



Published in final edited form as:

Science. 2014 December 19; 346(6216): 1533–1536. doi:10.1126/science.1255301.

## Promoter architecture dictates cell-to-cell variability in gene expression

Daniel L. Jones<sup>1,\*</sup>, Robert C. Brewster<sup>1,2,\*</sup>, and Rob Phillips<sup>1,2,†</sup>

<sup>1</sup>Department of Applied Physics, California Institute of Technology, Pasadena, CA 91125, USA

<sup>2</sup>Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

### Abstract

Variability in gene expression among genetically identical cells has emerged as a central preoccupation in the study of gene regulation; however, a divide exists between the predictions of molecular models of prokaryotic transcriptional regulation and genome-wide experimental studies suggesting that this variability is indifferent to the underlying regulatory architecture. We constructed a set of promoters in *Escherichia coli* in which promoter strength, transcription factor binding strength, and transcription factor copy numbers are systematically varied, and used messenger RNA (mRNA) fluorescence in situ hybridization to observe how these changes affected variability in gene expression. Our parameter-free models predicted the observed variability; hence, the molecular details of transcription dictate variability in mRNA expression, and transcriptional noise is specifically tunable and thus represents an evolutionarily accessible phenotypic parameter.

The single-molecule events underlying gene expression, such as transcription factor binding and unbinding or RNA polymerase (RNAP) open complex formation, are inherently stochastic—a stochasticity inherited by gene expression itself. Over the past decade, theorists have sought to elucidate how changes in molecular kinetic parameters such as transcription factor binding and unbinding rates affect variability in expression (1, 2), whereas experimentalists have measured variability in gene expression at both the mRNA and protein level in prokaryotes and eukaryotes (3–6). Possible phenotypic consequences (4, 7–9) include the intriguing hypothesis that transcriptional noise may increase the fitness of microbial populations by providing phenotypic variability in a population of genetically identical cells (10, 11).

Models of transcription hinge on the molecular details of the promoter architecture (where “promoter architecture” refers collectively to the locations and strengths of transcription

<sup>†</sup>Corresponding author. phillips@pboc.caltech.edu.

<sup>\*</sup>These authors contributed equally to this work.

#### SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/346/6216/1533/suppl/DC1](http://www.sciencemag.org/content/346/6216/1533/suppl/DC1)

Materials and Methods

Supplementary Text

Figs. S1 to S11

Tables S1 to S3

References (24–32)

factor and RNAP binding sites governing a particular gene) and make quantitative predictions for the dependence of the variability on these details. For example, two extremely common promoter architectures (12) are shown schematically in Fig. 1A. Here, each rate parameter ( $r$ ,  $k_{\text{off}}^R$ ,  $k_{\text{on}}^R$ , and  $\gamma$ ) has a physical interpretation (Fig. 1B) as an element that can be tuned independently by genetic manipulation. The effect of promoter architecture on mean levels of gene expression is well established in prokaryotes, where thermodynamic models successfully predict gene expression as a function of promoter architecture (13–15). However, the associated predictions for how transcriptional noise depends on these parameters remain untested in any systematic way. In direct contrast to such models, some high-throughput experiments have culminated in the assertion that the cell-to-cell variability in gene expression is “universal,” dictated solely by the mean level of expression and insensitive to the details of the promoter driving the expression (3, 5, 6).

To confront this divide, we constructed a library of synthetic promoters driving a LacZ reporter in *E. coli* and measured the resulting mRNA copy number distributions using single-molecule mRNA fluorescence in situ hybridization (FISH) (16). Our approach ensures that differences in promoter sequence between constructs have clear interpretations in terms of the molecular parameters underlying transcription (e.g., transcription factor unbinding rate, basal transcription rate). This allows us to directly compare predictions of models incorporating those parameters with experimentally observed mRNA distributions, and hence to directly link the molecular events underlying transcription with observed variability in gene expression.

For the case of constitutive expression, shown schematically in Fig. 1A, mRNA transcripts are produced and degraded stochastically at rates  $r$  and  $\gamma$ , respectively, with constant probability per unit time. It can be shown (17) that the resulting steady-state mRNA copy number distribution is given by a Poisson distribution with mean  $r/\gamma$ . In the following experimental results, we use the Fano factor, defined as the variance divided by the mean, to characterize variability in gene expression. This metric reports the fold change in the squared coefficient of variation ( $\text{CV}^2 = \text{variance}/\text{mean}^2$ ) with respect to a Poisson process, for which  $\text{CV}_{\text{Poisson}}^2 = 1/\text{mean}$ ; hence,  $\text{CV}^2/\text{CV}_{\text{Poisson}}^2 = \text{variance}/\text{mean}$ . Therefore, the predicted Fano factor for constitutive expression equals 1 identically. However, this analysis is incomplete: The schematics of Fig. 1A represent the dynamics of the stochastic processes (transcription factor binding and unbinding, mRNA degradation, transcription initiation) that contribute to so-called “intrinsic” variability in gene expression, but do not account for the fact that rate parameters such as the repressor binding rate  $k_{\text{on}}^R$  and transcription rate  $r$  are themselves subject to fluctuations due to cell-to-cell variability in repressor and RNAP copy numbers, respectively. Such effects, collectively termed “extrinsic variability,” tend to increase the measured variability (18).

One important contribution to extrinsic noise comes from variability in gene copy number due to chromosome replication (Fig. 2A, bottom panel). It can be shown (16) that the effect of gene copy number variation on the variability in expression is independent and additive to the variability predicted from transcriptional noise, such that

$$\text{Fano} = \underbrace{\frac{\langle m^2 \rangle_1 - \langle m \rangle_1^2}{\langle m \rangle_1}}_{\text{Transcription}} + \underbrace{\frac{f(1-f)}{1+f} \langle m \rangle_1}_{\text{Gene copy number}} \quad (1)$$

where  $\langle m \rangle_1$  is the mean mRNA copy number from a single gene copy, and  $f$  is the fraction of the time a cell has two copies of the gene of interest. The first term is simply the promoter architecture-dependent Fano factor of a single copy of a gene, whereas the second term is the contribution due to gene copy number variation.

To quantitatively test the predictions of the model for constitutive expression, we measured the mRNA copy number distribution using mRNA FISH for 18 unique constitutive promoters (19). In Fig. 2B, we plot the Fano factor versus mean expression for each of this set of promoters (see fig. S9 for full mRNA copy number distributions for each promoter). The solid black line is the prediction resulting from consideration of intrinsic noise alone. The shaded regions represent the effects of what we believe are the three most important additional sources of noise (16). The green shaded region, quantization error, is the variability introduced by our measurement and analysis process. The red shaded region covers the expected contribution from cell-to-cell differences in RNAP copy number; the blue region is the expected contribution from gene copy number variation. Note that the red shaded region is an indirect estimate based on literature sources, whereas the blue and green regions are supported by direct measurements (16). The data and theoretical predictions are in good accord, implying that the dynamics of constitutive transcription are Poissonian with some additional extrinsic noise. In Fig. 2C, we plot the Fano factor minus the predicted gene copy number contribution and observe a quantitative disagreement between the measured noise in expression and the prediction of the “universal” noise model as reported in (5). But to conclusively demonstrate the architecture dependence of the variability, we need to look at alternative regulatory architectures.

To that end, we consider an architecture in which transcription can be blocked by a repressor transcription factor. As shown in Fig. 1A, the promoter transitions from the transcriptionally active (repressor unbound) to inactive state (repressor bound) at rate  $k_{\text{on}}^R$ , and from the inactive to active state at rate  $k_{\text{off}}^R$ . The predicted mean expression and Fano factor depend on each of these rates (Fig. 1A, right half); we can tune  $k_{\text{on}}^R$  by changing the concentration of repressor in the cell, and can tune  $k_{\text{off}}^R$  by changing the repressor binding site sequence. Note that the predicted relationship between the mean and the Fano factor has a characteristic form depending on which of these rates is being tuned (Fig. 3, A and B, dashed lines).

To test the predicted effect of changing  $k_{\text{on}}^R$ , we took two of the constitutive promoters described above and placed them under simple repression via a LacI Oid binding site immediately downstream of the promoter (16). The difference in transcription rate for the two constructs is reflected in different values of  $r/\gamma$ . At the same time, we introduced into our cells a genetic circuit enabling inducible control of LacI expression, effectively permitting systematic changes in repressor number. In Fig. 3A, we plot the measured Fano factor as a function of the mean expression over LacI concentrations ranging from ~0 to 50

LacI molecules per cell, for both promoters. In addition, we plot the zero-free-parameter theoretical prediction for the Fano factor as a function of mean using the measured value of  $r/\gamma$  from the constitutive data and the LacI unbinding rate from (20).

Similarly, we vary  $k_{\text{off}}^R$  by altering the sequence of the LacI binding site. Holding the RNAP binding site constant (and thus  $r/\gamma$  constant), we created constructs corresponding to four different LacI binding sites (16). At constant repressor concentration (i.e., constant  $k_{\text{on}}^R$ ), tuning mean expression by altering  $k_{\text{off}}^R$  is predicted to yield a characteristic curve, whereas different repressor concentrations (and hence  $k_{\text{on}}^R$  values) correspond to distinct instances of this curve. In Fig. 3B, we plot the Fano factor resulting from changing  $k_{\text{off}}^R$  at each of three different repressor concentrations. We find agreement in the trends between theory and experiment, although this agreement is less good than in the case of tuning  $k_{\text{on}}^R$ . One possible explanation [reported in (20)] is that changing transcription factor–DNA binding affinity affects the transcription factor–DNA association rate  $k_{\text{on}}^R$  as well as the dissociation rate  $k_{\text{off}}^R$ , contrary to our assumption that  $k_{\text{on}}^R$  is constant along each curve in Fig. 3B. However, the most important outcome of this set of measurements is a demonstration of the qualitatively distinct variability profile when a different set of transcriptional parameters are controlled, illustrating once again the systematic dependence of variability on promoter architecture.

We have shown that transcriptional noise is well predicted by molecularly detailed models for the two most common promoter architectures in *E. coli* as the various genetic knobs are tuned. This agreement is not the result of fitting theory curves to data, because the predicted curves are generated using physical parameter values reported elsewhere in the literature and in that sense are zero-parameter predictions. Earlier reports of “bursty” transcription (5, 21) are based on the observation that the Fano factor is greater than 1 for constitutive mRNA production (as well as direct kinetic measurements). Various explanatory hypotheses have been proposed, including transcriptional silencing via DNA condensation by nucleoid proteins (22), negative supercoiling induced by transcription, or the formation of long-lived “dead-end” initiation complexes (23). Although our data do not rule out these hypotheses, we find that extrinsic noise is sufficient to explain the deviation from Fano = 1 in our constitutive expression data (Fig. 2B). Thus, we find no need to invoke alternative hypotheses to explain the observed “burstiness” of constitutive transcription.

Many interesting earlier experiments make it difficult to interpret differences between promoters and induction conditions in terms of distinct physical parameters because of the wide variety of promoter architectures in play as well as the diverse mechanisms of induction. We have instead taken a “synthetic biology” approach of building promoters from the ground up. By directly controlling aspects of the promoter architecture, our goal has been to directly relate changes in promoter architecture to changes in observed gene expression variability. We believe that this work has demonstrated that mutations in regulatory DNA can alter gene expression noise. This suggests that gene expression noise may be a tunable property subject to evolutionary selection pressure, as mutations in regulatory DNA could provide greater fitness by increasing (or decreasing) variability.

Demonstrating the relevance of this hypothesis in natural environments remains an ongoing challenge.

## Supplementary Material

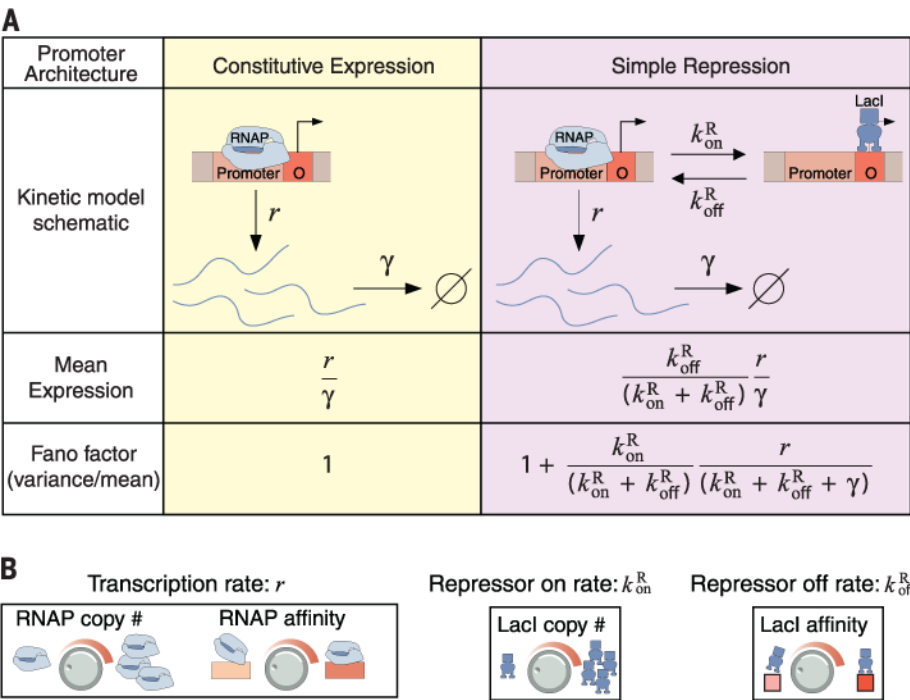
Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

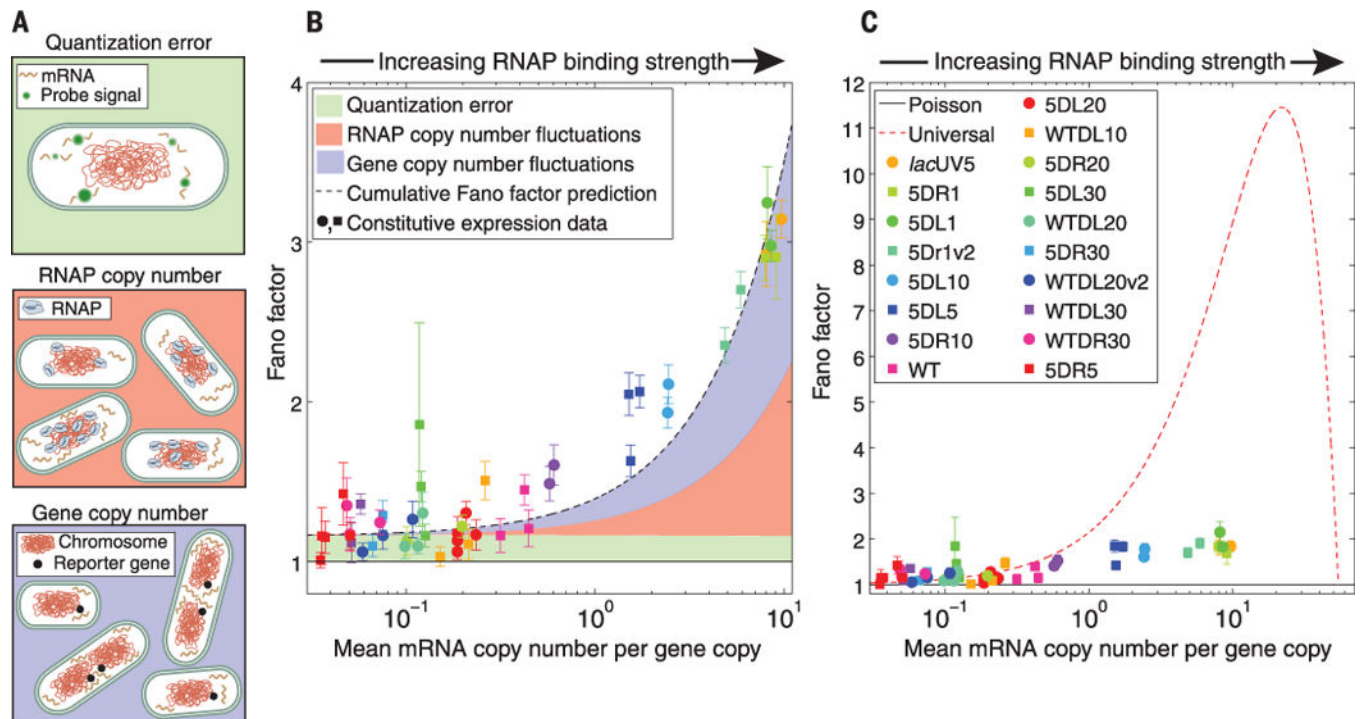
We thank H. J. Lee, C. Wiggins, Y. Lin, X. Zhu, F. Weinert, M. Rydenfelt, R. Milo, H. Garcia, N. Belliveau, and J. Sheung for useful discussions. Supported by NIH grants DP1 OD000217 (Directors Pioneer Award), R01 GM085286, and 1 U54 CA143869 (Northwestern PSOC Center); La Fondation Pierre Gilles de Gennes (R.P.); and the Donna and Benjamin M. Rosen Center for Bioengineering at Caltech (D.L.J.). Raw microscopy image data are archived in the Phillips laboratory at Caltech and are available upon request.

## REFERENCES AND NOTES

1. Paulsson J, Ehrenberg M. *Phys Rev Lett*. 2000; 84:5447–5450. [PubMed: 10990965]
2. Sanchez A, Garcia HG, Jones D, Phillips R, Kondev J. *PLOS Comput Biol*. 2011; 7:e1001100. [PubMed: 21390269]
3. Taniguchi Y, et al. *Science*. 2010; 329:533–538. [PubMed: 20671182]
4. Blake WJ, et al. *Mol Cell*. 2006; 24:853–865. [PubMed: 17189188]
5. So LH, et al. *Nat Genet*. 2011; 43:554–560. [PubMed: 21532574]
6. Salman H, et al. *Phys Rev Lett*. 2012; 108:238105. [PubMed: 23003996]
7. Maamar H, Raj A, Dubnau D. *Science*. 2007; 317:526–529. [PubMed: 17569828]
8. Eldar A, Elowitz MB. *Nature*. 2010; 467:167–173. [PubMed: 20829787]
9. Süel GM, Kulkarni RP, Dworkin J, Garcia-Ojalvo J, Elowitz MB. *Science*. 2007; 315:1716–1719. [PubMed: 17379809]
10. Thattai M, van Oudenaarden A. *Genetics*. 2004; 167:523–530. [PubMed: 15166174]
11. Kussell E, Leibler S. *Science*. 2005; 309:2075–2078. [PubMed: 16123265]
12. Salgado H, et al. *Nucleic Acids Res*. 2013; 41:D203–D213. [PubMed: 23203884]
13. Bintu L, et al. *Curr Opin Genet Dev*. 2005; 15:125–135. [PubMed: 15797195]
14. Kuhlman T, Zhang Z, Saier MH Jr, Hwa T. *Proc Natl Acad Sci USA*. 2007; 104:6043–6048. [PubMed: 17376875]
15. Vilar JM, Leibler S. *J Mol Biol*. 2003; 331:981–989. [PubMed: 12927535]
16. See supplementary materials on *Science* Online.
17. Sánchez A, Kondev J. *Proc Natl Acad Sci USA*. 2008; 105:5081–5086. [PubMed: 18353986]
18. Swain PS, Elowitz MB, Siggia ED. *Proc Natl Acad Sci USA*. 2002; 99:12795–12800. [PubMed: 12237400]
19. Brewster RC, Jones DL, Phillips R. *PLOS Comput Biol*. 2012; 8:e1002811. [PubMed: 23271961]
20. Hammar P, et al. *Science*. 2012; 336:1595–1598. [PubMed: 22723426]
21. Golding I, Paulsson J, Zawilski SM, Cox EC. *Cell*. 2005; 123:1025–1036. [PubMed: 16360033]
22. Sanchez A, Choubey S, Kondev J. *Annu Rev Biophys*. 2013; 42:469–491. [PubMed: 23527780]
23. Mitarai N, Dodd IB, Crooks MT, Sneppen K. *PLOS Comput Biol*. 2008; 4:e1000109. [PubMed: 18617999]

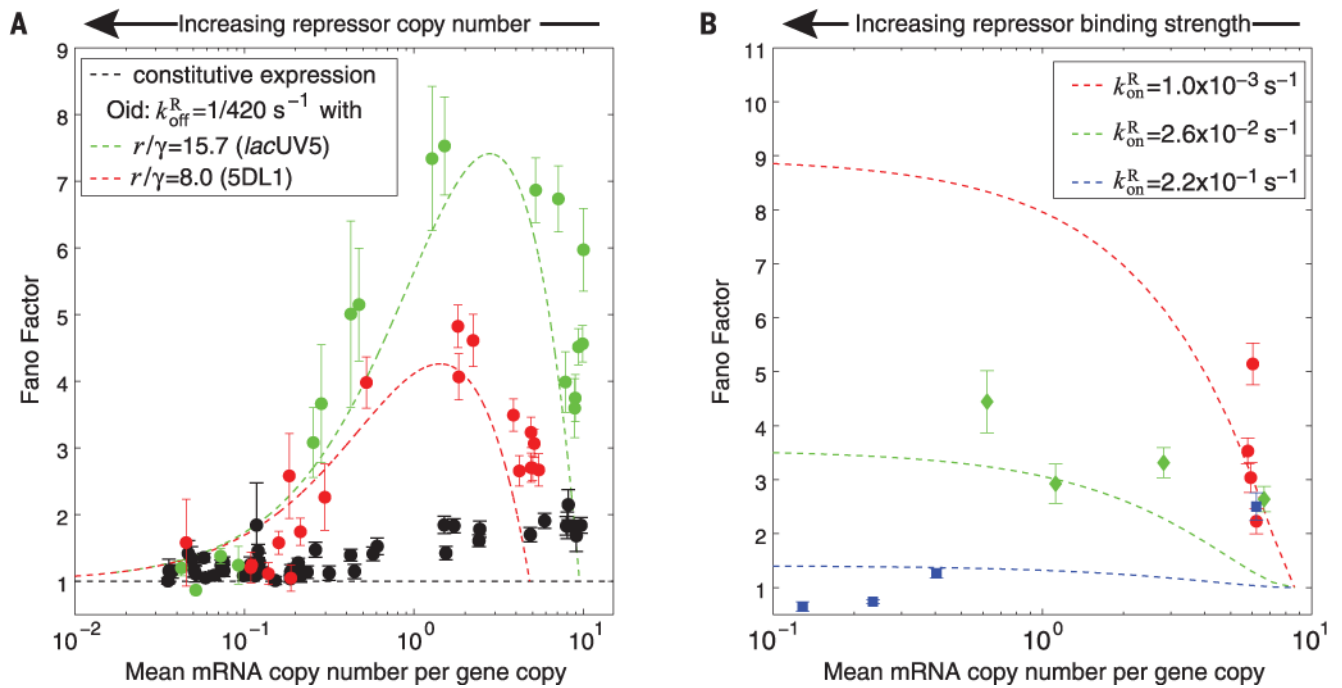


**Fig. 1. Schematics of the kinetics of transcription for two simple regulatory architectures**  
(**A**) Theoretical treatment of two common promoter architectures and the predicted expression (both mean and variability) as a function of the relevant rate parameters. (**B**) Examples of the experimental knobs available for tuning the various model rate parameters: Basal transcription rate  $r$  is tuned by RNAP copy number and RNAP binding site affinity (left); repressor binding rate  $k_{\text{on}}^R$  is tuned by repressor copy number (center); and repressor unbinding rate  $k_{\text{off}}^R$  is tuned by its binding site affinity (right).



**Fig. 2. Variability in gene expression for constitutive expression**

(A) Examples of additional noise sources (not accounted for in models of chemical kinetics) present in expression measurements. (B) Fano factor (gene copy number variation not subtracted) versus mean expression, plotted for each of 18 constitutive promoters along with estimates of the contributions shown schematically in (A). These factors can account for essentially the entirety of the deviation from Fano = 1. (C) Measured Fano factor for various promoters under constitutive expression, with gene copy number variation subtracted. For reference, the predictions of pure Poissonian production (black solid line) and the “universal noise” curve observed in (5) (red dashed) theories are shown. In (B) and (C), each strain is represented by a unique symbol, and each instance represents repeated measurements with error bars from bootstrap sampling expression measurements of individual cells.



**Fig. 3. Variability in gene expression for systematic tuning of repression**

(A) Fano factor versus mean mRNA copy number for two promoters (choices of  $r/\gamma$ ) while tuning  $k_{\text{on}}^R$  by inducing LacI to varying levels. For reference, the black data are the constitutive data from Fig. 2. (B) Fano factor versus mean mRNA copy number for *lacUV5* while tuning  $k_{\text{off}}^R$  by changing repressor binding site identity at fixed repressor copy number; each color represents a different induction condition from red (lowest LacI induction) to blue (highest LacI induction). For both (A) and (B), the parameter-free predictions from kinetic theory are shown as dashed lines in the corresponding color, holding promoter ( $r/\gamma$ ) and (A) repressor binding strength  $k_{\text{off}}^R$  or (B) repressor binding rate  $k_{\text{on}}^R$  constant. In both cases, the Fano factor at a given mean depends on the choice of molecular parameters and agrees with the expectations from theory. The effect of gene copy number variation was subtracted from all data points; error bars result from bootstrap sampling expression measurements of individual cells.