HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2011 December 01.

Published in final edited form as:

Nat Genet. 2011 June ; 43(6): 554–560. doi:10.1038/ng.821.

GENERAL PROPERTIES OF THE 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], and **Ido Golding**[1,2,5,6]

[1] Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas 77030, USA

[2] Department of Physics, University of Illinois, Urbana, Illinois 61801, USA

[3] Department of Life Sciences, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel

[4] Zlotowski Center for Neuroscience, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel

[5] Center for Biophysics and Computational Biology, University of Illinois, Urbana, Illinois 61801, USA

[6] Center for the Physics of Living Cells, University of Illinois, Urbana, Illinois 61801, USA

Abstract

Gene activity is described by the time-series of discrete, stochastic mRNA production events. This *transcriptional time-series* exhibits intermittent, bursty behavior. One consequence of this temporal intricacy is that gene expression can be tuned by varying different features of the time-series. What schemes for varying the transcriptional time-series are observed in the cell? Are the observed properties of these time-series optimized for cellular function? To address these questions, we characterize mRNA copy-number statistics at single-molecule resolution from multiple *Escherichia coli* promoters. We find that the degree of burstiness depends only on the gene expression level, while being independent of the details of gene regulation. The observed behavior is explained by the underlying variation in the duration of bursting events. Using information theory, we find that the properties of the transcriptional time series allow the cell to efficiently map the extracellular concentration of inducer molecules to intracellular levels of mRNA and proteins.

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 [2,3,4]. Furthermore, it generally cannot be described as a simple Poisson process. In other words,

Contact: Ido Golding, igolding@illinois.edu.

AUTHOR CONTRIBUTION

I.G., L.-h.S., and R.S. designed the project. L.-h.S. performed the majority of experiments and 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.

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 [4,5,6,7,8]. A suitable mathematical framework for describing gene activity data is the two-state model [8,9,10], 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", i.e. 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, this picture is still purely phenomenological. Despite considerable theoretical attention [2,13,14,15,16,17] we do not possess 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 times series [8], henceforth referred to as different "modulation schemes". This is demonstrated 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; see FIGURE 2A below). Changes in mRNA level can be obtained by modifying any of the three kinetic parameters characterizing mRNA production (FIGURE 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 mRNA's are produced during each transcription burst. (mRNA degradation naturally affects the expression level as well, and its rate can be modified as a regulatory mechanism [18,19,20]. 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) [2], defined as the ratio of the variance (σ^2) to the mean ($\langle n \rangle$) of mRNA copy-number. $b = 1$ corresponds to non-bursty (Poissonian) mRNA production. For short, rapid bursts, b is equal to the mRNA burst size [21]. In the more general case, b indicates how "bursty" the time-series is relative to a Poisson process [22] (FIGURE S1). 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 [8,9,10,23,24]. 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 $\eta^2 = \sigma^2/\langle n \rangle^2$ [25] (FIGURE 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 of [26], which we adapted for counting mRNA in *E. coli* at single-transcript resolution (see **ONLINE METHODS**). Briefly, a set of ~50–70 fluorescently labeled oligonucleotide probes, each 20 bases in length, was designed against the transcript-of-interest. Probes were hybridized to fixed cells and imaged using epifluorescence microscopy. To estimate the number of mRNA molecules from the gene-of-interest in a given cell, the total intensity of fluorescent foci in the cell was measured, yielding an estimate of the number of bound probes, in turn indicating the number of target mRNA molecules. This approach follows the one previously used in live-cell studies of mRNA kinetics using the MS2 system [2]. FIGURE 2 demonstrates the dynamic range and accuracy of measuring mRNA copy-numbers using smFISH, for the case of the P_{lac} promoter. mRNA levels covering ~3 orders of magnitude (~0.1–60 molecules/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 (FIGURES S2–S5). 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] (FIGURE 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 universal behavior across different 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} [36,37,38] (see TABLES S1–S3). In cases where promoter activity is regulated by growth conditions (e.g. the presence of a specific sugar or amino acid), a range of growth conditions was used so that the full range of mRNA levels could be achieved (see SUPPLEMENTARY NOTE). This ensemble of promoters allowed us to scan a range of expression levels (~0.01–60 mRNA/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]); see TABLE S3. All of these factors can conceivably affect the observed fluctuations in gene activity [40,41,42,43,44,45]. In total, >150 independent experiments were made, 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 data set, we plot (FIGURES 3A and 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. The expression levels were corrected for the differences in gene copy number (see FIGURES S6 and S7 and SUPPLEMENTARY NOTE) and mRNA lifetime (see TABLE S4, FIGURES S8 and S9 and SUPPLEMENTARY NOTE), so that the characteristics of mRNA production from a single-copy promoter can 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 $\langle n \rangle$ (FIGURE 3B), which is the typical behavior of intrinsic, but not of extrinsic, noise [46]. (ii) In the limit of low $\langle n \rangle$, $\frac{\sigma^2}{\langle n \rangle} \approx 1$ (FIGURE 3A), as expected for the intrinsic noise of a Poisson process. That transcription is Poissonian at very low expression level has been shown previously [47,48]. (iii) In the limit of high $\langle n \rangle$, η^2 decreases sharply rather than approach a plateau (FIGURE 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 FIGURES 3A and 3B is that b and η^2 exhibit 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 (e.g. the presence or absence of feedback). Instead, gene on/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 discussion below). Thus, all genes expressed at a given level exhibit 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, i.e. 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) exhibit a positive, non-zero covariance (FIGURE S10), suggesting that their temporal activity may indeed be correlated. As we discuss below, the observed universality in transcription 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 FIGURES 3A and 3B to the theoretical ones in FIGURES 1D and 1E, we note that the observed mRNA statistics is 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} (see **ONLINE METHODS**) [8,9,10,23,24]. As seen in FIGURE 3A, a good fit is obtained ($R^2 = 0.81$). 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 0.58, respectively). Moreover, the scenario of varying k_{off} yields a fit superior to the alternatives when compared on a promoter-by-promoter basis (FIGURES S11 and S12). 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 ($R^2 = 0.085$, FIGURE S13). 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 are in agreement, beyond the mere values of $\langle n \rangle$ and σ^2 (FIGURE S14). 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 find that 6 of 7 data sets exhibit a correlation coefficient above 0.85 between data and theory (FIGURE S15). The average deviation of a single-promoter data from the universal fit is ~33% (FIGURE S15).

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) and 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 3 orders of magnitude when expression level is varied (FIGURE 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 ($k_{TX} = 0.23 \pm 0.11\ s^{-1}$) is in good agreement with values from the literature [50,51,52]. We also note that k_{on} and k_{TX} exhibit a dependence on the bacterial growth rate (FIGURE S12).

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 the process of transcription was not directly observed. 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, IPTG and arabinose. mRNA production was followed in individual cells by measuring the intensity of fluorescent foci created when MS2-GFP binds to its RNA recognition sequence [2]. As expected, mRNA kinetics was found to consist of periods of activity, where 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 in turn 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.

The transcriptional time-series optimizes information representation by the cell

We have thus seen that the discrete time-series of gene activity exhibits universal properties; that is, the same kinetic parameters are common to 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 the approach of [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, i.e. mean mRNA number $\langle n \rangle$ as a function of stimulus c. The downstream production of protein was stochastically modeled using known parameters [47,48,52] (see SUPPLEMENTARY NOTE). Importantly, a calculation using three different promoters studied in this work (P_{lac}, P_{marII}, and $P_{bioBFCD}$) yielded almost identical results (FIGURE S16 and S17). 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, see 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 like 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 exhibited by the transcriptional time-series under the different modulation schemes (see FIGURE 1D above) and under a broad range of kinetic parameters; specifically, this form captures the $\sigma^2(\langle n \rangle)$ behavior seen in our experiments (FIGURE 4A). We next calculated the mutual information (maximized over possible inputs, see **ONLINE METHODS**) as a function of the parameters (κ,α) (FIGURE 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 $I(\kappa,\alpha)$ (FIGURE 4B). When plotting a histogram of I values obtained from a broad range of kinetic parameters (FIGURE 4C), one sees that the maximal mutual information of the actual time series (I~2.5 bits, or discrimination of >5 input levels), is significantly 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, which are observed in experiment, are superior to most other possible parameter sets, in the sense of allowing the cell to best represent information about its environment through the discrete activity of its genes.

DISCUSSION

Multiple studies in recent years have demonstrated that gene activity is often bursty rather than Poissonian [2,4,5,6,7,8], and can be described via a two-state model for mRNA production [8,9,10]. 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 non-bursty at low expression levels (at or below $\langle n \rangle \sim 1$ molecules/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 of the gene off-rate as the means to change the expression level, while maintaining the gene on-rate and transcription rate constant. 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, see FIGURE S11); rather, it was observed in the complete ensemble of promoters and stimuli examined. We note that a more complex scenario, where multiple kinetic parameters are simultaneously varied, i s a lso consistent with the observed smFISH data (see SUPPLEMENTARY NOTE and FIGURE S18). 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 (FIGURE 3D above).

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 displayed 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 [25] concluded that protein fluctuations were likely dominated by the mRNA species, as 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 a gene-independent behavior of the transcriptional burst size (FIGURE 3A above). The authors in [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 significantly 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 e.g. Figure S22 in [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 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 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 demonstrated how the properties of the transcriptional time-series itself, beyond merely 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 [66,67,68,69,70].

Two important limitations of our work need to be mentioned: First, when describing gene activity we centered on the mRNA species only, while leaving out the downstream production of proteins. In doing so we implicitly assumed that protein kinetics is 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), suggest 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 will have to reveal whether the fact that transcription burstiness appears in both kingdoms reflects a similarity in underlying mechanisms, or instead results form the selection of an advantageous phenotype in different systems.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

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 NIH (R01GM082837) and NSF (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).

References

1. Golding I, Cox EC. Eukaryotic transcription: what does it mean for a gene to be 'on'? Curr Biol. 2006; 16:R371–373. [PubMed: 16713947]
2. Golding I, Paulsson J, Zawilski SM, Cox EC. Real-time kinetics of gene activity in individual bacteria. Cell. 2005; 123:1025–1036. [PubMed: 16360033]
3. Raj A, van Oudenaarden A. Nature, nurture, or chance: stochastic gene expression and its consequences. Cell. 2008; 135:216–226. S0092-8674(08)01243-9 [pii]. 10.1016/j.cell.2008.09.050 [PubMed: 18957198]
4. Chubb JR, Liverpool TB. Bursts and pulses: insights from single cell studies into transcriptional mechanisms. Curr Opin Genet Dev. 2010; 20:478–484. S0959-437X(10)00116-4 [pii]. 10.1016/j.gde.2010.06.009 [PubMed: 20638837]
5. Chubb JR, Trcek T, Shenoy SM, Singer RH. Transcriptional pulsing of a developmental gene. Curr Biol. 2006; 16:1018–1025. S0960-9822(06)01426-6 [pii]. 10.1016/j.cub.2006.03.092 [PubMed: 16713960]
6. Pare A, et al. Visualization of individual Scr mRNAs during Drosophila embryogenesis yields evidence for transcriptional bursting. Curr Biol. 2009; 19:2037–2042. S0960-9822(09)01848-X [pii]. 10.1016/j.cub.2009.10.028 [PubMed: 19931455]
7. Yunger S, Rosenfeld L, Garini Y, Shav-Tal Y. Single-allele analysis of transcription kinetics in living mammalian cells. Nat Methods. 2010; 7:631–633. nmeth.1482 [pii]. 10.1038/nmeth.1482 [PubMed: 20639867]
8. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA synthesis in mammalian cells. PLoS Biol. 2006; 4:e309. 06-PLBI-RA-0625R2 [pii]. 10.1371/journal.pbio.0040309 [PubMed: 17048983]
9. Peccoud J, Ycart B. Markovian Modeling of Gene-Product Synthesis. Theoretical Population Biology. 1995; 48:222–234.
10. Shahrezaei V, Swain PS. Analytical distributions for stochastic gene expression. Proc Natl Acad Sci U S A. 2008; 105:17256–17261. 0803850105 [pii]. 10.1073/pnas.0803850105 [PubMed: 18988743]
11. Zenklusen D, Larson DR, Singer RH. Single-RNA counting reveals alternative modes of gene expression in yeast. Nat Struct Mol Biol. 2008; 15:1263–1271. nsmb.1514 [pii]. 10.1038/nsmb.1514 [PubMed: 19011635]
12. Tan RZ, van Oudenaarden A. Transcript counting in single cells reveals dynamics of rDNA transcription. Mol Syst Biol. 2010; 6:358. msb201014 [pii]. 10.1038/msb.2010.14 [PubMed: 20393578]
13. Mitarai N, Dodd IB, Crooks MT, Sneppen K. The generation of promoter-mediated transcriptional noise in bacteria. PLoS Comput Biol. 2008; 4:e1000109.10.1371/journal.pcbi.1000109 [PubMed: 18617999]
14. Dobrzynski M, Bruggeman FJ. Elongation dynamics shape bursty transcription and translation. Proc Natl Acad Sci U S A. 2009 0803507106 [pii]. 10.1073/pnas.0803507106
15. van Zon JS, Morelli MJ, Tanase-Nicola S, ten Wolde PR. Diffusion of transcription factors can drastically enhance the noise in gene expression. Biophys J. 2006; 91:4350–4367. biophysj.106.086157 [pii]. 10.1529/biophysj.106.086157 [PubMed: 17012327]
16. Fang FC. Sigma cascades in prokaryotic regulatory networks. Proc Natl Acad Sci U S A. 2005; 102:4933–4934. 0501417102 [pii]. 10.1073/pnas.0501417102 [PubMed: 15795367]
17. Blake WJ, MKA, Cantor CR, Collins JJ. Noise in eukaryotic gene expression. Nature. 2003; 422:633–637. nature01546 [pii]. 10.1038/nature01546 [PubMed: 12687005]

18. Bernstein JA, Khodursky AB, Lin PH, Lin-Chao S, Cohen SN. Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. Proc Natl Acad Sci U S A. 2002; 99:9697–9702. [pii]. 10.1073/pnas.112318199112318199 [PubMed: 12119387]

19. Selinger DW, Saxena RM, Cheung KJ, Church GM, Rosenow C. Global RNA half-life analysis in Escherichia coli reveals positional patterns of transcript degradation. Genome Res. 2003; 13:216–223.10.1101/gr.912603 [PubMed: 12566399]

20. Singh D, et al. Regulation of ribonuclease E activity by the L4 ribosomal protein of Escherichia coli. Proc Natl Acad Sci U S A. 2009; 106:864–869. 0810205106 [pii]. 10.1073/pnas.0810205106 [PubMed: 19144914]

21. Thattai M, van Oudenaarden A. Intrinsic noise in gene regulatory networks. Proc Natl Acad Sci U S A. 2001; 98:8614–8619. [PubMed: 11438714]

22. Goh KI, Barabasi A. Burstiness and memory in complex systems. Epl. 2008; 81 Artn 48002. 10.1209/0295-5075/81/48002

23. Kepler TB, Elston TC. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. Biophys J. 2001; 81:3116–3136. S0006-3495(01)75949-8 [pii]. 10.1016/S0006-3495(01)75949-8 [PubMed: 11720979]

24. Simpson ML, Cox CD, Sayler GS. Frequency domain chemical Langevin analysis of stochasticity in gene transcriptional regulation. J Theor Biol. 2004; 229:383–394. S0022519304001481 [pii]. 10.1016/j.jtbi.2004.04.017 [PubMed: 15234205]

25. Bar-Even A, et al. Noise in protein expression scales with natural protein abundance. Nat Genet. 2006; 38:636–643. ng1807 [pii]. 10.1038/ng1807 [PubMed: 16715097]

26. Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. Imaging individual mRNA molecules using multiple singly labeled probes. Nat Methods. 2008; 5:877–879. nmeth.1253 [pii]. 10.1038/nmeth.1253 [PubMed: 18806792]

27. Kuhlman T, Zhang Z, Saier MH Jr, Hwa T. Combinatorial transcriptional control of the lactose operon of Escherichia coli. Proc Natl Acad Sci U S A. 2007; 104:6043–6048. 0606717104 [pii]. 10.1073/pnas.0606717104 [PubMed: 17376875]

28. Paulsson J, Ehrenberg M. Random signal fluctuations can reduce random fluctuations in regulated components of chemical regulatory networks. Phys Rev Lett. 2000; 84:5447–5450. [PubMed: 10990965]

29. Tokeson JP, Garges S, Adhya S. Further inducibility of a constitutive system: ultrainduction of the gal operon. J Bacteriol. 1991; 173:2319–2327. [PubMed: 2007555]

30. Weickert MJ, Adhya S. The galactose regulon of Escherichia coli. Mol Microbiol. 1993; 10:245–251. [PubMed: 7934815]

31. Alekshun MN, Levy SB. The mar regulon: multiple resistance to antibiotics and other toxic chemicals. Trends Microbiol. 1999; 7:410–413. S0966-842X(99)01589-9 [pii]. [PubMed: 10498949]

32. Hernandez VJ, Bremer H. Guanosine tetraphosphate (ppGpp) dependence of the growth rate control of rrnB P1 promoter activity in Escherichia coli. J Biol Chem. 1990; 265:11605–11614. [PubMed: 2114400]

33. Potrykus K, et al. Antagonistic regulation of Escherichia coli ribosomal RNA rrnB P1 promoter activity by GreA and DksA. J Biol Chem. 2006; 281:15238–15248. M601531200 [pii]. 10.1074/jbc.M601531200 [PubMed: 16597620]

34. Abdel-Hamid AM, Cronan JE. 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. 2007; 189:369–376. JB.01373-06 [pii]. 10.1128/JB.01373-06 [PubMed: 17056747]

35. Barker DF, Campbell AM. Use of bio-lac fusion strains to study regulation of biotin biosynthesis in Escherichia coli. J Bacteriol. 1980; 143:789–800. [PubMed: 6782078]

36. Michalowski CB, Little JW. Positive autoregulation of cI is a dispensable feature of the phage lambda gene regulatory circuitry. J Bacteriol. 2005; 187:6430–6442. 187/18/6430 [pii]. 10.1128/JB.187.18.6430-6442.2005 [PubMed: 16159777]

37. Sauer RT, Jordan SR, Pabo CO. Lambda repressor: a model system for understanding protein-DNA interactions and protein stability. Adv Protein Chem. 1990; 40:1–61. [PubMed: 2195849]

38. Lim WA, Sauer RT. Alternative packing arrangements in the hydrophobic core of lambda repressor. Nature. 1989; 339:31–36.10.1038/339031a0 [PubMed: 2524006]
39. Shen-Orr SS, Milo R, Mangan S, Alon U. Network motifs in the transcriptional regulation network of Escherichia coli. Nat Genet. 2002; 31:64–68. ng881 [pii]. 10.1038/ng881 [PubMed: 11967538]
40. Garcia HG, Sanchez A, Kuhlman T, Kondev J, Phillips R. Transcription by the numbers redux: experiments and calculations that surprise. Trends Cell Biol. 2010 S0962–8924(10)00144-3 [pii]. 10.1016/j.tcb.2010.07.002
41. Kittisopikul M, Suel GM. Biological role of noise encoded in a genetic network motif. Proc Natl Acad Sci U S A. 2010; 107:13300–13305. 1003975107 [pii]. 10.1073/pnas.1003975107 [PubMed: 20616054]
42. Kaern M, Elston TC, Blake WJ, Collins JJ. Stochasticity in gene expression: from theories to phenotypes. Nat Rev Genet. 2005; 6:451–464. nrg1615 [pii]. 10.1038/nrg1615 [PubMed: 15883588]
43. Pedraza JM, van Oudenaarden A. Noise propagation in gene networks. Science. 2005; 307:1965–1969. 307/5717/1965 [pii]. 10.1126/science.1109090 [PubMed: 15790857]
44. Kittisopikul M, Suel GM. Biological role of noise encoded in a genetic network motif. Proc Natl Acad Sci U S A. 107:13300–13305. 1003975107 [pii]. 10.1073/pnas.1003975107 [PubMed: 20616054]
45. Cox CD, McCollum JM, Allen MS, Dar RD, Simpson ML. Using noise to probe and characterize gene circuits. Proc Natl Acad Sci U S A. 2008; 105:10809–10814. 0804829105 [pii]. 10.1073/pnas.0804829105 [PubMed: 18669661]
46. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. Science. 2002; 297:1183–1186. [pii]. 10.1126/science.1070919297/5584/1183 [PubMed: 12183631]
47. Cai L, Friedman N, Xie XS. Stochastic protein expression in individual cells at the single molecule level. Nature. 2006; 440:358–362. nature04599 [pii]. 10.1038/nature04599 [PubMed: 16541077]
48. Yu J, Xiao J, Ren X, Lao K, Xie XS. Probing gene expression in live cells, one protein molecule at a time. Science. 2006; 311:1600–1603. 311/5767/1600 [pii]. 10.1126/science.1119623 [PubMed: 16543458]
49. Taniguchi Y, et al. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science. 2010; 329:533–538. 329/5991/533 [pii]. 10.1126/science.1188308 [PubMed: 20671182]
50. Kennell D, Riezman H. Transcription and translation initiation frequencies of the Escherichia coli *lac* operon. J Mol Biol. 1977; 114:1–21. 0022-2836(77)90279-0 [pii]. [PubMed: 409848]
51. Liang S, et al. Activities of constitutive promoters in Escherichia coli. J Mol Biol. 1999; 292:19–37. S0022283699930565 [pii]. 10.1006/jmbi.1999.3056 [PubMed: 10493854]
52. Neidhardt, FC. Escherichia coli and Salmonella typhimurium: cellular and molecular biology. American Society for Microbiology; 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. 1997; 25:1203–1210. gka167 [pii]. [PubMed: 9092630]
54. Fusco D, et al. Single mRNA molecules demonstrate probabilistic movement in living mammalian cells. Curr Biol. 2003; 13:161–167. S0960982202014367 [pii]. [PubMed: 12546792]
55. Bertrand E, et al. Localization of ASH1 mRNA particles in living yeast. Mol Cell. 1998; 2:437–445. S1097-2765(00)80143-4 [pii]. [PubMed: 9809065]
56. Golding I, Cox EC. RNA dynamics in live Escherichia coli cells. Proc Natl Acad Sci U S A. 2004; 101:11310–11315. [pii]. 10.1073/pnas.0404431010404443101 [PubMed: 15277674]
57. Alon, U. An introduction to systems biology: design principles of biological circuits. Chapman & Hall/CRC; 2007.
58. Tkacik G, Callan CG Jr, Bialek W. Information flow and optimization in transcriptional regulation. Proc Natl Acad Sci U S A. 2008; 105:12265–12270. 0806077105 [pii]. 10.1073/pnas.0806077105 [PubMed: 18719112]
59. Tkacik G, Walczak AM, Bialek W. Optimizing information flow in small genetic networks. Phys Rev E Stat Nonlin Soft Matter Phys. 2009; 80:031920. [PubMed: 19905159]

60. Shannon CE. A mathematical theory of communication. Bell Syst Tech J. 1948; 27:379–423. 623–656.
61. Newman JR, et al. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature. 2006; 441:840–846. nature04785 [pii]. 10.1038/nature04785 [PubMed: 16699522]
62. Geva-Zatorsky N, et al. Protein dynamics in drug combinations: a linear superposition of individual-drug responses. Cell. 2010; 140:643–651. S0092-8674(10)00126-1 [pii]. 10.1016/j.cell.2010.02.011 [PubMed: 20211134]
63. Mao C, et al. Quantitative analysis of the transcription control mechanism. Mol Syst Biol. 2010; 6:431. msb201083 [pii]. 10.1038/msb.2010.83 [PubMed: 21081924]
64. Dekel E, Alon U. Optimality and evolutionary tuning of the expression level of a protein. Nature. 2005; 436:588–592. nature03842 [pii]. 10.1038/nature03842 [PubMed: 16049495]
65. Eldar A, Elowitz MB. Functional roles for noise in genetic circuits. Nature. 2010; 467:167–173. nature09326 [pii]. 10.1038/nature09326 [PubMed: 20829787]
66. Singh A, Razooky B, Cox CD, Simpson ML, Weinberger LS. Transcriptional bursting from the HIV-1 promoter is a significant source of stochastic noise in HIV-1 gene expression. Biophys J. 2010; 98:L32–34. S0006-3495(10)00304-8 [pii]. 10.1016/j.bpj.2010.03.001 [PubMed: 20409455]
67. Kaufmann BB, Yang Q, Mettetal JT, van Oudenaarden A. Heritable stochastic switching revealed by single-cell genealogy. PLoS Biol. 2007; 5:e239. 07-PLBI-RA-0874 [pii]. 10.1371/journal.pbio.0050239 [PubMed: 17803359]
68. Choi PJ, Cai L, Frieda K, Xie XS. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. Science. 2008; 322:442–446. 322/5900/442 [pii]. 10.1126/science.1161427 [PubMed: 18927393]
69. Zeng L, et al. Decision making at a subcellular level determines the outcome of bacteriophage infection. Cell. 2010; 141:682–691. S0092-8674(10)00352-1 [pii]. 10.1016/j.cell.2010.03.034 [PubMed: 20478257]
70. Zong C, So L-h, Sepulveda LA, Skinner SO, Golding I. Lysogen stability is determined by the frequency of activity bursts from the fate-determining gene. Mol Syst Biol. 2010; 6
71. Austin DW, et al. Gene network shaping of inherent noise spectra. Nature. 2006; 439:608–611. nature04194 [pii]. 10.1038/nature04194 [PubMed: 16452980]
72. Mooney RA, et al. Regulator trafficking on bacterial transcription units in vivo. Mol Cell. 2009; 33:97–108. S1097-2765(08)00891-5 [pii]. 10.1016/j.molcel.2008.12.021 [PubMed: 19150431]
73. Reppas NB, Wade JT, Church GM, Struhl K. The transition between transcriptional initiation and elongation in E. coli is highly variable and often rate limiting. Mol Cell. 2006; 24:747–757. S1097-2765(06)00738-6 [pii]. 10.1016/j.molcel.2006.10.030 [PubMed: 17157257]
74. Whitelaw NC, Chong S, Whitelaw E. Tuning in to noise: epigenetics and intangible variation. Dev Cell. 2010; 19:649–650. S1534-5807(10)00478-8 [pii]. 10.1016/j.devcel.2010.11.001 [PubMed: 21074715]
75. Gillespie DT. Exact Stochastic Simulation of Coupled Chemical-Reactions. Journal of Physical Chemistry. 1977; 81:2340–2361.

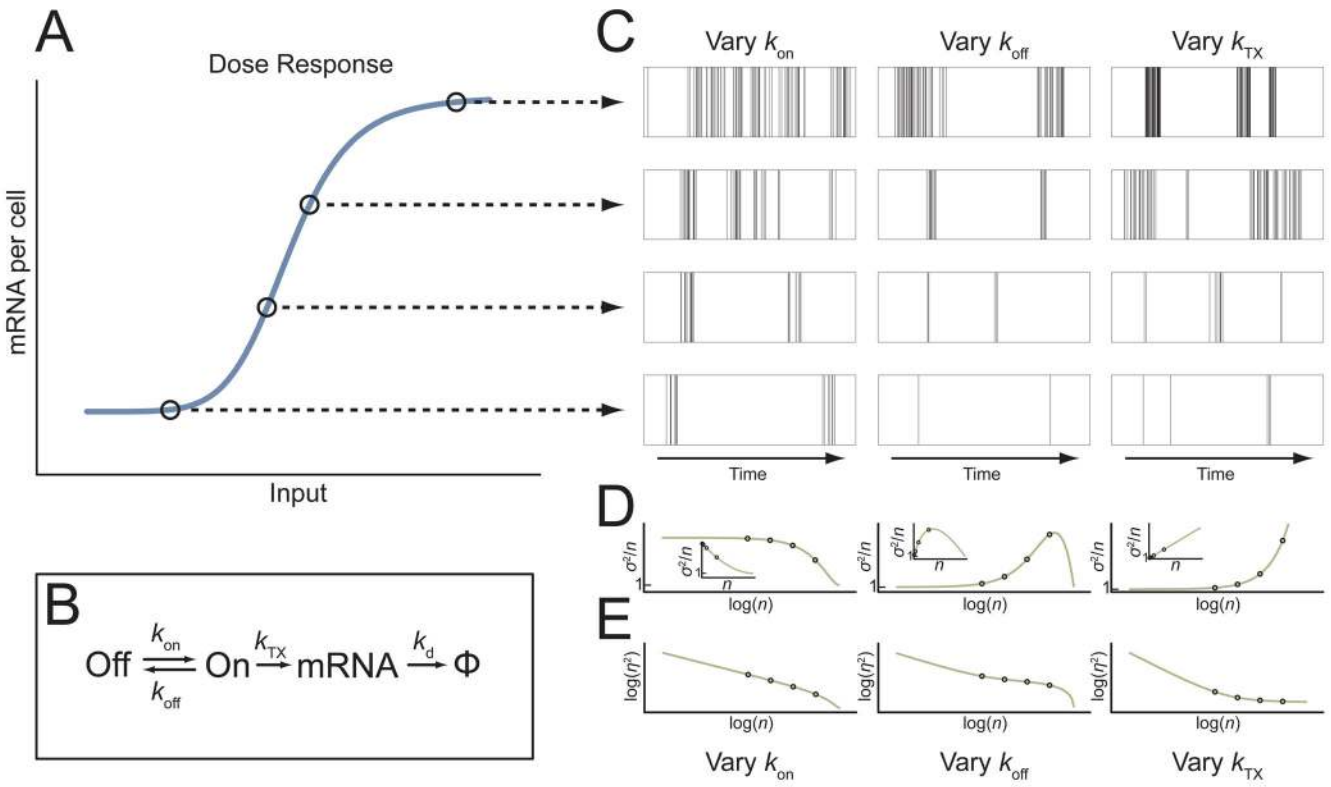

Figure 1. Different features of the transcriptional time-series can be modulated to vary gene expression level

(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 exhibited by many bacterial promoters [27,57], for example see FIGURE 2A below.

(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 panel (A). Each plot shows the time-series of mRNA production events (bars). Data was created by simulating the two-state model using the Gillespie method [75]. In each of the three cases shown, only a single parameter of gene activity was varied (k_{on}, left; k_{off}, middle; k_{TX}, right). All time-series in the same row produce 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, while the insets show the same data on a linear scale. $b(\langle n \rangle)$ was calculated 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 $\langle n \rangle$. $\eta^2(\langle n \rangle)$ was calculated analytically for the two-state model[9].

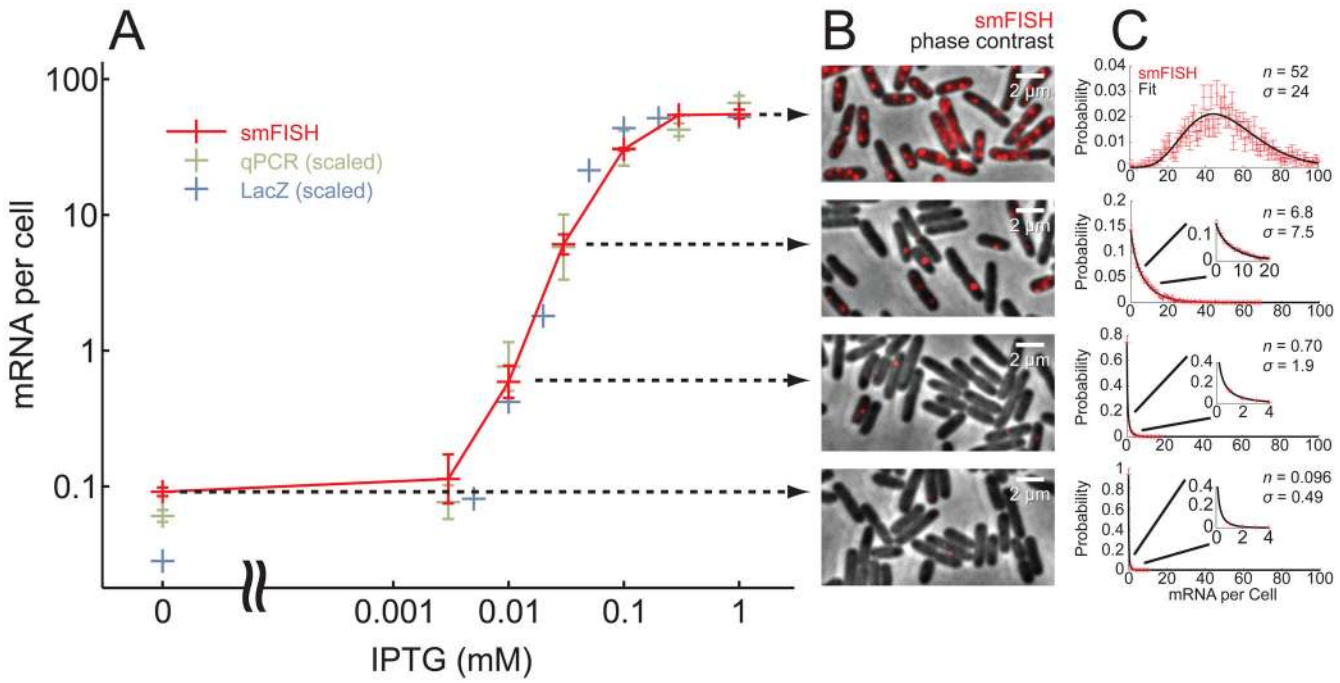

Figure 2. Single-molecule FISH used to characterize mRNA copy-number statistics

(A) Gene expression level (mRNA/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 single-molecule FISH (smFISH, average of 2 independent experiments) is shown, as well as the results of quantitative PCR (qPCR, average of 2 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 denote standard errors from duplicate experiments. The good agreement between the three assays, over ~3 orders of expression level, demonstrates 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.

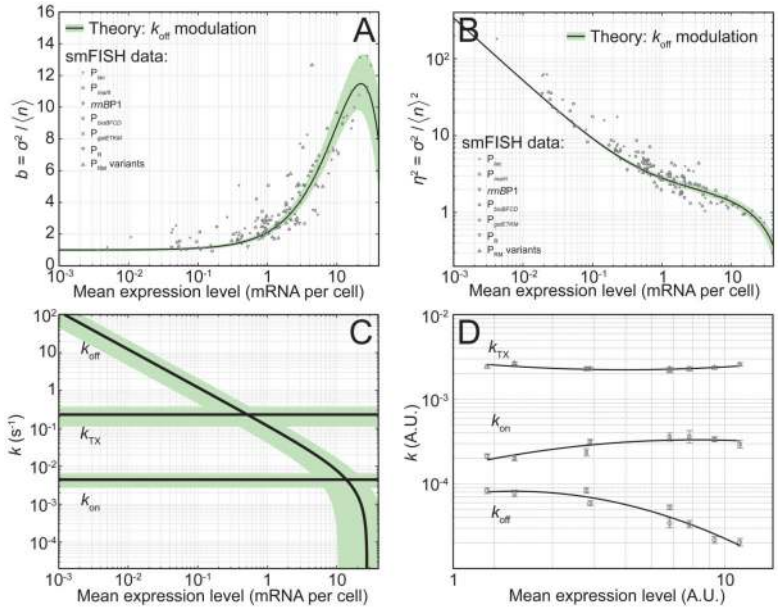

Figure 3. Gene expression level in *E. coli* is varied by changing the gene off-rate

(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}. The theoretical curve is obtained by solving analytically the expression for $b(\langle n \rangle)$ and then using k_{on} and k_{TX} as fitting parameters. 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 as in panel (A). The theoretical parameters (k_{on}, k_{TX}) extracted from fitting $b(\langle n \rangle)$ in panel (A) were used to plot the theoretical curve.

(C) The estimated rate parameters for gene activity in *E. coli*. These were obtained from fitting $b(\langle n \rangle)$ in panel (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 (FIGURE S11). The error in k_{off} (green shade) is calculated from the resulting fit.

(D) Direct measurement of the two-state rate parameters in individual living cells. mRNA production from the promoter $P_{lac/ara}$ was quantified using the MS2-GFP method [2]. Data (markers) is from 9 independent experiments (>400 cells). Error bars represent standard errors within each experiment. Solid lines are fits to second degree polynomials.

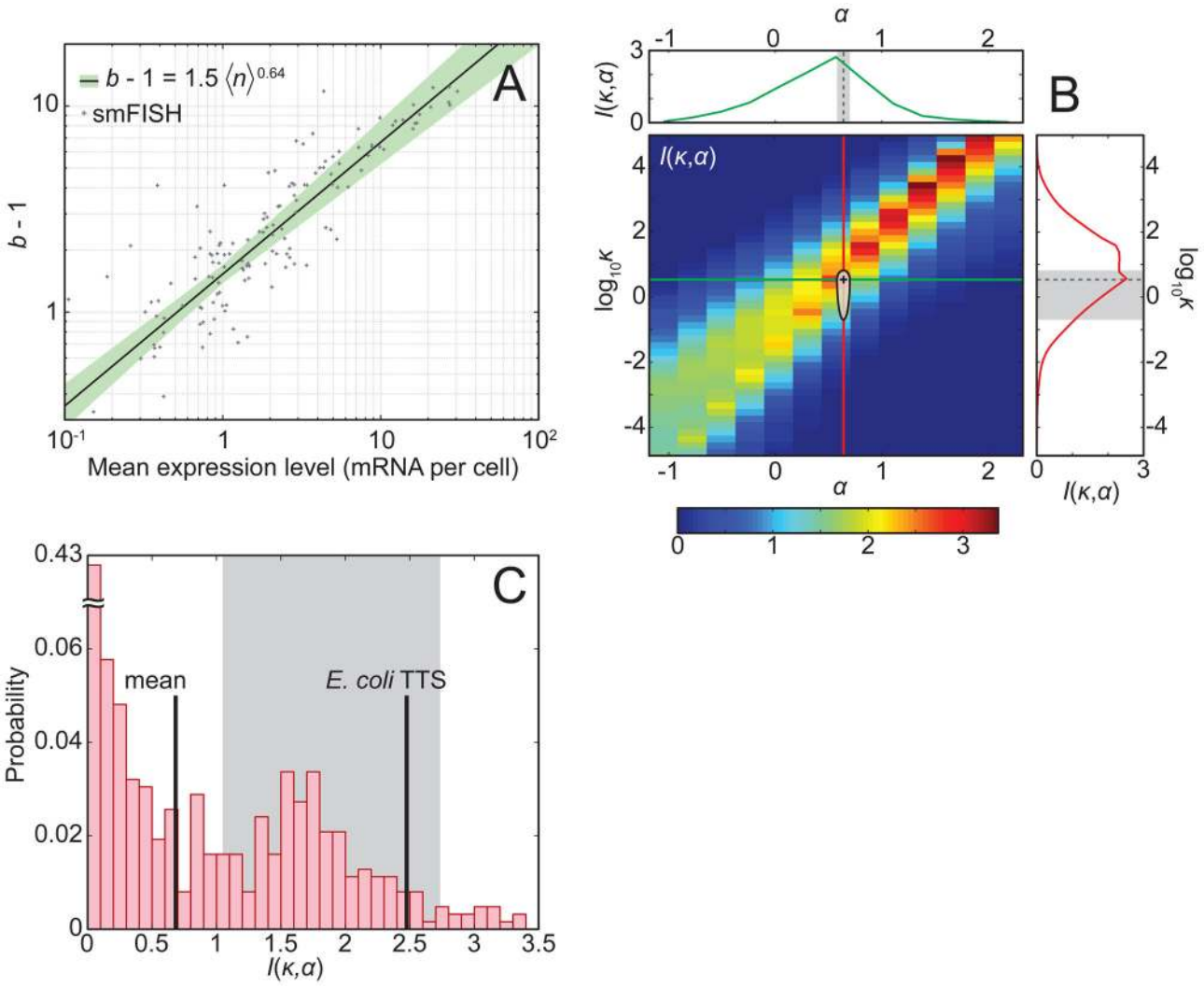

Figure 4. The transcriptional time-series optimizes information representation by the cell

(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 (same data set as in FIGURE 3A above). Solid line, fit to a power law $\sigma^2/\langle n\rangle - 1 = \langle n\rangle^{\alpha}/\kappa$. The power law yields a good fit ($R^2 = 0.76$) in the range $\langle n\rangle \approx$ 0.3–40, and allows an estimation of the parameters κ and α.

(B) The calculated mutual information I between 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 $b(\langle n\rangle)$ is 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 the multiple sources: κ and α estimation from the fit in panel (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 panel (B). The *E. coli* transcriptional time-series exhibits a mutual information value ($I \approx 2.5$) that is much higher than the average performance by all

possible modulation schemes ($I = 0.68$). The shaded area corresponds to the experimental error estimate for κ and α, as in Panel (B).