Homeostasis of protein and mRNA concentrations in growing cells

Jie Lin[1] and Ariel Amir[1]

[1] *School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA*

Many experiments show that the concentrations of protein and mRNA fluctuate but on average are constant in growing cells, independent of the genome copy number. However, models of stochastic gene expression often assume constant gene expression rates that are proportional to the gene copy numbers, which are therefore incompatible with experiments. Here, we construct a minimal gene expression model to fill this gap. We show that (1) because the ribosomes translate all proteins, the concentrations of proteins are regulated in an exponentially growing cell volume; (2) the competition between genes for the RNA polymerases makes the gene expression rate independent of the genome number; (3) the fluctuations in ribosome level and cell density can generate a global extrinsic noise in protein concentrations; (4) correlations between mRNA and protein levels can be quantified.

Despite the noisy nature of gene expression [1–6], various aspects of single cell dynamics, such as volume growth, are effectively deterministic. Recent single-cell measurements show that the growth of cell volumes is often exponential. These include bacteria [7–10], budding yeast [10–14] and mammalian cells [10, 15]. Moreover, the mRNA and protein numbers are on average proportional to the cell volume throughout the cell cycle [16–20]. Therefore, the homeostasis of mRNA concentration and protein concentration is maintained in an exponentially growing cell volume. The exponential growths of mRNA and protein number indicate dynamical transcription and translation rates proportional to the cell volume, and also independent of the genome copy number. However, current gene expression models often assume a fixed cell volume with constant transcription and translation rates [1, 5, 21–23], which are proportional to the gene copy number. Therefore, fixed cell volume models fail to reveal how cells keep constant mRNA and protein concentrations as the cell volume grows and the genome is replicated.

The homeostasis of protein and mRNA concentrations imply that there must be a regulatory mechanism in place to prevent the accumulation of noise over time and to maintain a bounded distribution of concentrations. The goal of this work is to identify such a mechanism by developing a genome-wide coarse-grained model taking into account explicitly cell volume growth. We will consider an idealized cell in which genes are constitutively expressed for simplicity. The ubiquity of homeostasis suggests that the global machineries of gene expression, RNA polymerases (RNAPs) and ribosomes, should play a central role within the model. Indeed, as we will show later, the exponential growth of cell volume, protein and mRNA number originates from the auto-catalytic nature of ribosomes, the limiting factor in the translational process [24–26]. The bounded distributions of concentrations are a result of the fact that ribosomes make all proteins. The independence of the mRNA concentration of the genome copy number is a natural result of the limiting nature of RNAP in the transcriptional process in which genes are competing for this global resource. Furthermore, we attempt to identify candidates for the global extrinsic noise, which sets the lower bound of noise in protein concentrations. Within our model, the only two possibilities are the fluctuations in ribosome levels and cell volume growth rate. We show that these two mechanisms lead to distinct correlation patterns between protein levels, therefore providing a method to determine the dominant contribution to the global extrinsic noise from experimental data.

In the following, we first introduce the model and derive the exponential growth of mRNA and protein numbers. We consider the scenario wherein the cell volume is proportional to the total mass of proteins, based on the fact that cellular dry mass density is well regulated [27]. We show that this model produces mRNA and protein concentrations fluctuating around a constant value throughout the cell cycle with bounded fluctuations. We verify our theoretical results using numerical simulations. We next study the possible sources of global extrinsic noise that affect all proteins. We show that the fluctuation in the the ribosome level can generate the global extrinsic noise, and the ribosome concentration is negatively correlated with constitutively expressed proteins – providing a method to experimentally test the relative contribution of ribosome level fluctuations to the global extrinsic noise. We also show that within a more refined model in which cell volume is determined by the activity of e.g., cell wall proteins, the volume growth rate fluctuations can also generate a systematic global extrinsic noise. Other potential sources of global extrinsic noise are considered, e.g. finite duration of DNA replication, the fluctuations in RNAP number, mRNA degradation time, yet none of them can generate a significant global extrinsic noise.

Importantly, our model allows us to distinguish the number from the concentration since we consider the cell volume explicitly. In Ref. [28], the reported correlation coefficients between mRNA and protein expression levels are much smaller than the predictions of fixed cell volume models [29], which has been a puzzle in the field of

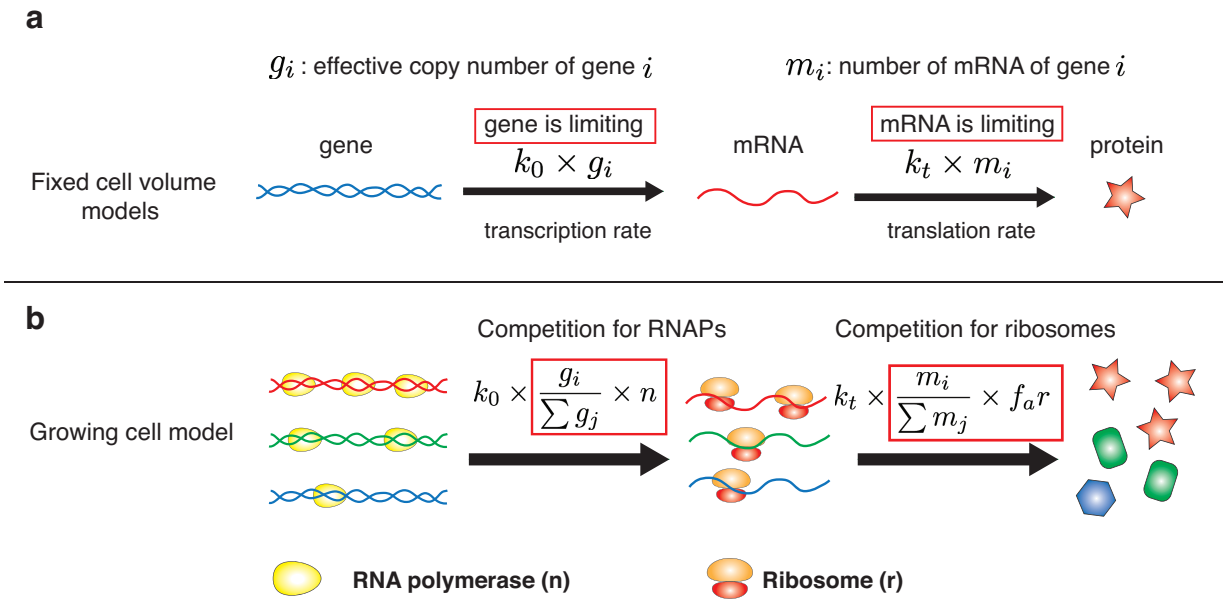

Figure 1. **The growing cell model of stochastic gene expression in comparison with fixed cell volume models.** (**a**) In fixed cell volume models, the transcription rate is proportional to the gene copy number, and the translation rate is proportional to the mRNA number. These assumptions imply that the gene number and mRNA number are the limiting factors in gene expression.
(**b**) In the genome-wide model (growing cell model), we introduce as limiting factors RNA polymerases (RNAP) and ribosomes. Genes with different colors are transcribed with different rates. Here k_0 is a constant and the gene regulation is coarse-grained into the gene allocation fraction $\phi_i = g_i / \sum_j g_j$. g_i is the effective copy number of gene i (also accounting for the promoter strength). n is the total number of RNA polymerases. Translation rates of mRNA depend on the number of active ribosomes ($f_a r$), the translation rate k_t, and the fraction of mRNA i in the total pool of mRNA.

stochastic gene expression. We show that the small correlations between mRNA and protein are due to the different volume-normalizations protocols used for mRNA and protein, thus reconciling the apparent discrepancy.

RESULTS

Model of stochastic gene expression

In fixed cell volume models, one usually focuses on one single gene, the transcription rate is proportional to the gene copy number, and the translation rate is proportional to the number of mRNAs [30–32]. Fixed cell volume models imply that the gene number is the limiting factor in the transcription and the mRNA number is the limiting factor in the translation. However, the first assumption is incompatible with the proportionality of mRNA numbers with cell size, with no dependence on the ploidy [16–18, 20], since fixed cell volume models will predict a doubled transcription rate after the replication of the whole genome.

Considering translation, various experiments have shown that the number of ribosomes is the limiting factor rather than the number of mRNAs. The most direct evidence is the growth law: the growth rate of cells is proportional to the fraction of ribosomal proteins in the total proteome (with a constant factor depending on the growth condition) [33] both for bacterial cells [24, 34] and budding yeast cells [25]. This means a constant fraction of ribosomes are actively translating mRNAs. These results suggest that in general cells are far below the saturation limit in which there are too many ribosomes that the mRNAs can bind. We will therefore assume the typical biological situation in which mRNAs in the cell compete for the limiting resource of actively working ribosomes, therefore the translation rate of one type of mRNA is proportional to the number of active ribosomes times its fraction in the total pool of mRNAs.

Experiments have also shown that mutants of fission yeasts altered in cell size regulated global transcription to maintain similar transcription rates per cell volume regardless of the cellular DNA content. The changes in total transcription correlated with coordinated changes in gene occupancy by RNA polymerases [35]. These results suggest that the number of RNAP is the actual limiting factor in transcription rather than the gene number, and similar evidence is shown in bacterial cells [36] and mammalian cells [17, 37]. To address the limiting nature of RNAP, we define an effective gene copy number g_i for each gene to account for its copy number and the binding strength of promoters, which determines its ability

to compete for RNAPs. The transcription rate for one specific gene i is proportional to the fraction of RNAPs that are working on its gene(s), $\phi_i = g_i/\sum_j g_j$. Gene regulation is thus coarse-grained into the gene allocation fraction ϕ_i.

In the following, m, p, r, n represents mRNA, protein, ribosome and RNA polymerase, respectively. Proteins (p) also include RNAPs (n) and ribosomes (r) [24]. We consider the degradation of mRNA with degradation time τ for all genes. The protein number decreases only through cell divisions. The key assumptions of this work are summarized in the following sets of reactions and Figure 1,

$$m_i \xrightarrow{k_0(g_i/\sum g_j)n} m_i + 1, \tag{1}$$

$$m_i \xrightarrow{m_i/\tau} m_i - 1, \tag{2}$$

$$p_i \xrightarrow{k_t(m_i/\sum_j m_j)f_a r} p_i + 1. \tag{3}$$

Here k_0, k_t are constants, characterizing the transcription (translation) rate of a single RNAP (ribosome). f_a is the fraction of active ribosomes. In our model, the gene allocation fraction ϕ_i is invariant upon replicating the entire genome, therefore the independence of the genome copy number is automatically satisfied. For simplicity, we first assume a *constant* set of value ϕ_i. This can be formally thought of as corresponding to an instantaneous replication of the genome. In reality, the finite duration of DNA replication would lead to ϕ_i's that change during the cell cycle. We later analyze a more complete version of the model which includes these gene dosage effects, but we first consider the simplified scenario of constant ϕ_i that will capture the essential features of the problem.

Due to the fast degradation of mRNA compared with the cell cycle duration [38, 39], the mRNA number can be well approximated as being in steady state. We can express the ensemble-averaged number of mRNA of gene i as

$$\langle m_i(t)\rangle = k_0\phi_i\langle n(t)\rangle\tau. \tag{4}$$

Equation (3) then leads to the time-dependence of average ribosome number,

$$\frac{d\langle r\rangle}{dt} = k_t f_a \phi_r \langle r\rangle. \tag{5}$$

Our model reproduces the autocatalytic nature of ribosome production and the growth rate

$$\mu = k_t f_a \phi_r, \tag{6}$$

determined by the relative abundance of active ribosomes in the proteome [24, 25]. We remark that the competition between mRNAs for ribosomes is crucial to obtain the exponential growth of ribosomes; increasing transcription rate does not affect the translation rate due to the cancellation in the factor $m_i/\sum_j m_j$.

Similarly, we can find the number of protein i grows as $d\langle p_i\rangle/dt = k_t f_a \phi_i \langle r\rangle$. As the cell grows and divides, the dynamics becomes insensitive to the initial conditions, so the protein number will grow exponentially as well. The ratio between the average of two protein numbers in the steady state is set by the ratio of their production rate, and therefore $\langle p_i\rangle/\langle p_j\rangle = \phi_i/\phi_j$. The average number of mRNA traces the number of RNA polymerases according to Equation (4), and therefore also grows exponentially. Throughout the cell cycle we have

$$\langle m_i(t)\rangle = m_b(i)\exp(\mu t), \tag{7}$$

$$\langle p_i(t)\rangle = p_b(i)\exp(\mu t), \tag{8}$$

where $m_b(i)$ ($p_b(i)$) is the number of mRNA (protein) of gene i at cell birth.

Theoretical models of stochastic gene expression have seldom considered cell division explicitly, and those which did considered a random generation time drawn from a preassigned distribution [40, 41], or did not consider the time-dependence of transcription and translation rate [30, 42]. The random generation time assumption is incompatible with the exponential growth of cell volume and protein numbers, because in the presence of noise it would lead to the divergence of cell size [43–45]. Here we take explicitly cell division into account and, for concreteness, use the "adder" model for cell division. We consider an initiator protein I, which accumulates from 0 after cell birth, and triggers the cell division once I reaches the division threshold I_c and is then destroyed (or "reset", e.g., after initiation of DNA replication in bacteria, the ATP-bound DnaA is dephosphorylated to the ADP-bound form) [46–48]. During a division event, we assume proteins and mRNAs are divided between the two daughter cells following a binomial distribution [49]. The initiator protein sets the scale of absolute protein number, and the average number of proteins produced in one cell cycle is equal to $\Delta(i) = I_c\phi_i/\phi_I$ [47]. Since the protein number grows twofold during one cell cycle, the average protein number of gene i at cell birth is $p_b(i) = I_c\phi_i/\phi_I$ and the corresponding average mRNA number at cell birth is $m_b(i) = k_0 I_c \tau \phi_i \phi_n/\phi_I$. We remark that the exact molecular mechanism of cell division is not important, and does not affect our results.

Constant cell density model

In the simplest scenario, one may assume that the dry density of cells (denoted below as "density") is strongly regulated, and therefore the cell volume is approximately proportional to the total protein mass, i.e., $V \propto M = \sum_j p_j$, which is a reasonable approximation for bacteria [27]. We assume each protein has the same mass and set the cell density as 1 to simplify the following formulas. We define the fraction of mRNA i in the total mRNA pool

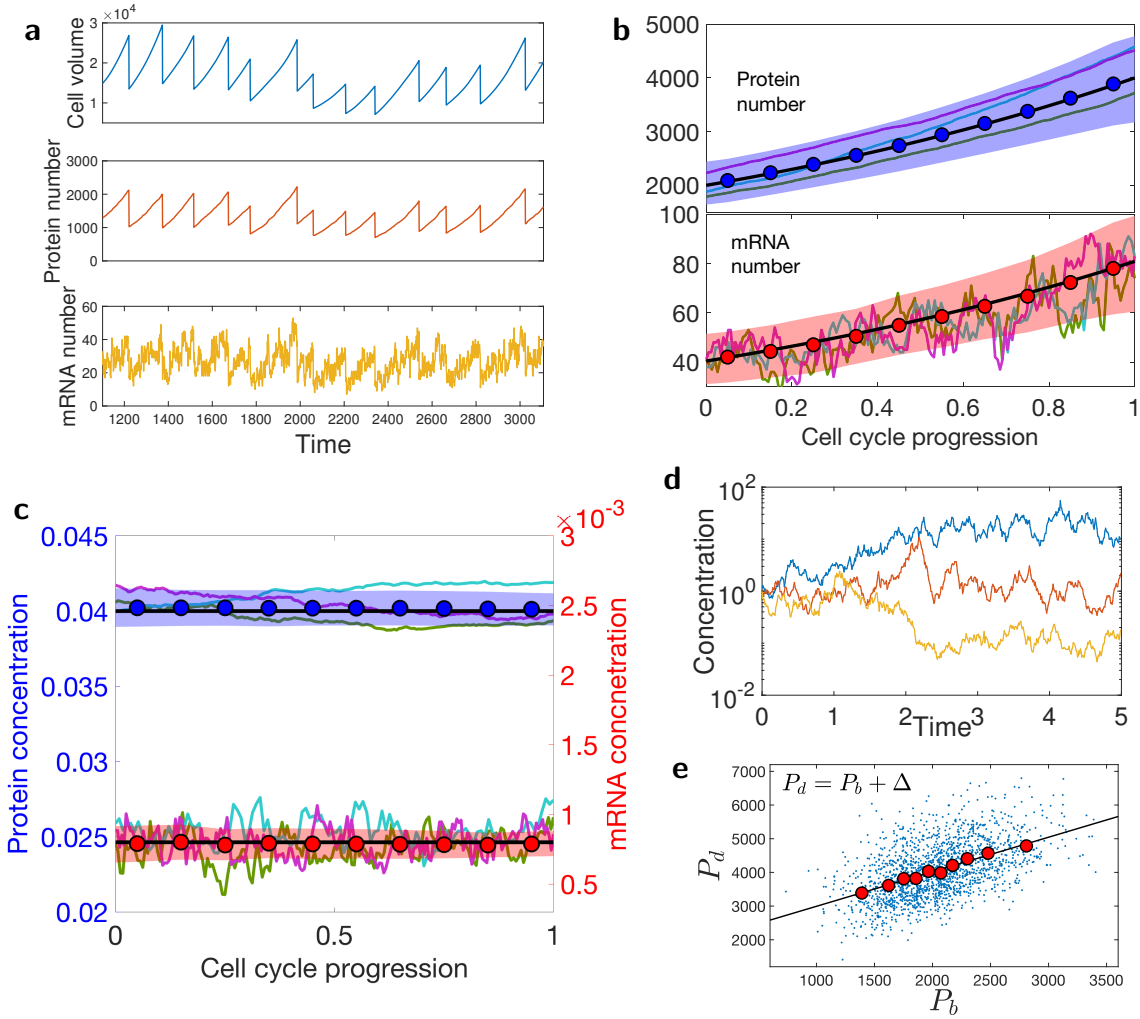

Figure 2. **Exponential growth of the cell volume, protein number, mRNA number; the homeostasis of protein and mRNA concentrations throughout the cell cycle.**
(**a**) Within the constant density model, the cell volume is proportional to the sum of all protein copy numbers. Typical trajectories of cell volume, protein number, and mRNA number are shown ($\phi_i = 0.018$).
(**b**) The averaged values of protein and mRNA number of a highly expressed gene ($\phi_i = 0.04$), are shown (circles) with 3 single trajectories in the background. The black lines are theoretical predictions of Equations (7, 8). The average is over 130 cell cycles. The color band represents the standard deviation (same for c).
(**c**) The averaged values of protein and mRNA concentrations of the same gene as in (b) are shown (circles). The black lines are theoretical predictions of Equations (12, 13). Three single trajectories are shown in the background.
(**d**) Three trajectories of diverging concentrations if the protein number and cell volume grow independently. See the numerical details in Methods.
(**e**) The scatter plot of the protein numbers at cell division (P_d) *v.s.* the protein numbers at cell birth (P_b). The circles are binned data. The black line is a linear fit of the binned data with slope 1.03, consistent with the adder correlations.

as $f_i = m_i / \sum_j m_j$, and the concentration of mRNA and protein of gene i as $cm_i = m_i/V$, $c_i = p_i/V$. We denote the RNAP and ribosome concentration as c_n and c_r. According to Equations (1-3), the deterministic equations of the above variables then become

$$\frac{df_i}{dt} = \frac{k_0 n}{\sum_j m_j}(\phi_i - f_i) \tag{9}$$

$$\frac{dc_i}{dt} = k_t c_r f_a (f_i - c_i) \approx \mu(f_i - c_i). \tag{10}$$

$$\frac{dcm_i}{dt} = \frac{1}{\tau}(k_0 \phi_i c_n \tau - (1 + \mu\tau) cm_i). \tag{11}$$

Using the condition that mRNA degradation time is much smaller than the doubling time ($\mu\tau \ll 1$), we find the fixed points for the dynamics of f_i, ϕ_i, and cm_i. These are, respectively, ϕ_i, f_i, and $k_0\phi_i c_n\tau$. The stability of the fixed point is due to the global nature of RNAPs and ribosomes: any noises arising from the copy number of RNAPs (ribosomes) equally affect all mRNAs (proteins), and therefore leave the relative fraction of mRNAs (proteins) invariant. The average concentrations of mRNA and protein of gene i become

$$\langle c_i\rangle = \phi_i, \tag{12}$$

$$\langle cm_i\rangle = k_0\tau\langle c_n\rangle\phi_i = k_0\tau\phi_n\phi_i, \tag{13}$$

and the results are independent of the cell volume and genome copy number as expected [16–18, 20].

Because of the short lifetime of mRNAs, the production of proteins can be approximated by instantaneous bursts [22]. We introduce the burst size parameter b_0 as the average number of proteins made per burst, $b_0 = k_t f_a\langle r(t)\rangle/\langle\sum_j m_j\rangle\times\tau \approx k_t f_a\phi_r/(k_0\phi_n)$, independent of the cell volume. ϕ_i for $N = 200$ proteins are uniformly sampled in logarithmic space, with the sum over ϕ_i (including ribosomes and RNAPs) constraint to be precisely one. We choose the parameters to be biologically relevant for bacteria: the doubling time $T = \ln(2)/\mu = 150$ min, $r_b = 10^4$, $n_b = 10^3$, $b_0 = 0.8$, $I_c = 20$, $\phi_r = 0.2$, $f_a = 0.7$ and $\tau = 3.5$ min, see other numerical details in Methods. Our conclusions are independent of the specific choice of parameters.

In Figure 2a, we show the typical trajectories from our simulations of cell volume, protein number and mRNA number for the same gene over multiple generations. To verify the exponential growth of protein and mRNA, we average the protein and mRNA numbers given a fixed relative phase in the cell cycle progression, which is normalized by the generation time and changes from 0 to 1. The averaged values of protein and mRNA numbers (circles) are nicely predicted by the exponential growth, Equations (7, 8) (black lines) without fitting parameters, as shown in Figure 2b with 3 single trajectories in the background.

The corresponding trajectories of protein and mRNA concentrations are shown in Figure 2c, with bounded fluctuations around the predicted averaged values (black lines). We remark that the fact that fluctuations of concentrations are bounded is a key result of our model. In contrast, if the protein number and cell volume grow exponentially but independently, the ratio between them will diverge as the effects of noise accumulate. An intuitive way to see this divergence is to consider the logarithm of concentration, which exhibits a random walk behavior (Figure 2d). Within our model, we may also study the protein number dynamics: how does the protein number at cell division correlate with that at cell birth? We find that the correlations follow an "adder" (i.e. the number of new *proteins* added is uncorrelated with the number at birth), as shown in Figure 2e. While this has been quantified in various organisms with respect to cell volume, checking correlations between protein content at cell birth and division has received significantly less attention [50]. Related to this, we study the auto-correlation function of protein concentration in time. We find that the auto-correlation function is approximately exponential, with a correlation time bounded from below by the doubling time (Figure S1). Both of these results are experimentally testable predictions.

Noise in ribosomal gene allocation fraction contributes to the global extrinsic noise

In this section, we study the noise in gene expression and focus on the contribution of protein concentration fluctuations. We first remark that to meaningfully quantify the noise in gene expression, the relevant quantity to look at is the protein concentration rather than the absolute number since the protein number grows exponentially throughout the cell cycle. We simulate an asynchronous population with 5000 cells and following previous works, compute the normalized protein number as the protein concentration multiplied by the average cell volume [28]. We quantify the noise as the square of coefficient of variation (CV), which is the ratio between the standard deviation (σ_p) and the mean (μ_p) of the population.

To generate a global extrinsic noise that could affects the concentrations of all proteins, we introduce a noise in the gene allocation fraction of ribosome, ϕ_r [28]. The cell-to-cell variability in ϕ_r is in practice unavoidable in a growing population due to the fluctuation in local environment and abundance of nutrients. We add a noise term to ϕ_r for each cell and keep the normalization of ϕ_i,

$$\phi_r \to \phi_r(1+\xi_r),$$
$$\phi_i \to \frac{\phi_i}{1+\phi_r\xi_r}. \tag{14}$$

Here ξ_r follows a Gaussian distribution with a standard deviation σ_r. We show the resulting noise curves in Figure 3a. The noise curves decay as $CV^2 = (1+b_0)/\mu_p$ for low copy number proteins which are dominated by the intrinsic noise. This modified Poissonian behavior due to the bursty protein production respects the prediction from fixed cell volume models. As we increase the variability in ϕ_r, the deviation of the noise curves from the intrinsic noise becomes more significant. We quantify the magnitude of global extrinsic noise as the CV of the most expressed proteins and compare it with the CV of ϕ_r, as shown in the inset of Figure 3a. As expected from Equation 14, the global extrinsic noise is smaller than the noise in ϕ_r and satisfies $CV(ext)/CV(\phi_r) \approx \phi_r$

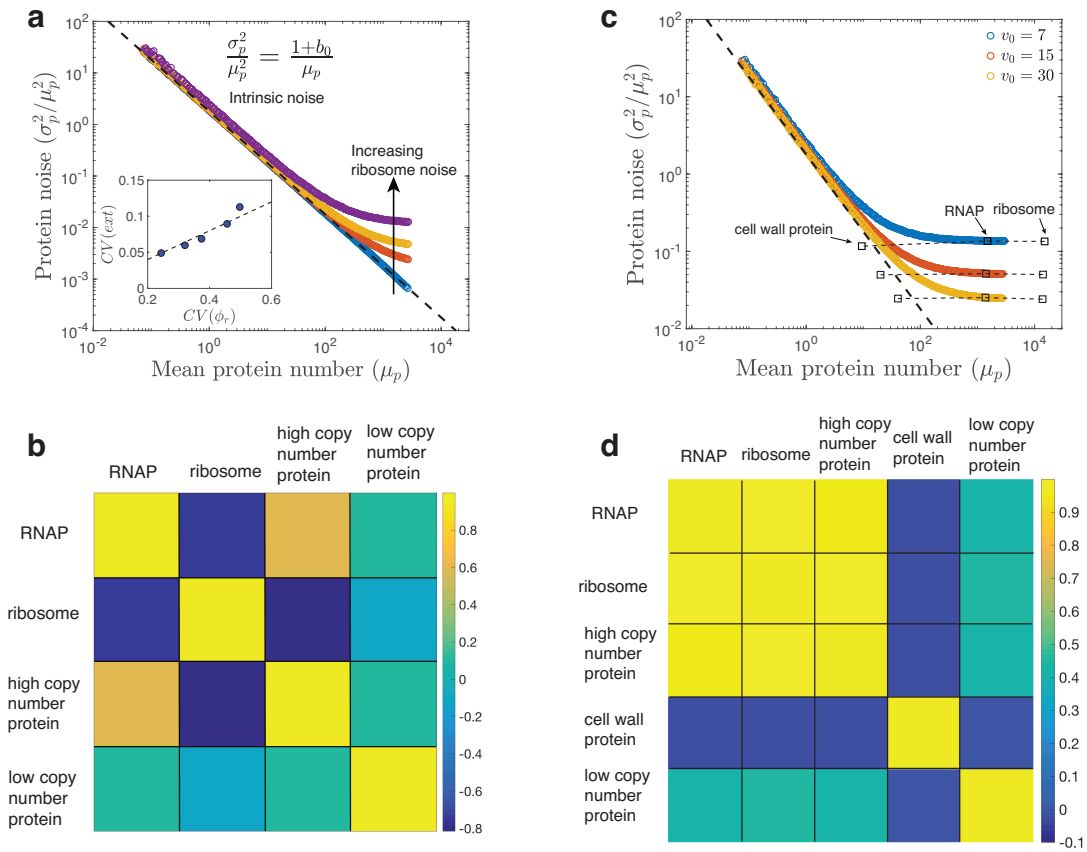

Figure 3. **Noises and correlations in protein concentrations.**
(**a**) The noise curves with fluctuations in the gene allocation fraction of ribosomes (ϕ_r). As we increase the noise in ϕ_r, the global extrinsic noise increases. The lower (blue) curve corresponds to the scenario without noise in ϕ_r and only exhibits intrinsic noise. The black dashed line is the theoretical prediction from fixed cell volume models. In the inset, we show $CV(ext)$ *v.s.* $CV(\phi_r)$ and their ratio is close to the prediction ϕ_r (dashed line).
(**b**) Pearson correlation coefficients between protein concentrations in the case of fluctuating ribosome levels. The high copy number protein has the mean protein number $\mu_p \approx 2000$ ($\phi_i \approx 0.03$), and the low copy number protein has $\mu_p \approx 20$ ($\phi_i \approx 3 \times 10^{-4}$). While proteins are generally positively correlated, the ribosome has a negative correlation with other proteins.
(**c**) The noise curves for different expression level of cell wall proteins, quantified by $v_0 = I_c\phi_v/\phi_I$ as indicated (the average cell wall protein at cell birth). For each v_0, the noises of the cell wall proteins, RNAPs and ribosomes are also shown as squares from left to right. The dashed line shows the intrinsic noise.
(**d**) Pearson correlation coefficients between protein concentrations in the case of fluctuating cell volume growth rate ($v_0 = 30$). The high and low copy number proteins have the same μ_p as (b). While proteins are generally positively correlated, the cell wall proteins have a weak negative correlation with other proteins.

(Methods). We note that to generate the global extrinsic noise, $CV(ext) \approx 0.3$ as reported in Ref. [28], one need an abnormally large $CV(\phi_r)$. Within our model, as is also observed experimentally, highly expressed proteins are positively correlated. However, since ribosomes and RNAPs play a special role within our model, it is not clear *a priori* whether they will exhibit the same correlation pattern. We tested this numerically, and indeed found that the correlation coefficients among them are atypical, as shown in Figure 3b. Constitutively expressed proteins including RNAP are generally positively correlated because of the global extrinsic noise set by ribosome, while ribosome shows a negative correlation with other proteins. This is due to the competition between ribosomal genes and other genes for the limiting sources of RNAP, and mathematically due to the normalization of the sum of ϕ_i. We summarize the correlation pattern in Figure 3b, which provides a method to experimentally study the relative contribution of ribosomal level fluctuations to the global extrinsic noise.

Other potential sources of global extrinsic noise

In this section, we consider other potential sources of the global extrinsic noise. First, we relax the condition

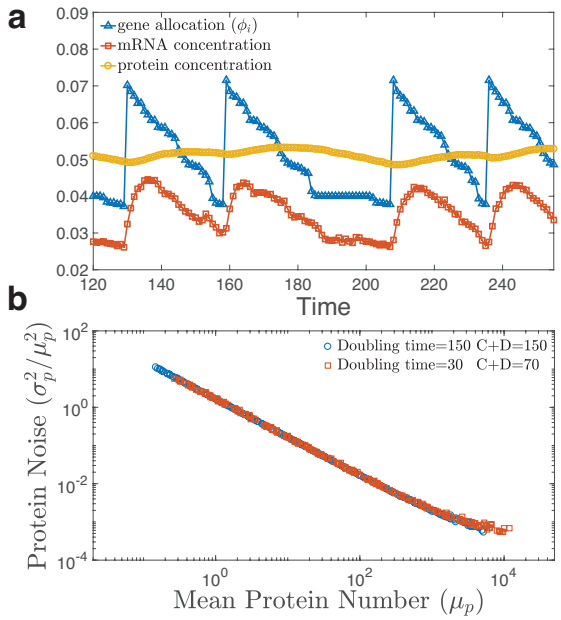

Figure 4. **Effects of finite duration of DNA replication.** (**a**) The time trajectory of gene allocation fraction, mRNA concentration and protein concentration of a high copy number protein ($\mu_p \approx 10^4$). The doubling time $T = 30$ min , and we use the values of the C and D periods from Ref. [51], namely, $C = 35$ min and $D = 35$ min. In this situation, the cell undergoes DNA replication throughout the cell cycle. Nevertheless, the noise in ϕ_i does not propagate to the noise in protein concentration significantly. The value of mRNA concentration is 5 times amplified for clarity.
(**b**) The noise curves when gene dosage effects are considered due to DNA replication, do not show significant global extrinsic noises. Two different doubling times are considered.

that the cell density is strictly constant, and consider a model in which the cell volume growth is governed by cell wall proteins.

We simplify the problem by considering cells that grow anisotropically so that their surface area:volume (SA:V) ratio is constant (Methods). An important example is rod-shaped cells. The model presented here should be equally applicable to exponentially growing mammalian cells, where the role of cell wall proteins in defining cellular geometry would be taken by other surface-area-defining proteins (e.g., membrane proteins). The volume growth rate is proportional to the surface area growth rate, which is set by the number of cell wall proteins [52],

$$\frac{dV}{dt} = k_v p_v, \tag{15}$$

with k_v constant and p_v the number of cell wall proteins. We note that the cell volume growth rate is proportional to the concentration of the cell wall protein, $\dot{V}/V = k_v c_v$. Within this variant of the model, even without active cell volume regulation, the cell density will be auto-regulated: an increase in the concentration of cell wall protein induces a faster cell volume growth which subsequently suppresses the cell wall protein concentration. We define ϕ_v as the gene allocation fraction of the cell wall proteins and its characteristic number is thus $v_0 = I_c\phi_v/\phi_I$. Detailed derivations on the dynamics of cell density are in Methods. In this model, the global extrinsic noise of protein concentrations comes from the fluctuating growth of cell volume due to the finite copy number of cell wall proteins. The global extrinsic noise increases as the number of cell wall proteins decreases (Figure 3c) and equal to the fluctuation of cell density (Figure S2de). Furthermore, it is the cell wall proteins that show slightly negative correlations with other proteins, while other proteins including ribosomes exhibit positive correlation (Figure 3d). We conclude that the correlation coefficients between ribosomes, cell wall proteins and constitutively expressed proteins will inform us of the dominant source of global extrinsic noise.

We also study other potential sources of global extrinsic noise, including the fluctuation in the allocation fraction of RNAP (ϕ_n), mRNA degradation time, and division threshold of initiator proteins. We find that none of them contributes significantly to the global extrinsic noise (Methods).

Effect of gene replication

So far, we considered a constant ϕ_i throughout the cell cycle assuming an instantaneous replication of the genome. In this section, we relax this condition and study the effects of finite DNA replication time. We consider the bacterial model of DNA replication, specifically, *Escherichia coli* , for which the mechanism of DNA replication is well characterized [51]. The duration of DNA replication is constant as is defined as the C period, and the corresponding cell division follows after an approximately constant duration known as the D period. Details of the DNA replication model are in Methods. In Figure 4a, we show the time trajectories of gene allocation fraction, mRNA concentration and protein concentration for a doubling time of $T = 30$ min with $C + D = 70$ min. The DNA replication introduces a cell cycle dependent modulation of ϕ_i, which propagates to the mRNA concentration, which in turn affects the protein concentration. However, this cell cycle modulation only contributes a small global extrinsic noise: the CV of the most highly expressed proteins is about 0.02 (Figure 4b). Therefore, within our growing cell model, the finite duration of DNA replication does not significantly affect the protein level fluctuations. However, in fixed cell volume models, the modulation of protein concentrations due to DNA replications turns out to be much stronger since the transcription rate is directly proportional to the gene number [30, 31].

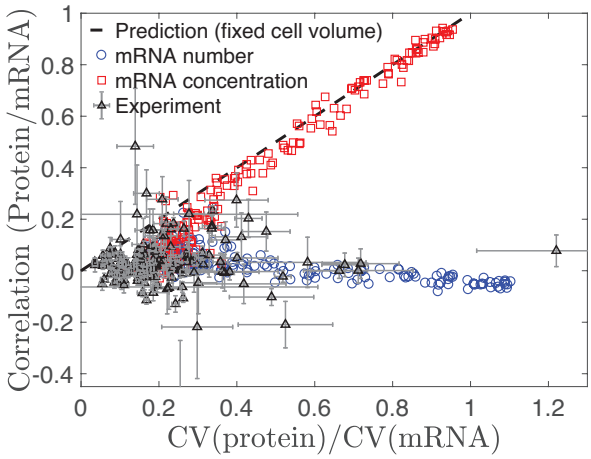

Figure 5. **Correlations between protein and mRNA.** The y axis is the Pearson Correlation coefficient between protein concentration and (i) mRNA number (blue circles), (ii) mRNA concentration (red squares). The x axis is the ratio of CV between protein concentrations and mRNA numbers (concentrations). Each data point is a single gene. The prediction from the fixed cell volume model is shown as the dashed line. Experimental data from Taniguchi *et al.* [28] are shown as gray triangles. Here we take the model with cell wall proteins to generate the global extrinsic noise in protein concentrations ($v_0 = 7$) and the results are independent of the specific mechanisms of the global extrinsic noise.

Correlations between mRNA and protein levels

Ref. [29] has shown that there must be a positive correlation between mRNA and protein expression levels within fixed cell volume models, and a simple relation connects the correlation coefficient (CC) between protein and mRNA levels and their CVs,

$$CC(protein/mRNA) = \frac{CV(protein)}{CV(mRNA)} > 0. \qquad (16)$$

Indeed, from the asynchronous populations simulated from our model, if we *normalize* both the mRNA and protein by the cell volume, the correlation coefficients between the mRNA and protein concentrations are positive and follow well the prediction of [29], shown as the red squares in Figure 5.

In contrast, negligible correlations were observed experimentally [28]. Ref. [29] suggested that the small correlation between mRNA and protein can emerge due to a strong anti-correlation between the transcription and translation rates. Here, we show that the puzzling vanishing correlation is a matter of normalization, which can be accounted for only when the growing cell volume is considered. To appreciate this, note that the authors of Ref. [28] correlated the mRNA *absolute number* and protein concentration, a distinction beyond the fixed cell volume model. Based on our model, we indeed find a nearly zero correlation coefficient between mRNA number and protein concentration, much below the prediction from fixed cell volume models [29, 53], but consistent with the experimental results, shown in Figure 5. To understand this, note that the average values of protein concentrations are constant throughout the cell cycle (Figure 2c), therefore given a protein concentration in one cell, the corresponding cell cycle progression is uniformly distributed between 0 and 1. In contrast, the average mRNA number increases twofold throughout the cell cycle (Figure 2b), therefore leading to a small correlation between mRNA number and protein concentration.

CONCLUSION

In this work, we propose a coarse-grained model of stochastic gene expression incorporating the cell cycle effects including cell volume growth and cell division. The limiting nature of ribosomes in the translation process leads to the exponential growth of protein numbers, and similarly, the limiting nature of RNA polymerases leads to the exponential growth of mRNA numbers. Homeostasis of protein concentrations originates from the fact that ribosomes make all proteins. Homeostasis of mRNA concentration comes from the resulting bounded concentrations of RNAPs. Our model is consistent with the constancy of mRNA concentration as the genome copy number varies since the transcription rate depends on the relative fraction of genes in the genome rather than its absolute number.

Predictions of fixed cell volume models on the intrinsic noise of protein concentrations remain valid; our model allows us to study the possible sources of global extrinsic noise. Because the cell-to-cell variability in the ribosomal gene allocation fraction can never be avoided in real experiments, we consider the noise in ϕ_r as the main source of global extrinsic noise. While we also show that the fluctuation in the cell volume growth due to a finite copy number of cell wall proteins can in principle also generate a global extrinsic noise. However, these two different mechanisms show distinct correlation patterns between proteins, which can be tested experimentally. Explicitly considering growing cell volume also enables us to reconcile the experimental observations of low correlations between protein and mRNA [28], with the predicted positive correlations based on fixed cell volume models [29].

In this work, we only consider constitutively expressed genes, and the only interactions between them are the competitions for ribosomes and RNAPs. It is possible to add additional specific interactions between genes, e.g., introducing the transcription factors to regulate a subset of genes. These lead to transcriptional bursts in mRNA productions [6, 49, 54, 55], which so far are coarse-grained into the effective copy number of genes. It will be interesting to study the interplay between the global interactions and local interactions between genes.

We note that the assumption that RNA polymerases and ribosomes are the limiting factors in transcription and translation can be violated in some cases. In reality, there is an upper bound of the number of RNAPs (ribosomes) that a single gene (mRNA) can hold, and it would be interesting to explore these limits. When genes get saturated, the transcription rate becomes independent of the RNAP number, and becomes proportional to the gene copy number. When mRNAs get saturated, the translation rate becomes independent of the ribosome number and proportional to mRNA number. Experimentally, it appears that the DNA saturation limit is achieved by forcing cells to grow abnormally large [35]. Studying the transition between the two regimes of growth limitations, experimentally and theoretically, would be an interesting future direction.

METHODS

Simulations of independent growth model

In the growth model corresponding to Figure 2d, we assume the protein number and cell volume grow exponentially and independently,

$$\frac{dp}{dt} = (1 + \xi_p(t))p \tag{17}$$

$$\frac{dV}{dt} = (1 + \xi_V(t))V. \tag{18}$$

Here, $\xi_p(t)$, $\xi_V(t)$ are white noise terms, with the autocorrelation function, $\langle \xi_{p,V}(0)\xi_{p,V}(t)\rangle = A_{p,V}\delta(t)$. In Figure 2d of the main text, we choose $A_p = A_V = 1$.

Simulations of growing cell model

We simulated Equations (1,2,3), fixing r_b, n_b, b_0, ϕ_r, f_a, I_c, τ as well as the growth rate μ. Other parameters are inferred given the above values, *e.g.*, $\phi_n = n_b\phi_r/r_0$, $k_t = \mu/(\phi_r f_a)$, $k_0 = k_t f_a r_b/(b_0 n_b)$. We fix the time step δt so that the probability for one event to happen during a time step is smaller than 0.1.

The global extrinsic noise *v.s.* the noise in ribosomal gene allocation fraction

In Equation 14 of the main text, we introduce the noise in ϕ_r and here we expand this Equation around small noise,

$$\phi_i \to \phi_i(1 - \phi_r\xi_r), \tag{19}$$

$$\phi_r \to \phi_r(1 + \xi_r)(1 - \phi_r\xi_r). \tag{20}$$

Since $1 + \xi_r$ must be positive, we draw ξ_r from a truncated Gaussian distribution whose support is $(-1, \inf)$. It is straightforward to find the global extrinsic noise, which is the CV of highly expressed protein as $CV(ext) = \phi_r\sigma_r$. Here σ_r is the standard deviation of ξ_r. The CV of ϕ_r is approximately $(1 - \phi_r)\sigma_r$, and therefore we obtain $CV(ext) = \phi_r/(1 - \phi_r)CV(\phi_r) \approx \phi_r CV(\phi_r)$. Here we neglect higher order terms of ϕ_r.

Dynamics of cell density in the model with cell wall proteins

The dynamics of cell density $\rho = M/V$ follows

$$\frac{d\rho}{dt} = \frac{k_v p_v}{M}\left(\frac{k_t f_a r}{k_v p_v} - \rho\right). \tag{21}$$

Again, we find ρ is attracted to the fixed point $k_t f_a r/(k_v p_v)$, due to the auto-regulation of cell wall proteins which prevents the density from diverging. According to the previous results, we can find the average cell density in the steady state as $\langle\rho\rangle = k_t f_a \phi_r/(k_v\phi_v)$. The protein and mRNA concentrations become

$$\langle c_i\rangle = \frac{\langle p_i\rangle}{M}\langle\rho\rangle = \phi_i\langle\rho\rangle, \tag{22}$$

$$\langle cm_i\rangle = \frac{\langle m_i\rangle}{M}\langle\rho\rangle = k_o\phi_n\tau\phi_i\langle\rho\rangle. \tag{23}$$

These equations show that for cells with constant SA:V ratio, the homeostasis of mRNA and protein concentration is achieved even if cell wall proteins are not explicitly regulated. Equation 22 shows that the global extrinsic noise is directly proportional to the fluctuation of cell density (Figure S2d) and therefore decreases as the number of cell wall protein increases.

Potential sources of the global extrinsic noise

We can introduce noise in the gene allocation fraction of RNAP, in a similar fashion to the way we addressed ribosomal noise. Similar considerations as for the ribosome case show that the global extrinsic noise generated by these fluctuations in ϕ_n satisfies $CV(ext) \approx \phi_n CV(\phi_n)$. However, we expect the relative contribution of RNAP level fluctuations to the global extrinsic noise to be negligible compared to that due to e.g., ribosome level fluctuations, for two reasons: first, the number of RNAP is typically much smaller than that of ribosomes (e.g., about 10 times smaller in bacteria [56]), suggesting that $\phi_n \ll \phi_r$. Furthermore, while the "growth law" (the proportionability of the growth rate and ribosomal fraction in the proteome) implies that environmental variability will translate to variability in ϕ_r, it is plausible that the resulting variability in ϕ_n will be smaller. For these reasons, we expect RNAP level fluctuations to contribute insignificantly to the global extrinsic noise.

We have tested the effects of fluctuations in RNAP number on the gene expression noise numerically. We introduce a random noise to the RNAP fraction of the proteome as $\phi_n \to \phi_n(1+\xi_n)$, and maintain the normalization. Motivated by experimental data on bacteria [24, 56], we choose $\phi_r = 0.2$, $\phi_n = 0.1\phi_r$. The noise of the RNAP number is independent for each cell in the asynchronous population. As we show in Figure S2a, the noise curve of normalized protein number in the presence of RNAP noise shows negligible extrinsic noise.

Extending our model, we can allow the degradation time of mRNAs to be time-dependent *e.g.*, due to the noise of ribonuclease abundance. One expects that a longer mRNA life time will generate more proteins and therefore may induce a global extrinsic noise. However, as can be seen from Equations 10 of the main text, the life times of mRNA do not affect the protein dynamics. We add a time-dependent noise to the mRNA life time, $\tau \to \tau(1+\xi_\tau(t))$ (the same noise for all genes but different across cells), The noise satisfies the Ornstein-Uhlenbeck process, with an auto-correlation function $\langle \xi_\tau(0)\xi_\tau(t)\rangle = 0.09\exp(-t/150)$. We numerically confirm that a fluctuating mRNA life time cannot generate the global extrinsic either (Figure S2b).

We then consider the fluctuations in the division threshold I_c of the initiator proteins. Replacing the strict threshold I_c by a fluctuating variable broadens the doubling time distribution. However, since we normalize the protein number by the cell volume, it does not affect the protein concentration noise either (Figure S2c).

Gene dosage effects

In reality, the gene allocation fraction ϕ_i changes during the cell cycle due to the finite duration of DNA replication. In this section we introduce the modified version of the gene expression model incorporating DNA replication. Although our model is general, we focus on DNA replication in bacteria for concreteness, specifically *E. coli* where this process is very well characterized. We expect our conclusions to be generally valid. Furthermore, we refine our model for cell division, assuming that the initiator protein triggers the initiation of DNA replication rather than cell division, with the threshold I_c proportional to the number of origins of replication [51, 57] (the number of which doubles at each initiation). We assume that the cell division takes place a fixed time $C+D$ after initiation of the DNA replication, where C, D are respectively the time for DNA replication and the time between the completion of DNA replication and cell division. The number of origins reduce by half at each cell division. Other details are the same as in the main text. Each gene doubles its copy number during the C period, and we choose this gene replication time to be randomly and uniformly distributed across all genes. When a gene i replicates,

$$\phi_i \to 2\phi_i \tag{24}$$

$$\phi_j \to \frac{\phi_j}{\sum_k \phi_k}, \tag{25}$$

where the second equation accounts for the normalization of the gene allocation fraction. We choose the experimentally reported C and D and cell doubling time from Ref. [51].

[1] Johan Paulsson, "Models of stochastic gene expression," Physics of Life Reviews **2**, 157 – 175 (2005).

[2] Mads Kærn, Timothy C Elston, William J Blake, and James J Collins, "Stochasticity in gene expression: from theories to phenotypes," Nature Reviews Genetics **6**, 451–464 (2005).

[3] Juan M Pedraza and Alexander van Oudenaarden, "Noise propagation in gene networks," Science **307**, 1965–1969 (2005).

[4] Arjun Raj and Alexander van Oudenaarden, "Nature, nurture, or chance: stochastic gene expression and its consequences," Cell **135**, 216–226 (2008).

[5] Mukund Thattai and Alexander Van Oudenaarden, "Intrinsic noise in gene regulatory networks," Proceedings of the National Academy of Sciences **98**, 8614–8619 (2001).

[6] Alvaro Sanchez and Ido Golding, "Genetic determinants and cellular constraints in noisy gene expression," Science **342**, 1188–1193 (2013).

[7] Ping Wang, Lydia Robert, James Pelletier, Wei Lien Dang, Francois Taddei, Andrew Wright, and Suckjoon Jun, "Robust growth of escherichia coli," Current Biology **20**, 1099 – 1103 (2010).

[8] Manuel Campos, Ivan V Surovtsev, Setsu Kato, Ahmad Paintdakhi, Bruno Beltran, Sarah E Ebmeier, and Christine Jacobs-Wagner, "A constant size extension drives bacterial cell size homeostasis," Cell **159**, 1433–1446 (2014).

[9] Sattar Taheri-Araghi, Serena Bradde, John T Sauls, Norbert S Hill, Petra Anne Levin, Johan Paulsson, Massimo Vergassola, and Suckjoon Jun, "Cell-size control and homeostasis in bacteria," Current Biology **25**, 385–391 (2015).

[10] Nathan Cermak, Selim Olcum, Francisco Feijo Delgado, Steven C Wasserman, Kristofor R Payer, Mark A Murakami, Scott M Knudsen, Robert J Kimmerling, Mark M Stevens, Yuki Kikuchi, Arzu Sandikci, Masaaki Ogawa, Vincent Agache, Francois Baleras, David M Weinstock, and Scott R Manalis, "High-throughput measurement of single-cell growth rates using serial microfluidic mass sensor arrays," Nature Biotechnology **34**, 1052–1059 (2016).

[11] SG Elliott and CS McLaughlin, "Rate of macromolecular synthesis through the cell cycle of the yeast saccharomyces cerevisiae," Proceedings of the National Academy of Sciences **75**, 4384–4388 (1978).

[12] Stephen Cooper, "Distinguishing between linear and exponential cell growth during the division cycle: single-cell studies, cell-culture studies, and the object of cell-cycle

research," Theoretical Biology and Medical Modelling **3**, 10 (2006).

[13] Stefano Di Talia, Jan M Skotheim, James M Bean, Eric D Siggia, and Frederick R Cross, "The effects of molecular noise and size control on variability in the budding yeast cell cycle," Nature **448**, 947–951 (2007).

[14] Michel Godin, Francisco Feijó Delgado, Sungmin Son, William H Grover, Andrea K Bryan, Amit Tzur, Paul Jorgensen, Kris Payer, Alan D Grossman, Marc W Kirschner, *et al.*, "Using buoyant mass to measure the growth of single cells," Nature methods **7**, 387–390 (2010).

[15] Warren K Sinclair and Dennis W Ross, "Modes of growth in mammalian cells," Biophysical journal **9**, 1056–1070 (1969).

[16] Hermannus Kempe, Anne Schwabe, Frédéric Crémazy, Pernette J Verschure, and Frank J Bruggeman, "The volumes and transcript counts of single cells reveal concentration homeostasis and capture biological noise," Molecular biology of the cell **26**, 797–804 (2015).

[17] Olivia Padovan-Merhar, Gautham P Nair, Andrew G Biaesch, Andreas Mayer, Steven Scarfone, Shawn W Foley, Angela R Wu, L Stirling Churchman, Abhyudai Singh, and Arjun Raj, "Single mammalian cells compensate for differences in cellular volume and dna copy number through independent global transcriptional mechanisms," Molecular cell **58**, 339–352 (2015).

[18] Robert Ietswaart, Stefanie Rosa, Zhe Wu, Caroline Dean, and Martin Howard, "Cell-size-dependent transcription of flc and its antisense long non-coding rna coolair explain cell-to-cell expression variation," Cell Systems **4**, 622 – 635.e9 (2017).

[19] Naama Brenner, Erez Braun, Anna Yoney, Lee Susman, James Rotella, and Hanna Salman, "Single-cell protein dynamics reproduce universal fluctuations in cell populations," The European Physical Journal E **38**, 1–9 (2015).

[20] Xiao-yu Zheng and Erin K O'Shea, "Cyanobacteria maintain constant protein concentration despite genome copy-number variation," Cell reports **19**, 497–504 (2017).

[21] Nir Friedman, Long Cai, and X. Sunney Xie, "Linking stochastic dynamics to population distribution: An analytical framework of gene expression," Phys. Rev. Lett. **97**, 168302 (2006).

[22] Vahid Shahrezaei and Peter S Swain, "Analytical distributions for stochastic gene expression," Proceedings of the National Academy of Sciences **105**, 17256–17261 (2008).

[23] Long Cai, Nir Friedman, and X Sunney Xie, "Stochastic protein expression in individual cells at the single molecule level," Nature **440**, 358–362 (2006).

[24] M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, and T. Hwa, "Interdependence of Cell Growth and Gene Expression: Origins and Consequences," Science **330**, 1099 (2010).

[25] Eyal Metzl-Raz, Moshe Kafri, Gilad Yaakov, Ilya Soifer, Yonat Gurvich, and Naama Barkai, "Principles of cellular resource allocation revealed by condition-dependent proteome profiling," eLife **6**, e28034 (2017).

[26] Shlomi Reuveni, Måns Ehrenberg, and Johan Paulsson, "Ribosomes are optimized for autocatalytic production," Nature **547**, 293–297 (2017).

[27] Markus Basan, Manlu Zhu, Xiongfeng Dai, Mya Warren, Daniel Sévin, Yi-Ping Wang, and Terence Hwa, "Inflating bacterial cells by increased protein synthesis," Molecular Systems Biology **11** (2015), 10.15252/msb.20156178.

[28] Yuichi Taniguchi, Paul J. Choi, Gene-Wei Li, Huiyi Chen, Mohan Babu, Jeremy Hearn, Andrew Emili, and X. Sunney Xie, "Quantifying e. coli proteome and transcriptome with single-molecule sensitivity in single cells," Science **329**, 533–538 (2010).

[29] Andreas Hilfinger, Thomas M Norman, and Johan Paulsson, "Exploiting natural fluctuations to identify kinetic mechanisms in sparsely characterized systems," Cell systems **2**, 251–259 (2016).

[30] John A. Cole and Zaida Luthey-Schulten, "Careful accounting of extrinsic noise in protein expression reveals correlations among its sources," Phys. Rev. E **95**, 062418 (2017).

[31] Daniel L Jones, Robert C Brewster, and Rob Phillips, "Promoter architecture dictates cell-to-cell variability in gene expression," Science **346**, 1533–1536 (2014).

[32] Noreen Walker, Philippe Nghe, and Sander J Tans, "Generation and filtering of gene expression noise by the bacterial cell cycle," BMC biology **14**, 11 (2016).

[33] Frederick Carl Neidhardt, John L Ingraham, and Moselio Schaechter, *Physiology of the bacterial cell: a molecular approach*, Vol. 20 (Sinauer Sunderland, 1990).

[34] Sheng Hui, Josh M Silverman, Stephen S Chen, David W Erickson, Markus Basan, Jilong Wang, Terence Hwa, and James R Williamson, "Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria," Molecular systems biology **11**, 784 (2015).

[35] Jacob Zhurinsky, Klaus Leonhard, Stephen Watt, Samuel Marguerat, Jürg Bähler, and Paul Nurse, "A coordinated global control over cellular transcription," Current biology **20** (2010).

[36] MG Chandler and RH Pritchard, "The effect of gene concentration and relative gene dosage on gene output inescherichia coli," Molecular and General Genetics MGG **138**, 127–141 (1975).

[37] Edward E Schmidt and Ueli Schibler, "Cell size regulation, a mechanism that controls cellular rna accumulation: consequences on regulation of the ubiquitous transcription factors oct1 and nf-y and the liver-enriched transcription factor dbp." The Journal of cell biology **128**, 467–483 (1995).

[38] Jonathan A Bernstein, Arkady B Khodursky, Pei-Hsun Lin, Sue Lin-Chao, and Stanley N Cohen, "Global analysis of mrna decay and abundance in escherichia coli at single-gene resolution using two-color fluorescent dna microarrays," Proceedings of the National Academy of Sciences **99**, 9697–9702 (2002).

[39] Yulei Wang, Chih Long Liu, John D Storey, Robert J Tibshirani, Daniel Herschlag, and Patrick O Brown, "Precision and functional specificity in mrna decay," Proceedings of the National Academy of Sciences **99**, 5860–5865 (2002).

[40] Otto G Berg, "A model for the statistical fluctuations of protein numbers in a microbial population," Journal of theoretical biology **71**, 587–603 (1978).

[41] Mohammad Soltani, Cesar A Vargas-Garcia, Duarte Antunes, and Abhyudai Singh, "Intercellular variability in protein levels from stochastic expression and noisy cell cycle processes," PLoS Comput Biol **12**, e1004972 (2016).

[42] François Bertaux, Samuel Marguerat, and Vahid Shahrezaei, "Impact of division rate and cell size on gene expression noise," bioRxiv , 209593 (2017).

[43] Ariel Amir, "Cell size regulation in bacteria," Phys. Rev. Lett. **112**, 208102 (2014).
[44] Naama Brenner, C. M. Newman, Dino Osmanović, Yitzhak Rabin, Hanna Salman, and D. L. Stein, "Universal protein distributions in a model of cell growth and division," Phys. Rev. E **92**, 042713 (2015).
[45] Ariel Amir, "Is cell size a spandrel?" eLife **6**, e22186 (2017).
[46] L Sompayrac and O Maaløe, "Autorepressor model for control of dna replication," Nature **241**, 133–135 (1973).
[47] Khem Raj Ghusinga, Cesar A Vargas-Garcia, and Abhyudai Singh, "A mechanistic stochastic framework for regulating bacterial cell division," Scientific Reports **6**, 30229 (2016).
[48] François Bertaux, Julius Von Kügelgen, Samuel Marguerat, and Vahid Shahrezaei, "A unified coarse-grained theory of bacterial physiology explains the relationship between cell size, growth rate and proteome composition under various growth limitations," bioRxiv , 078998 (2016).
[49] Ido Golding, Johan Paulsson, Scott M Zawilski, and Edward C Cox, "Real-time kinetics of gene activity in individual bacteria," Cell **123**, 1025–1036 (2005).
[50] Yu Tanouchi, Anand Pai, Heungwon Park, Shuqiang Huang, Rumen Stamatov, Nicolas E Buchler, and Lingchong You, "A noisy linear map underlies oscillations in cell size and gene expression in bacteria," Nature **523**, 357–360 (2015).
[51] Mats Wallden, David Fange, Ebba Gregorsson Lundius, Özden Baltekin, and Johan Elf, "The synchronization of replication and division cycles in individual e. coli cells," Cell **166**, 729–739 (2016).
[52] Felix Wong, Lars D. Renner, Gizem Özbaykal, Jayson Paulose, Douglas B. Weibel, Sven van Teeffelen, and Ariel Amir, "Mechanical strain sensing implicated in cell shape recovery in escherichia coli," Nature Microbiology **2**, 17115 EP – (2017).
[53] Andreas Hilfinger, Thomas M. Norman, Glenn Vinnicombe, and Johan Paulsson, "Constraints on fluctuations in sparsely characterized biological systems," Phys. Rev. Lett. **116**, 058101 (2016).
[54] Alvaro Sanchez, Sandeep Choubey, and Jane Kondev, "Regulation of noise in gene expression," Annual review of biophysics **42**, 469–491 (2013).
[55] Shasha Chong, Chongyi Chen, Hao Ge, and X Sunney Xie, "Mechanism of transcriptional bursting in bacteria," Cell **158**, 314–326 (2014).
[56] Ron Milo and Rob Phillips, *Cell biology by the numbers* (Garland Science, 2015).
[57] Po-Yi Ho and Ariel Amir, "Simultaneous regulation of cell size and chromosome replication in bacteria," Frontiers in Microbiology **6**, 662 (2015).

AUTHOR CONTRIBUTIONS

All authors conceived the work, carried out the work, and jointly wrote the manuscript.

ACKNOWLEDGMENTS

We thank Naama Barkai, Ido Golding, Andreas Hilfinger, Po-Yi Ho, Andrew Murray, Johan Paulsson, Leonardo A. Sepúlveda, Yuichi Taniguchi and Sven van Teeffelen for useful discussions related to this work. AA thanks the A.P. Sloan foundation, the Milton Fund, the Volkswagen Foundation and Harvard Dean's Competitive Fund for Promising Scholarship for their support. JL was supported by the George F. Carrier fellowship and the National Science Foundation through the Harvard Materials Research Science and Engineering Center (DMR-1420570).