Regulation of Noise in Gene Expression

Alvaro Sanchez,[1] Sandeep Choubey,[2] and Jane Kondev[2]

[1]Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; email: alvaros@mit.edu

[2]Department of Physics, Brandeis University, Waltham, Massachusetts 02453; email: sandeep@brandeis.edu; kondev@brandeis.edu

Annu. Rev. Biophys. 2013. 42:469–91

First published online as a Review in Advance on March 21, 2013

The *Annual Review of Biophysics* is online at biophys.annualreviews.org

This article's doi: 10.1146/annurev-biophys-083012-130401

Keywords

phenotypic heterogeneity, transcriptional regulation, gene expression noise, mRNA fluctuations, promoter architecture, noise-mean scaling

Abstract

The biochemical processes leading to the synthesis of new proteins are random, as they typically involve a small number of diffusing molecules. They lead to fluctuations in the number of proteins in a single cell as a function of time and to cell-to-cell variability of protein abundances. These in turn can lead to phenotypic heterogeneity in a population of genetically identical cells. Phenotypic heterogeneity may have important consequences for the development of multicellular organisms and the fitness of bacterial colonies, raising the question of how it is regulated. Here we review the experimental evidence that transcriptional regulation affects noise in gene expression, and discuss how the noise strength is encoded in the architecture of the promoter region. We discuss how models based on specific molecular mechanisms of gene regulation can make experimentally testable predictions for how changes to the promoter architecture are reflected in gene expression noise.

Contents

FLUCTUATIONS AND THE CENTRAL DOGMA OF MOLECULAR BIOLOGY

The central dogma of molecular biology describes information flow from DNA to protein via messenger RNA (mRNA). DNA stores information about the proteins that the cell needs to produce to achieve its functions. At the same time, DNA functions as a molecular computer that integrates a variety of chemical inputs that describe the conditions within the cell and in its environment in order to produce a particular chemical output in the form of mRNA. These regulatory computations are inherently random, as they are achieved by a small number of molecules undergoing Brownian motion. This means that for a fixed set of molecular inputs the output of the transcriptional and translational machinery, namely the number of proteins produced from a particular gene, is random. This can be observed in a clonal population of cells exposed to the same environmental conditions in which the amount of gene expressed varies from cell to cell. Similarly, in single-cell experiments that measure the amount of gene expression with time, even in steady state the number of mRNAs and proteins fluctuates from one time instance to the next. Because the activity of proteins and mRNAs is dictated by their copy number within the cell, these fluctuations can potentially affect the functioning of genetic circuits, sometimes leading to phenotypic variability in a population of genetically identical cells. A number of reviews have focused on different aspects of this phenomenon (3, 23, 27). Here we review the experimental evidence for the important role of transcriptional regulation in determining the size of these fluctuations, and the theoretical ideas that have been put forth to explain the experimental findings. Together, these studies are beginning to provide a quantitative understanding of how the promoter architecture affects the amount of gene expression noise. An exciting possibility suggested by these ideas is that the amount of gene expression noise is under the direct control of regulatory DNA and therefore subject to evolutionary forces.

Gene expression noise: fluctuations in the copy number of mRNA or protein molecules corresponding to a particular gene

Promoter architecture: list of all nucleosome binding sites and *cis*-regulatory elements that control gene expression

PHYSIOLOGICAL EFFECTS OF GENE EXPRESSION NOISE

An emerging view is that gene expression noise is not simply a necessary evil of the molecular computations of the central dogma, but is at times utilized by cells to achieve certain functions. In other words, evidence is mounting for specific physiological roles for gene expression noise, and a number of recent reviews have addressed this fascinating topic.

In gene expression, stochasticity is the driving force that generates phenotypic heterogeneity in microbial populations. In fact, population heterogeneity in the expression of even one protein can have important consequences for a population of cells. For instance, Blake et al. (7) found that cell-to-cell variability in the expression of a single antibiotic-resistant protein in *Saccharomyces cerevisiae* can lead to pronounced effects in the response of the whole population to an antibiotic challenge; at intermediate antibiotic concentrations, the strain with the largest phenotypic heterogeneity was best able to survive. Interestingly, the level of heterogeneity could be tuned genetically by modifying the sequence of regulatory DNA, such as the TATA box of the promoter.

In some instances, the effect of noise can be amplified by the presence of multi-stability in genetic networks. This can lead to multiple phenotypes coexisting in a cell population. Individual cells can make transitions between those phenotypes driven by fluctuations in the expression of certain key genes in the network (10). An example is provided by the behavior of individual *Bacillus subtilis* cells when they are subject to stress; the cells must decide whether to enter a competent state or to start forming spores (50, 51). The two phenotypes are mutually exclusive, and the decision between them is determined by transient values of the concentration of key proteins in the network, which can stochastically exceed a threshold value and thereby force the cells to assume the competence phenotype (28).

Not all gene expression noise, however, is beneficial. Some processes, such as the development of a multicellular organism, rely on precise spatial and temporal transmission of genetic information, and in these cases noise in gene expression needs to be kept to a minimum (1). Recent experiments have found that noise suppression mechanisms exist at the level of gene networks that control development (39). These studies, as well as many others, demonstrate that noise in gene expression is an important biological trait, and as such, it is important to understand how noise is affected by the DNA sequence of the gene and the sequence of its regulatory region.

Phenotypic heterogeneity: cell-to-cell variability in the adoption of a particular phenotype or expression of a particular protein

TATA box: a *cis*-regulatory element recognized by the TATA binding protein, which helps assemble the preinitiation complex on eukaryotic promoters

Nucleosome binding site: DNA sequence around the promoter that is associated with a high nucleosome occupancy

EFFECT OF PROMOTER ARCHITECTURE ON GENE EXPRESSION NOISE IN EUKARYOTES

The specific sequence of the regulatory DNA that controls the amount of expression from a particular gene has a large effect on gene expression noise in eukaryotic cells. This has been demonstrated in studies using natural and synthetic promoters, which differed from each other by the presence or absence of proximal nucleosome binding sites, the presence of a TATA box with variable strength, and the number of transcription factor binding sites, their strength, and their location on the promoter.

Some of these promoter architectural effects on noise were found by analyzing high-throughput data on gene expression noise in wild-type yeast and mammalian cells (4, 26, 33). Other studies were performed with synthetic or mutated promoters, and they were based on comparing noise levels from identical genes under the control of promoters that differed by only one particular architectural feature, for example, the number of enhancers (22, 32, 38, 52, 57). Experiments with synthetic promoters allow for a direct and systematic comparison between different architectural features of the promoter and their contribution to gene expression noise. Below we discuss the main lines of

evidence in support of the idea that transcriptional regulation is a major source of variability in gene expression, and that this variability can be controlled by tuning the regulatory DNA sequence.

ON-OFF promoter: type of promoter that switches stochastically between one transcriptionally active (ON) and one inactive (OFF) state

Transcriptional bursting: type of transcriptional dynamics in which transcriptional events are rare but clustered

Nucleosome Occupancy

Positioning of nucleosomes in eukaryotic chromosomes affects practically all genomic functions, including transcription. Here we review experimental evidence for the role of nucleosome occupancy in determining gene expression noise. We also describe a simple mechanistic model based on occlusion of promoter DNA by nucleosomes, which can account for the experimental observations.

Experimental evidence. The first investigations of gene expression noise in eukaryotic cells revealed that gene silencing by nucleosome occlusion of the promoter and its activation by chromatin remodeling are key factors affecting noise (40). DNA wrapping around nucleosomes is a stochastic process governed by diffusion (35), as is the binding of chromatin remodeling factors to the promoter DNA. It is believed that promoter DNA wrapped around nucleosomes is very stable and has a typical lifetime that is longer than the timescale of transcription. This hypothesis explained the results of a systematic study of noise generated by the PHO5 promoter in yeast cells, which is regulated by chromatin remodeling (40). It was found that removal either of particular components of the chromatin remodeling complex, or of the binding sites for the transcription factors that recruit the chromatin remodeling complex to the promoter, results in enhanced noise by reducing the activation rate of the promoter (40). Thus, promoter-DNA condensation into chromatin may lead to long-lived, silenced, or OFF, promoter states, which are followed by rapid, short-lived initiation events, leading to transcriptional bursting. Transcriptional bursting leads to transcriptional noise and cell-to-cell variability in gene expression.

Due to the importance of chromatin silencing and remodeling as a source of transcriptional noise, those genes that are in open chromatin regions, or whose promoters are depleted of nucleosomes, are expected to display lower levels of cell-to-cell variability than those genes whose promoters are covered by nucleosomes. Indeed, computational analyses of genomic nucleosome occupation patterns revealed the presence of two broad classes of promoters: one class containing nucleosome binding sites close to the transcription start site, and a second class containing a nucleosome-free region in the vicinity of the transcription start site (54, 55). Interestingly, the first class of promoters was associated with both large cell-to-cell variability in gene expression at the single-cell level and large promoter plasticity (i.e., large variability in mean gene expression levels across environmental conditions). In contrast, the second class of promoters was characterized by low variability and low plasticity (54, 55). This correlation between plasticity and noise was related to the architecture of the promoter. Still, a correlation between these two quantities does not mean they cannot be differentially regulated; additional bioinformatic analyses of promoter sequences have suggested that evolution might have tuned promoter architecture so that a promoter may exhibit a large gene expression dynamic range and low noise simultaneously (2, 26).

Biological importance. It has been hypothesized that essential genes, whose over- or underexpression may lead to cell death, would tend to exhibit low noise levels. Consistent with this hypothesis, computational analyses have revealed that these genes tend to be clustered in nucleosome-depleted open-chromatin regions on the chromosome (5), where noise is low (15).

Mechanistic explanation. The lack of nucleosomes in the proximity of the transcription start site is expected to reduce the probability of gene inactivation by the condensation of promoter DNA

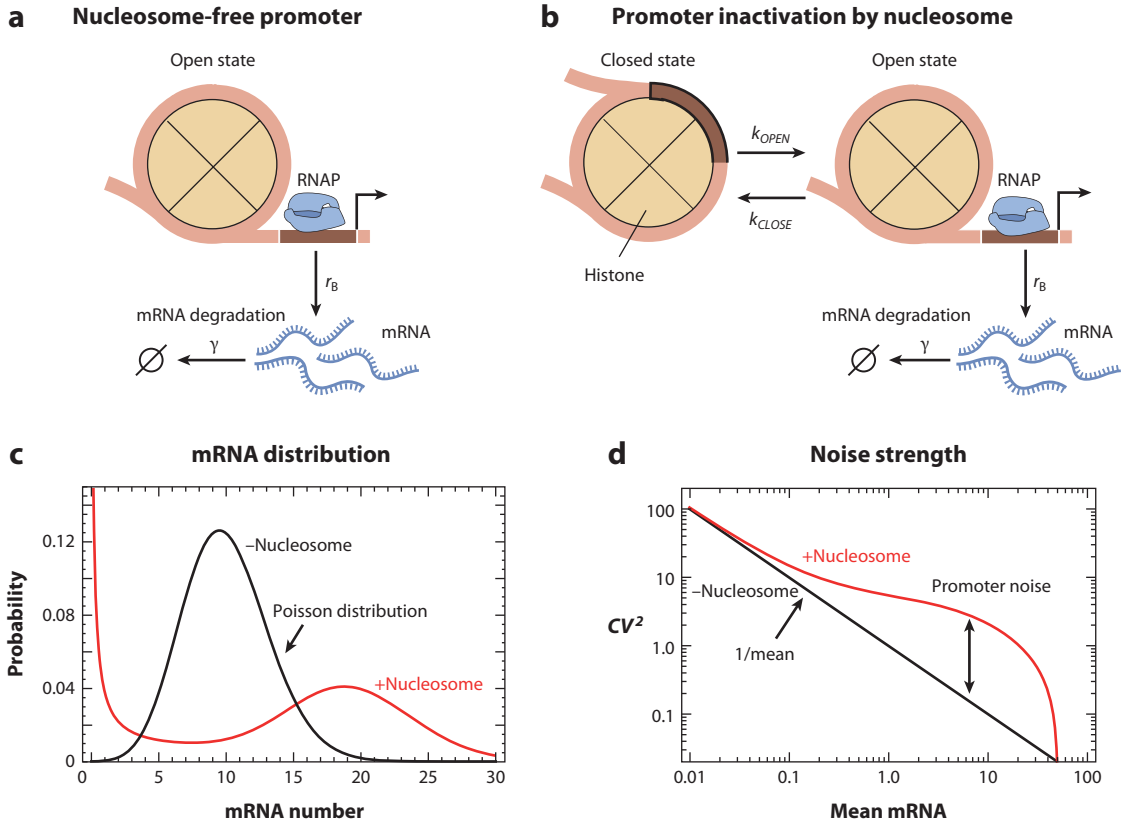


Figure 1

Transcriptional dynamics of nucleosome-mediated repression. (*a*) A nucleosome-free unregulated promoter is characterized by a constant probability per unit time of mRNA synthesis r_B. mRNA is degraded with a constant probability γ per unit time per molecule. (*b*) A promoter inactivated by a proximal nucleosome. The promoter switches between two states, an open state, in which transcription occurs with a constant probability per unit time r_B, and a closed state, in which transcription does not initiate owing to the nucleosome-mediated promoter inactivation (here represented by steric occlusion). The promoter switches from the open state to the closed state with probability per unit time k_{CLOSE}, and from the closed state to the open state with probability per unit time k_{OPEN}. mRNA is degraded with a constant probability per unit time per molecule (γ). (*c*) The steady-state mRNA distribution (*solid red line*: $r_B = 20$, $k_{CLOSE} = 0.01$, $k_{OPEN} = 0.01$; all the rates are in units of γ) predicted by the nucleosome model differs significantly from the nucleosome-free, unregulated promoter, which is described by a Poisson distribution with the same mean (*black line*: $r_B = 10$). (*d*) The predicted CV^2 of the mRNA distribution as a function of the mean for the unregulated (*black line*) and the two-state nucleosome model (*solid red line*: $r_B = 50$, $k_{OPEN} = 0.01$, k_{CLOSE} is variable). The promoter noise is the contribution to the noise (as measured by CV^2) by the nucleosome-mediated transcription regulation process.

into chromatin. We demonstrate in **Figure 1*a,b*** how the presence or absence of a nucleosome affects gene expression noise. We consider a gene that is inactivated by nucleosomes, together with another gene whose promoter is nucleosome free. Using the procedure described in the sidebar, Method for Computing the Gene Expression Mean and Noise, we can compute the gene expression noise predicted by both models. As shown in **Figure 1*c,d***, the promoter whose activity is switched off by the presence of a nucleosome displays a non-Poisson mRNA distribution, characterized by noise strength greater than that of the promoter that is always active. In contrast, the nucleosome-depleted gene shows a Poisson mRNA distribution, characterized by noise strength, defined by the coefficient of variation squared (CV^2 = variance/mean^2), equal to 1/mean.

METHOD FOR COMPUTING THE GENE EXPRESSION MEAN AND NOISE

To describe the temporal fluctuations of the number of mRNAs or proteins in a cell, we adopt a mathematical description in terms of a chemical master equation. This equation describes the time variation of the probability distribution that keeps track of the number of mRNAs or proteins in the cell and the state of the promoter that regulates the expression of the corresponding gene. The goal is to use this equation to compute the mean and the variance of the steady-state distribution of the number of mRNAs or proteins. For this we have developed a general algorithm (49), which we illustrate here by considering the example of a three-state promoter and the mRNA distribution it generates.

1. Make a list of all possible promoter states and of the rates that specify the stochastic transitions between the states. Then write down the chemical master equation for the time evolution of the joint probability of having the promoter in one of the possible states and there being m mRNAs in the cell. The master equations for the three-state promoter are

$$\frac{d}{dt}p(1,m) = k_{21}p(2,m) - k_{12}p(1,m) + \gamma(m+1)p(1,m+1) - \gamma m p(1,m),$$
$$\frac{d}{dt}p(2,m) = k_{32}p(3,m) + k_{12}p(1,m) - k_{21}p(2,m) - k_{23}p(2,m) + \gamma(m+1)p(2,m+1) - \gamma m p(2,m),$$
$$\frac{d}{dt}p(1,m) = k_{23}p(2,m) - k_{32}p(3,m) + \gamma(m+1)p(3,m+1) - \gamma m p(3,m) + r p(3,m-1) - r p(3,m).$$

   Here k_{ij} is the rate of transition from state i to state j, r is the rate of mRNA production (which only occurs in state 3), and γ is the rate of degradation (per molecule). These rates are the parameters of the model, some of which can be experimentally tuned by changing promoter sequence, transcription factor concentration, or other aspects of the promoter architecture.

2. Write down the above set of master equations in matrix form by defining the following matrices:

$$\vec{p}(m) = \begin{bmatrix} p(1,m) \\ p(2,m) \\ p(3,m) \end{bmatrix}; \quad \hat{K} = \begin{bmatrix} -k_{12} & k_{21} & 0 \\ k_{12} & -(k_{21}+k_{23}) & k_{32} \\ 0 & k_{23} & -k_{32} \end{bmatrix}; \quad \hat{R} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & r \end{bmatrix}; \quad \hat{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

   By setting the left-hand side of the equations in Step 1 to zero, and by using this matrix notation, we arrive at the equation for the steady-state probability distribution,

$$\left[\hat{K} - \hat{R} - m\gamma\hat{I}\right]\vec{p}(m) + \hat{R}\vec{p}(m-1) + (m+1)\gamma\hat{I}\vec{p}(m+1) = 0.$$

3. To compute the first two moments of the steady-state mRNA probability distribution, multiply both sides of the equation in Step 2 by m and m^2, respectively, and then sum over all values of m from 0 to ∞. The result of this calculation can be expressed in terms of the following partial moment vectors:

$$\vec{m}_{(0)} = \sum_{m=0}^{\infty} m^0 \vec{p}(m); \quad \vec{m}_{(1)} = \sum_{m=0}^{\infty} m\vec{p}(m).$$

Namely, from the final equation in Step 2 we obtain the equations

$$\hat{K}\vec{m}_{(0)} = 0,$$
$$\left[\hat{K} - \gamma\hat{I}\right]\vec{m}_{(1)} - \hat{R}\vec{m}_{(0)} = 0,$$

for the first two partial moments. This is a set of linear equations from which the first and the second moment of the mRNA distribution can be computed:

$$\langle m\rangle = \frac{\vec{r}\vec{m}_{(0)}}{\gamma},$$
$$\langle m^2\rangle = \langle m\rangle + \frac{\vec{r}\vec{m}_{(1)}}{\gamma}.$$

Here the vector $\vec{r}$ contains the ordered list of rates of transcription initiation for each promoter state.

TATA Box Strength

The TATA box is a DNA sequence recognized by TATA-box binding proteins that recruit the preinitiation complex, which then helps initiate multiple rounds of transcription before it stochastically disassembles. The strength of the TATA box is thought to determine the lifetime of the preinitiation complex and thus the number of rounds of transcription that result from each activation event (7). Here we review experiments that have demonstrated a role for the TATA box in determining gene expression noise and we describe a simple model that accounts for the experimental observations.

Experimental evidence. The previously mentioned study by Raser & O'Shea demonstrated that mutations that weaken the strength of the TATA box of the PHO5 gene in yeast cells result in a reduction in gene expression noise (40). A second study by Blake et al. (7) confirmed that mutations that decrease the strength of the TATA box of a different yeast promoter (a synthetic variant of GAL1) dramatically reduce cell-to-cell variability in gene expression.

A subsequent whole-genome noise analysis found that noisy genes were enriched in strong TATA boxes (33). However, most promoters that contain TATA boxes also contain nucleosomes in the vicinity of the transcriptional start site, so one possible explanation for the correlation between the presence of a TATA box and enhanced gene expression noise is the correlation between nucleosome positioning and TATA sequences, rather than a direct effect of the TATA box. A more recent study has directly examined this question (22). The authors studied a specific yeast promoter (HT2X) that contains both a TATA box and a proximal nucleosome-occupied site, and introduced mutations that abolished the TATA box. The authors observed a large reduction in gene expression noise as measured by CV^2. Similar results were observed when several TATA-box-containing promoters were subject to mutagenesis: Those mutations that eliminated the TATA box tended to significantly reduce the noise, particularly in those promoters with nucleosome-occupied proximal sites, where the TATA box was found to influence the burst size (22). Interestingly, the effect of the TATA box on the burst size was not observed in promoters lacking a proximal nucleosome, which suggests that the TATA box may be interacting with the nucleosome (22).

Burst size: average number of mRNAs synthesized during a burst of transcription

Gene expression plasticity: a measure of the dynamic range of expression of a particular gene

Operator: a specific transcription factor binding site in bacteria; here we generalize it to eukaryotic cells as well

Mechanistic explanation. Observed correlations between TATA box strength and gene expression noise can be captured by a simple model shown in **Figure 2*a***. We consider a promoter that can exist in two different states: without the full preinitiation complex assembled and with the preinitiation complex assembled. We plot in **Figure 2*b*** the predicted noise (measured by CV^2) as a function of the mean for two different TATA box strengths, assuming that TATA box strength affects the stability of the preinitiation complex and thus the rate at which it disassembles (7). In the plot, we assume that the mean expression can be controlled by tuning the assembly rate, again following the model described in Reference 7. This simple model predicts that strong TATA boxes may lead to larger noise than weak TATA boxes. However, if the rate of disassembly of the preinitiation complex is fast compared to the lifetime of the fluorescent reporter (**Figure 2*c***), or if the probability of reinitiation is low, the effect of the TATA box strength on noise may be too small to be observed experimentally when other regulatory elements, such as promoter inactivation by a nucleosome, are absent (22).

A very simple mechanism, shown in **Figure 3*a***, based on the interplay between the TATA box and nucleosome occupancy offers a possible explanation for the observed correlation between cell-to-cell variability in gene expression and gene expression plasticity (26, 54, 55). The model is based on the observation that this correlation is strongest in promoters containing both nucleosomes near the transcriptional start site and the TATA box (13, 26). In order to recapitulate this observation, we have generated 1,000 different promoters, by randomly sampling all the rates in the mechanism outlined in **Figure 3*a***, and computed its plasticity (defined as the dynamic range for the mean number of mRNAs) and the noise (CV^2 of the mRNA distribution) using the methods described in References 43 and 44 and summarized in the sidebar (see above). In **Figure 3*b***, we show that the experimentally observed correlation is reproduced by the simple model. Furthermore, we show that removal of the TATA box (i.e., eliminating the active state from the mechanism), or removal of the nucleosome (i.e., eliminating the closed state from the mechanism), has the effect of significantly reducing the correlation, as shown in **Figure 3*c,d***, mimicking experimental observations, and offering further support for the idea that plasticity and noise can be decoupled by specific changes to the promoter architecture. Although the mechanism shown in **Figure 3*a*** is a gross oversimplification of the very complex set of biochemical reactions that lead to the production of a new mRNA molecule, it is the minimal mechanism that captures stochastic gene silencing by nucleosomes and stochastic assembly and disassembly of the preinitiation complex at the TATA box.

Number of Transcription Factor Binding Sites

Experiments have shown that the number of binding sites for transcription factors can significantly affect gene expression noise. Although the mechanism is still not understood, we review the experiments and speculate about possible mechanisms.

Experimental evidence. In a recent study on yeast cells, To & Maheshri (57) demonstrated that multiple transcription factor binding sites (operators) increased gene expression noise. Here the authors varied the operator copy number for the tet-transcriptional activator (tTA) within a synthetic yeast promoter. They compared the noise strength for promoters with both one (1X) and seven (7X) copies of the tTA binding sites, and observed a pronounced increase in cell-to-cell variability of gene expression for the 7X promoter relative to the 1X promoter. These results are consistent with a previous study by Raj et al. (38), who also found that operator multiplicity leads to larger promoter noise in mammalian cells.

Similarly, Suter et al. (52) have reported strong effects of operator number (CCAAT boxes binding the transcriptional activator NF-Y) on mRNA and protein fluctuations driven by

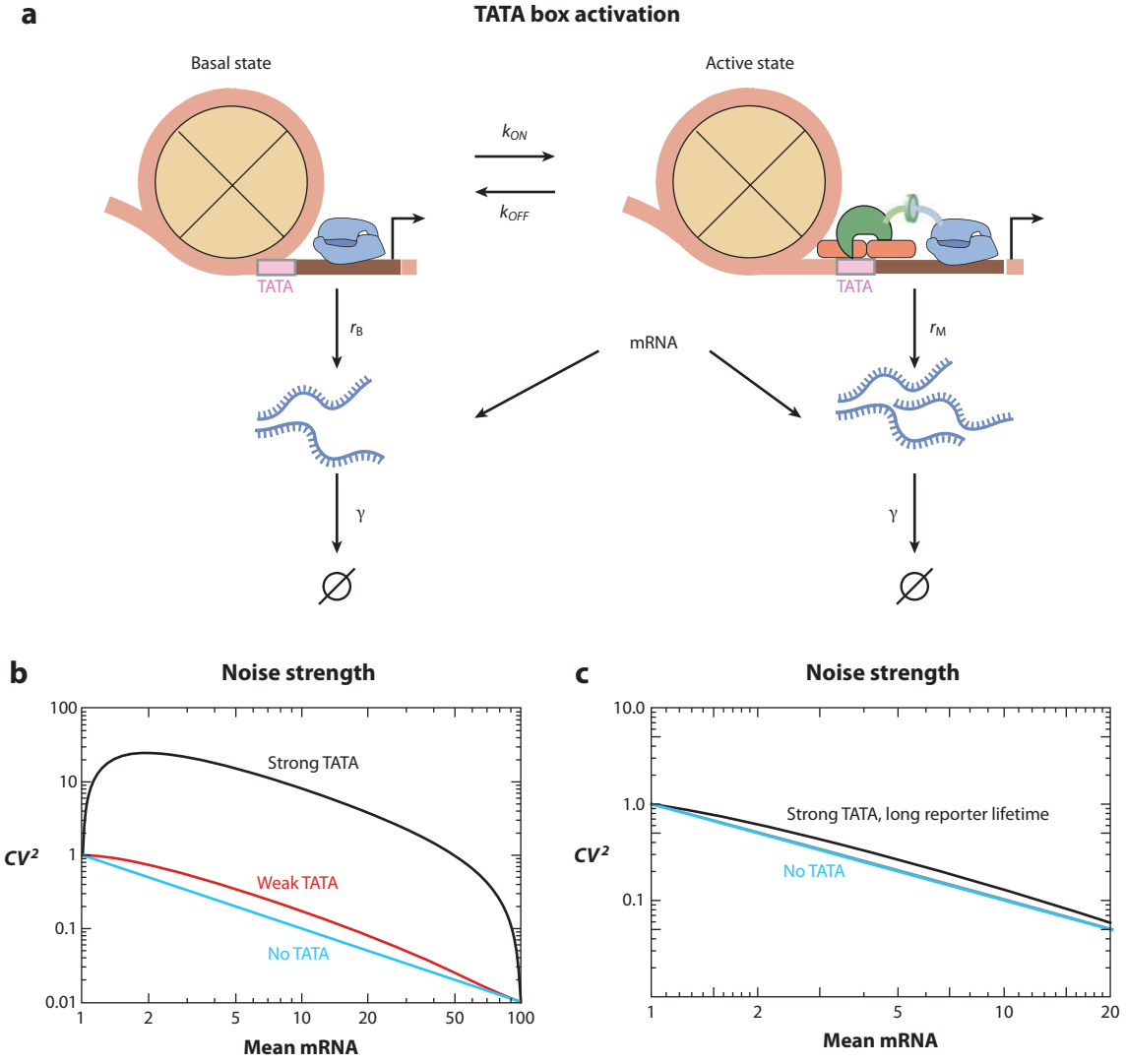


Figure 2

Transcriptional dynamics of TATA-box-mediated activation. (*a*) The TATA box activates the promoter by helping assemble the preinitiation complex. The rates of formation and dissociation of the preinitiation complex are given by k_{ON} and k_{OFF}, respectively, and the rates of mRNA production for the basal and active states are r_B and r_M, respectively. The mRNA degradation rate is γ. (*b*) Prediction for CV^2 as a function of the mean. Noise is enhanced with increasing strength of the TATA box; we have assumed that TATA box strength affects the stability of the preinitiation complex, which is quantified by the rate k_{OFF}. The parameters used to generate the plots for the solid black line are $r_B = 1$, $r_M = 100$, and $k_{OFF} = 0.01$, while k_{ON} is varied and sets the mean (all rates are in units of γ); for the solid red line the parameters are $r_B = 1$, $r_M = 100$, $k_{OFF} = 100$, and k_{ON} is variable. For the solid blue line, i.e., for a promoter without a TATA box, $CV^2 = 1/\langle \text{mRNA} \rangle$. (*c*) When the lifetime of the reporter is larger than the dissociation time of the preinitiation complex, the effect of the TATA box is small (parameters: $r_B = 1$, $r_M = 50$, $k_{OFF} = 200$, and k_{ON} is variable).

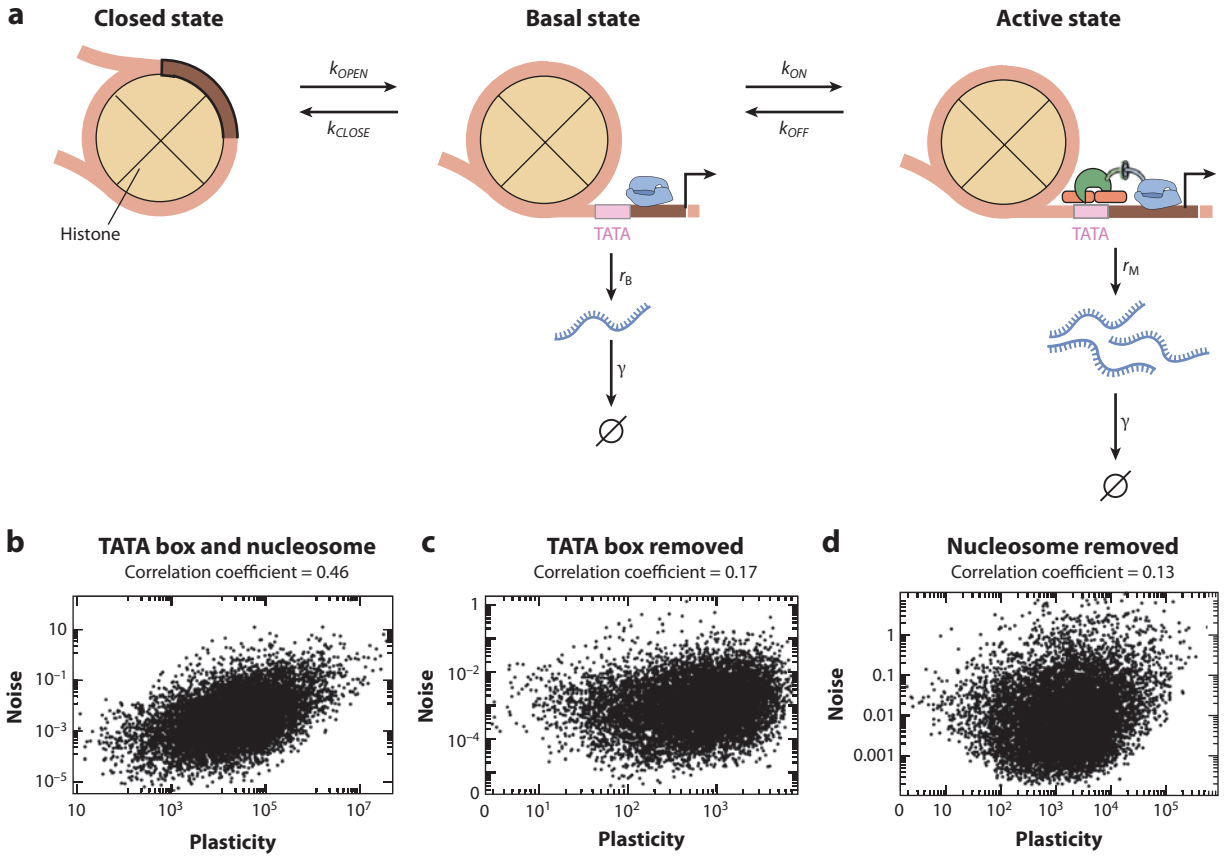


Figure 3

Transcriptional dynamics of nucleosome-inactivated, TATA-box-containing promoters. (*a*) A three-state model based on the interplay between the TATA box and the nucleosome can be used to explain the observed correlation between cell-to-cell variability and plasticity in gene expression. The three promoter states are a nucleosome-occluded promoter state (closed), an open promoter state with a basal transcription rate (basal), and an active state with the preinitiation complex formed at the TATA box (active). (*b–d*) Theoretical noise versus plasticity plots. Analytical expressions for the mean and the noise of gene expression, which were obtained using the method described in the sidebar, Method for Computing the Gene Expression Mean and Noise, were applied to the three-state model. The rates in the model were randomly sampled, producing 1,000 different parameter sets, which served as inputs for the analytical equations for the mean and noise. Parameters were sampled within the following ranges: $k_{ON} = 0.01–10{,}000$, $k_{OFF} = 0.01–10{,}000$, $k_{OPEN} = 0.01–10{,}000$, $k_{CLOSE} = 0.01–10{,}000$, $r_B = 0.01–10{,}000$, $r_M = 0.01–1{,}000{,}000$; all rates are given in units of γ. (*b*) When both the TATA box and the nucleosome are present, we see a high correlation between noise and plasticity. Removing either (*c*) the TATA box (by repeating the above parameter sampling procedure, but setting $k_{ON} = 0$) or (*d*) the nucleosome (by repeating the above parameter sampling procedure, but setting $k_{CLOSE} = 0$) significantly reduces the correlation between noise and plasticity.

mammalian promoters. This study used a luciferase assay to directly measure transcriptional dynamics in live cells. Their luciferase was short-lived and was expressed from an unstable mRNA whose transcription was directed by a library of natural and synthetic mammalian promoters with different architectures. In particular, the authors compared transcriptional stochasticity for promoters containing either one or two CCAAT boxes and found that doubling the number of CCAAT boxes resulted in a marked increase in the activation rate and the rate of transcription in the active state; however, the rate of inactivation was roughly unaffected by operator number (52).

Mechanistic explanation. So far, no general mechanistic model has been proposed to explain these experimental results. Qualitatively, an increase in noise is expected as we increase the number of transcription factor binding sites, solely due to the activator stochastically binding and falling off the promoter (43). Other sources of cell-to-cell variability, such as diffusion of transcription factors to their targets, have been put forth to explain activator-mediated noise in fly embryos (56) and in yeast cells (25). Because the number of binding sites does play a role in diffusion noise (56) and the time of capture of transcription factors by their binding sites (21), we suggest that the contribution of diffusion noise should perhaps be considered in connection with the experiments described above. It is also possible that cell-to-cell differences in activator concentrations may explain the observed effect of operator copy number on noise (34).

EFFECT OF PROMOTER ARCHITECTURE ON GENE EXPRESSION NOISE IN BACTERIA

Ironically, although our understanding of transcriptional regulation at the molecular level is much richer and more detailed for bacteria than for eukaryotes, the role that transcription factors play in transcriptional dynamics in live bacteria is still poorly understood. Decades of careful biochemical experiments have provided us with mechanisms by which different transcription factors exert control over gene expression (8, 42). We understand which steps on the biochemical pathway to mRNA synthesis are affected by a large number of transcriptional activators and repressors in vitro. Furthermore, many of these mechanisms have been tested in vivo by measuring the average gene expression level as a function of transcription factor concentration, and by comparing those measurements with mathematical models of the transcription process on the basis of specific mechanisms (6, 12, 17, 18, 24, 31). Typically, these mathematical models are constructed by listing all the possible states that the promoter can adopt. These promoter states are distinguished by the occupancies of the different operators in the promoter region by the corresponding transcription factors, as well as the occupancy of the promoter by RNA polymerase (RNAP) (6, 47). Then, the probability of different promoter states is computed using equilibrium thermodynamics, and the average rate of transcription (which is identified with the mean amount of gene expression) is obtained by computing the average of the transcription rates associated with those promoter states from which RNAP can initiate transcription (6, 47) (see **Figure 4**). These mathematical models allow us to compute, for a given mechanism of gene regulation, the shape of the input-output function for the promoter, e.g., the amount of gene expression as a function of transcription factor concentration. Direct comparison between these predictions and bulk in vivo data offers a way to test specific mechanisms of gene regulation. These in vivo tests of transcription-factor-mediated regulation have been very successful; the agreement between the predictions made by the models and the in vivo patterns of gene expression is excellent, and the precise effect that architectural features such as operator sequence and operator location have on the mean expression level is well understood. Therefore, although the effects of other regulatory processes, such as the condensation of DNA by nucleoid proteins (9), the modulation of DNA supercoiling (36), and the regulation of RNAP elongation along the gene by elongation factors (30, 41), have been long known and appreciated, the view that transcriptional regulation is to a large extent dominated by transcription factors and the small molecules (inducers) that modulate their activity emerged as a successful paradigm in the field (37).

The mechanistic models described above, however, have only been tested in the context of mean expression levels, which are independent of the transcriptional dynamics. In light of this, the simplest models of transcriptional dynamics can be constructed by extending the thermodynamic models to include stochastic transitions between the different promoter states. A constitutive

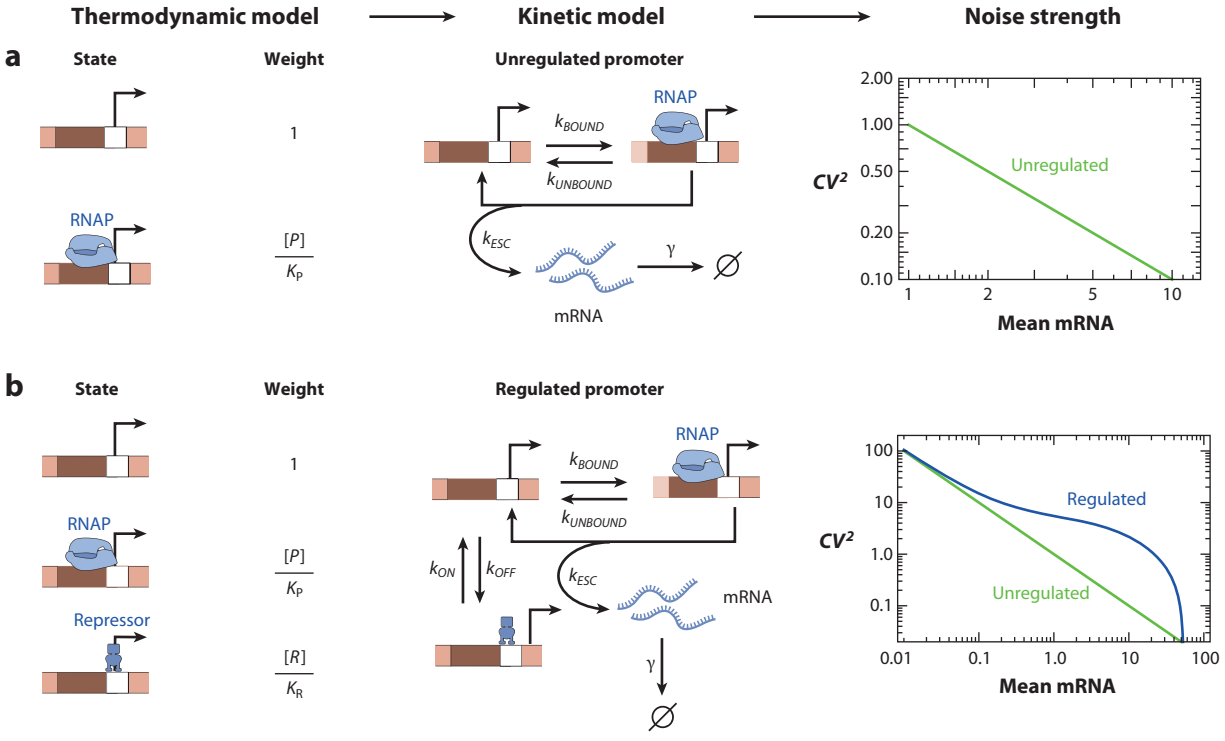


Figure 4

Thermodynamic and kinetic models of transcription-factor-mediated regulation of gene expression in bacteria. (*a*) A simple thermodynamic model of an unregulated promoter considers two states: one with RNA polymerase (RNAP) bound at the promoter and one with the promoter free of RNAP. The thermodynamic weights (un-normalized probabilities) for each of the states are shown. K_P represents the dissociation constant for RNAP binding to the promoter, and the intracellular concentration of RNAP is denoted by $[P]$. The kinetic model based on this simple thermodynamic model is built by connecting the states with arrows denoting stochastic kinetic transitions between them, which are parameterized by rate constants. The rates of RNAP binding and unbinding are given by k_{BOUND} and $k_{UNBOUND}$. The rate at which an RNAP molecule clears the promoter and initiates transcription is k_{ESC}. Prediction for CV^2 as a function of the mean is shown for this simple model, assuming that $k_{ESC} \gg k_{UNBOUND}$. In this limit we recover a Poisson distribution, and $CV^2 = 1/\langle \text{mRNA} \rangle$. (*b*) Thermodynamic model of a promoter regulated by a transcriptional repressor, whose intracellular concentration is denoted by $[R]$ and dissociation constant is given by K_R. The kinetic model is constructed in the same way as in panel *a*, by assigning stochastic transitions between the different states, which now include the rates of repressor association (k_{OFF}) and dissociation (k_{ON}). The prediction for CV^2 as a function of the mean is computed assuming $k_{ON} = 0.01$, $k_{BOUND} = 50$, $k_{UNBOUND} = 0.001$, $k_{ESC} = 200$, and k_{OFF} is variable; all rates are in units of γ.

promoter has by assumption only two promoter states, one with RNAP bound and one with the promoter free of RNAP. Transcriptional dynamics then describe RNAP forming an open complex with the promoter DNA, and its subsequent escape from the promoter. This model is represented in schematic form in **Figure 4*a***.

The expected gene expression noise from the kinetic model of a constitutive promoter can be computed using the tools described in the sidebar (see above) (43, 44). In the limit where the residence time of RNAP on the promoter is much shorter than the average transcription rate, a reasonable assumption in light of recent experimental observations (21), the mRNA distribution predicted by this mechanism is approximately Poisson, with $CV^2 = 1/\langle \text{mRNA} \rangle$. This approach can be extended to promoters that are regulated by transcription factors. For instance, in **Figure 4*b*** we show a simple mechanism for transcriptional repression by steric exclusion of RNAP binding,

which is known to be operational in the *lac* promoter in *Escherichia coli* (45, 46). The calculated CV^2 as a function of the mean expression level (which we identify with the mean number of mRNA molecules) for this mechanism is plotted in **Figure 4*b***. In this model, stochastic binding and unbinding of the repressor to the operator within the promoter region causes the promoter to randomly switch between inactive and active states, which results in an increase in mRNA number fluctuations. The noise-to-mean relationship for an arbitrary promoter, with any pattern of transcription factor binding sites, can be calculated using the methods described in the sidebar (see above), and it takes the form:

$$CV^2 = \frac{1}{\langle \mathrm{mRNA} \rangle} + CV^2_{promoter}.$$

This simple model, where transcription factors cause the promoter to switch between different states as they bind and fall off the promoter, makes the prediction that different promoters should have different noise-to-mean profiles (19, 43), given by $CV^2_{promoter}$, which describes the noise above the Poisson baseline given by $1/\langle \mathrm{mRNA} \rangle$.

A Case for Promoter-Architecture-Independent Noise

The relationship between the noise and the mean of gene expression was analyzed in a recent *E. coli* study for a set of seven different promoters, all driving the expression of the same gene (49). These constructs were inserted into *E. coli* strains and grown under different conditions. In total, about 120 different conditions were studied (the data are reproduced in **Figure 5*a***). The biochemical mechanisms of repression and activation for all these promoters are known to differ widely. However, the authors found that the noise-to-mean relationship is described by a universal function $CV^2 = (1/\langle \mathrm{mRNA} \rangle) \times (1 + 1.5\langle \mathrm{mRNA} \rangle^{0.64})$. A similar study (53), also using single-molecule FISH to detect and count individual mRNA molecules, analyzed 137 genes and also determined the noise-to-mean relationship. These two studies are in reasonable agreement with each other (**Figure 5*a***).

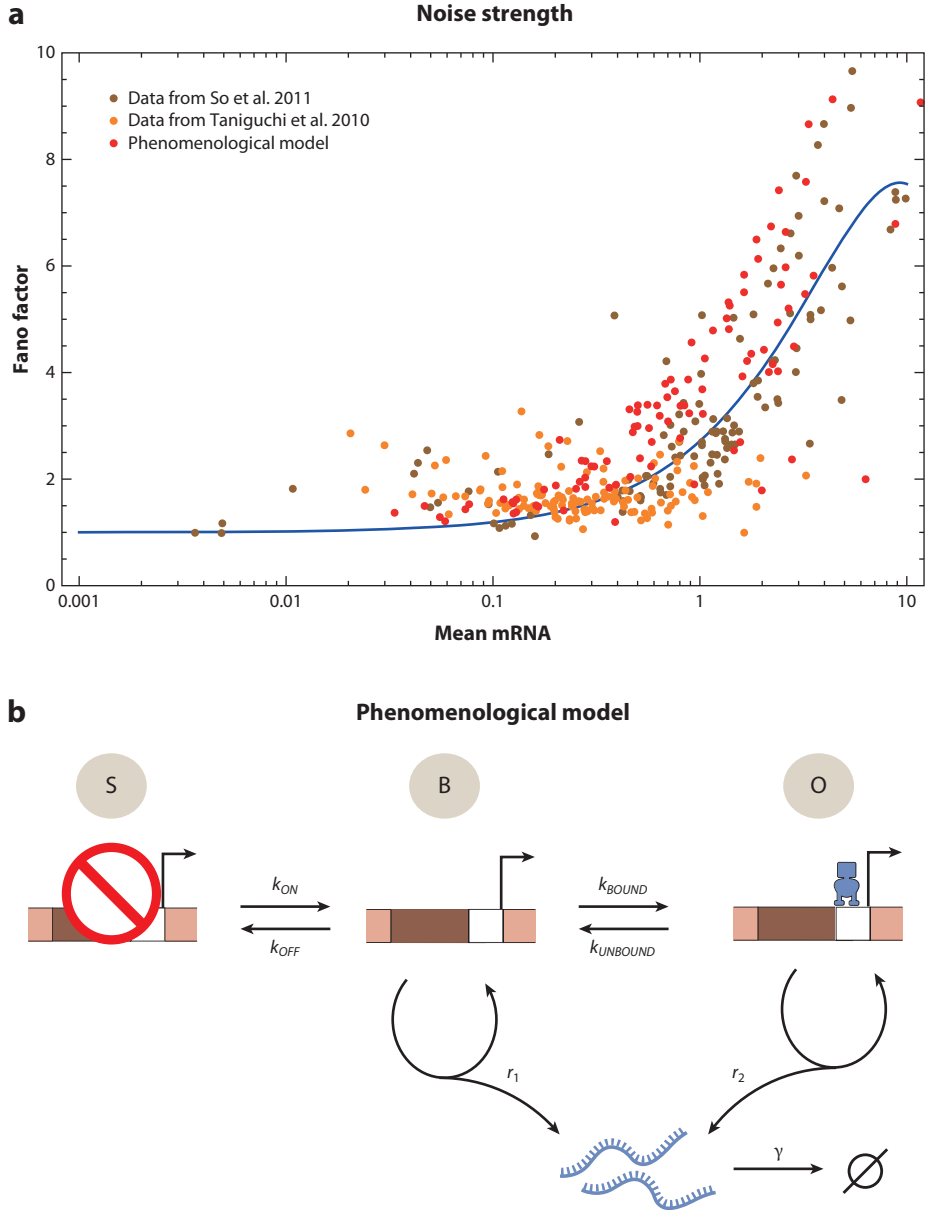
The universal trend that is followed by different promoters with different architectures and regulated by different transcription factors can be interpreted as evidence that the particular mechanistic details of transcription-factor-mediated gene regulation are not relevant for the noise-to-mean relationship. This is certainly possible, and it might well be that regulation by transcription factors does not affect mRNA fluctuations in *E. coli*. If this is the case, then what is causing the observed universal trend?

In earlier experiments, Golding et al. (20) measured transcriptional dynamics in live *E. coli* cells directly. In these experiments, mRNA molecules were detected by their association with fluorescently labeled proteins (MS2-GFP) that bound to the nascent transcript. The authors found that even at fully induced conditions, the promoter they studied (the $P_{lac/ara}$ promoter) switched stochastically between an active and an inactive state. The biological agent responsible for this bursting behavior remains unknown, but it is tempting to assume that whatever is causing the $P_{lac/ara}$ promoter to stochastically switch on and off may be also responsible for the observed universal mean-to-noise patterns. Because the $P_{lac/ara}$ promoter was fully induced (20), it seems like stochastic association and dissociation of its activators (araC) and repressors (LacI) would be an unlikely cause for the observed bursting. Many different possible origins for this bursting behavior have been proposed (29), but the question remains unresolved.

Possible Mechanism for Universal Noise

Gene expression noise measurements in bacteria have to be considered with some caution because single-molecule experiments in live or fixed cells are still in their infancy. However, the results

from the different techniques used so far, both single-molecule FISH and MS2 detection, are consistent with each other and strongly suggest that something other than transcription factors, or at least not transcription factors that are specific to individual promoters, can turn genes off and affect the absolute levels of gene expression. It has been suggested (49) that the observed trend could be explained by a phenomenological model where all genes in *E. coli* are switching between an ON and an OFF state, with similar activation rates (k_{ON}) and similar transcription rates in the ON state (r), and gene regulation is primarily achieved by varying the inactivation rate k_{OFF}. Thus, all transcription factor regulation would achieve, regardless of the specific mechanisms used, would be shortening or prolonging the time that the gene spends in the ON state (49).



How can we reconcile this idea with the overwhelming evidence from in vitro biochemistry and population-averaged in vivo gene expression measurements, that different transcription factors employ different mechanisms to regulate transcription? Although within the two-state, ON-OFF model, k_{OFF} regulation offers a quantitative explanation for the universal trend line (49), alternative phenomenological models can also explain the data. For instance, in **Figure 5*b*** we consider a model where all genes in *E. coli* switch slowly between one silenced state (S) and one basal state (B) by a transcription-factor-independent process, e.g., due to DNA condensation by nucleoid proteins or changes in DNA supercoiling. Only when the genes are in the basal state can transcription factors stochastically bind and fall off. This possibility is included in the model by considering a third state (O) in which the transcription factors are bound. When the gene is not silenced, we assume that transcription factors act by recruiting or excluding RNAP. As soon as the promoter switches back to the silenced state no transcription-factor-mediated regulation can occur. Using this phenomenological model, we have computed the means and Fano factors (Fano factor = variance/mean) for a collection of promoters, generated by randomly sampling the rates of switching between the O and B states, as well as the rates of transcription and mRNA degradation, as described in **Figure 5**. The data generated in this way show a correlation between noise and mean very similar to the one observed by Golding and colleagues (49), and these data can also be fitted by the k_{OFF} modulation model (**Figure 5*a***). This is not a refutation of the k_{OFF} modulation model; however, it offers an alternative explanation of the data, one that captures the essential feature of the direct observation of ON-OFF switching (20), and can accommodate known biochemical mechanisms of transcriptional regulation.

The observed universal trend between the noise and the mean might also be caused, at least in part, by a collection of many independent sources of noise, including yet unidentified biological sources as well as experimental artifacts or experimental error in the determination of mRNA copy numbers using FISH. In spite of this possibility, the observation of the universal trend is very intriguing and it deserves further experimental and theoretical exploration. Finally, the experimental observation of a universal noise-mean relationship in *E. coli* has not yet been replicated in

Figure 5

Universal noise-mean relationship for *Escherichia coli* promoters. (*a*) Noise (here quantified by the Fano factor = variance/mean) of the mRNA distribution plotted as a function of the mean mRNA for a set of seven promoters exposed to different growth conditions and genomic backgrounds. Data from So et al. (49) is shown as brown dots. A separate set of data from Taniguchi et al. (53) is also plotted (*orange dots*). It corresponds to an experiment that determined mRNA distributions for more than 100 different genes in *E. coli*. Both sets of data were acquired using single-molecule FISH. The blue line corresponds to the fit to a two-state ON/OFF phenomenological model, where the rate of promoter activation (k_{ON}) and the rate of transcription in the ON state (r) are assumed to be constant, and only k_{OFF} changes from promoter to promoter. This model is referred to as the k_{OFF} modulation model and it was proposed by So et al. (49). The red dots correspond to 100 different values of mRNA mean and noise, obtained by randomly sampling all the rates in a phenomenological model shown in panel *b*. This model assumes that the promoter is inactivated by a transcription-factor-independent mechanism leading to an inactive or silenced state (S). When the promoter is not inactive, transcription factors bind to and fall off the promoter, leading to transitions between the operator-occupied state (O) and the operator-free basal state (B), respectively. Bound transcription factors either activate or inactivate transcription by recruiting or inhibiting RNA polymerase (RNAP) binding. Rates of transitions between the S and B states are k_{ON} and k_{OFF}, and between B and O they are k_{BOUND} and $k_{UNBOUND}$. The rate of transcription from states B and O are r_1 and r_2, respectively. The model parameters are chosen at random, so both activation and repression are possible; activation occurs when $r_2 > r_1$, and repression occurs when $r_2 < r_1$. The different rates were randomly sampled from the intervals $r_1 = (0.01–1)$, $r_2 = (0.0001–10)$, $k_{BOUND} = (0.1–10)$, and $k_{UNBOUND} = (0.1–10)$, while $k_{ON} = 1/37$ and $k_{OFF} = 1/6$ were taken to be the experimentally measured rates (20); all the rates are in units of $\gamma = 1\ \text{min}^{-1}$.

other bacteria. It is possible that transcriptional dynamics may differ from organism to organism, and other bacteria may show different noise-to-mean profiles.

Evidence that Promoter Architecture Affects Noise

Contrary to the conclusions reached by the experiments described in the previous section, there is also evidence that promoter architecture may play a role in generating cell-to-cell variability in gene expression, although this is less convincing in bacteria than in eukaryotic cells. We review genome-wide and single-gene experiments that point to this possibility.

Genome-wide experiments. Beyond the universal trends discussed above, other experiments have found that different promoters can confer different levels of noise in gene expression for the same mean. The clearest result comes from a genome-wide analysis of noise in *E. coli* using a genomic library in which the GFP gene is fused to all promoters in the genome (48). *E. coli* was transformed with multicopy plasmids containing the GFP gene under the control of specific promoters, and then single-cell gene expression was measured by flow cytometry. In spite of the low resolution of gene expression measurements resulting from flow cytometry (i.e., it does not allow for single-molecule counting), the fact that GFP has a long lifetime, therefore buffering stochastic transitions between promoter states, and the fact that multicopy plasmids were employed (all of which reduce noise), differences in gene expression noise were observed from promoter to promoter relative to the baseline noise, which scaled as $CV^2 \sim 1/\langle \mathrm{GFP} \rangle$. The authors found that, as expected, promoters associated with housekeeping genes conferred low levels of noise. In contrast, promoters subject to regulation, such as those controlling the expression of genes responsible for adaptation to stress or for carbon metabolism, showed large levels of noise. Because, in this experiment, the only difference from plasmid to plasmid was the promoter sequence itself (both the plasmids and the fluorescent reporter were identical in all cases), the authors concluded that promoter-mediated noise could vary significantly from gene to gene. The difference in noise from promoter to promoter might be even larger at the mRNA level, or for proteins with a short lifetime, or if the authors had considered the genomic promoter in the absence of extrachromosomal copies on a plasmid. On the other hand, it could be that the variability observed by the authors comes from phenotypic multi-stability rather than from stochastic promoter dynamics.

The idea that noise associated with gene regulation can be significant in bacteria has found experimental support in organisms other than *E. coli*. A study by the Ackermann lab (16) used a clever strategy to identify genes in *Salmonella* whose expression fluctuates widely over short timescales and thus have high levels of noise. The authors created a genomic plasmid library of promoters fused to GFP. Then they used fluorescence-activated cell sorting to select noisy genes. The cell sorter was first used to select cells with high expression levels (within the highest 5%). Then the selected cells were grown overnight, and at mid-exponential phase the cell sorter was used to select for low expression levels (within the lowest 5%). This approach was iterated seven times, and the resulting cells were sequenced. The cells selected were those that had been transformed with flagellar promoters, and the expression of these promoters was bimodal (16). The authors controlled for the possibility that the observed bimodality was due to genetic differences between different individual cells, and concluded that it had its origin in stochasticity in gene expression (16). As before, it is possible that the larger noise in the bimodal promoters had more to do with a bistable phenotypic switch (of which the GFP-expressing promoter becomes a mere reporter) than with the promoter architecture itself.

Single-gene experiments. The one example we are aware of, in which the mechanistic origin of noise in gene expression in *E. coli* has been explored, is the work by Choi et al. (11), who investigated the effect of changing the architecture of the *lac* promoter on its gene expression stochasticity. The authors compared the induction profiles for the wild-type *lac* promoter and a mutated promoter carrying a deletion of both auxiliary operators that bind the Lac repressor. The occupation of these operators does not lead to repression; instead they stabilize (via DNA looping) the Lac repressor when it is bound to the main operator, which is responsible for repression of transcription. Gene expression from the *lac* promoter is bimodal at low and intermediate induction levels. However, the authors found that the deletion of these weak auxiliary operators led to unimodal induction (11). Choi et al. (11) hypothesized that DNA looping might reduce noise by leading to rapid association and dissociation from the promoter, which had been suggested previously on the basis of simulation results (58). We have computationally analyzed the mechanism proposed by Choi et al. using the tools described in the sidebar (see above), and found that, given our current understanding of the kinetics of DNA looping from experiments in vitro, looping should not lower noise. Instead, we expect looping to increase noise (43). We find that DNA looping leads to rapid rebinding of the repressor to the main operator once it dissociates from the main operator. This effectively decreases the rate of dissociation, slowing down the switching between active and inactive states, and therefore increases mRNA fluctuations (43).

Finally, a recent article investigated the behavior of three different promoters (P_{gapB}, P_{cggR}, and P_{pckA}) involved in the physiological switch between glycolysis and gluconeogenesis in *B. subtilis* (14). The authors used a novel two-photon fluorescence fluctuation spectroscopy technique to quantify GFP expression in individual cells with very high resolution and constructed histograms of the cell-to-cell variability. They found large differences in the mode of induction and repression as they shifted the media from malate to glucose. Under repressed conditions (in glucose media), P_{pckA} and P_{gapB} had a lower frequency of activation and a lower GFP production in the active state. P_{cggR} also had a lower frequency of activation under repressed conditions (which, for this promoter, corresponded to growth in malate media), but the burst size actually increased. This created a small subpopulation of repressed cells that had nearly as much expression as induced cells, due to rare but very strong bursts of expression. This study suggests that there may be promoter-specific differences in transcriptional dynamics in *B. subtilis*, and that these differences may lead to large gene expression heterogeneity, with some repressed cells displaying expression levels characteristic of induced cells. Remarkably, the authors found that the characteristic pattern of transcriptional bursting observed for P_{cggR} at repressed conditions was consistent with the specific mechanism of repression of this promoter, which is based on roadblocking the RNAP during elongation. The authors proposed and analyzed a simple model in which bursting was caused by spontaneous dissociation of the repressor (which has a very large affinity for its operator), which then lets through a train of RNAP molecules that were stalled by the presence of the roadblock provided by the repressor (14). Although further studies are needed to confirm this mechanism, this study suggests that, for these promoters, the molecular mechanism of transcription-factor-mediated repression is sufficient to explain bursting behavior and heterogeneity in gene expression.

The examples described above attest that the promoter architecture, and in particular transcription factor dynamics at the promoter, may at least in some instances have a direct effect on gene expression noise. More importantly, we believe these examples make a strong case for the need to perform in bacteria the same kind of experiments that have been done in yeast and mammalian cells, where promoter architectures are systematically perturbed and the effects of these perturbations on the mean and noise in gene expression are characterized.

IS GENE EXPRESSION NOISE UNIVERSAL?

In the previous section we described seemingly conflicting views of gene expression noise that arose from single-cell measurements of gene expression outputs from bacteria. In particular, the existence of a noise-to-mean trend line, which is followed by a large number of genes, seems to be in conflict with the idea that gene-specific mechanisms of transcriptional regulation (for which there is an abundance of in vitro biochemical evidence) can lead to a characteristic noise signature, quantified by the dependence of the noise on the mean. Thus, the existence of such a universal relationship between noise and mean might appear to disprove the idea that gene expression noise is a trait that can be regulated at the promoter level or dictated by the DNA sequence of the regulatory region. In this section we further elaborate on the experimental evidence for universal noise by reviewing recent experiments in yeast cells in which a universal relationship between the mean and the noise in gene expression was discovered to be similar to the relationship found in bacteria. These experiments also found that certain promoters produce noise that significantly deviates from the universal behavior, providing further support for the idea that promoter-architecture-specific elements can play an important role in determining noise.

Universal Scaling Between Noise and Mean in Yeast Cells

The first evidence for a universal scaling law between noise in gene expression and the mean expression level came from a genome-wide study in yeast cells (33), as well as a study published simultaneously (4) that analyzed noise in 43 different yeast genes that were exposed to different environmental conditions. In these studies the authors used a genomic GFP fusion library to measure cell-to-cell distributions of GFP fluorescence. The authors plotted the CV^2 of the GFP distributions as a function of their means and found a scaling relation $CV^2 \sim 1/\langle \text{GFP} \rangle$ at intermediate gene expression values (at high expression values, the noise became decoupled from the mean). However, they also found many genes that significantly deviate from this trend. Those tend to be stress-response genes and often have the type of architecture that leads to large levels of noise, including TATA boxes and occupied proximal nucleosomes.

This point has been further examined by analyzing the noise-to-mean relationship for 22 genes in yeast cells, as well as constructing a mutant library for each of those genes (22). The results of this last study were revealing; namely, each mutagenized promoter occupied a characteristic region of the noise-mean plot. Those genes that lacked a TATA box and were nucleosome depleted followed the same trend line. However, promoters containing a TATA box and an occupied proximal nucleosome exhibited much larger noise and were found to lie above the trend line. In other words, specific differences in promoter architecture were identified with significant differences in the noise-mean curve. The effect of promoter architecture on the noise-mean relationship is further supported by the observation that those mutants of the high-noise promoters that, by accident, lost their TATA box fell down to the trend line, followed by the low-noise promoters and their mutants (see **Figure 6**).

Another test of the idea that deviations from universal gene expression noise may be caused by promoter architecture was performed by analyzing mRNA distributions in single yeast cells using FISH, both for a SAGA-regulated, TATA-box-containing gene that deviates from the universal trend line (PDR5) and for a constitutive gene that falls within the trend line (MND1) (59). As expected, the PDR5 gene displayed distributions of nascent and cytoplasmic mRNA that were consistent with bursting. The constitutive gene, on the other hand, was described by a Poisson distribution of cytoplasmic mRNA and a distribution of nascent mRNA that was also consistent

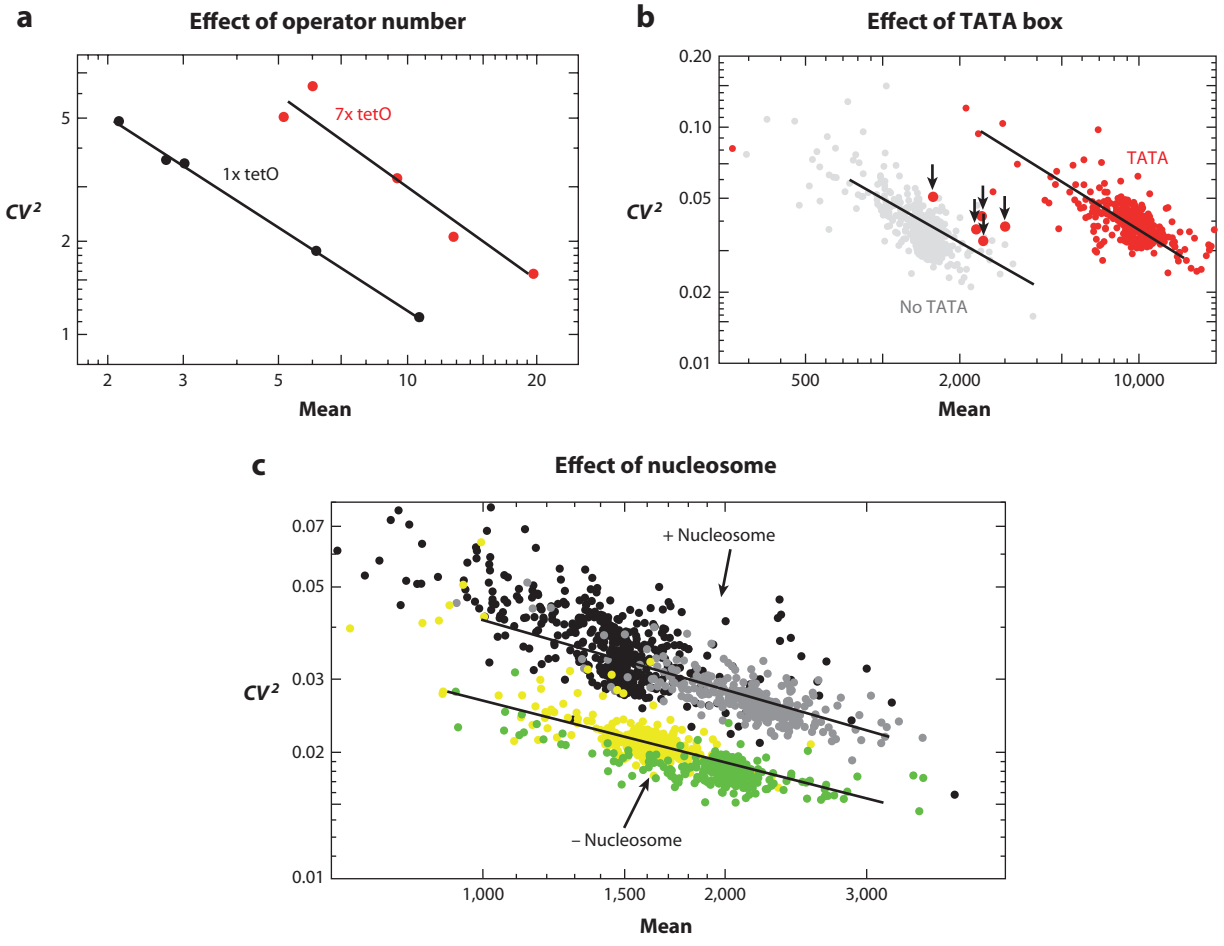


Figure 6

Summary of experimental evidence that promoter architecture affects noise in gene expression. (*a*) Noise-to-mean relationships for two promoters that differ in the copy number of tetO binding sites for the tTA activator. The noise is consistently higher for the promoter with seven (7X) operators than for the promoter with just one (1X). (Data courtesy of Reference 57.) (*b*) Noise-to-mean relationship for a library of mutants of the HXT2 promoter (*red*) as well as mutants of ADH3 (*light gray*). The HXT2 promoter has an occupied proximal nucleosome and a TATA box. The ADH3 mutated promoters, in contrast, have an occupied proximal nucleosome but lack a TATA box. The TATA-box-containing mutant promoters have a larger gene expression noise than the two promoters lacking the TATA box, with the exception of five mutants (*small arrows*) whose noise-mean relationship followed the same trend as the TATA-less promoters. When those HXT2 mutants were sequenced, they were found to have TATA-box-abolishing mutations, which confirmed that the presence of a TATA box increases the noise for nucleosome-occluded promoters. (Data courtesy of Reference 22.) (*c*) Noise-to-mean relationship for two nucleosome-depleted promoters (VMA7 and ARO2, *green and yellow*, respectively) and two nucleosome-containing promoters (ADH3 and ERG11, *black and gray*, respectively). All promoters lack a TATA box. Data shown in panel *c* represent a subset of all the data collected by Hornung et al. (22) and are representative of the trends between nucleosome occupancy and noise observed by these authors. (The black lines in all the data plots are not fits but serve as guides for the eye.)

with random independent activation (59). The trend line is consistent with the idea that a majority of genes have low promoter noise, so that $CV^2_{promoter}$ is small and $CV^2 \sim 1/\langle \text{mRNA} \rangle$. In contrast the genes that rise above this trend line are regulated with large promoter noise, so that $CV^2 \sim 1/\langle \text{mRNA} \rangle + CV^2_{promoter}$.

SUMMARY AND OUTLOOK

We have examined the role of gene regulation, and the architecture of transcriptional promoters in particular, in generating noise in gene expression. We believe that there is clear evidence that the process of gene regulation can strongly affect cell-to-cell variability in gene expression in eukaryotic cells. The evidence for this is summarized in **Figure 6**, which shows that architectural features such as the number of transcription factor binding sites, the presence of a TATA box, or the propensity of a promoter to be occluded by nucleosomes all have clear effects on mRNA and protein fluctuations in a population of genetically identical cells.

The precise mechanisms involved in the regulation of gene expression in eukaryotes are very complex, often involving dozens of different regulators, coregulators, nucleosomes, and long-range regulation by DNA looping. Owing to this intrinsic complexity, the task of connecting microscopic models of gene regulation with the observed noise patterns has remained out of reach, and many experimental observations, such as the increase in burst size as the operator number is increased (52, 57), or the correlation between the probability that a gene is active and the rate of activation (52), remain disconnected from the molecular mechanisms believed to be responsible for gene activation and inactivation. The development of a quantitative connection between mechanistic molecular-scale models of gene regulation and single-cell measurements of gene expression represents a frontier in our understanding of gene regulation in eukaryotic cells, and one where the combination of modeling and careful, quantitative experiments can be particularly fruitful.

The effect of gene regulation on mRNA and protein fluctuations in bacteria is much less well understood. On the one hand, there is evidence of universal trends at the protein and mRNA levels that correlate noise with the mean in gene expression. On the other hand, certain promoters exhibit larger noise than would be expected based on the mean, and, at least in the case of protein noise, those promoters tend to fall within certain functional categories. This last fact leads us to suspect that there may be a role for gene regulation in generating noise in bacteria. The observation of transcriptional bursting even in fully induced promoters suggests the existence of an unidentified mechanism by which promoters turn off. Identifying this mechanism, its properties, and its relationship with regulation by transcription factors are all important open questions in the biophysics of gene regulation. Their resolution has the potential to shed new light on important issues such as the nature of the regulatory code, namely how regulatory DNA sequence translates into a specific input-output function for gene expression. A better understanding of the regulatory code may in turn provide clues about the evolution of regulatory DNA. With the wealth of quantitative techniques for examining gene regulation at the single-cell and single-molecule levels at our disposal, we believe that these questions will be answered in the near future.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We wish to thank Rob Phillips, Hernan Garcia, Jeff Gelles, Timothy Harden, Daniel Jones, Rob Brewster, and Paul Wiggins for years of stimulating discussions and shared ideas about transcriptional dynamics and the biological mechanisms underlying them. We also want to express our gratitude to Narendra Maheshri, T.L. To, Ido Golding, Lok-Hang So, Naama Barkai, and Gil Hornung for sharing their data and discussing their results with us, some of them before they

were published. A.S. wishes to thank Jeff Gore for his generous support while writing this review. This work was supported by the NSF through grant DMR-1206146 to J.K.

LITERATURE CITED

1. Arias AM, Hayward P. 2006. Filtering transcriptional noise during development: concepts and mechanisms. *Nat. Rev. Genet.* 7(1):34–44
2. Bajic D, Poyatos JF. 2012. Balancing noise and plasticity in eukaryotic gene expression. *BMC Genomics* 13(1):343
3. Balázsi G, van Oudenaarden A, Collins JJ. 2011. Cellular decision making and biological noise: from microbes to mammals. *Cell* 144(6):910–25
4. Bar-Even A, Paulsson J, Maheshri N, Carmi M, O'Shea E, et al. 2006. Noise in protein expression scales with natural protein abundance. *Nat. Genet.* 38(6):636–43
5. Batada NN, Hurst LD. 2007. Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nat. Genet.* 39(8):945–49
6. Bintu L, Buchler NE, Garcia HG, Gerland U, Hwa T, et al. 2005. Transcriptional regulation by the numbers: applications. *Curr. Opin. Genet. Dev.* 15(2):125–35
7. **Blake WJ, Balázsi G, Kohanski MA, Isaacs FJ, Murphy KF, et al. 2006. Phenotypic consequences of promoter-mediated transcriptional noise. *Mol. Cell* 24(6):853–65**
8. Browning DF, Busby SJ. 2004. The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.* 2(1):57–65
9. Browning DF, Grainger DC, Busby SJ. 2010. Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. *Curr. Opin. Microbiol.* 13(6):773–80
10. Chalancon G, Ravarani CNJ, Balaji S, Martinez-Arias A, Aravind L, et al. 2012. Interplay between gene expression noise and regulatory network architecture. *Trends Genet.* 28(5):221–32
11. **Choi PJ, Cai L, Frieda K, Xie XS. 2008. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science* 322(5900):442–46**
12. Dodd IB. 2004. Cooperativity in long-range gene regulation by the CI repressor. *Genes Dev.* 18(3):344–54
13. Dong D, Shao X, Deng N, Zhang Z. 2011. Gene expression variations are predictive for stochastic noise. *Nucleic Acids Res.* 39(2):403–13
14. **Ferguson ML, Le Coq D, Jules M, Aymerich S, Radulescu O, et al. 2012. Reconciling molecular regulatory mechanisms with noise patterns of bacterial metabolic promoters in induced and repressed states. *Proc. Natl. Acad. Sci. USA* 109(1):155–60**
15. Field Y, Kaplan N, Fondufe-Mittendorf Y, Moore IK, Sharon E, et al. 2008. Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS Comput. Biol.* 4(11):e1000216
16. Freed NE, Silander OK, Stecher B, Böhm A, Hardt W-D, Ackermann M. 2008. A simple screen to identify promoters conferring high levels of phenotypic noise. *PLoS Genet.* 4(12):e1000307
17. Garcia HG, Phillips R. 2011. Quantitative dissection of the simple repression input–output function. *Proc. Natl. Acad. Sci. USA* 108:12173–78
18. Garcia HG, Sanchez A, Boedicker JQ, Osborne M, Gelles J, et al. 2012. Operator sequence alters gene expression independently of transcription factor occupancy in bacteria. *Cell Rep.* 2(1):150–61
19. Garcia HG, Sanchez A, Kuhlman T, Kondev J, Phillips R. 2010. Transcription by the numbers redux: experiments and calculations that surprise. *Trends Cell Biol.* 20(12):723–33
20. **Golding I, Paulsson J, Zawilski SM, Cox EC. 2005. Real-time kinetics of gene activity in individual bacteria. *Cell* 123(6):1025–36**
21. Hammar P, Leroy P, Mahmutovic A, Marklund EG, Berg OG, Elf J. 2012. The *lac* repressor displays facilitated diffusion in living cells. *Science* 336(6088):1595–98
22. **Hornung G, Bar-Ziv R, Rosin D, Tokuriki N, Tawfik DS, et al. 2012. Noise-mean relationship in mutated promoters. *Genome Res.* 22:2409–17**
23. Kaern M, Elston TC, Blake WJ, Collins JJ. 2005. Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* 6(6):451–64

7. Demonstrates that noise determined by promoter architecture can affect the fitness of microbial organisms.

11. Presents evidence that promoter architecture in *E. coli* affects gene expression noise.

14. Provides evidence that transcription-factor-mediated regulatory mechanisms in *B. subtilis* affect gene expression noise.

20. Demonstrates transcriptional bursting in bacteria.

22. Demonstrates that promoter architecture affects gene expression noise in yeast cells.

24. Kuhlman T, Zhang Z, Saier MH Jr, Hwa T. 2007. Combinatorial transcriptional control of the lactose operon of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 104(14):6043–48
25. Larson DR, Zenklusen D, Wu B, Chao JA, Singer RH. 2011. Real-time observation of transcription initiation and elongation on an endogenous yeast gene. *Science* 332(6028):475–78
26. Lehner B. 2010. Conflict between noise and plasticity in yeast. *PLoS Genet.* 6(11):e1001185
27. Losick R, Desplan C. 2008. Stochasticity and cell fate. *Science* 320(5872):65–68
28. Maamar H, Raj A, Dubnau D. 2007. Noise in gene expression determines cell fate in *Bacillus subtilis*. *Science* 317(5837):526–29
29. Mitarai N, Dodd IB, Crooks MT, Sneppen K. 2008. The generation of promoter-mediated transcriptional noise in bacteria. *PLoS Comput. Biol.* 4(7):e1000109
30. Mooney RA, Artsimovitch I, Landick R. 1998. Information processing by RNA polymerase: recognition of regulatory signals during RNA chain elongation. *J. Bacteriol.* 180(13):3265–75
31. Müller J, Oehler S, Müller-Hill B. 1996. Repression of *lac* promoter as a function of distance, phase and quality of an auxiliary *lac* operator. *J. Mol. Biol.* 257(1):21–29
32. Murphy KF, Balázsi G, Collins JJ. 2007. Combinatorial promoter design for engineering noisy gene expression. *Proc. Natl. Acad. Sci. USA* 104(31):12726–31
33. Newman JRS, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, et al. 2006. Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* 441(7095):840–46
34. Ochab-Marcinek A, Tabaka M. 2010. Bimodal gene expression in noncooperative regulatory systems. *Proc. Natl. Acad. Sci. USA* 107(51):22096–101
35. Polach KJ, Widom J. 1995. Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation. *J. Mol. Biol.* 254(2):130–49
36. Pruss GJ, Drlica K. 1989. DNA supercoiling and prokaryotic transcription. *Cell* 56(4):521–23
37. Ptashne M. 2004. *A Genetic Switch: Phage Lambda Revisited*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. 3rd ed.
38. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. 2006. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* 4(10):e309
39. Raj A, Rifkin SA, Andersen E, van Oudenaarden A. 2010. Variability in gene expression underlies incomplete penetrance. *Nature* 463(7283):913–18

40. Demonstrates the role of chromatin remodeling in transcriptional stochasticity.

40. Raser JM, O'Shea EK. 2004. Control of stochasticity in eukaryotic gene expression. *Science* 304(5678):1811–14

41. Roberts JW, Shankar S, Filter JJ. 2008. RNA polymerase elongation factors. *Annu. Rev. Microbiol.* 62:211–33
42. Rojo F. 2001. Mechanisms of transcriptional repression. *Curr. Opin. Microbiol.* 4(2):145–51

43. Gives a systematic computational investigation of how promoter architecture is expected to affect gene expression noise in bacteria.

43. Sánchez A, Garcia HG, Jones D, Phillips R, Kondev J. 2011. Effect of promoter architecture on the cell-to-cell variability in gene expression. *PLoS Comput. Biol.* 7(3):e1001100

44. Sánchez A, Kondev J. 2008. Transcriptional control of noise in gene expression. *Proc. Natl. Acad. Sci. USA* 105(13):5081–86
45. Sánchez A, Osborne ML, Friedman LJ, Kondev J, Gelles J. 2011. Mechanism of transcriptional repression at a bacterial promoter by analysis of single molecules. *EMBO J.* 30(19):3940–46
46. Schlax PJ, Capp MW, Record MT Jr. 1995. Inhibition of transcription initiation by *lac* repressor. *J. Mol. Biol.* 245(4):331–50
47. Shea MA, Ackers GK. 1985. The OR control system of bacteriophage lambda. A physical-chemical model for gene regulation. *J. Mol. Biol.* 181(2):211–30
48. Silander OK, Nikolic N, Zaslaver A, Bren A, Kikoin I, et al. 2012. A genome-wide analysis of promoter-mediated phenotypic noise in *Escherichia coli*. *PLoS Genet.* 8(1):e1002443

49. The authors found a universal trend between gene expression noise and mean in *E. coli*.

49. So L-H, Ghosh A, Zong C, Sepúlveda LA, Segev R, Golding I. 2011. General properties of transcriptional time series in *Escherichia coli*. *Nat. Genet.* 43(6):554–60

50. Süel GM, Garcia-Ojalvo J, Liberman LM, Elowitz MB. 2006. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature* 440(7083):545–50
51. Süel GM, Kulkarni RP, Dworkin J, Garcia-Ojalvo J, Elowitz MB. 2007. Tunability and noise dependence in differentiation dynamics. *Science* 315(5819):1716–19

52. Suter DM, Molina N, Gatfield D, Schneider K, Schibler U, Naef F. 2011. Mammalian genes are transcribed with widely different bursting kinetics. *Science* 332(6028):472–74
53. Taniguchi Y, Choi PJ, Li G-W, Chen H, Babu M, et al. 2010. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 329(5991):533–38
54. Tirosh I, Barkai N. 2008. Two strategies for gene regulation by promoter nucleosomes. *Genome Res.* 18(7):1084–91
55. Tirosh I, Barkai N, Verstrepen KJ. 2009. Promoter architecture and the evolvability of gene expression. *J. Biol.* 8(11):95
56. Tkacik G, Gregor T, Bialek W. 2008. The role of input noise in transcriptional regulation. *PLoS ONE* 3(7):e2774
57. To T-L, Maheshri N. 2010. Noise can induce bimodality in positive transcriptional feedback loops without bistability. *Science* 327(5969):1142–45
58. Vilar JMG, Leibler S. 2003. DNA looping and physical constraints on transcription regulation. *J. Mol. Biol.* 331(5):981–89
59. Zenklusen D, Larson DR, Singer RH. 2008. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat. Struct. Mol. Biol.* 15(12):1263–71

57. Demonstrates that operator number increases gene expression noise in yeast cells.

Annual Review of Biophysics
Volume 42, 2013

Contents

Index

Errata

An online log of corrections to *Annual Review of Biophysics* articles may be found at http://biophys.annualreviews.org/errata.shtml