Disentangling intrinsic and extrinsic gene expression noise in growing cells

Jie Lin[1,2] and Ariel Amir[2]

[1] *Center for Quantitative Biology and Peking-Tsinghua Joint Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China*
[2] *John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA*

(Dated: January 30, 2021)

Gene expression is a stochastic process. Despite the increase of protein numbers in growing cells, the protein concentrations are often found to be confined within small ranges throughout the cell cycle. Generally, the noise in protein concentration can be decomposed into an intrinsic and an extrinsic component, where the former vanishes for high expression levels. Considering the time trajectory of protein concentration as a random walker in the concentration space, an effective restoring force (with a corresponding "spring constant") must exist to prevent the divergence of concentration due to random fluctuations. In this work, we prove that the magnitude of the effective spring constant is directly related to the fraction of intrinsic noise in the total protein concentration noise. We show that one can infer the magnitude of intrinsic, extrinsic, and measurement noises of gene expression solely based on time-resolved data of protein concentration, without any a priori knowledge of the underlying gene expression dynamics. We apply this method to experimental data of single-cell bacterial gene expression. The results allow us to estimate the average copy numbers and the translation burst parameters of the studied proteins.

INTRODUCTION

Gene expression in all forms of life is subject to noise [1–7]. Experimentally, stochastic gene expression has been intensively studied, mostly in growing cells with exponentially growing cell volume [8–12] in which the copy numbers of mRNAs and proteins in general double on average during the cell cycle, as widely observed in bacterial and eukaryotic cells [8, 13–15]. To reduce cell cycle effects, a more biologically relevant protocol to quantify the stochastic degree of gene expression is to calculate the variability of *concentration* because most genes in proliferating cells exhibit approximately constant protein concentrations throughout the cell cycle over multiple generations [13, 16–21]. In yeast and mammalian cells, most genes also exhibit approximately constant mRNA concentrations throughout the cell cycle [14, 22, 23].

Considering the time trajectory of protein concentration as a one dimensional random walker in the space of concentration, it must be subject to an effective restoring force to prevent the divergence of concentration in the long time limit (note that cell growth contributes to this restoring force via the effect of dilution, as discussed extensively in Ref. [19]). However, little is known about how the strength of this restoring force is related to the stochastic nature of protein concentration. In this work we show that one can in fact infer the contribution of intrinsic and extrinsic noise (which we will define later) to the total gene expression noise from the properties of the restoring force. Previous works on solving this challenge often rely on particular models of the underlying dynamics of gene expression [24–27]. Here we develop a novel protocol which is, in contrast, insensitive to many of the details of the gene expression dynamics, and is thus applicable to a broad class of models. The protocol only relies on analysis of time-series data of protein concentrations. We expect it to be applicable to exponentially growing cells such as bacteria, yeast and cancer cells [8–12].

In the following, we first introduce a general framework to study the variability of mRNA and protein concentrations in growing cells. Within the framework, the initiation rates of transcription and translation can be age-dependent (here, we define age as the elapsed time since cell birth), *e.g.*, due to gene dosage effects as well as more complex cell cycle dependencies [15]. We show that independent of the details of the gene expression dynamics, the variances of mRNA and protein concentrations can always be decomposed into an extrinsic component and an intrinsic component. In the large cell volume limit, the intrinsic noise vanishes while the extrinsic noise remains finite [28]. We then introduce our protocol to extract the fraction of intrinsic noise, extrinsic noise and measurement noise in the total noise of protein concentrations and finally apply the method to experimental data of bacterial gene expression.

Decomposition of noise.—For simplicity, we consider a cell growing exponentially at a constant growth rate μ with a constant doubling time $T = \ln(2)/\mu$, see Fig. 1. When the cell divides, the cell volume divides symmetrically, therefore the molecules are assumed to be segregated binomially and symmetrically between the two daughter cells [3]. Since for both bacterial and eukaryotic cells the degradation times of many proteins are longer than the cell cycle duration [29], we consider a non-degradable protein in the main text. Our results are equally valid for proteins with a finite degradation rate after some slight modifications (Supplementary Information, SI A) [30]. Our results are also robust against

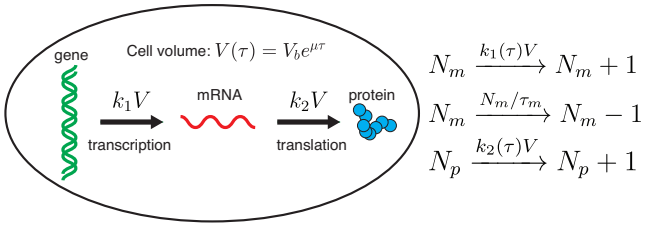


FIG. 1. The cell volume V grows exponentially in time with a growth rate μ and τ is the cell age. k_1 and k_2 are the transcription rate and translation rate per cell volume which can be age-dependent. The chemical reactions of gene expression are summarized on the right. N_m and N_p are the absolute mRNA and protein copy numbers respectively. τ_m is the lifetime of mRNA using which one can define the translation burst parameter $\beta\tau_m$ (the average number of proteins produced in the lifetime of a single mRNA).

fluctuating growth rates and doubling times as we show in SI B. We allow the initiation rates of transcription and translation per cell volume, k_1, k_2, to be time dependent and, for example, they can exhibit stochastic dynamics. One can further express $k_2 = \beta m$ where m is the mRNA concentration and β is the initiation rate of translation per mRNA. Mechanistically β is determined by the binding rate of ribosomes to mRNAs and largely determined by the concentration of ribosomes, which is roughly constant throughout the cell cycle [20].

Consider an experiment where one tracks a single lineage of cells over multiple generations, records the data of protein concentrations p uniformly in time with resolution Δt, and finally computes the resulting variance of concentrations based on all collected data. We find that the resulting variance of protein concentration σ_p^2 can be generally decomposed into three components (SI A):

$$\sigma_p^2 = \underbrace{\frac{\mathrm{cov}(k_2, p)}{\mu}}_{\text{Upstream noise}} + \underbrace{\left\langle \frac{k_2}{2\mu V} \right\rangle}_{\text{Poisson noise}} + \underbrace{\frac{\overline{p(T)}}{4\ln(2) V_b}}_{\text{Partitioning noise}} . \quad (1)$$

Here $\mathrm{cov}(k_2, p) = \langle k_2 p \rangle - \langle k_2 \rangle \langle p \rangle$ and $\langle \cdot\cdot \rangle$ represents average over time. The first part represents the noise due to a fluctuating upstream factor, namely, the initiation rate of translation per cell volume. One important source of upstream noise is the fluctuation in mRNA copy number [28]. The second term represents the noise due to the stochastic production process which we denote as Poisson noise here. The last term stems from the random partitioning during cell division where $T = \ln 2/\mu$ is the doubling time. The Poisson noise and the partitioning noise scale with the inverse of cell volume and their contributions to the square of the coefficient of variation (variance/mean^2) vanish for highly expressed proteins. In contrast, the upstream noise stems from the fluctuation in the translation rate per cell volume and it does not vanish in the large cell volume limit. We therefore define the sum of the Poisson noise and the partitioning noise as intrinsic and the upstream noise as extrinsic, consistent with previous works [28, 31]. We numerically confirm the validity of the noise decomposition for multiple gene expression dynamics including stochastic transcription and translation rate (SI B, Fig. S1).

We remark that the definition of extrinsic noise in our framework is different from the extrinsic noise inferred from the dual-reporter setup [1, 32], which is defined as the correlated noise of two identical genes controlled by the same promoters. The possible sources of extrinsic noise in the dual-reporter setup belong to a subset of those of the extrinsic noise in our framework which includes all possible upstream factors correlated or not across genes. Therefore, the extrinsic noise from the dual-reporter method is typically smaller than the extrinsic noise defined in our current framework, as we will discuss further later.

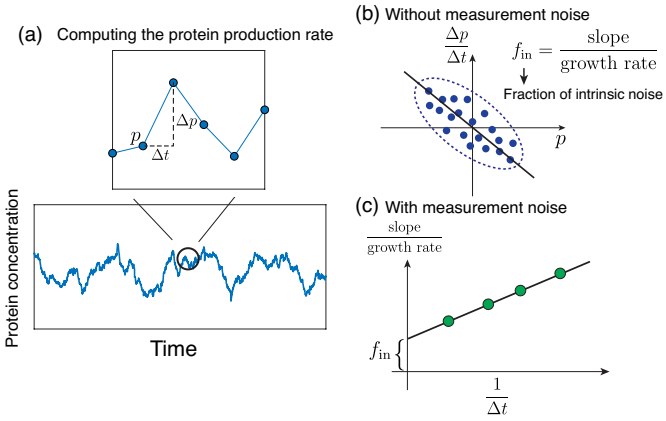


FIG. 2. (a) Given a time series of protein concentration, we first compute the discrete time derivative of protein concentration $\Delta p/\Delta t$ with a time interval Δt. (b) Next, we perform a linear fit of $\Delta p/\Delta t$ against the current protein concentration p and consider the absolute value of the fitted slope. In the case of negligible measurement noise, the fraction of intrinsic noise is the ratio between the slope and growth rate. (c) For experimental data with measurement noise, we compute $\Delta p/\Delta t$ for multiple time intervals Δt and repeat the protocol in (b) for each time interval. Finally, we perform a linear fit of the normalized slopes against $1/\Delta t$ and infer the fraction of intrinsic noise from the intercept.

Extracting the fraction of intrinsic and extrinsic noise.—In the following, we discuss a protocol to disentangle the contribution of intrinsic and extrinsic noise to the total noise based on the time trajectory of concentration [Fig. 2(a, b)]. We consider a discrete increment of protein concentration over a small time window, $\Delta p(t) = p(t + \Delta t) - p(t)$, which can be expressed as

$$\Delta p(t) = \frac{\Delta N_p(t)}{V(t)} - \mu p(t) \Delta t, \quad (2)$$

where $\Delta N_p(t)$ is a random variable from a Poisson distribution with mean $k_2(t) V \Delta t$ assumed constant within

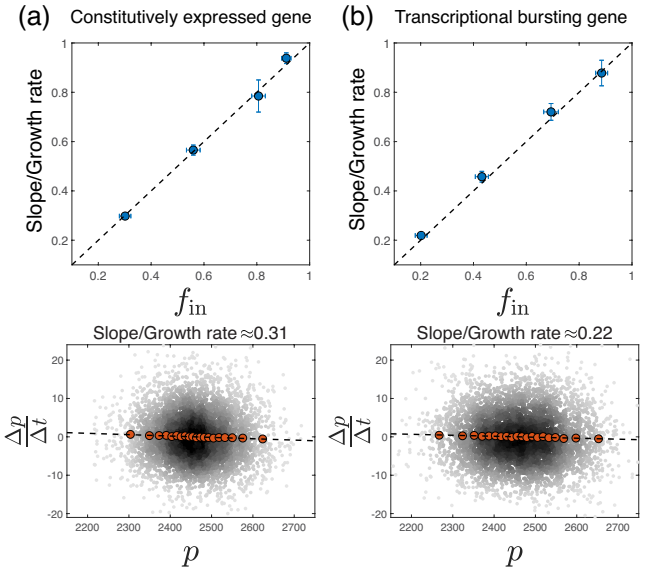


FIG. 3. (a) Simulation of a constitutively expressed gene. (Upper) We compare the predicted fraction of intrinsic noise (y axis) to the measured value (x axis). (Bottom) An example of the raw numerical data with the binned data shown as well (red circles). The dashed line is the linear fit of the raw data. The same analysis also applies to panel (b). Here $k_1 = 10$. (b) Simulation of a transcriptional bursting gene with $k_{\rm on} = 10$, $k_{\rm off} = 10$, $k_1 = 20$. In all upper panels, the doubling time $T = 60$, $\tau_m = 10$, and β is varied so that $\log_{10}\beta = -2, -1.5, -1, -0.5$. In all bottom panels, $\log_{10}\beta = -0.5$. We compute the time-derivative of protein concentration with a time interval $\Delta t = 0.5$. The errorbars are computed as the standard deviation of 5 independent simulations and in each simulation, 10^3 cell cycles are tracked.

the small time interval Δt. The second term on the right side arises from dilution due to cell growth. The protein concentration fluctuates but does not diverge in the long time limit, therefore we can make an analogy with a Brownian particle attracted to a fixed point with a linear restoring force equal to $-kx$ where k is the spring constant and x is the particle position relative to its equilibrium point. In the case of a Brownian particle, one can find the spring constant of the restoring force as the slope in the linear fitting of the discrete velocity $\Delta x/\Delta t$ *vs.* x. In the case of protein concentration, one can do a similar analysis by linearly fitting the discrete time derivative of protein concentration $\Delta p/\Delta t$ *vs.* p. Considering a least square linear fitting, the slope of the linear fitting is found to be

$$S \equiv -\frac{\mathrm{cov}\left(\frac{\Delta p(t)}{\Delta t}, p(t)\right)}{\sigma_p^2} = \mu - \frac{\mathrm{cov}(k_2(t), p(t))}{\sigma_p^2}. \qquad (3)$$

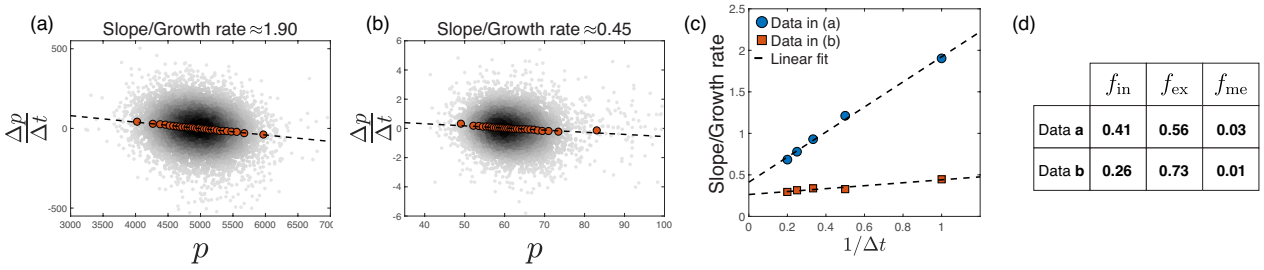
where we have used Eq. (2). If the covariance between the translation rate and protein concentration vanishes, the spring constant of the restoring force is simply the growth rate. Combined with Eq. (1), we find that the slope is proportional to the growth rate and the proportional constant is precisely the fraction of intrinsic noise in the total protein concentration noise variance:

$$S = \mu\left(1 - \frac{\mathrm{cov}(k_2(t), p(t))}{\mu\sigma_p^2}\right) = \mu f_{\rm in}. \qquad (4)$$

The above equation shows that we can extract the fraction of intrinsic noise $f_{\rm in}$ in the total noise by linearly fitting the time derivative of the protein concentration against the current protein concentration without any *a priori* knowledge of the underlying gene expression dynamics. Extrinsic noise reduces the slope in the linear fitting which precisely equals the growth rate μ in the absence of extrinsic noise. An extended discussion along with an intuitive argument on the effects of extrinsic noise based on a Langevin equation is provided in SI F. We remark that our protocols are also applicable to nongrowing cells with a constant cell volume given the lifetime of the studied protein is known (SI A).

Analysis of synthetic data.—We test Eq. (4) on synthetic data, first considering a constitutively expressed gene where the initiation rate of transcription per cell volume k_1 is constant as is the initiation rate of translation per mRNA β. This assumption corresponds to the case in which both RNA polymerase and ribosomes are limiting for gene expression, as discussed in detail in Ref. [19]. We compute $f_{\rm in}$ numerically using Eq. (1) and compare it with the prediction from Eq. (4), finding excellent agreement [Fig. 3(a)]. To test the robustness of our protocol, we also verify our theoretical results on various other gene expression dynamics: (1) the scenario of transcriptional bursting where a gene switches from "off" state to "on" state with rate $k_{\rm on}$ and vice versa with rate $k_{\rm off}$ [Fig. 3(b)]; (2) a gene with a constant transcription rate proportional to the gene number which doubles in the middle of the cell cycle [Fig. S2(a)]; this scenario corresponds to the situation when the gene copy number is the sole limiting factor of transcription [19]; (3) a gene with a transcription rate modulated throughout the cell cycle due to a finite period of DNA replication [Fig. S2(b), see details in SI E]; (4) a gene with a fluctuating transcription rate [Fig. S2(c)]; (5) a gene with a fluctuating translation rate per mRNA [Fig. S2(d)]. In all cases, the predicted fractions of intrinsic noise match the actual values well. We also find that in all cases increasing the translation rate per mRNA β increases the fraction of extrinsic noise as the effects of upstream noise are amplified, consistent with the analytical results of constitutively expressed genes (SI C, D). We have also confirmed the robustness of our results against the number of cell cycles sampled and the effects of fluctuating growth rates and division volumes (Fig. S3). Note that in the case of a fluctuating growth rate one also has to account for the correlation between the protein concentration and growth rate, as discussed in SI B.

In our framework the extrinsic noise is extracted from the time trajectory of the protein concentration of a sin-



	$f_{\rm in}$	$f_{\rm ex}$	$f_{\rm me}$
Data **a**	**0.41**	**0.56**	**0.03**
Data **b**	**0.26**	**0.73**	**0.01**

FIG. 4. (a) We compute the time derivative of protein concentration as a function of the current protein concentration using data from Ref. [33] and the measured slope normalized by the growth rate is 1.90. The time interval used is $\Delta t = 1$ min and the growth rate is $\mu = 0.0213$ min^{-1}. (b) We repeat the analysis using another data from Ref. [8] where the measured slope normalized by the growth rate is 0.45. Here $\Delta t = 1$ min and $\mu = 0.0327$ min^{-1}. (c) We adjust the time interval to compute the time derivative of protein concentration and compute the slope in the linear fit of $\Delta p/\Delta t$ *vs.* p. The normalized slope is linearly fitted as a function of the inverse of the time interval. The fraction of intrinsic noise in the total noise can be calculated from the intercept of the linear fit. We also infer the fraction of measurement noise in the total noise from the slope of the linear fit. (d) We summarize the calculated fractions of different noise for the two data sets. $f_{\rm in}$: the fraction of intrinsic noise. $f_{\rm ex}$: the fraction of extrinsic noise. $f_{\rm me}$: the fraction of measurement noise.

gle gene, which is distinct from that of the dual-reporter method. If the two genes in the dual-reporter setup share the same fluctuating translation rate $k_2(t)$, the two definitions of extrinsic noise will coincide [SI G, Fig. S6(a)]. However, if the correlated noise between the two genes is at the transcriptional level, the extrinsic noise inferred from the dual-reporter will be smaller than the one extracted from our protocol, which we confirm numerically [Fig. S6(b)].

Analysis of experimental data.—Experimentally, the measured protein concentration is always augmented by measurement noise. To model the effects of measurement noise, we assume the measured protein concentration at time t to equal

$$p(t) = p_0(t) + \eta(t) \tag{5}$$

where $p_0(t)$ is the actual protein concentration and $\eta(t)$ is the measurement noise term assumed uncorrelated between different measurements. We will revisit this assumption later on and show that the datasets we analyzed are consistent with it. The covariance between $\Delta p/\Delta t$ and p becomes $\mathrm{cov}(\frac{\Delta p}{\Delta t}, p) = \mathrm{cov}(\frac{\Delta p_0}{\Delta t}, p_0) - \sigma_\eta^2/\Delta t$. Compared with Eq. (4), the slope in the linear fitting of $\Delta p/\Delta t$ *vs.* p is modified to

$$S \equiv -\frac{\mathrm{cov}(\frac{\Delta p}{\Delta t}, p)}{\sigma_p^2} = \mu\big(f_{\rm in} + \frac{\sigma_\eta^2}{\mu\sigma_p^2\Delta t}\big). \tag{6}$$

We confirm Eq. (6) using numerical simulations with artificial measurement noise. In this case since σ_η^2 is assigned and $f_{\rm in}$ is known, we can directly compare the left and right sides of Eq. (6), obtaining good agreement (SI H, Fig. S7). Experimentally, the fluorescence level may not accurately reflect the instantaneous protein number due to a finite maturation time of the fluorescent protein. We have confirmed that the effects of a finite maturation time does not affect our results for experimentally relevant values of the maturation times [34] (SI I, Fig. S8).

We analyze two datasets of *E. coli* growth. In both, cells are exponentially growing and a fluorescent protein is constitutively expressed [8, 33]. A single lineage of cells is tracked for about 100 generations with cell volume and fluorescence level measured simultaneously. In both cases, the time interval between two consecutive data points is 1 min. To compute $f_{\rm in}$, we increase the time interval to compute $\Delta p/\Delta t$ and find the slopes in the linear fitting of $\Delta p/\Delta t$ *vs.* p for each time interval [see examples for $\Delta t = 1$ min in Fig. 4(a, b)]. We then linearly fit the resulting slopes as a function of $1/\Delta t$ [Fig. 2(c)] and the results agree well with the prediction of Eq. (6) [Fig. 4(c)]. Notably, this allows us to infer both $f_{\rm in}$ as the intercept of the linear fit, and the fraction of measurement noise from the slope. The results are summarized in Fig. 4(d). To justify the assumption of uncorrelated measurement noise, we show that the scaling with Δt in Eq. (6) is violated for correlated measurement noise (SI H, Fig. S7).

In this way we find that the ratio between the measurement noise and the total noise in the two data sets are respectively 17% and 10% in terms of their standard deviations, which are the square roots of the numbers in Figure 4(d). We can further use our analytic results for constitutively expressed genes as used in these experiments to estimate the average copy numbers of proteins at cell birth and the translation burst parameter $\beta\tau_m$ (see Eqs. S28, S29 in SI C) [31]. We find that $N_p \approx 230$ at cell birth, $\beta\tau_m \approx 1.37$ for Data in Fig. 4(a), and $N_p \approx 210$ at cell birth, $\beta\tau_m \approx 2.81$ for Data in Fig. 4(b). The differences between the two data sets are presumably due to the different strains and promoters. We note that if the normalization constant to convert the fluorescence level to protein number is known, one can also compute the partitioning noise based on Eq. (1) and the Poisson noise

as the remaining component of the intrinsic noise, which is confirmed using the synthetic data [Fig. S1(g)].

Summary and outlook.—In this work, we start from a general framework of stochastic gene expression in exponentially growing cells. Our approach allows us to take into account the cell growth and division explicitly and study the variability in protein *concentrations*, directly relevant to experiments on proliferating cells such as bacteria, yeast or cancer cells. We derive a broadly applicable decomposition of the protein concentration noise, finding that the total noise can be expressed as the sum of the noise due to upstream factors, the Poisson noise due to the random process of production and degradation, and the noise due to random partitioning during cell division. These results are independent of the underlying details of the particular dynamics of mRNA and protein synthesis. Given a time trajectory of protein concentration, one may linearly fit the discrete time derivative of protein concentration as a function of the protein concentration. We find that the slope of the fit, normalized by the growth rate, equals the fraction of intrinsic noise in the total protein concentration noise in the absence of measurement noise. We verify our theoretical framework on synthetic data of protein concentrations for genes with various underlying gene expression dynamics.

Importantly, we generalize our protocol to analyze experimental data of *E. coli* gene expression and show how a generalization of the method can simultaneously reveal the fraction of *measurement noise* in addition to that of intrinsic and extrinsic noise. Our framework predicts that the slope in the linear fitting of the time derivative of protein concentration *vs.* the current protein concentration has a linear dependence on the inverse of the time interval used to compute the time derivative, which agrees well with the experimental results. Assuming a model of a constitutively expressed protein as used in these experiments, our approach also allows us to infer the average copy numbers of proteins at cell birth as well as the translation burst parameter.

The generality of our approach and the agreement between experiments and theoretical predictions suggests that the method should be broadly applicable and will serve as a useful tool for gene expression analysis including mammalian cells and other non-microbial eukaryotes as long as a sufficient number of cell cycles are sampled. Our protocol to extract the intrinsic and extrinsic noise relies only on the time trajectory of protein concentration of a single gene, in contrast to the dual-reporter protocol which relies on measuring protein concentrations of two identical genes. Combing our method with the dual-reporter method, one can further decompose the extrinsic noise into correlated and uncorrelated components. Theoretically, our work paves the way to further studies on the nature of the widely-observed yet poorly understood extrinsic noise in gene expression.

We thank Ido Golding and Lydia Robert for useful discussions and feedback. We also thank the anonymous reviewers for their comments. A.A. was supported by NSF CAREER grant 1752024 and the Harvard Dean's Competitive Fund. A.A. and J.L. thank support from Harvard's MRSEC (DMR-1420570).

[1] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Science **297**, 1183 (2002).
[2] J. M. Raser and E. K. O'shea, Science **304**, 1811 (2004).
[3] I. Golding, J. Paulsson, S. M. Zawilski, and E. C. Cox, Cell **123**, 1025 (2005).
[4] J. R. Newman, S. Ghaemmaghami, J. Ihmels, D. K. Breslow, M. Noble, J. L. DeRisi, and J. S. Weissman, Nature **441**, 840 (2006).
[5] A. Raj and A. van Oudenaarden, Cell **135**, 216 (2008).
[6] Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie, Science **329**, 533 (2010).
[7] N. Eling, M. D. Morgan, and J. C. Marioni, Nature Reviews Genetics **20**, 536 (2019).
[8] P. Wang, L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. Wright, and S. Jun, Current Biology **20**, 1099 (2010).
[9] M. Godin, F. F. Delgado, S. Son, W. H. Grover, A. K. Bryan, A. Tzur, P. Jorgensen, K. Payer, A. D. Grossman, M. W. Kirschner, *et al.*, Nature Methods **7**, 387 (2010).
[10] M. Campos, I. V. Surovtsev, S. Kato, A. Paintdakhi, B. Beltran, S. E. Ebmeier, and C. Jacobs-Wagner, Cell **159**, 1433 (2014).
[11] S. Taheri-Araghi, S. Bradde, J. T. Sauls, N. S. Hill, P. A. Levin, J. Paulsson, M. Vergassola, and S. Jun, Current Biology **25**, 385 (2015).
[12] N. Cermak, S. Olcum, F. F. Delgado, S. C. Wasserman, K. R. Payer, M. A Murakami, S. M. Knudsen, R. J. Kimmerling, M. M. Stevens, Y. Kikuchi, A. Sandikci, M. Ogawa, V. Agache, F. Baleras, D. M. Weinstock, and S. R. Manalis, Nature Biotechnology **34**, 1052 (2016).
[13] H. A. Crissman and J. A. Steinkamp, The Journal of Cell Biology **59**, 766 (1973).
[14] O. Padovan-Merhar, G. P. Nair, A. G. Biaesch, A. Mayer, S. Scarfone, S. W. Foley, A. R. Wu, L. S. Churchman, A. Singh, and A. Raj, Molecular Cell **58**, 339 (2015).
[15] M. Wang, J. Zhang, H. Xu, and I. Golding, Nature microbiology **4**, 2118 (2019).
[16] S. Elliott and C. McLaughlin, Proceedings of the National Academy of Sciences **75**, 4384 (1978).
[17] N. Brenner, E. Braun, A. Yoney, L. Susman, J. Rotella, and H. Salman, The European Physical Journal E **38**, 1 (2015).
[18] N. Walker, P. Nghe, and S. J. Tans, BMC Biology **14**, 11 (2016).
[19] J. Lin and A. Amir, Nature Communications **9**, 4496 (2018).
[20] G. E. Neurohr, R. L. Terry, J. Lengefeld, M. Bonney, G. P. Brittingham, F. Moretto, T. P. Miettinen, L. P. Vaites, L. M. Soares, J. A. Paulo, *et al.*, Cell **176**, 1083 (2019).
[21] N. Nordholt, J. H. van Heerden, and F. J. Bruggeman, Current Biology **30**, 2238 (2020).
[22] H. Kempe, A. Schwabe, F. Crémazy, P. J. Verschure, and F. J. Bruggeman, Molecular Biology of the Cell **26**, 797

(2015).
[23] X.-M. Sun, A. Bowman, M. Priestman, F. Bertaux, A. Martinez-Segura, W. Tang, C. Whilding, D. Dormann, V. Shahrezaei, and S. Marguerat, Current Biology **30**, 1217 (2020).
[24] D. L. Jones, R. C. Brewster, and R. Phillips, Science **346**, 1533 (2014).
[25] R. D. Dar, B. S. Razooky, L. S. Weinberger, C. D. Cox, and M. L. Simpson, PLoS One **10**, 1 (2015).
[26] O. Lenive, P. D. Kirk, and M. P. Stumpf, BMC Systems Biology **10**, 81 (2016).
[27] M. Soltani, C. A. Vargas-Garcia, D. Antunes, and A. Singh, PLoS Comput Biol **12**, e1004972 (2016).
[28] J. Paulsson, Nature **427**, 415 (2004).
[29] R. Milo and R. Phillips, *Cell biology by the numbers* (Garland Science, 2015).
[30] See Supplemental Material [url] for detailed discussions, which includes Refs. [35–40].
[31] J. Paulsson, Physics of Life Reviews **2**, 157 (2005).
[32] P. S. Swain, M. B. Elowitz, and E. D. Siggia, Proceedings of the National Academy of Sciences **99**, 12795 (2002).
[33] Y. Tanouchi, A. Pai, H. Park, S. Huang, N. E. Buchler, and L. You, Scientific Data **4**, 170036 (2017).
[34] E. Balleza, J. M. Kim, and P. Cluzel, Nature Methods **15**, 47 (2018).
[35] D. Huh and J. Paulsson, Nature Genetics **43**, 95 (2011).
[36] E. Powell, Microbiology **15**, 492 (1956).
[37] J. Lin and A. Amir, Cell Systems **5**, 358 (2017).
[38] E. H. Simpson, Journal of the Royal Statistical Society: Series B (Methodological) **13**, 238 (1951).
[39] D. T. Gillespie, The Journal of Chemical Physics **113**, 297 (2000).
[40] J. Lin, M. Manhart, and A. Amir, Genetics **215** (2020).

Supplemental Material

Jie Lin
Center for Quantitative Biology and Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China and John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA

Ariel Amir
John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA
(Dated: February 10, 2021)

A. Derivation of the noise decomposition

For any chemical reaction in which the number of particle i changes $d_{\alpha,i}$ in reaction α with rate r_α [1], the time dependence of the covariance of the numbers of particle i and j is

$$\frac{d\mathrm{cov}(x_i,x_j)}{dt}=\sum_\alpha d_{\alpha,i}(\overline{x_j r_\alpha}-\overline{x_j}\ \overline{r_\alpha})+d_{\alpha,j}((\overline{x_i r_\alpha}-\overline{x_i}\ \overline{r_\alpha})+d_{\alpha,i}d_{\alpha,j}\overline{r_\alpha}. \tag{S1}$$

Here $\mathrm{cov}(x,y)=\overline{xy}-\overline{x}\ \overline{y}$. Based on the chemical reactions of gene expression we introduced in Fig. 1 of the main text,

$$N_m \xrightarrow{k_1 V} N_m+1, \tag{S2}$$

$$N_m \xrightarrow{N_m/\tau_m} N_m-1, \tag{S3}$$

$$N_p \xrightarrow{k_2 V} N_p+1, \tag{S4}$$

one can find the time-dependence of the variance of protein concentration using $p=N_p/V(\tau)$:

$$\frac{d\sigma_p^2(\tau)}{d\tau}=2\mathrm{cov}(k_2(\tau),p)+\frac{\overline{k_2}(\tau)}{V(\tau)}-2\mu\sigma_p^2(\tau). \tag{S5}$$

Here we use $\sigma_p^2(\tau)$ to represent the *instantaneous* variances of protein concentrations as functions of the age (where the variance is taken over an ensemble of cell cycles conditioned on the same age). The average $\overline{x}(\tau)$ is over different cell cycles conditioned on the same age, $\sigma_x^2(\tau)=\overline{x^2}(\tau)-\overline{x}^2(\tau)$, and $\mathrm{cov}(x,y)(\tau)=\overline{x(\tau)y(\tau)}-\overline{x}(\tau)\ \overline{y}(\tau)$. We integrate Eq. (S5) within a cell cycle from 0 to T and what is left on the left side of Eq. (S5) is simply proportional to $\sigma_p^2(T)-\sigma_p^2(0)$ where T is the doubling time. Because of the random partitioning of molecules, the variances of protein concentrations at cell division and at cell birth are related by $\sigma_p^2(0)=\sigma_p^2(T)+\overline{p(T)}/2V_b$ [1]. In the following, we will use $\langle\cdot\cdot\rangle_\tau$ to represent averaging over age uniformly within a cell cycle. Finally, we obtain the general expression of the variance of protein concentration:

$$\langle\sigma_p^2(\tau)\rangle_\tau=\underbrace{\left\langle\frac{\mathrm{cov}(k_2(\tau),p(\tau))}{\mu}\right\rangle_\tau}_{\text{Upstream noise}}+\underbrace{\left\langle\frac{\overline{k_2}(\tau)}{2\mu V}\right\rangle_\tau}_{\text{Poisson noise}}+\underbrace{\frac{\overline{p(T)}}{4\ln(2)V_b}}_{\text{Partitioning noise}}. \tag{S6}$$

Here $\langle\sigma_p^2(\tau)\rangle_\tau=\int_0^T\sigma_p^2(\tau)d\tau/T=\langle\overline{p^2}(\tau)\rangle_\tau-\langle\overline{p}^2(\tau)\rangle_\tau$ is the age-averaged $\sigma_p^2(\tau)$, and the first two terms on the right side of Eq. (S6) are defined similarly.

In practice, in order to use Eq. (S6) one needs to accurately know the age of each cell within the cell cycle. We will now derive a variant of Eq. (S6) where the variance is taken over all collected data points uniformly sampled in time. This version is easier to use on experimental data, and in the main text all of our analysis of synthetic and experimental data are performed using this "pooled average" rather than the age-average discussed above. For example, the protein variance would take the simple form $\sigma_p^2=\langle p^2\rangle-\langle p\rangle^2$ where $\langle\cdot\cdot\rangle$ represents average over all data uniformly sampled in time over multiple cell cycles (note that this protocol is not valid for data from a snapshot of a population of growing cells in which the age distribution is non-uniform [2, 3]).

Since uniform in time sampling of a trajectory can be thought of as integrating over age, this implies that $\langle \overline{f}(\tau)\rangle_\tau = \langle f\rangle$ for any random variable. This immediately tells us that the last two terms on the RHS of Eq. (S6) are identical in the age-average and the pooled average. The subtlety arises in the first term on the RHS and the LHS, where we shall shortly see the two averaging methods are distinct - and nonetheless the structure of the equation is maintained. Consider first the LHS. $\langle \sigma_p^2(\tau)\rangle_\tau = \langle \overline{p^2}(\tau)\rangle_\tau - \langle \overline{p}^2(\tau)\rangle_\tau$ while $\sigma_p^2 = \langle p^2\rangle - \langle p\rangle^2$. The first term clearly is identical by our previous logic (where the random variable considered is p^2). The second term is different though, and we find that $\sigma_p^2 = \langle \sigma_p^2(\tau)\rangle_\tau + \langle \overline{p}^2(\tau)\rangle_\tau - \langle \overline{p}(\tau)\rangle_\tau^2$. In a similar fashion we find that $\mathrm{cov}(k_2,p) = \langle \mathrm{cov}(k_2(\tau),p(\tau)\rangle_\tau + \langle \overline{k_2}(\tau)\overline{p}(\tau)\rangle_\tau/\mu - \langle \overline{k_2}(\tau)\rangle_\tau\langle \overline{p}(\tau)\rangle_\tau/\mu$ where $\mathrm{cov}(k_2,p) = \langle k_2 p\rangle - \langle k_2\rangle\langle p\rangle$ is calculated using the pooled average. According to Eq. (S4) and the exponential growth of cell volume, it is straightforward to find that $d\overline{p}(\tau)/d\tau = \overline{k_2}(\tau) - \mu\overline{p}(\tau)$ and therefore, $d\overline{p}(\tau)^2/d\tau = 2\overline{k_2}(\tau)\overline{p}(\tau) - 2\mu\overline{p}^2$. By integrating the time derivative of $\overline{p}(\tau)^2$ from 0 to T and using the boundary condition $\overline{p}(0) = \overline{p}(T)$, we find that $\langle \overline{k_2}(\tau)\overline{p}(\tau)\rangle_\tau/\mu = \langle \overline{p}(\tau)^2\rangle_\tau$. Similarly, it is straightforward to find that $\langle \overline{p}(\tau)\rangle_\tau = \langle \overline{k_2}(\tau)\rangle_\tau/\mu$. Therefore, Eq. (S6) is valid for the pooled average as well,

$$\sigma_p^2 = \underbrace{\frac{\mathrm{cov}(k_2,p)}{\mu}}_{\text{Upstream noise}} + \underbrace{\left\langle \frac{k_2}{2\mu V}\right\rangle}_{\text{Poisson noise}} + \underbrace{\frac{\overline{p(T)}}{4\ln(2)V_b}}_{\text{Partitioning noise}} . \tag{S7}$$

Note that Eq. (S6) and Eq. (S7) are equivalent for proteins with $\overline{p}(\tau)$ constant throughout the cell cycle and the above derivation is also valid for proteins with a finite lifetime.

Using Eq. (S1), we also obtain the time-dependence of the covariance of mRNA and protein concentrations, and the variance of mRNA concentration using $m = N_m/V(\tau)$, $p = N_p/V(\tau)$:

$$\frac{d\mathrm{cov}(m,p)(\tau)}{d\tau} = \mathrm{cov}(k_2(\tau),m) + \mathrm{cov}(k_1(\tau),p) - (2\mu + \frac{1}{\tau_m})\mathrm{cov}(m,p)(\tau), \tag{S8}$$

$$\frac{d\sigma_m^2(\tau)}{d\tau} = 2\mathrm{cov}(k_1(\tau),m) + \frac{\overline{m}}{\tau_m V(\tau)} + \frac{\overline{k_1}}{V(\tau)} - 2(\mu + \frac{1}{\tau_m})\sigma_m^2(\tau). \tag{S9}$$

Here we use $\sigma_m^2(\tau)$ to represent the *instantaneous* variance of mRNA concentration as a function of the age. The average $\overline{x}(\tau)$ is over different cell cycles conditioned on the same age, $\sigma_x^2(\tau) = \overline{x^2} - \overline{x}^2$, and $\mathrm{cov}(x,y)(\tau) = \overline{xy} - \overline{x}\ \overline{y}$.

Using a similar argument for mRNAs, we integrate Eq. (S9) from 0 to T within a cell cycle. Using the boundary condition of mRNA concentration, we obtain

$$\sigma_m^2 = \underbrace{\frac{\mathrm{cov}(k_1,m)}{\mu + \frac{1}{\tau_m}}}_{\text{Upstream noise}} + \underbrace{\left\langle \frac{\frac{m(t)}{\tau_m} + k_1(t)}{2(\mu + \frac{1}{\tau_m})V(t)}\right\rangle}_{\text{Poisson noise}} + \underbrace{\frac{\langle m(T)\rangle}{4(\mu + \frac{1}{\tau_m})V_b T}}_{\text{Partitioning noise}} . \tag{S10}$$

Similar logic leads to the following equation for proteins with a finite lifetime τ_p,

$$\sigma_p^2 = \underbrace{\frac{\mathrm{cov}(k_2,p)}{\mu + \frac{1}{\tau_p}}}_{\text{Upstream noise}} + \underbrace{\left\langle \frac{\frac{p(t)}{\tau_p} + k_2(t)}{2(\mu + \frac{1}{\tau_p})V(t)}\right\rangle}_{\text{Poisson noise}} + \underbrace{\frac{\langle p(T)\rangle}{4(\mu + \frac{1}{\tau_p})V_b T}}_{\text{Partitioning noise}} . \tag{S11}$$

Our protocol to extract the fraction of intrinsic noise is equally valid for a degradable protein if its lifetime τ_p is known. In this case, one should replace μ by $\mu + 1/\tau_p$ in Eqs. 4, 6 in the main text. If the cell does not grow, $\mu = 0$ and the intrinsic noise only includes the Poisson noise.

B. Numerical confirmation of the noise decomposition and extraction of intrinsic noise

We test the decomposition of noise for protein concentration, Eq. (1) in the main text, by simulating genes with various underlying dynamics. In all panels of Fig. S1, $T = 60$, $\tau_m = 10$, and $k_1 = 10$ if not specified. In Fig. S1(a), we simulate a constitutively expressed gene with constant transcription rate k_1 and constant translation rate per mRNA β. In Fig. S1(b), we simulate a gene with transcriptional bursting such that transcription only occurs in the "on" state. $k_{\mathrm{on}} = 10$, $k_{\mathrm{off}} = 10$, $k_1 = 20$. In Fig. S1(c), we simulate a gene with a transcription rate proportional to the gene copy number which doubles in the middle of the cell cycle. The transcription rate changes from k_1 to $2k_1$ in the middle of the cell cycle and k_1 is chosen such that the average mRNA number at cell birth is the same as

the constitutively expressed gene. In Fig. S1(d), we simulate a gene with a transcription rate that depends on the cell cycle due to a finite period of DNA replication (section E). In Fig. S1(e), we simulate a gene with a fluctuating transcription rate such that $k_1(t) = \langle k_1 \rangle + \xi_1(t)$ and the autocorrelation function of the noise decays exponentially in time, $\langle \xi_1(t)\xi_1(t')\rangle = A_1 \exp(-|t-t'|/\tau_1)$. We take $A_1 = 0.01k_1^2$ and $\tau_1 = T/2$ in the simulation where T is the cell cycle duration. In Fig. S1(f), we simulate a gene with a fluctuating translation rate per mRNA such that $\beta = \langle\beta\rangle + \xi_2(t)$ where $\langle \xi_2(t)\xi_2(t')\rangle = A_2 \exp(-|t-t'|/\tau_2)$. We take $A_2 = 0.01\langle\beta\rangle^2$ and $\tau_2 = T/2$ in the simulation. $\log_{10}\langle\beta\rangle = -1$ in the upper panel of Fig. S1(f). $\log_{10}\langle\beta\rangle = -2, -1.5, -1, -0.5$ in the bottom panel of Fig. S1(f). The same simulations discussed above are used to confirm the validity of extraction of intrinsic noise, Eq. (4) in the main text, shown in Fig. S2. In Figure S1(g), we consider an alternative method to decompose the noise directly from the data by first extracting the total intrinsic noise from the slope S in the linear fitting of $\Delta p/\Delta t$ *vs.* p, and then computing the partitioning noise using the last term of Eq. (S7). Since the sum of Poisson and partitioning noise is the total intrinsic noise, which is known, we can next infer the magnitude of the Poisson noise. This decomposition of noise agrees with our calculation of noise based on Eq. (S7), as shown in Figure S1(g).

In most of our simulations, 10^3 cell cycles are tracked. We have also tested the effects of the number of sampled cell cycles on the robustness of our methods. We find that as long as the number of sampled cell cycles is above or comparable to 50, the data statistics appears to suffice to test our methods [Fig. S3(a)].

Within our model we assume that the growth rate is constant and the cell division is symmetric. We also tested the effects of fluctuating growth rates and division volumes. Note that the above two noises both contribute to fluctuations in the generation times. To add noise in division volumes, we consider the adder model so that the division volume is $V_d = V_b + \Delta_v + \xi_V$ where Δ_v is constant and ξ_V is a Gaussian noise with a standard deviation equal to $0.1\Delta_v$ (similar to the value observed in nature for *E. coli* [6]). We find that our prediction for the fraction of intrinsic noise works well [Fig. S3(b)].

We also consider the scenario in which the growth rate fluctuates. We introduce a Gaussian noise in the growth rate so that the growth rate of each cell cycle can be different and we set the CV of growth rates as 0.1, motivated by the magnitude of the fluctuations in *E. coli* [6]. We considered the scenario that the growth rate fluctuation is uncorrelated with the translation rate per mRNA, as would be the case if the growth rate fluctuations are associated with those in the numbers of cell wall growth proteins. We find that in this case, the growth rate will be negatively correlated with the protein concentration as a higher growth rate dilutes the protein concentration faster. We extend our model to this scenario by taking the correlation between the growth rate and the protein concentration into account. In this scenario, the slope in the linear fit between the time derivative of the protein concentration and the current protein concentration becomes

$$S = -\frac{\mathrm{cov}(\frac{\Delta p}{\Delta t}, p)}{\sigma_p^2} = \frac{\mathrm{cov}(\mu p, p)}{\sigma_p^2} - \frac{\mathrm{cov}(k_2, p)}{\sigma_p^2} = \frac{\mathrm{cov}(\mu p, p)}{\sigma_p^2} - f_{\mathrm{ext}}, \tag{S12}$$

where f_{ext} is the fraction of extrinsic noise due to the upstream factors in the total noise of protein concentration. We confirm this prediction in Figure S3(c). Note that the first term on the right hand side can be directly extracted from experimental data where protein levels and cell volume are measured, hence our method for extracting the fraction of intrinsic and extrinsic noise from time-series data remains intact. Finally, we also checked the correlation between the growth rate and protein concentration for the experimental data and found that the Pearson correlation coefficients are very small for both data sets [Figure S3(d, e)]. This result suggests that our analysis of the experimental data in the main text remains valid.

C. Mathematical derivations for a constitutively expressed gene

Using the fact that k_1 is constant for a constitutively expressed gene, Eqs. (S9, S8, S5) are simplified to

$$\frac{d\sigma_m^2}{d\tau} = \frac{\langle m\rangle}{\tau_m V(\tau)} + \frac{k_1}{V(\tau)} - 2(\mu + \frac{1}{\tau_m})\sigma_m^2, \tag{S13}$$

$$\frac{d\mathrm{cov}(m,p)}{d\tau} = \beta\sigma_m^2 - (2\mu + \frac{1}{\tau_m})\mathrm{cov}(m,p), \tag{S14}$$

$$\frac{d\sigma_p^2}{d\tau} = 2\beta\mathrm{cov}(m,p) + \frac{\beta\langle m\rangle}{V(\tau)} - 2\mu\sigma_p^2. \tag{S15}$$

For a constitutively expressed gene, the instantaneous averaged mRNA concentration follows $d\overline{m}/dt = k_1\overline{m} - (\mu + 1/\tau_m)\overline{m}$ according to Eqs. (S2, S3), therefore $\overline{m}$ in the steady state is constant and equal to to the time-averaged

value $\langle m\rangle = k_1/(\mu + 1/\tau_m)$. We solve Eq. (S13) relying on the exponential growth of cell volume $V(\tau) = V_b e^{\mu\tau}$ and find

$$\sigma_m^2(\tau) = \sigma_m^2(0)e^{-2(\mu+\frac{1}{\tau_m})\tau} + \frac{k_1 + \frac{\langle m\rangle}{\tau_m}}{V_b}\frac{1}{\mu + \frac{2}{\tau_m}}(e^{-\mu\tau} - e^{-2(\mu+\frac{1}{\tau_m})\tau}). \tag{S16}$$

We now take $\tau = \log(2)/\mu$ and use the boundary condition $\sigma_m^2(0) - \sigma_m^2(T) = \langle m\rangle/(2V_b)$ to obtain

$$\sigma_m^2(T) = \sigma_m^2(0)2^{-2(1+\frac{1}{\mu\tau_m})} + \frac{k_1 + \frac{\langle m\rangle}{\tau_m}}{V_b}\frac{1}{\mu + \frac{2}{\tau_m}}\Big(\frac{1}{2} - \big(\frac{1}{2}\big)^{2+\frac{2}{\mu\tau_m}}\Big) = \sigma_m^2(0) - \frac{\langle m\rangle}{2V_b}, \tag{S17}$$

from which we obtain

$$\sigma_m^2(0) = \frac{\frac{k_1+\frac{\langle m\rangle}{\tau_m}}{V_b}\frac{1}{\mu+\frac{2}{\tau_m}}\Big(\frac{1}{2} - \big(\frac{1}{2}\big)^{2+\frac{2}{\mu\tau_m}}\Big) + \frac{\langle m\rangle}{2V_b}}{1 - 2^{-2(1+\frac{1}{\mu\tau_m})}}. \tag{S18}$$

Using the expression of $\langle m\rangle$, it is straightforward to verify that $\frac{k_1+\frac{\langle m\rangle}{\tau_m}}{\mu+\frac{2}{\tau_m}} = \langle m\rangle$, which leads to

$$\sigma_m^2(\tau) = \frac{\langle m\rangle}{V_b}e^{-\mu\tau}. \tag{S19}$$

Given the time dependence of $\sigma_m^2(\tau)$ we can further solve for $\mathrm{cov}(m,p)$ using Eq. (S14) and find the general solution as

$$\mathrm{cov}(m,p) = \mathrm{cov}(m,p)(\tau=0)e^{-(2\mu+\frac{1}{\tau_m})\tau} + \frac{\beta\langle m\rangle}{V_b}\frac{1}{\mu+\frac{1}{\tau_m}}\Big(e^{-\mu\tau} - e^{-(2\mu+\frac{1}{\tau_m})\tau}\Big). \tag{S20}$$

Using the boundary condition $\mathrm{cov}(m,p)(\tau=0) = \mathrm{cov}(m,p)(\tau=T)$, we find that

$$\mathrm{cov}(m,p)(\tau=0) = \frac{\beta\langle m\rangle}{V_b}\frac{1}{\mu+\frac{1}{\tau_m}}\frac{\frac{1}{2} - (\frac{1}{2})^{2+\frac{1}{\mu\tau_m}}}{1 - (\frac{1}{2})^{2+\frac{1}{\mu\tau_m}}}. \tag{S21}$$

Therefore,

$$\begin{aligned}\mathrm{cov}(m,p) &= \frac{\beta\langle m\rangle}{V_b}\frac{1}{\mu+\frac{1}{\tau_m}}\frac{\frac{1}{2} - (\frac{1}{2})^{2+\frac{1}{\mu\tau_m}}}{1 - (\frac{1}{2})^{2+\frac{1}{\mu\tau_m}}}e^{-(2\mu+\frac{1}{\tau_m})\tau} + \frac{\beta\langle m\rangle}{V_b}\frac{1}{\mu+\frac{1}{\tau_m}}\Big(e^{-\mu\tau} - e^{-(2\mu+\frac{1}{\tau_m})\tau}\Big) \\ &= -\frac{\beta\langle m\rangle}{V_b}\frac{1}{\mu+\frac{1}{\tau_m}}\frac{2^{1+\frac{1}{\mu\tau_m}}}{2^{2+\frac{1}{\mu\tau_m}} - 1}e^{-(2\mu+\frac{1}{\tau_m})\tau} + \frac{\beta\langle m\rangle}{V_b}\frac{1}{\mu+\frac{1}{\tau_m}}e^{-\mu\tau} \\ &= Ae^{-(2\mu+\frac{1}{\tau_m})\tau} + Be^{-\mu\tau},\end{aligned} \tag{S22}$$

with $A = -\frac{\beta\langle m\rangle}{V_b}\frac{1}{\mu+\frac{1}{\tau_m}}\frac{2^{1+\frac{1}{\mu\tau_m}}}{2^{2+\frac{1}{\mu\tau_m}}-1}$ and $B = \frac{\beta\langle m\rangle}{V_b}\frac{1}{\mu+\frac{1}{\tau_m}}$. Given the time dependence of $\mathrm{cov}(m,p)$ we can rewrite Eq. (S15) as

$$\frac{d\sigma_p^2}{dt} = 2\beta Ae^{-(2\mu+\frac{1}{\tau_m})\tau} + 2\beta Be^{-\mu\tau} + \frac{\beta\langle m\rangle}{V_b}e^{-\mu\tau} - 2\mu\sigma_p^2. \tag{S23}$$

Its general solution is

$$\sigma_p^2(\tau) = \sigma_p^2(0)e^{-2\mu\tau} + 2\beta A\tau_m(e^{-2\mu\tau} - e^{-(2\mu+\frac{1}{\tau_m})\tau}) + \frac{2\beta B + \frac{\beta\langle m\rangle}{V_b}}{\mu}(e^{-\mu\tau} - e^{-2\mu\tau}). \tag{S24}$$

We now take $\tau = \log(2)/\mu$ and use the boundary condition $\sigma_p^2(0) - \sigma_p^2(T) = \langle p\rangle/(2V_b)$ so that

$$\sigma_p^2(T) = \frac{\sigma_p^2(0)}{4} + 2\beta A\tau_m\Big(\frac{1}{4} - \big(\frac{1}{2}\big)^{2+\frac{1}{\mu\tau_m}}\Big) + \frac{2\beta B + \frac{\beta\langle m\rangle}{V_b}}{4\mu} = \sigma_p^2(0) - \frac{\langle p\rangle}{2V_b}. \tag{S25}$$

We find that

$$\sigma_p^2(0) = \frac{2\langle p\rangle}{3V_b} + \frac{2\beta A\tau_m}{3}\Big(1-\big(\frac{1}{2}\big)^{\frac{1}{\mu\tau_m}}\Big) + \frac{2\beta B + \frac{\beta\langle m\rangle}{V_b}}{3\mu} = \frac{\langle p\rangle}{V_b} + \frac{2\beta A\tau_m}{3}\Big(1-\big(\frac{1}{2}\big)^{\frac{1}{\mu\tau_m}}\Big) + \frac{2\beta B}{3\mu} \tag{S26}$$

where we have used $\langle p\rangle = \beta\langle m\rangle/\mu$.

We now compute the upstream noise, the Poisson noise and the partitioning noise for a constitutively expressed gene using Eq. (S7). In this case, the Poisson noise is equal to the partitioning noise and

$$\sigma^2_{p,\text{poisson}} = \sigma^2_{p,\text{partitioning}} = \frac{\langle p\rangle}{4\ln(2)V_b}. \tag{S27}$$

In the limit $\mu\tau_m \ll 1$, we find

$$\text{cov}(m,p) = \frac{\beta\langle m\rangle\tau_m}{V_b}e^{-\mu\tau}. \tag{S28}$$

Therefore, the upstream noise becomes

$$\sigma^2_{p,\text{upstream}} = \frac{\beta\int_0^T \text{cov}(m,p)dt}{T\mu} = \frac{\langle p\rangle}{2\ln(2)}\frac{\beta\tau_m}{V_b}. \tag{S29}$$

We can also rewrite Eqs. (S27, S29) in terms of CV^2 (variance/mean^2) as

$$\text{CV}^2_{\text{intrinsic}} = \frac{1}{2\ln(2)\langle N_{p,b}\rangle}, \tag{S30}$$

$$\text{CV}^2_{\text{extrinsic}} = \frac{\beta\tau_m}{2\ln(2)\langle N_{p,b}\rangle}, \tag{S31}$$

here $\langle N_{p,b}\rangle$ is the average protein number at cell birth. The above calculation can be easily generalized to a constitutively expressed protein with a finite lifetime τ_p. Eqs. (S14, S15) are modified as

$$\frac{d\text{cov}(m,p)}{d\tau} = \beta\sigma_m^2 - (2\mu + \frac{1}{\tau_m} + \frac{1}{\tau_p})\text{cov}(m,p), \tag{S32}$$

$$\frac{d\sigma_p^2}{d\tau} = 2\beta\text{cov}(m,p) + \frac{\beta\langle m\rangle}{V(\tau)} - 2(\mu + \frac{1}{\tau_p})\sigma_p^2. \tag{S33}$$

Therefore the solution of Eq. (S14) is valid for Eq. (S32) as well after replacing $1/\tau_m$ by $1/\tau_m + 1/\tau_p$. In the limit $\mu\tau_m \ll 1$, we find

$$\text{cov}(m,p) = \frac{\beta\langle m\rangle\tau_m\tau_p}{V_b(\tau_m+\tau_p)}e^{-\mu\tau}. \tag{S34}$$

The upstream noise becomes

$$\sigma^2_{p,\text{upstream}} = \frac{\beta\int_0^T \text{cov}(m,p)dt}{T(\mu+\frac{1}{\tau_p})} = \frac{\langle p\rangle}{2\ln(2)(1+\frac{1}{\mu\tau_p})}\frac{\beta\tau_m\tau_p}{V_b(\tau_m+\tau_p)}. \tag{S35}$$

Similarly, it is straightforward to find that in this case

$$\sigma^2_{p,\text{poisson}} = \frac{\langle p\rangle(1+\frac{2}{\mu\tau_p})}{4\ln(2)(1+\frac{1}{\mu\tau_p})V_b}, \tag{S36}$$

$$\sigma^2_{p,\text{partitioning}} = \frac{\langle p\rangle}{4\ln(2)(1+\frac{1}{\mu\tau_p})V_b}. \tag{S37}$$

D. Noise strength of a constitutively expressed gene

To quantify the noise strength of protein concentration, one can either use the CV^2 (variance/mean^2) or the Fano factor (variance/mean). While the CV^2 is dimensionless, the Fano factor has dimensions of concentration. To make the Fano factor of protein concentration dimensionless, one needs to multiply the Fano factor of protein concentration by an arbitrary volume scale $\tilde{V}$. A common way is to use the average cell volume $\overline{V}$ [7]. For uniform-time sampling, the average cell volume is simply $\overline{V} = \int_0^T V_b e^{\mu t} dt/T = V_b/\ln 2$. Assuming a non-degradable protein and $\mu\tau_m \ll 1$, we obtain

$$\mathrm{CV}^2 = \frac{1+\beta\tau_m}{2\ln(2)\langle p\rangle V_b}, \tag{S38}$$

$$\mathrm{Fano} = \frac{1+\beta\tau_m}{2\ln 2 V_b}\overline{V} = \frac{1+\beta\tau_m}{2(\ln 2)^2} \approx 1.04(1+\beta\tau_m). \tag{S39}$$

Experimentally, it is also common to sample the data by taking a snapshot of a population of cells. In this protocol, the distribution of age is non-uniform and decays exponentially, $P(\tau) = 2\mu e^{-\mu\tau}$. The average cell volume is $\overline{V} = \int_{V_b}^{2V_b} V \frac{2}{V^2} dV = 2\ln 2 V_b$. Using Eq. (S24), we find that

$$\mathrm{CV}^2 = \frac{13}{18}\frac{\frac{27}{26}+\beta\tau_m}{\langle p\rangle V_b} \approx 0.72\frac{1.04+\kappa_2\tau_1}{\langle p\rangle V_b}, \tag{S40}$$

$$\mathrm{Fano} = \frac{13}{18}\frac{\frac{27}{26}+\beta\tau_m}{V_b}\overline{V} \approx 1.04+\beta\tau_m. \tag{S41}$$

Interestingly, the Fano factors derived here are very close to the Fano factor $1+\beta\tau_m$ one would obtain from the constant cell volume model [8].

E. Cell cycle with a finite DNA replication period

We relax the assumption of instantaneous DNA replication in the main text to take into account the effects of a finite DNA replication period on the transcription rate per cell volume k_1. We assume a doubling time of 60 mins. The gene is constitutively expressed and replicated in the middle of the cell cycle at $\tau = 30$ mins. DNA replication starts from $\tau = 20$ mins with a duration 20 mins. Because of the competition between genes for the limiting resource such as RNA polymerase [9, 10], the transcription rate of the gene under consideration decreases during DNA replication with a jump right after the gene is duplicated (Fig. S4). Our predictions regarding the decomposition of the noise of protein concentration and its magnitude [Eq. (1) in the main text], and the extraction of intrinsic noise [Eq. (4) in the main text] are nicely confirmed in this case as well [Fig. S1(d) and Fig. S2(b)].

F. Simplified model based on the Langevin equation

To better understand the effects of extrinsic noise, we consider a simplified version of the model in the main text by taking a time interval such that $k_2 V\delta t \gg 1$ but also small enough that the change of k_2 is negligible. On such a time scale, the chemical reaction Eq. (S4) can be approximated by a discrete Langevin equation [12]:

$$\delta p = (k_2 - \mu p)\delta t + \sqrt{\frac{k_2\delta t}{V}}W. \tag{S42}$$

Here the noise stemming from the random production becomes white noise where W is a Gaussian random variable with variance 1. For simplicity, we have neglected the partitioning noise. Since the time interval δt can be arbitrarily small in the large volume limit, the upstream noise in the translation rate k_2 should be considered as a continuous random variable (i.e., it has a finite correlation time). Finally, we obtain the following continuous Langevin equation as an approximation for the dynamics of protein concentration:

$$\dot{x} = -\mu x + \eta + \xi, \tag{S43}$$

where x is the deviation of the protein concentration from its average. ξ is a white noise term such that $\langle\xi(t)\xi(t')\rangle = 2D_\xi\delta(t-t')$. η corresponds to the upstream noise in the translation rate which is a continuous random variable. For concreteness, we assume the autocorrelation function of η decays exponentially so that $\langle\eta(t)\eta(t')\rangle = A_\eta e^{-|t-t'|/\tau_\eta}$.

Before we move on to the theoretical analysis of the above equation, we first present an intuitive argument why the extrinsic noise reduces the slope in the linear fitting of $\Delta x/\Delta t$ vs. x. Consider the limit $\tau_\eta \gg 1/\mu$ so that η can be approximated as a constant for many generations of cell cycle. For a constant η the slope is simply the growth rate μ. Combining multiple sets of data with different η's, it is evident that the slope becomes smaller than μ [Fig. S5(a)]. Note that this effect is essentially due to "Simpson's paradox", where correlations can dramatically change when pooling together different sub-populations [11, 13]. While this argument is based on the assumption of a very slow time dependence of η, the conclusion is generally valid as we show in the following.

From Eq. (S43), the solution of x can be formally written as

$$x = \int_0^t e^{-\mu(t-t')}[\eta(t') + \xi(t')]dt', \tag{S44}$$

where the term associated with the initial condition is neglected since we are interested in the steady state. It is straightforward to find the variance of x as

$$\sigma_x^2 = \frac{A_\eta \tau_\eta}{\mu}\frac{1}{1+\mu\tau_\eta} + \frac{D_\xi}{\mu}, \tag{S45}$$

where the first term can be considered as the upstream noise in the main text and the second term can be considered as the intrinsic noise. We now calculate the covariance between $\frac{\Delta x}{\Delta t}$ and x:

$$\mathrm{cov}(\frac{\Delta x}{\Delta t}, x) = -\mu\sigma_x^2 + \mathrm{cov}(\eta, x). \tag{S46}$$

Therefore, the slope (S) in the linear regression of $\frac{\Delta x}{\Delta t}$ *vs.* x becomes

$$S \equiv -\frac{\mathrm{cov}(\frac{\Delta x}{\Delta t}, x)}{\sigma_x^2} = \mu - \frac{\mathrm{cov}(\eta, x)}{\sigma_x^2}. \tag{S47}$$

We now calculate $\mathrm{cov}(\eta, x)$

$$\mathrm{cov}(\eta, x) = \int_0^t e^{-\mu(t-t')}\langle\eta(t)\eta(t')\rangle dt' = \frac{A_\eta\tau_\eta}{1+\mu\tau_\eta}.$$

Combining with Eq. (S45, S47), we find that the relative slope indeed tells us the relative fraction of intrinsic noise in the total noise

$$\frac{S}{\mu} = \frac{D_\xi/\mu}{\sigma_x^2}. \tag{S48}$$

Note that when $D_\eta > 0$ and $D_\xi = 0$ the noise of x is all coming from η and $S = 0$, while the variance of x is still finite. The above calculation is valid even in the limit $\tau_\eta \ll 1/\mu$, which is beyond the simple argument assuming $\tau_\eta \gg 1/\mu$. We numerically test this situation and find a zero slope in the linear fitting of $\Delta x/\Delta t$ *v.s.* x [Fig. S5(b)].

G. Comparison with the dual-reporter setup

As we discuss in the main text, the extrinsic noise inferred from the dual-reporter setup is in general smaller than the one inferred from the time trajectory of protein concentration based on our protocol. We first consider two identical genes that share the same fluctuating translation rates $k_2(t)$ and compute the uncorrelated noise as $\sigma^2_{p,\mathrm{uncorrelated}} = \langle(p_1(t) - p_2(t))^2\rangle/2$ (which is referred to as the intrinsic noise in the dual-reporter setup). We then compute the fraction of correlated noise as $f_{\mathrm{ex,\ dual}} = (\sigma_p^2 - \sigma^2_{p,\mathrm{uncorrelated}})/\sigma_p^2$ (which is referred to as the extrinsic noise in the dual-reporter setup). We compare it with the one inferred from the slope in the linear fitting of $\Delta p/\Delta t$ *vs.* p, $f_{\mathrm{ex,\ slope}}$ and find excellent agreement [Fig. S6(a)].

We also consider two identical genes that share the same fluctuating transcription rates $k_1(t)$. In this scenario, the translation rates $k_2(t)$ of the two genes are correlated but not identical due to the randomness of mRNA production and degradation. Therefore, $f_{\mathrm{ex,\ dual}} < f_{\mathrm{ex,\ slope}}$ as we confirm numerically [Fig. S6(b)].

H. Numerical simulation with artificial measurement noise

We simulate a constitutively expressed gene and add measurement noise to the protein concentration. We first consider the case of uncorrelated measurement noise. We vary the time interval Δt to compute $\Delta p/\Delta t$ and compare the measured slopes in the linear fit of $\Delta p/\Delta t$ *vs.* p with the theoretical prediction, Eq. (6) in the main text. The prediction is nicely confirmed [Fig. S7]. We also consider the case of correlated noise and assume the autocorrelation function of the measurement noise decays exponentially in time with a decay time τ_η. In this case, the simulation results do not agree with Eq. (6) in the main text. The agreement of Eq. (6) in the main text and the experimental data therefore supports our assumption of uncorrelated measurement noise.

I. Effects of finite maturation time of fluorescent protein

We consider the effects of a finite maturation time of fluorescent protein to our theoretical predictions. We will consider two models for the maturation process: a Poisson process, following Ref. [14], and later on, a model in which the maturation process takes a constant (fixed) time. For the former, we modify Eqs. 1-3 in the main text and define N_u as the number of immature fluorescent proteins and N_f as the number of maturated fluorescent proteins:

$$N_m \xrightarrow{k_1 V} N_m + 1, \tag{S49}$$

$$N_m \xrightarrow{N_m/\tau_m} N_m - 1, \tag{S50}$$

$$N_u \xrightarrow{k_2 V} N_u + 1, \tag{S51}$$

$$N_u, N_f \xrightarrow{RN_u} N_u + 1, N_f - 1. \tag{S52}$$

Here R is the maturation rate. In experiments, what one can actually measure is the concentration of maturated fluorescent proteins, f. We compute the inferred fraction of intrinsic noise from the maturated fluorescent proteins as

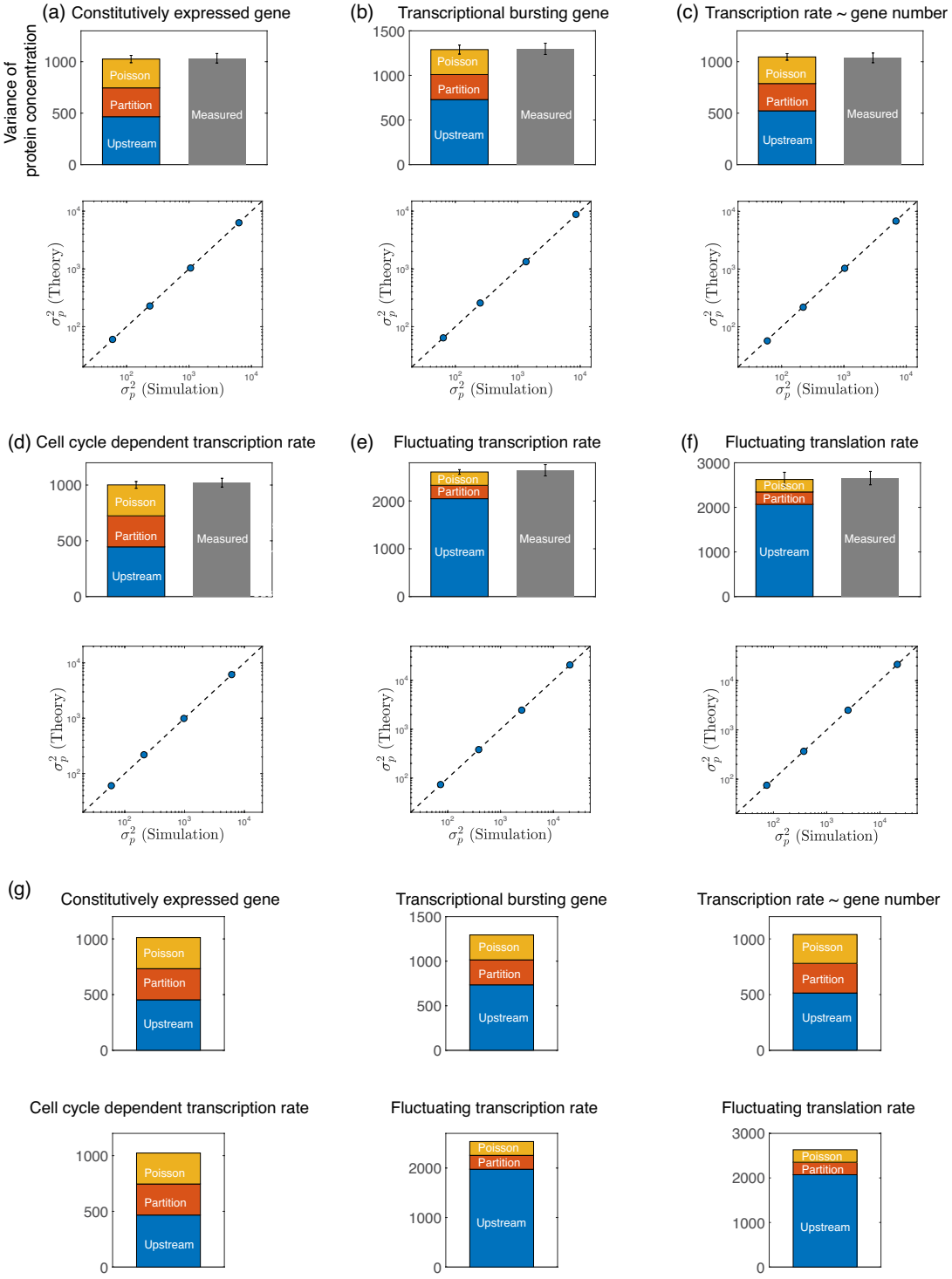
$$\frac{S_f}{\mu} \equiv -\frac{\mathrm{cov}\left(\frac{\Delta f(t)}{\Delta t}, f(t)\right)}{\mu \sigma^2_{f,\mathrm{time}}}, \tag{S53}$$

and compare it to the result of Eq. (3) in the main text using the concentration of total protein p. We compare the two slopes (S_f *vs.* S) using synthetic data where both can be accessed. As expected, when the maturation rate is large (corresponding to a maturation time short compared with the cell cycle duration), they are approximately equal and as the maturation rate decreases, the fraction of intrinsic noise inferred from the maturated fluorescent protein deviates from its true value, and is larger than it [Fig. S8(a)]. Typical maturation times range from several minutes to several tens of minutes [14]. The results of Fig. S8(a) therefore suggest that our protocol should provide accurate results for the majority of experimental scenarios.

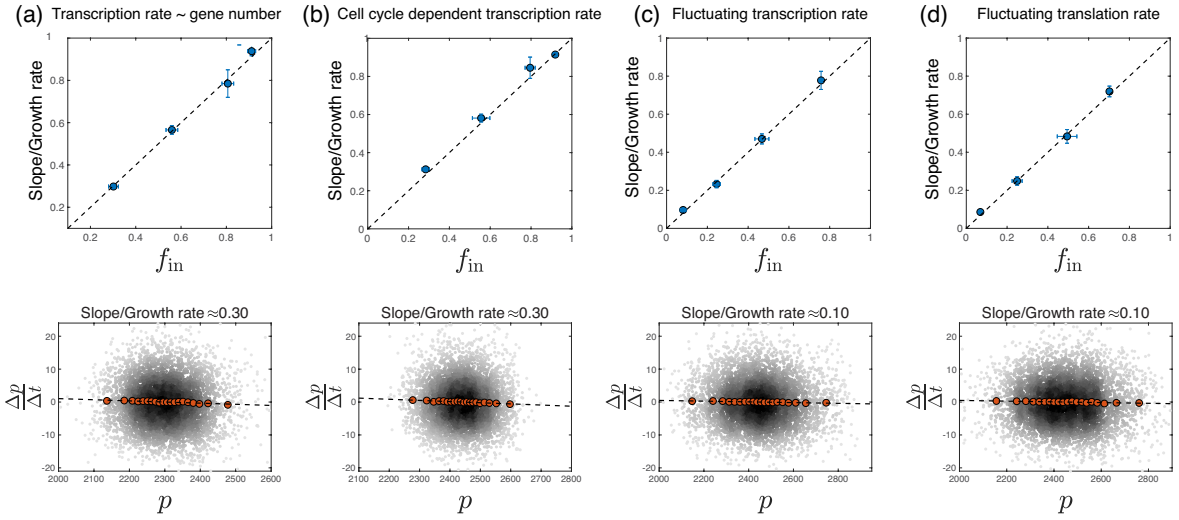
We also consider an alternative model of maturation process, assuming that the fluorescent proteins maturate after a fixed amount of time τ [Fig. S8(b)]. We find that in this case, the inferred fractions of intrinsic noise obtained using the maturated fluorescent protein levels match closely the true values even for large maturation times. Taken together with the results for the Poisson model of protein maturation, we conclude that maturation times should not significantly affect our analysis.

[1] D. Huh and J. Paulsson, Nature Genetics **43**, 95 (2011).
[2] E. Powell, Microbiology **15**, 492 (1956).
[3] J. Lin and A. Amir, Cell Systems **5**, 358 (2017).
[4] Y. Tanouchi, A. Pai, H. Park, S. Huang, N. E. Buchler, and L. You, Scientific Data **4**, 170036 (2017).
[5] P. Wang, L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. Wright, and S. Jun, Current Biology **20**, 1099 (2010).
[6] S. Taheri-Araghi, S. Bradde, J. T. Sauls, N. S. Hill, P. A. Levin, J. Paulsson, M. Vergassola, and S. Jun, Current Biology **25**, 385 (2015).
[7] Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie, Science **329**, 533 (2010).
[8] J. Paulsson, Physics of Life Reviews **2**, 157 (2005).
[9] J. Lin and A. Amir, Nature Communications **9**, 4496 (2018).
[10] X.-M. Sun, A. Bowman, M. Priestman, F. Bertaux, A. Martinez-Segura, W. Tang, C. Whilding, D. Dormann, V. Shahrezaei, and S. Marguerat, Current Biology **30**, 1217 (2020).

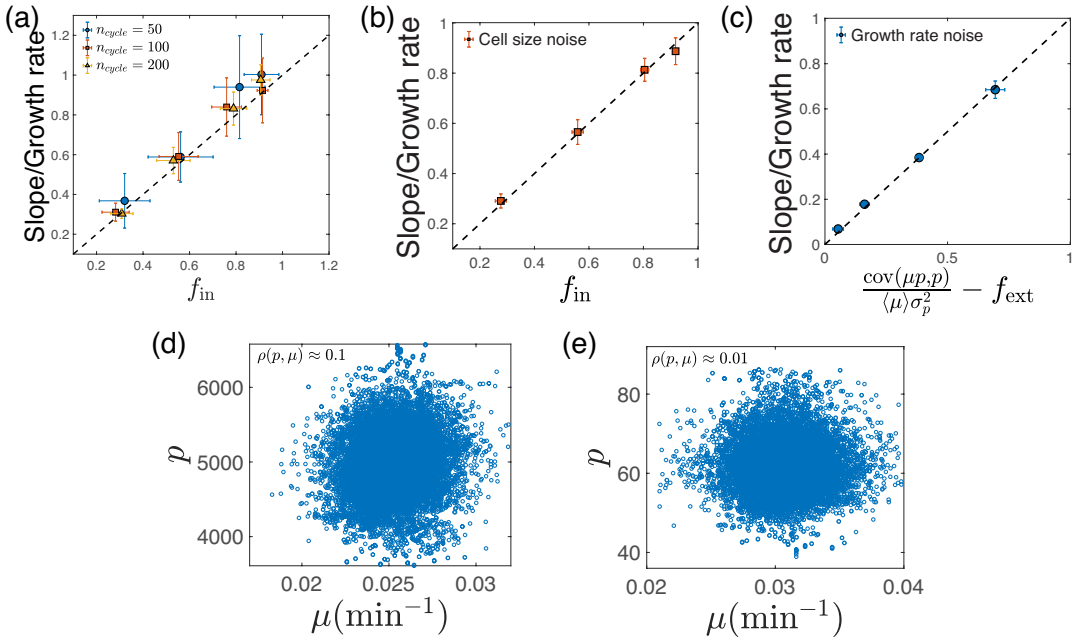
[11] E. H. Simpson, Journal of the Royal Statistical Society: Series B (Methodological) **13**, 238 (1951).
[12] D. T. Gillespie, The Journal of Chemical Physics **113**, 297 (2000).
[13] J. Lin, M. Manhart, and A. Amir, Genetics **215** (2020).
[14] E. Balleza, J. M. Kim, and P. Cluzel, Nature Methods **15**, 47 (2018).



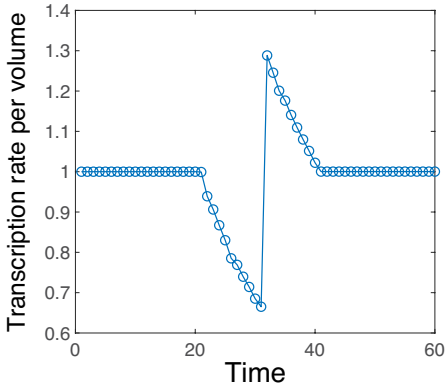
Supplementary Figure S1 Decomposition of the noise of protein concentration.
(a) Simulation of a constitutively expressed gene. (Upper) Total measured noise of protein concentration and the three sources of noise, which are calculated using Eq. (S7). (Bottom) The translation rate per mRNA β is varied and the predicted variance of protein concentration (Eq. (1) in the main text) is compared with the measured value. The same analysis applies to the following panels. (b) Simulation of a transcriptional bursting gene. (c) Simulation of a scenario where the transcription rate is proportional to the gene copy number. (d) Simulation of a gene with transcription rate modulated throughout the cell cycle. (e) Simulation of a gene with a fluctuating transcription rate such that $k_1(t) = \langle k_1 \rangle + \xi_1(t)$ where $\xi_1(t)$ is the noise term. (f) Simulation of a gene with a fluctuating translation rate per mRNA such that $\beta(t) = \langle \beta \rangle + \xi_2(t)$ where $\xi_2(t)$ is the noise term. In all panels, $T = 60$, $\tau_m = 10$, and $k_1 = 10$ if not specified. In all upper panels, $\beta = 0.1$ and in all bottom panels, β is varied so that $\log_{10} \beta = -2, -1.5, -1, -0.5$. Other simulation details are explained in the text. The errorbars are computed as the population standard deviation of 5 independent simulations. (g) An alternative decomposition of noise. We first compute the total intrinsic noise from the slope in the linear fitting of $\Delta p / \Delta t$ *vs.* p. We then compute the partitioning noise from Eq. (S7), which immediately informs us of the contribution of the Poisson noise as well. The results quantitatively agree with those shown in Fig. S1(a)-(f).



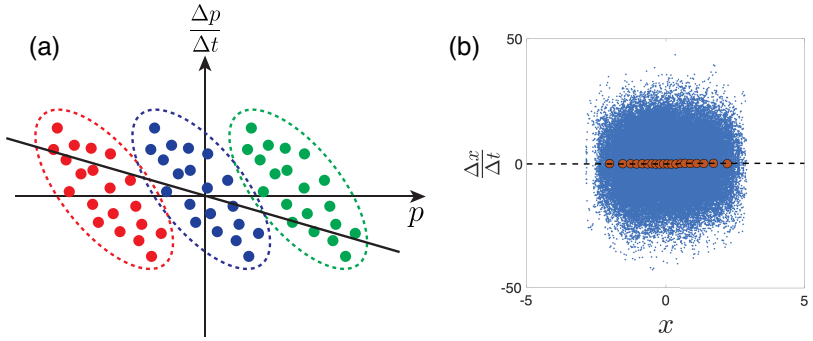
Supplementary Figure S2 Extraction of the fraction of intrinsic noise based on synthetic data. (a) Simulation of a scenario where the transcription rate is proportional to the gene copy number. (b) Simulation of a gene with transcription rate modulated throughout the cell cycle. (c) Simulation of a gene with a fluctuating transcription rate $k_1(t)$. (d) Simulation of a gene with a fluctuating translation rate per mRNA $\beta(t)$. In all upper panels, β is varied so that $\log_{10}\beta = -2, -1.5, -1, -0.5$. In all bottom panels, $\log_{10}\beta = -0.5$. We compute the time-derivative of protein concentration with a time interval $\Delta t = 0.5$. Other simulation details are explained in Methods. The errorbars are computed as the population standard deviation of 5 independent simulations.



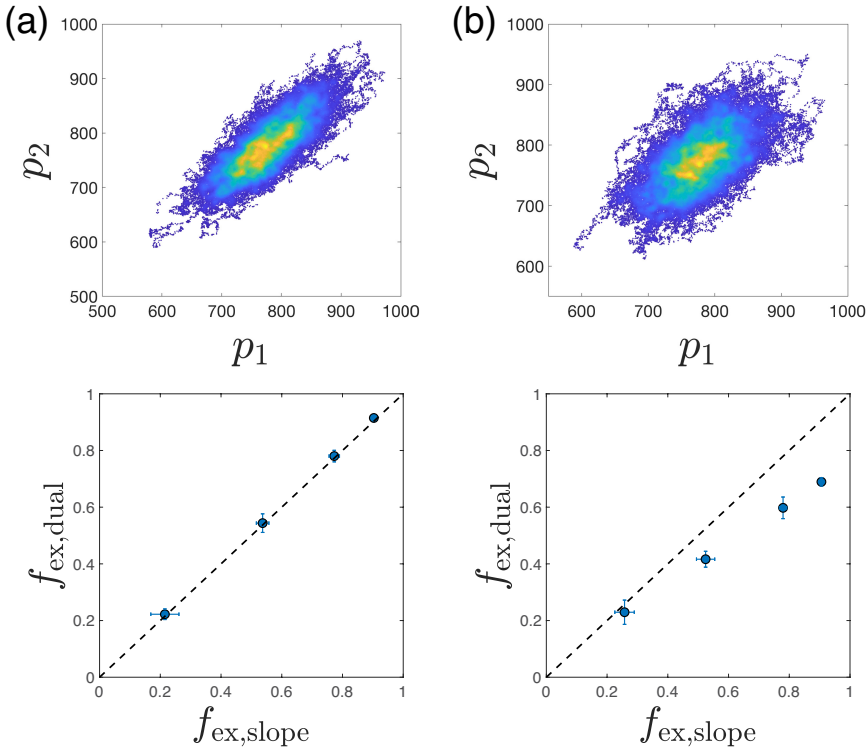
Supplementary Figure S3 Effects of finite number of cell cycles, and effects of fluctuating growth rates and division volumes.
(a) A constitutively expressed gene is simulated, with the results averaged over 10 independent simulations. In each simulation, $n_{\rm cycle}$ cell cycles are tracked. We compare the predicted fraction of intrinsic noise (y axis) to the measured value (x axis). The errorbars are computed as the population standard deviation of 10 independent simulations. (b) Simulation of a constitutively expressed gene where fluctuations in division volumes are considered. The errorbars are computed as the population standard deviation of 5 independent simulations. (c) Simulation of a constitutively expressed gene where fluctuations in growth rates are considered. The errorbars are computed as the population standard deviation of 5 independent simulations. The results agree well with Eq. (S12). (d, e) We find small Pearson correlation coefficients $\rho(\mu, p)$ between the growth rate and the protein concentration for both data sets. (d) is for data from Ref. [4]. (e) is for data from Ref. [5]. For each cell cycle, we fit the cell size as an exponential function of time to find the growth rate.



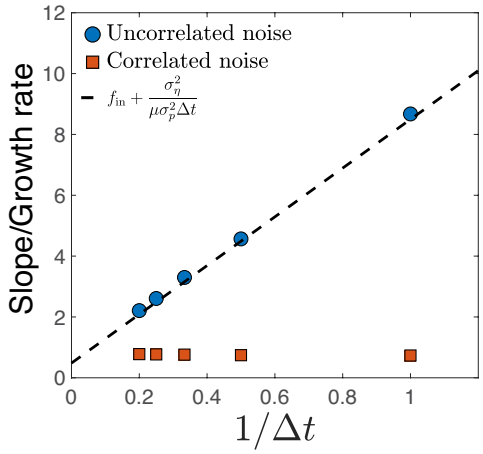
Supplementary Figure S4 Numerical simulation of a cell cycle with a finite DNA replication period.
The time dependence of the transcription rate per cell volume k_1.



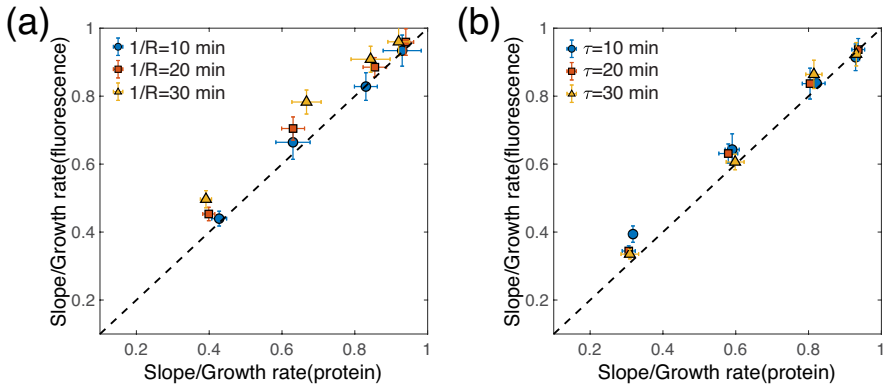
Supplementary Figure S5 Simplified model based on the Langevin equation.
(a) Schematic illustration of the lowering of the slope due to extrinsic noise; In essence, this is due to Simpson's paradox [11]. (b) Numerical test of the simplified model. Here, $\mu = 1$, $\tau_\eta = 0.01$, $A_\eta = 100$, $D_\xi = 0$ and the time interval of the simulation $\Delta t = 0.001$. The red circles are binned data and the dashed line has a zero slope.



Supplementary Figure S6 Numerical test of dual-reporter setup.
(a) We simulate two identical genes that share the same fluctuating translation rate $k_2(t)$ and set the translation rate per mRNA as $\beta = \langle\beta\rangle + \xi_2(t)$ where $\xi_2(t)$ is the noise term. We assume $\langle\xi_2(t)\xi_2(t')\rangle = A_2 \exp(-|t-t'|/\tau_2)$ with $A_2 = 0.01\beta^2$ and $\tau_2 = T/2$. $T = 60$, $k_1 = 10$ and $\tau_m = 10$. (Upper) We show the raw synthetic data from simulations of two identical genes with $\langle\beta\rangle = 0.1$. (Bottom) We compare the fractions of extrinsic noise inferred from the dual-reporter setup and from the slope in the linear fitting of $\Delta p/\Delta t$ *vs.* p. β is varied so that $\log_{10}\langle\beta\rangle = -2, -1.5, -1, -0.5$. The errorbars are computed as the population standard deviation of 5 independent simulations. (b) We simulate two identical genes that share the same fluctuating transcription rate $k_1(t)$ and set the transcription rate as $k_1(t) = \langle k_1\rangle + \xi_1(t)$ where $\xi_1(t)$ is the noise term. We assume $\langle\xi_1(t)\xi_1(t')\rangle = A_1 \exp(-|t-t'|/\tau_1)$ with $A_1 = 0.01k_1^2$ and $\tau_1 = T/2$. $T = 60$, $\langle k_1\rangle = 10$ and $\tau_m = 10$. (Upper) we show the original data of the two identical genes with $\beta = 0.1$. (Bottom) β is varied so that $\log_{10}\beta = -2, -1.5, -1, -0.5$. The errorbars are computed as the population standard deviation of 5 independent simulations.



Supplementary Figure S7 Numerical test of synthetic data with artificial measurement noise. We simulate a constitutively expressed gene and add Gaussian noise to every recorded protein concentration (circles). $T = 60$, $k_1 = 10$, $\beta = 0.1$, $\tau_m = 10$. The time interval Δt is varied from 1 to 5 and for each time interval we compute the slope from the linear fitting of $\Delta p/\Delta t$ *vs.* p. The dashed line is the theoretical prediction (Eq. (6) in the main text). $f_{\rm in} \approx 0.50$, $\sigma_\eta^2 = 100$, and $\sigma_p^2 \approx 1.14 \times 10^3$. We repeat the analysis for correlated measurement noise (squares). The autocorrelation function of the measurement noise decays exponentially with a decay time $\tau_\eta = T/2$ with other parameters kept the same as uncorrelated noise.



Supplementary Figure S8 Numerical simulation of a fluorescent protein with a finite maturation time. (a) We simulate a constitutively expressed gene with $T = 30$ min, $\tau_m = 10$ min, $k_1 = 10$ min^{-1} and adjust β so that $\log_{10}\beta = -2, -1.5, -1, -0.5$ min^{-1}. We model the maturation process as a Poisson process and compare the inferred fractions of intrinsic noise (Eq. (3) in the main text) and Eq. (S53) here for three different maturation rates. The dashed line is the $y = x$ line. The errorbars are computed as the population standard deviation of 5 independent simulations. (b) The same analysis for a maturation process with a fixed maturation time τ. The errorbars are computed as the population standard deviation of 5 independent simulations.