation. The observed behavior can be explained by the underlying variation in the duration of bursting events. Using Shannon's mutual information function, we estimate the mutual information transmitted between an outside stimulus, such as the extracellular concentration of inducer molecules, and intracellular levels of mRNA. This suggests that the outside stimulus transmits information reflected in the properties of transcriptional time series.

A gene's activity can be described by the discrete time series of mRNA production events[1,2]. This transcriptional time series is stochastic rather than deterministic[2–4]. Furthermore, it generally cannot be described as a simple Poisson process. In other words, mRNA molecules are not produced with a constant probability per unit time; instead, mRNA production is often bursty (pulsatile) in both bacteria[2] and higher organisms[4–8]. A suitable mathematical framework for describing gene activity data is the two-state model[8–10], where a gene stochastically fluctuates between 'off' and 'on' states, and mRNA is produced stochastically only in the on state. This scenario can lead to the occurrence of transcription 'bursts', periods of intense activity separated by periods of quiescence. Measured mRNA kinetics[2,5] and copy-number statistics[2,8,11,12] have been shown to be consistent with the two-state picture in a variety of model systems. However, despite considerable theoretical attention[2,13–17], we do not have a biophysical understanding of the nature of the on and off states and what governs the transitions between them.

An important consequence of the temporal intricacy of the transcriptional time series is that the expression level of a gene (defined here as the mean mRNA copy number per cell, $\langle n \rangle$) does not uniquely determine the parameters of the time series. In other words, an ensemble of many different time series can produce the same mRNA level. Similarly, a change in the level of expression (as, for example, in response to different stimulus levels) can in principle occur by varying different properties of the transcriptional time series[8], henceforth referred to as different 'modulation schemes'. This is shown in **Figure 1** for a hypothetical bacterial promoter. In the example shown, the mean mRNA level $\langle n \rangle$ is tuned over a 30-fold range in response to a change in environmental stimulus (for example, the concentration of a specific sugar in the growth medium; **Fig. 2a**). Changes in mRNA level can be obtained by modifying any of the three kinetic parameters characterizing mRNA production (**Fig. 1b**), thereby modulating different properties of the transcriptional time series: k_{on}, the rate of switching to the on state ('on rate'), which determines the rate of transcription bursts; k_{off}, the rate of switching back to the off state ('off rate'), which determines the duration of transcription bursts; and k_{TX}, the rate of producing mRNA while in the on state, which determines how many mRNAs are produced during each transcription burst. mRNA degradation naturally affects expression level as well, and its rate can be modified as a regulatory mechanism[18–20]. In our analysis below, we decouple such effects from variations in mRNA production parameters by correcting for differences in mRNA lifetimes. We also assume that only a single kinetic parameter is altered when changing expression level (for the alternative, see **Supplementary Note**). As seen in **Figure 1c**, each of the modulation schemes creates a transcriptional time series of different characteristics at a given gene expression level. Even though these different time series produce the same mean level of mRNA, the different characteristics of the time series are in turn reflected in the degree of cell-to-cell variability in mRNA numbers. This effect can be quantified using the Fano factor (b) (ref. 2), defined as the ratio of the variance (σ^2) to the mean ($\langle n \rangle$) of mRNA copy number. $b = 1$ corresponds to nonbursty (Poissonian) mRNA production. For short, rapid bursts, b is equal to the mRNA

[1]Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas, USA. [2]Department of Physics, University of Illinois, Urbana, Illinois, USA. [3]Department of Life Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel. [4]Zlotowski Center for Neuroscience, Ben-Gurion University of the Negev, Beer-Sheva, Israel. [5]Center for Biophysics and Computational Biology, University of Illinois, Urbana, Illinois, USA. [6]Center for the Physics of Living Cells, University of Illinois, Urbana, Illinois, USA. Correspondence should be addressed to I.G. (igolding@illinois.edu).

Received 10 December 2010; accepted 5 April 2011; published online 1 May 2011; doi:10.1038/ng.821

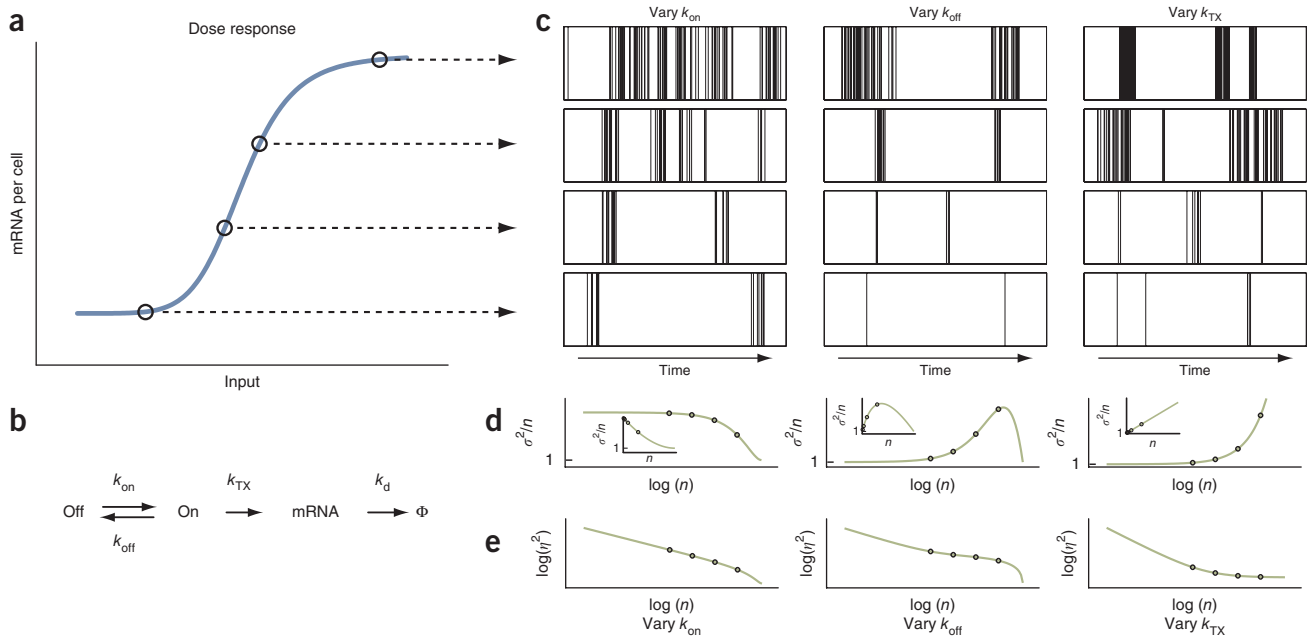

Figure 1 Different features of the transcriptional time series can be modulated to vary gene expression level. (**a**) Schematic representation of the gene-activity curve for a typical bacterial promoter. The expression level (mean number of mRNAs per cell, $\langle n \rangle$) as a function of the external stimulus is shown. The curve is arbitrary but is typical of the sigmoidal response shown by many bacterial promoters[27,57] (for example, see **Fig. 2a**). (**b**) The kinetic parameters governing mRNA production and annihilation in the two-state model. (**c**) Different modulation schemes of the transcriptional time series, all capable of creating the gene activity curve in **a**. Each plot shows the time series of mRNA production events (bars). We created the data by simulating the two-state model using the Gillespie method[75]. In each of the three cases shown, we varied only a single parameter of gene activity (k_{on}, left; k_{off}, middle; k_{TX}, right). All time series in the same row produced the same mean mRNA level $\langle n \rangle$. (**d**) The effect of the different modulation schemes on the observed mRNA copy-number statistics. The burstiness, $b = \sigma^2 / \langle n \rangle$, is plotted as a function of the mean mRNA number $\langle n \rangle$. The main panel shows $b(\langle n \rangle)$ on a semilog scale, and the insets show the same data on a linear scale. We calculated $b(\langle n \rangle)$ analytically for the two-state model[9]. (**e**) The noise, $\eta^2 = \sigma^2 / \langle n \rangle^2$, as a function of the mean mRNA number $\langle n \rangle$. We calculated $\eta^2(\langle n \rangle)$ analytically for the two-state model[9]. For more details see Online Methods.

burst size[21]. In the more general case, b indicates how 'bursty' the time series is relative to a Poisson process[22] (**Supplementary Fig. 1**). For simplicity, we refer to b as the 'burstiness' of the transcriptional time series. The two-state transcription model allows us to calculate $\langle n \rangle$ and σ, and therefore b, for any set of kinetic parameters[8–10,23,24]. As seen in **Figure 1d**, each of the modulation schemes described above yields a typical curve for b as a function of the mean mRNA level $\langle n \rangle$. These curves are distinct from each other; thus, measuring $b(\langle n \rangle)$ experimentally would in principle allow us to discriminate among the different scenarios and identify which kinetic parameter of the transcriptional time series is varied. Similar analysis can be performed on the 'noise' in the time series, quantified by the squared coefficient of variation $\eta^2 = \sigma^2/\langle n \rangle^2$ (ref. 25) (**Fig. 1e**).

RESULTS

Quantifying mRNA statistics at single-molecule resolution

We quantified the copy-number statistics of endogenous mRNA using single-molecule fluorescence *in situ* hybridization (smFISH) following the method given in reference 26, which we adapted for counting mRNA in *E. coli* at single-transcript resolution (Online Methods). Briefly, we designed a set of ~50–70 fluorescently labeled oligonucleotide probes, each 20 bases in length, against the transcript of interest (**Supplementary Table 1**). We hybridized the probes to fixed cells and imaged them using epifluorescence microscopy. To estimate the number of mRNA molecules from the gene of interest in a given cell, we measured the total intensity of fluorescent foci in the cell, yielding an estimate of the number of bound probes, in turn indicating the number of target mRNA molecules. This approach follows that previously used in live-cell studies of mRNA kinetics using the MS2 system[2]. **Figure 2** shows the dynamic range and accuracy of measuring mRNA copy numbers using smFISH for the case of the P_{lac} promoter. mRNA levels covering approximately three orders of magnitude (~0.1–60 molecules per cell) could be measured. The smFISH-based estimation of mRNA numbers was in excellent agreement with measurements using quantitative PCR (qPCR) as well as with data from the literature[27]. A similar comparison made in four other promoters yielded good agreement between smFISH and other assays (**Supplementary Figs. 2–5**). The smFISH-based measurements allowed us to obtain the copy-number statistics of mRNA transcripts from a gene of interest under a given growth condition. The mRNA histograms were well described by a negative binomial distribution[28] (**Fig. 2c**), consistent with the prediction of the two-state model[8,10]. In particular, the smFISH data allowed us to accurately measure both the mean ($\langle n \rangle$) and variance (σ^2) of mRNA copy number and therefore calculate the burstiness parameter $b = \sigma^2/\langle n \rangle$ characterizing the transcriptional time series.

Burstiness exhibits similar behavior across genes and conditions

We used smFISH to quantify mRNA statistics from 20 promoters: P_{lac}[27], $P_{galETKM}$[29,30], P_{marII}[31], *rrnB*P1[32,33], $P_{bioBFCD}$[34,35], bacteriophage λ promoter P_R and 13 variants of the bacteriophage λ promoter P_{RM}[36–38] (**Supplementary Tables 2–4**). In cases where promoter activity is regulated by growth conditions (for example, the presence of a specific sugar or amino acid), we used a range of growth conditions so that the full range of mRNA levels could be obtained (**Supplementary Note**). This ensemble of promoters

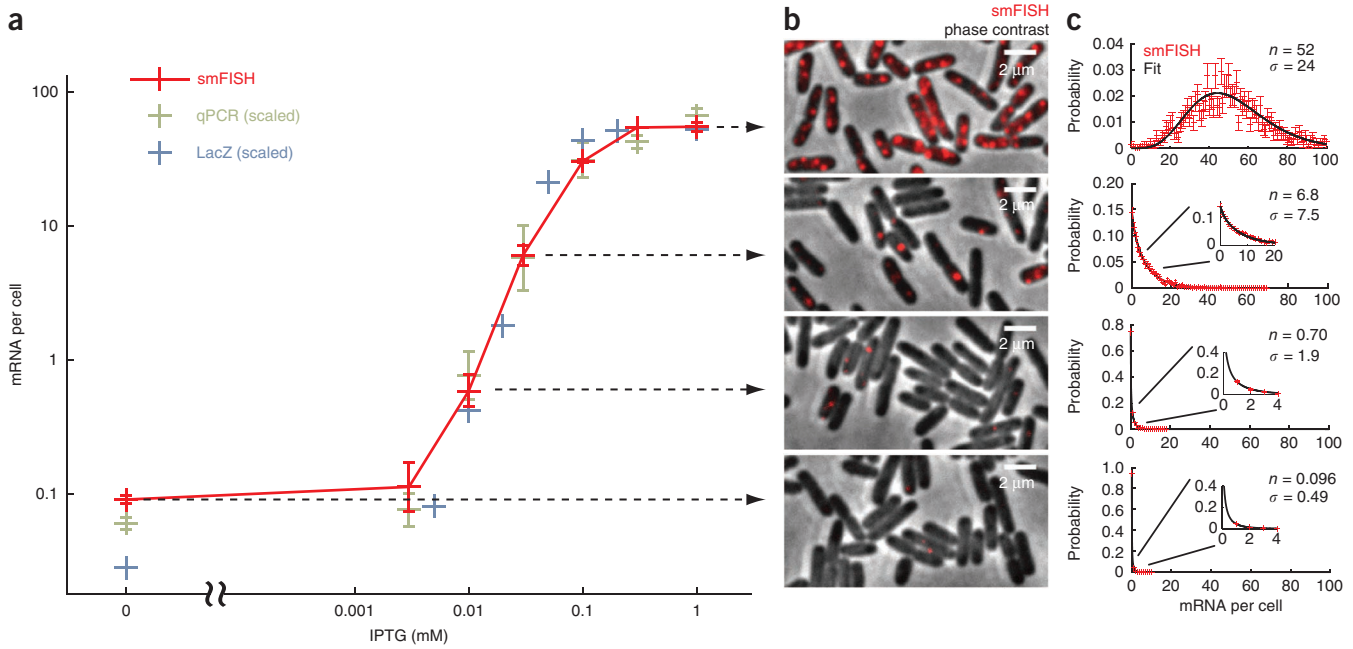

Figure 2 Single-molecule FISH (smFISH) used to characterize mRNA copy-number statistics. (**a**) Gene expression level (mRNA per cell) from the P_{lac} promoter as a function of inducer (isopropyl β-D-1-thiogalactopyranoside, IPTG) concentration. The mean mRNA number per cell as measured by smFISH (average of two independent experiments) is shown, as well as the results of quantitative PCR (qPCR; average of two independent experiments; normalized by the mean smFISH level) and β-galactosidase activity assay, as reported in the literature[27] (normalized by the mean smFISH level). Error bars represent the standard errors from duplicate experiments. The good agreement between the three assays, over approximately three orders of expression level, shows the accuracy and dynamic range of the smFISH method. (**b**) Typical images of smFISH-labeled cells at different induction levels. An overlay of the phase contrast (grayscale) and smFISH probes targeting the *lacZ* gene (red) is shown. Each image corresponds to the expression level designated by the horizontal arrow. (**c**) *lacZ* mRNA copy-number histograms obtained from smFISH at different induction levels. The experimental data (red) and the fit to a negative binomial distribution (black) are shown, as well as the estimated values for mean mRNA number $\langle n \rangle$ and standard deviation σ in that sample. Each plot corresponds to the expression level designated by the horizontal arrow. See Online Methods for more information.

allowed us to scan a range of expression levels (~0.01–60 mRNA per cell), different molecular mechanisms of transcription regulation (activation, repression and combinations thereof) and topologies of gene networks controlling gene activity (such as the presence or absence of feedback[39]; **Supplementary Table 4**). All of these factors can conceivably affect the observed fluctuations in gene activity[40–45]. In total, we performed >150 independent experiments, with each one yielding the distribution of mRNA copy number from a given gene at a given stimulus level.

To characterize the transcriptional time series in the complete dataset, we plotted (**Fig. 3a**,**b**) the burstiness b and the noise η^2 from each experiment as a function of the mean expression level $\langle n \rangle$ at that condition. We corrected the expression levels for the differences in gene copy number (**Supplementary Figs. 6**,**7** and **Supplementary Note**) and mRNA lifetime (**Supplementary Tables 5**,**6**, **Supplementary Figs. 8**,**9** and **Supplementary Note**) so that the characteristics of mRNA production from a single-copy promoter could be examined. We first note that the cell-to-cell variability in mRNA numbers is dominated by the inherent fluctuations of the two-state process ('intrinsic noise') rather than by cell-to-cell difference in parameter values ('extrinsic noise'). This is suggested by the following observations: (i) the noise η^2 decreases monotonically with $\langle n \rangle$ (**Fig. 3b**), which is the typical behavior of intrinsic, but not of extrinsic, noise[46]. (ii) In the limit of low $<n>$, $\sigma^2/<n> \approx 1$ (**Fig. 3a**), as expected for the intrinsic noise of a Poisson process. That transcription is Poissonian at very low expression levels has been shown previously[47,48]. (iii) In the limit of high $\langle n \rangle$, η^2 decreases sharply rather than approaching a plateau (**Fig. 3b**). Such a plateau would be expected in the presence of extrinsic noise[46,49]. The observed dominance of intrinsic noise in mRNA number fluctuations is consistent with previous observations that extrinsic noise is an important factor at the level of the protein species[46,49] but not mRNA[2].

The most striking feature in **Figure 3a** and **b** is that b and η^2 show gene-independent behavior; that is, the values from different genes and growth conditions show a clear trend, with a dependence on the expression level $\langle n \rangle$ alone. Thus, the properties of the time series seem to depend primarily on the mean mRNA level, not on the specific gene or stimulus (this observation is made more quantitative below). The gene-independent behavior immediately suggests that the rate parameters in the two-state picture are not determined by the details of molecular regulation of an individual promoter (such as the binding and unbinding kinetics of a specific transcription factor) or the topology of the individual gene network (for example, the presence or absence of feedback). Instead, gene on and off switching is dominated by a process that acts in a similar manner on different genes, possibly exerting its influence at a genome-wide level (see the discussion section below). Thus, all genes expressed at a given level have a similar transcriptional time series. Note that this similarity in time-series characteristics does not necessarily mean that the actual activity of different genes is coordinated in time (that is, that genes turn on and off in unison). It is interesting to note, however, that multiple copies of the same gene (present when the bacterial chromosome replicates) have a positive, nonzero covariance (**Supplementary Fig. 10**), suggesting that their temporal activity may indeed be correlated. As we discuss below, the observed universality in transcription

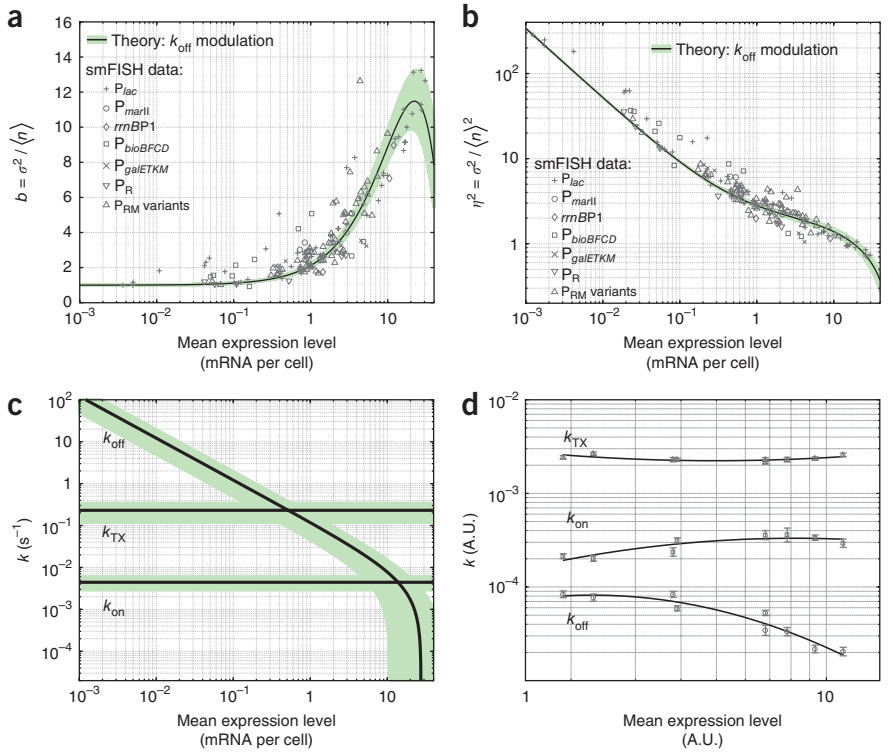

Figure 3 Gene expression level in *E. coli* is varied by changing the gene off rate. (**a**) The burstiness b as a function of the mean expression level $\langle n \rangle$. Markers, smFISH data. Solid line, theoretical prediction for the case of varying only k_{off}. We obtained the theoretical curve by solving analytically the expression for $b(\langle n \rangle)$ and then using k_{on} and k_{TX} as fitting parameters. The shaded green area designates the 95% confidence interval of the fit. (**b**) The noise η^2 as a function of the mean expression level $\langle n \rangle$. Notations are as in **a**. We used the theoretical parameters (k_{on} and k_{TX}) extracted from fitting $b(\langle n \rangle)$ in **a** to plot the theoretical curve. (**c**) The estimated rate parameters for gene activity in *E. coli*. We obtained these parameters from fitting $b(\langle n \rangle)$ in **a** to the case of varying k_{off} in the two-state model. The errors in k_{on} and k_{TX} (green shade) are based on the variability in estimates between individual promoters (**Supplementary Fig. 11**). We calculated the error in k_{off} (green shade) from the resulting fit. (**d**) Direct measurement of the two-state rate parameters in individual living cells. We quantified mRNA production from the promoter $P_{lac/ara}$ using the MS2-GFP method[2]. Data (markers) are from nine independent experiments (>400 cells). Error bars represent standard errors within each experiment. Solid lines are fits to second degree polynomials. See Online Methods for more information.

burstiness readily explains previous observations made at the protein level[49] (and similar findings in yeast[25]).

Expression level is varied by modulating the gene off rate

We next used the experimental data of $b(\langle n \rangle)$ and $\eta^2(\langle n \rangle)$ to ask what property of the transcriptional time series is modulated as gene expression level is varied. When comparing the experimental plots in **Figure 3a** and **b** to the theoretical plots in **Figure 1d** and **e**, we note that the observed mRNA statistics are consistent with the assumption that expression level is changed by varying the rate at which the gene switches back to the off state (off rate, k_{off}), or, in other words, the duration of transcription bursts. Specifically, note that $b(\langle n \rangle)$ starts with a Poisson-like behavior ($b \sim 1$) and then increases as a sub-linear function of $\langle n \rangle$. The observation can be made quantitative by fitting the experimental data for $b(\langle n \rangle)$ and $\eta^2(\langle n \rangle)$ to the analytical expressions for the two-state model under the scenario of varying k_{off} (Online Methods)[8–10,23,24]. As seen in **Figure 3a**, we obtained a good fit ($R^2 = 0.81$). For comparison, trying to fit the observed data with the two alternative scenarios, modulating the gene on rate k_{on} or the transcription rate k_{TX}, yielded inferior fits ($R^2 = -7.9 \times 10^{-6}$ and $R^2 = 0.58$, respectively). Moreover, the scenario of varying k_{off} yielded a fit superior to the alternatives when compared on a promoter-by-promoter basis (**Supplementary Figs. 11** and **12**). As a control, trying to fit a simulated collection of promoters with randomly selected kinetic parameters using the k_{off} modulation description also yielded a poor fit ($R^2 = 0.085$; **Supplementary Fig. 13**). As an additional test for the validity of our parameter estimation, we performed detailed stochastic simulations of mRNA kinetics and verified that the theoretical and experimental copy-number histograms were in agreement beyond the values of $\langle n \rangle$ and σ^2 (**Supplementary Fig. 14**). The theoretical fit allows us to make the observation of gene independence more quantitative: when comparing the data from individual promoters to the universal fit, we found that six of seven datasets had a correlation coefficient above 0.85 between the data and theory (**Supplementary Fig. 15**). The average deviation of single-promoter data from the universal fit is ~33% (**Supplementary Fig. 15**).

Fitting the experimental data to the scenario of k_{off} modulation allowed us to estimate the values of the three kinetic parameters governing mRNA production: k_{on} (the rate of switching to the on state, which determines the frequency of bursts), k_{TX} (the rate of producing mRNA while the gene is on)—both of which are approximately constant for different genes and expression levels—and k_{off} (the rate of switching back to the off state, which determines the duration of bursts), which changes over more than three orders of magnitude when expression level is varied (**Fig. 3c**). We note that, of these three parameters, the only one which has been estimated in the past is k_{TX}, which corresponds to the maximal transcription initiation rate possible (when a gene is constantly on). The value obtained from our single-cell measurements ($k_{TX} = 0.23 \pm 0.11$ per second (s.e.m.)) is in good agreement with values from the literature[50–52]. We also note that k_{on} and k_{TX} have a dependence on the bacterial growth rate (**Supplementary Fig. 12**).

The examination of mRNA number statistics, though strongly indicating that k_{off} alone is varied to control expression level, is limited by the fact that we did not directly observe the process of transcription. To overcome this limitation and gain further support for the observation of k_{off} modulation, we quantified the kinetics of mRNA production from one promoter, $P_{lac/ara}$[53], in individual living cells. We used the MS2-GFP system[54,55], previously used to demonstrate transcriptional bursting in *E. coli*[2,56]. Briefly, cells were grown under the microscope in the presence of different levels of the inducers isopropyl β-D-thiogalactopyranoside (IPTG) and arabinose. We followed mRNA production in individual cells by measuring the intensity of fluorescent foci created when MS2-GFP binds to its RNA recognition sequence[2]. As expected, we found mRNA kinetics to consist

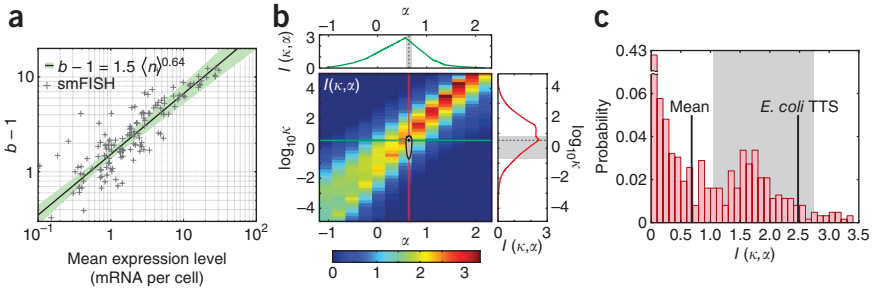

Figure 4 The transcriptional time series optimizes information representation by the cell. (**a**) The plot shows $b - 1 = \sigma^2 / \langle n \rangle - 1$ as a function of the mean expression level $\langle n \rangle$. Markers designate experimental data (we used the same dataset here as in **Fig. 3a**). Solid line, fit to a power law $\sigma^2 / \langle n \rangle - 1 = \langle n \rangle^{\alpha} / \kappa$. The power law yielded a good fit ($R^2 = 0.76$) in the range $\langle n \rangle \approx 0.3 - 40$ and allowed an estimation of the parameters κ and α. (**b**) The calculated mutual information I between the outside stimulus and the transcriptional time series (scaled to represent the protein species) is plotted for a typical bacterial promoter. A power-law behavior of $b(\langle n \rangle)$ was assumed, $b - 1 = \langle n \rangle^{\alpha} / \kappa$, and I is plotted as a function of the parameters κ and α. As seen from the plots to the right and above, the values of κ and α corresponding to the experimental data lie very close to the 'ridge' in $I(\kappa,\alpha)$. The shaded region around the experimental data point (+) represents the error estimate based on multiple sources: κ and α estimation from the fit in **a**; the number of protein molecules produced from each mRNA[47,52]; mRNA lifetime[18]; and cell doubling time. (**c**) The histogram of mutual information (I) values is plotted for the different (κ,α) combinations examined in **b**. The *E. coli* transcriptional time series has a mutual information value ($I \approx 2.5$) that is much higher than the average performance by all possible modulation schemes ($I \approx 0.68$). The shaded area corresponds to the experimental error estimates for κ and α, as in **b**. See Online Methods for more information.

of periods of activity, in which a random number of transcripts are produced, separated by periods of inactivity[2]. Measuring the mean durations of off and on periods, as well as the amount of mRNA produced within each on period, allowed us then to estimate k_{on}, k_{off} and k_{TX} at a given gene activity level. As seen in **Figure 3d**, the behavior of these kinetic parameters is consistent with the observations above: changing the level of mRNA $\langle n \rangle$ is achieved by varying k_{off} while k_{on} and k_{TX} are kept approximately constant.

Information representation by transcriptional time series

We have thus seen that the discrete time series of gene activity have general properties; that is, the kinetic parameters are similar between different genes and environmental conditions. It is then natural to ask, can the specific choice of kinetic parameters optimize some function of the living cell and therefore be subject to evolutionary selection[57]? To address this question, we followed an approach previously used[58,59] and considered the way gene activity is used by the cell to represent information about its environment. For example, the activity of the lactose promoter can be thought of as 'telling' the cell how much lactose is present in its environment. We quantified the efficiency of information representation by the cell using Shannon's mutual information[60], $I(p,c)$, a function that measures how much information is transmitted to the output (protein level, p) about changes in the input stimulus, c (for example, sugar concentration). In calculating $I(p,c)$, we used the experimentally measured dose response of the promoter, that is, the mean mRNA number $\langle n \rangle$ as a function of stimulus c. We modeled the downstream production of protein using known parameters[47,48,52] (**Supplementary Note**). Importantly, a calculation using three different promoters studied in this work (P_{lac}, P_{marII} and $P_{bioBFCD}$) yielded almost identical results (**Supplementary Figs. 16,17**). The mutual information I depends critically on the way the variance of mRNA copy numbers, σ^2, changes with the mean $\langle n \rangle$ (the statistics of the protein species follows the same scaling relations, up to a calculable factor; **Supplementary Note**). To examine how the mutual information varies as a function of time-series parameters, we wrote σ^2 in the phenomenological form $\sigma^2/\langle n \rangle = 1 + \langle n \rangle^{\alpha}/\kappa$ (such that the deviation of the burstiness b from the Poisson case goes as the mean $\langle n \rangle$ to the power α). By varying the parameters κ and α, this functional form allowed us to approximate the behavior shown by the transcriptional time series under the different modulation schemes (**Fig. 1d**) and under a broad range of kinetic parameters; specifically, this form captures the $\sigma^2/\langle n \rangle$ behavior seen in our experiments (**Fig. 4a**). We next calculated the mutual information (maximized over possible inputs; Online Methods) as a function of the parameters (κ,α) (**Fig. 4b**), thus exploring the efficiency of information representation over the space of possible time-series characteristics. We found that the parameters describing the actual transcriptional time series ($\kappa = 3.5 \pm 3.2$, $\alpha = 0.64 \pm 0.06$) are close to optimal: they lie on a 'ridge' in the map of $I(\kappa,\alpha)$ (**Fig. 4b**). When plotting a histogram of I values obtained from a broad range of kinetic parameters (**Fig. 4c**), one sees that the maximal mutual information of the actual time series ($I \approx 2.5$ bits, or discrimination of >5 input levels) is markedly higher than the mean performance obtained by randomly choosing the time-series parameters (≈0.68 bits). In other words, the specific parameters of the transcriptional time series show higher optimization than most other possible parameter sets in the sense of allowing the cell to transmit information about its environment.

DISCUSSION

Multiple studies in recent years have shown that gene activity is often bursty rather than Poissonian[2,4–8] and can be described by a two-state model for mRNA production[8–10]. In this work, we have extended and generalized these observations by describing how the transcriptional time series in *E. coli* is modulated when gene expression level is varied. We found that promoter activity tends to be nonbursty at low expression levels (at or below $\langle n \rangle \sim 1$ molecule per cell); the degree of burstiness, as characterized by the Fano factor $b = \sigma^2/\langle n \rangle$, then rises in a sub-linear manner with increasing gene activity. This behavior is consistent with varying the gene off rate as the means to change the expression level while maintaining a constant gene on rate and transcription rate. In other words, the duration of the transcription bursts is the main feature that changes as expression level is varied. Importantly, this behavior is not gene or input specific (although it can also be observed when examining a single gene; **Supplementary Fig. 11**); rather, it was observed through the set of promoters and stimuli examined. We note that a more complex scenario, in which multiple kinetic parameters are simultaneously varied, is also consistent with the observed smFISH data (**Supplementary Note** and **Supplementary Fig. 18**). However, such a scenario does not need to be invoked in order to explain the experimental data. The multi-parameter modulation scenario also appears inconsistent with the live-cell data (**Fig. 3d**).

A number of past studies have characterized the noise level of multiple genes using a library of fluorescent protein fusions[25,49,61,62].

A study in yeast[25] found that the squared coefficient of variation η^2 showed a genome-wide trend of power-law dependence on mean expression level (a similar trend was recently observed when examining different mutants of a single yeast promoter[63]). This gene-independent behavior is consistent with our findings here. Moreover, by modeling the underlying kinetics, the authors in reference 25 concluded that protein fluctuations were likely dominated by the mRNA species, as was assumed in our work. A recent genome-wide study in *E. coli*[49] found that the Fano factor increased monotonically with mean protein level. This observation is most easily explained by our findings of gene-independent behavior of the transcriptional burst size (**Fig. 3a**). The authors in reference 49 also performed measurement of mRNA levels in single cells, which they analyzed using the assumption of Poissonian kinetics[49]. The measured values of mRNA numbers per cell, as well as the range of expression levels, were substantially smaller than those observed in our study. In addition, the authors found no correlation between mRNA and protein numbers from a given gene in individual cells. It is possible that the use of a single fluorescent probe per gene limited the accuracy of their measurement (see for example, Supplementary Figure 22 in ref. 49) and thus did not allow a quantitative characterization of cell-to-cell variability in mRNA numbers.

From an evolutionary point of view, we note that the expression level of a gene has been shown to be a phenotype that is subject to selection[64]. More recently, a number of studies have suggested that, beyond the mean expression level, the degree of population heterogeneity (noise) in gene expression may also be subject to selection[65]. Here we estimated the mutual information between an external stimulus and the transcriptional time series and showed that the specific modulation scheme chosen by the cell is efficient in the sense of reliably representing, through the transcriptional time series, the environment in which the cell resides. In quantifying this efficiency, we showed how the properties of the transcriptional time series itself, beyond just the mean expression level, emerge as a meaningful phenotype subject to selection. We note that this new observation also extends and generalizes previous works in which the burstiness of gene expression was suggested to affect the cellular phenotype[66–70].

Two important limitations of our work need to be mentioned: first, when describing gene activity, we centered on the mRNA species only and left out the downstream production of proteins. In doing so, we implicitly assumed that protein kinetics are enslaved by mRNA kinetics to a sufficient degree such that the discrete, stochastic time series of mRNA production governs cell-to-cell heterogeneity[2,25]. This assumption is supported by the observation that protein-number statistics[49] closely reflect the properties of mRNA statistics, as found here. Second, by mainly using *in situ* hybridization to count mRNA, we were able to obtain snapshots of cell populations but were naturally unable to follow the time course of gene activity in individual cells (with the exception of a single promoter). This limitation prevented us from examining temporal correlations in the transcriptional time series. Correlations in the gene-activity trajectories of individual cells have been shown to contain important information about the underlying gene regulatory network[45,71]. Such correlations are likely to be affected by the bursty behavior described here. Extending the use of the MS2-based system[2] to multiple promoters should allow the characterization of such temporal effects in the future.

At this stage, there is no mechanistic, molecular-level understanding of what gives rise to the bursty behavior of gene activity in bacteria; specifically, what the physiological nature of the gene on and off states is and, therefore, also how the rates of switching between states can be varied in the individual cell or over the time course of evolution. The most common theoretical model used to explain two-state gene activity in bacteria involves the binding and unbinding of transcription factors at the promoter[2,13,23,24,40]. However, our finding here that the properties of the transcriptional time series are gene independent (rather than gene specific) suggests that the observed two-state kinetics involves gene-nonspecific features such as DNA topology, RNA polymerase dynamics or regulation by broad-target DNA-binding proteins[13,72,73]. Interestingly, these types of mechanisms are reminiscent of those suggested to underlie non-Poissonian transcription kinetics in eukaryotes, where burstiness is broadly ascribed to chromatin modifications[8,17,74]. Future studies are needed to consider whether observation of transcription burstiness in both kingdoms reflects a similarity in underlying mechanisms or results from the selection of an advantageous phenotype in different systems.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We are grateful to S. Adhya, W. Boos, M. Cashel, V. Chakravartty, L. Chubiz, D. Court, J. Cronan, M. Dreyfus, M. Elowitz, Y. Feng, H. Garcia, T. Hwa, J. Imlay, S. Jang, T. Kuhlman, J. Little, A. van Oudenaarden, A. Raj, C. Rao, R. Milo, V. Shahrezaei, A. Sokac and P. Swain for advice and for providing reagents. We thank members of the Golding laboratory for providing help with experiments. We thank H. Garcia, J. Kondev, R. Phillips, A. Raj, Á. Sánchez, S. Sawai and G. Tkačik for commenting on earlier versions of the manuscript. Work in the Golding laboratory was supported by grants from US National Institutes of Health (R01GM082837) and the National Science Foundation (082265, PFC: Center for the Physics of Living Cells). Work in the Segev and Golding laboratories was supported by a joint grant from the Human Frontier Science Program (RGY 70/2008).

AUTHOR CONTRIBUTIONS

I.G., L.-h.S. and R.S. designed the project. L.-h.S. performed the majority of experiments and the theoretical analysis of gene activity. L.A.S. and C.Z. performed additional experiments and developed analysis tools for gene activity. R.S. and A.G. performed the information theory analysis. I.G., L.-h.S., L.A.S., R.S. and A.G. wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Golding, I. & Cox, E.C. Eukaryotic transcription: what does it mean for a gene to be 'on'? *Curr. Biol.* **16**, R371–R373 (2006).
2. Golding, I., Paulsson, J., Zawilski, S.M. & Cox, E.C. Real-time kinetics of gene activity in individual bacteria. *Cell* **123**, 1025–1036 (2005).
3. Raj, A. & van Oudenaarden, A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**, 216–226 (2008).
4. Chubb, J.R. & Liverpool, T.B. Bursts and pulses: insights from single cell studies into transcriptional mechanisms. *Curr. Opin. Genet. Dev.* **20**, 478–484 (2010).
5. Chubb, J.R., Trcek, T., Shenoy, S.M. & Singer, R.H. Transcriptional pulsing of a developmental gene. *Curr. Biol.* **16**, 1018–1025 (2006).
6. Paré, A. *et al.* Visualization of individual Scr mRNAs during *Drosophila* embryogenesis yields evidence for transcriptional bursting. *Curr. Biol.* **19**, 2037–2042 (2009).
7. Yunger, S., Rosenfeld, L., Garini, Y. & Shav-Tal, Y. Single-allele analysis of transcription kinetics in living mammalian cells. *Nat. Methods* **7**, 631–633 (2010).
8. Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y. & Tyagi, S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* **4**, e309 (2006).
9. Peccoud, J. & Ycart, B. Markovian modeling of gene-product synthesis. *Theor. Popul. Biol.* **48**, 222–234 (1995).
10. Shahrezaei, V. & Swain, P.S. Analytical distributions for stochastic gene expression. *Proc. Natl. Acad. Sci. USA* **105**, 17256–17261 (2008).

11. Zenklusen, D., Larson, D.R. & Singer, R.H. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat. Struct. Mol. Biol.* **15**, 1263–1271 (2008).
12. Tan, R.Z. & van Oudenaarden, A. Transcript counting in single cells reveals dynamics of rDNA transcription. *Mol. Syst. Biol.* **6**, 358 (2010).
13. Mitarai, N., Dodd, I.B., Crooks, M.T. & Sneppen, K. The generation of promoter-mediated transcriptional noise in bacteria. *PLOS Comput. Biol.* **4**, e1000109 (2008).
14. Dobrzynski, M. & Bruggeman, F.J. Elongation dynamics shape bursty transcription and translation. *Proc. Natl. Acad. Sci. USA* **106**, 2583–2588 (2009).
15. van Zon, J.S., Morelli, M.J., Tanase-Nicola, S. & ten Wolde, P.R. Diffusion of transcription factors can drastically enhance the noise in gene expression. *Biophys. J.* **91**, 4350–4367 (2006).
16. Fang, F.C. Sigma cascades in prokaryotic regulatory networks. *Proc. Natl. Acad. Sci. USA* **102**, 4933–4934 (2005).
17. Blake, W.J., Kaern, M., Cantor, C.R. & Collins, J.J. Noise in eukaryotic gene expression. *Nature* **422**, 633–637 (2003).
18. Bernstein, J.A., Khodursky, A.B., Lin, P.H., Lin-Chao, S. & Cohen, S.N. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl. Acad. Sci. USA* **99**, 9697–9702 (2002).
19. Selinger, D.W., Saxena, R.M., Cheung, K.J., Church, G.M. & Rosenow, C. Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. *Genome Res.* **13**, 216–223 (2003).
20. Singh, D. *et al.* Regulation of ribonuclease E activity by the L4 ribosomal protein of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **106**, 864–869 (2009).
21. Thattai, M. & van Oudenaarden, A. Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. USA* **98**, 8614–8619 (2001).
22. Goh, K.I. & Barabasi, A. Burstiness and memory in complex systems. *Epl* **81**, 48002 (2008).
23. Kepler, T.B. & Elston, T.C. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys. J.* **81**, 3116–3136 (2001).
24. Simpson, M.L., Cox, C.D. & Sayler, G.S. Frequency domain chemical Langevin analysis of stochasticity in gene transcriptional regulation. *J. Theor. Biol.* **229**, 383–394 (2004).
25. Bar-Even, A. *et al.* Noise in protein expression scales with natural protein abundance. *Nat. Genet.* **38**, 636–643 (2006).
26. Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).
27. Kuhlman, T., Zhang, Z., Saier, M.H. Jr. & Hwa, T. Combinatorial transcriptional control of the lactose operon of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **104**, 6043–6048 (2007).
28. Paulsson, J. & Ehrenberg, M. Random signal fluctuations can reduce random fluctuations in regulated components of chemical regulatory networks. *Phys. Rev. Lett.* **84**, 5447–5450 (2000).
29. Tokeson, J.P., Garges, S. & Adhya, S. Further inducibility of a constitutive system: ultrainduction of the gal operon. *J. Bacteriol.* **173**, 2319–2327 (1991).
30. Weickert, M.J. & Adhya, S. The galactose regulon of *Escherichia coli*. *Mol. Microbiol.* **10**, 245–251 (1993).
31. Alekshun, M.N. & Levy, S.B. The mar regulon: multiple resistance to antibiotics and other toxic chemicals. *Trends Microbiol.* **7**, 410–413 (1999).
32. Hernandez, V.J. & Bremer, H. Guanosine tetraphosphate (ppGpp) dependence of the growth rate control of rrnB P1 promoter activity in *Escherichia coli*. *J. Biol. Chem.* **265**, 11605–11614 (1990).
33. Potrykus, K. *et al.* Antagonistic regulation of *Escherichia coli* ribosomal RNA rrnB P1 promoter activity by GreA and DksA. *J. Biol. Chem.* **281**, 15238–15248 (2006).
34. Abdel-Hamid, A.M. & Cronan, J.E. Coordinate expression of the acetyl coenzyme A carboxylase genes, accB and accC, is necessary for normal regulation of biotin synthesis in *Escherichia coli*. *J. Bacteriol.* **189**, 369–376 (2007).
35. Barker, D.F. & Campbell, A.M. Use of bio-lac fusion strains to study regulation of biotin biosynthesis in *Escherichia coli*. *J. Bacteriol.* **143**, 789–800 (1980).
36. Michalowski, C.B. & Little, J.W. Positive autoregulation of cI is a dispensable feature of the phage lambda gene regulatory circuitry. *J. Bacteriol.* **187**, 6430–6442 (2005).
37. Sauer, R.T., Jordan, S.R. & Pabo, C.O. Lambda repressor: a model system for understanding protein-DNA interactions and protein stability. *Adv. Protein Chem.* **40**, 1–61 (1990).
38. Lim, W.A. & Sauer, R.T. Alternative packing arrangements in the hydrophobic core of lambda repressor. *Nature* **339**, 31–36 (1989).
39. Shen-Orr, S.S., Milo, R., Mangan, S. & Alon, U. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.* **31**, 64–68 (2002).
40. Garcia, H.G., Sanchez, A., Kuhlman, T., Kondev, J. & Phillips, R. Transcription by the numbers redux: experiments and calculations that surprise. *Trends Cell Biol.* **20**, 723–733 (2010).
41. Kittisopikul, M. & Suel, G.M. Biological role of noise encoded in a genetic network motif. *Proc. Natl. Acad. Sci. USA* **107**, 13300–13305 (2010).
42. Kaern, M., Elston, T.C., Blake, W.J. & Collins, J.J. Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* **6**, 451–464 (2005).
43. Pedraza, J.M. & van Oudenaarden, A. Noise propagation in gene networks. *Science* **307**, 1965–1969 (2005).
44. Kittisopikul, M. & Suel, G.M. Biological role of noise encoded in a genetic network motif. *Proc. Natl. Acad. Sci. USA* **107**, 13300–13305 (2010).
45. Cox, C.D., McCollum, J.M., Allen, M.S., Dar, R.D. & Simpson, M.L. Using noise to probe and characterize gene circuits. *Proc. Natl. Acad. Sci. USA* **105**, 10809–10814 (2008).
46. Elowitz, M.B., Levine, A.J., Siggia, E.D. & Swain, P.S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
47. Cai, L., Friedman, N. & Xie, X.S. Stochastic protein expression in individual cells at the single molecule level. *Nature* **440**, 358–362 (2006).
48. Yu, J., Xiao, J., Ren, X., Lao, K. & Xie, X.S. Probing gene expression in live cells, one protein molecule at a time. *Science* **311**, 1600–1603 (2006).
49. Taniguchi, Y. *et al.* Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538 (2010).
50. Kennell, D. & Riezman, H. Transcription and translation initiation frequencies of the *Escherichia coli lac* operon. *J. Mol. Biol.* **114**, 1–21 (1977).
51. Liang, S. *et al.* Activities of constitutive promoters in *Escherichia coli*. *J. Mol. Biol.* **292**, 19–37 (1999).
52. Neidhardt, F.C. Rererer. *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology* (American Society for Microbiology, Washington, D.C., 1987).
53. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1–I2 regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210 (1997).
54. Fusco, D. *et al.* Single mRNA molecules demonstrate probabilistic movement in living mammalian cells. *Curr. Biol.* **13**, 161–167 (2003).
55. Bertrand, E. *et al.* Localization of ASH1 mRNA particles in living yeast. *Mol. Cell* **2**, 437–445 (1998).
56. Golding, I. & Cox, E.C. RNA dynamics in live *Escherichia coli* cells. *Proc. Natl. Acad. Sci. USA* **101**, 11310–11315 (2004).
57. Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Chapman & Hall/CRC, Boca Raton, Florida, USA, 2007).
58. Tkacik, G., Callan, C.G. Jr. & Bialek, W. Information flow and optimization in transcriptional regulation. *Proc. Natl. Acad. Sci. USA* **105**, 12265–12270 (2008).
59. Tkacik, G., Walczak, A.M. & Bialek, W. Optimizing information flow in small genetic networks. *Phys. Rev. E* **80**, 031920 (2009).
60. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423, 623–656 (1948).
61. Newman, J.R. *et al.* Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* **441**, 840–846 (2006).
62. Geva-Zatorsky, N. *et al.* Protein dynamics in drug combinations: a linear superposition of individual-drug responses. *Cell* **140**, 643–651 (2010).
63. Mao, C. *et al.* Quantitative analysis of the transcription control mechanism. *Mol. Syst. Biol.* **6**, 431 (2010).
64. Dekel, E. & Alon, U. Optimality and evolutionary tuning of the expression level of a protein. *Nature* **436**, 588–592 (2005).
65. Eldar, A. & Elowitz, M.B. Functional roles for noise in genetic circuits. *Nature* **467**, 167–173 (2010).
66. Singh, A., Razooky, B., Cox, C.D., Simpson, M.L. & Weinberger, L.S. Transcriptional bursting from the HIV-1 promoter is a significant source of stochastic noise in HIV-1 gene expression. *Biophys. J.* **98**, L32–L34 (2010).
67. Kaufmann, B.B., Yang, Q., Mettetal, J.T. & van Oudenaarden, A. Heritable stochastic switching revealed by single-cell genealogy. *PLoS Biol.* **5**, e239 (2007).
68. Choi, P.J., Cai, L., Frieda, K. & Xie, X.S. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science* **322**, 442–446 (2008).
69. Zeng, L. *et al.* Decision making at a subcellular level determines the outcome of bacteriophage infection. *Cell* **141**, 682–691 (2010).
70. Zong, C., So, L.-h., Sepulveda, L.A., Skinner, S.O. & Golding, I. Lysogen stability is determined by the frequency of activity bursts from the fate-determining gene. *Mol. Syst. Biol.* **6**, 440 (2010).
71. Austin, D.W. *et al.* Gene network shaping of inherent noise spectra. *Nature* **439**, 608–611 (2006).
72. Mooney, R.A. *et al.* Regulator trafficking on bacterial transcription units *in vivo*. *Mol. Cell* **33**, 97–108 (2009).
73. Reppas, N.B., Wade, J.T., Church, G.M. & Struhl, K. The transition between transcriptional initiation and elongation in *E. coli* is highly variable and often rate limiting. *Mol. Cell* **24**, 747–757 (2006).
74. Whitelaw, N.C., Chong, S. & Whitelaw, E. Tuning in to noise: epigenetics and intangible variation. *Dev. Cell* **19**, 649–650 (2010).
75. Gillespie, D.T. Exact stochastic simulation of coupled chemical-reactions. *J. Phys. Chem.* **81**, 2340–2361 (1977).

ONLINE METHODS

Growth media and conditions. All strains were grown in M9 minimal media supplemented with thiamine, casamino acids and glucose (Teknova, #M8010) unless otherwise stated. All strains were grown at 37 °C with shaking unless otherwise stated (**Supplementary Note**).

Strains. Bacterial strains used are listed in **Supplementary Table 2**. All of the strains are *E. coli* K-12 derivatives. BW14894 was used as a negative control for smFISH experiments using *lacZ* probes. MG1655 was used as a negative control for smFISH experiments using *cI* and *cro* probes. MG1655, JL5902 and JL2497 were used as hosts for bacteriophage λ in smFISH experiments using *cI* probes to study the P_{RM} promoter. Phage strains used are listed in **Supplementary Table 3**. **Supplementary Table 4** provides additional details of the promoters used in this study.

Fluorescence *in situ* hybridization (FISH). The procedures are based on reference 26 and our recent adaptation of the protocol for use with *E. coli*[70] (**Supplementary Note**).

Microscopy. We pipetted 1 μl of sample onto a 24 × 50 mm #1 coverslip (Fisher Scientific, #12-545F). A 1 mm thick 1.5% agarose gel pad (in 1× PBS) was laid on the sample. A 22 × 22 mm #1 coverslip (Fisher Scientific, #12-545B) was placed on top of the agarose gel pad. The sample was imaged using an inverted epifluorescence microscope (Nikon Instruments Eclipse TE2000-E) and a cooled EMCCD camera (Photometrics Cascade 512). A 100× numerical aperture 1.40 oil immersion phase contrast objective (Nikon Instruments Plan Apo) was used in conjunction with a 2.5× lens in front of the camera. The microscope and camera were controlled using the Metamorph software (Molecular Devices). A TexasRed filter set (Nikon Instruments, #96365) was used for imaging the smFISH probes, and a DAPI filter set (Nikon Instruments, #96310) was used to image DNA stained by DAPI. *z*-stacks with nine slices at 250 nm spacing were acquired for phase contrast and TexasRed images. Each sample was imaged at multiple slide positions to obtain a total of at least 1,000 cells.

Data analysis. All image processing and data analysis were performed using MATLAB (MathWorks).

Cell recognition was performed on phase-contrast images of cells using the *Schnitzcell* MATLAB module (gift of M. Elowitz, California Institute of Technology). The output was checked and corrected using the manual interface offered by *Schnitzcell*. The output consisted of label matrices representing areas occupied by cells, which were then used for further analysis.

A spot recognition program developed in our lab was used to automatically identify and quantify localized fluorescence signals. A Gaussian filter was first applied to smooth out noise, and spots were recognized by the presence of a local maximum in both the *x* and *y* directions. This was done at each *z* position in the stack of images, and each spot was quantified at the *z* position where it had the highest fluorescence intensity (that is, where the spot is in focus).

The fluorescence intensity corresponding to a single mRNA needed to be estimated in each smFISH experiment so that the total fluorescence intensity in a cell could be normalized to give the absolute number of mRNAs. The typical intensity of false positives in an experiment was first estimated from the histogram of individual spot intensities of a negative control. Histograms of individual spot intensities from low-expression samples were then examined. Because most spots in these samples were expected to contain a single mRNA, the first peak that emerged above the false positive range in each of these histograms served as an estimate for the intensity of a single mRNA. The mean intensity of the first peaks from multiple such histograms was taken as the single mRNA intensity for that particular experiment. The sum of intensities of all spots in each cell was then normalized by that value to give the absolute number of mRNAs in the cell.

In the approximation that mRNA production occurs in short rapid bursts, transcription kinetics is characterized by two parameters, the burst size *b* and burst frequency *r*[8,10]. These parameters were estimated in two different ways: first, the histograms of mRNA copy number were fitted to a negative binomial distribution using the MATLAB Curve Fitting Toolbox, and *b* and *r* were calculated from the fitting parameters[8,10]. Second, *b* and *r* were estimated using the relations $b = \sigma^2 / \langle n \rangle$ and $(b - 1)r = \langle n \rangle$, where σ^2 and $\langle n \rangle$ are the variance and mean of mRNA copy number. The two methods gave values of *b* and *r* that were in good agreement (data not shown).

The measured values of *n* and *b* were corrected for differences in the gene copy number and mRNA lifetime between different genes and growth conditions (**Supplementary Note**).

Gillespie simulation of the two-state transcription model. The Gillespie algorithm[75] was used to simulate the stochastic kinetics of the two-state model[8,10] (**Fig. 1**). In this model, each copy of a gene can either be in an active (on) or inactive (off) state. It switches from on to off with rate k_{off} and from off to on with rate k_{on}. When the gene is 'on', mRNA is produced at a rate k_{TX}. mRNA is degraded at a rate k_d. The probability per unit time of each reaction occurring was calculated from these reaction rates, and the reaction trajectory was simulated stochastically. Cell division and gene replication were incorporated as optional features in the simulations. Reactant species segregated according to binomial statistics upon cell division. Gene copy number doubled at a specified time. To mimic smFISH experiments, a random time point was chosen from each of 1,000 trajectories, at which the number of mRNA was 'measured'.

Estimation of transcription parameters. The analytical expressions for variance and mean of mRNA numbers, arising from the two-state model[9], are:

$$\sigma^2 = \frac{k_{on}k_{TX}}{(k_{on}+k_{off})k_d} + \frac{k_{on}k_{off}}{(k_{on}+k_{off})^2}\frac{k_{TX}^2}{k_d(k_{on}+k_{off}+k_d)} \quad (1)$$

and

$$\langle n \rangle = \frac{k_{on}k_{TX}}{(k_{on}+k_{off})k_d} \quad (2)$$

These expressions were used to write $\sigma^2/\langle n \rangle$ as a function of $\langle n \rangle$. If we assume that k_{on} is modulated, we get (by eliminating k_{on} from equation 1 and equation 2)

$$\frac{\sigma^2}{\langle n \rangle} = 1 + \frac{(k_{TX} - k_d\langle n \rangle)^2}{k_{TX}k_{off} + k_d(k_{TX} - k_d\langle n \rangle)} \quad (3)$$

Similarly, if we assume that k_{off} is modulated, we get

$$\frac{\sigma^2}{\langle n \rangle} = 1 + \frac{k_d(k_{TX} - k_d\langle n \rangle)\langle n \rangle}{k_{on}k_{TX} + k_d^2\langle n \rangle} \quad (4)$$

and if we assume k_{TX} is modulated, we get

$$\frac{\sigma^2}{\langle n \rangle} = 1 + \frac{k_d k_{off}\langle n \rangle}{k_{on}(k_{on}+k_{off}+k_d)} \quad (5)$$

The MATLAB Curve Fitting Toolbox was used to fit these functions to data obtained from smFISH experiments. As described in the main text, the fit to k_{off} modulation was significantly better than the alternative ones, supporting the hypothesis that k_{off} is modulated.

Calculating the mutual information between outside stimulus and gene activity. The mutual information $I(m,c)$ gives an estimate of how much information is transmitted to the output level of mRNAs (*m*, or similarly the protein level; **Supplementary Note**) about changes in the input concentration, *c*. We take the approximation that the input values are drawn from a Gaussian distribution $P(c)$ with mean μ and variance σ^2. From experiments (for example, **Fig. 2a**), we also know the dependence of mean mRNA number on *c*, $n = f(c)$, so by knowing $P(c)$ we can evaluate the distribution of *n* as function of the input *c* using the formula

$$P(n) = P(c)\left|\frac{\partial f(c)}{\partial c}\right|^{-1} \quad (6)$$

Equivalently we can compute the mutual information between the output expression level *m* and the mean *n*,

$$I(m,n) = \int dm \int dn\, P(m,n)\log_2\left[\frac{P(m,n)}{P(m)P(n)}\right] \quad (7)$$

The experimental mRNA (and therefore protein) statistics follows the relation $\sigma^2/\langle n\rangle = 1 + \langle n\rangle^{\alpha}/\kappa$. We can thus write the probability distribution above in terms of n alone. It is interpreted as the conditional distribution $P(m|n)$, the probability of obtaining m mRNAs having their mean fixed at n.

Using the identity $P(m,n) = P(m \mid n)P(n)$, it is straightforward to numerically compute $I(m,n)$ for varying mean μ and variance σ^2 of the input distribution and obtain the pair of $\{\mu,\sigma^2\}$ that maximizes the mutual information for this specific form of input distribution (Gaussian). The information transfer is maximal if the input is Gaussian and the channel or the conditional distribution is also Gaussian. For the negative binomial distribution, it is not known what input distribution will maximize the mutual information; however, the Gaussian input is a good approximation in the limit of large n. The maximum of the mutual information thus estimated for different choice of $\{\kappa,\alpha\}$ is shown in **Figure 4**.

doi:10.1038/ng.821