Published in final edited form as:

Nat Genet. 2011 February ; 43(2): 95–100. doi:10.1038/ng.729.

Non-genetic heterogeneity from random partitioning at cell division

Dann Huh^{1,2} and Johan Paulsson^{1,3}

¹ Department of Systems Biology, Harvard University, Boston, MA 02115, USA

² Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA

Abstract

Gene expression involves many inherently probabilistic steps that create fluctuations in protein abundances between cells. A collection of in-depth analyses and genome-scale surveys have suggested how such noise arises and spreads through genetic networks, often as expected from stochastic models that predict statistical properties in terms of the rates of gene activation, transcription and translation. But noise also arises at cell division when molecules are partitioned stochastically between the two daughter cells. Here we mathematically demonstrate how such partitioning errors contribute to the non-genetic heterogeneity in a population. Our results show that partitioning errors are hard to correct, and that the resulting noise profiles closely mimic those of gene expression noise, making it remarkably difficult to separate between the two. By applying the results to previous experimental studies and distinguishing between actual creation versus mere transmission of noise we surprisingly hypothesize that much of the cell-to-cell heterogeneity that has been attributed to various aspects of gene expression instead comes from random segregation at cell division. We also propose experiments to separate between the two types of noise, and discuss future directions.

Random births and deaths of individual molecules can produce substantial fluctuations in low-abundance components that spread through reaction networks and produce large variation even in high-abundance components. The same principles apply to cell division. Segregation of molecules is always probabilistic to some extent^{1–6}, and the resulting partitioning errors between daughter cells can cause persistent fluctuations in all dependent processes. The observed noise in RNAs and proteins thus arise both from gene expression and partitioning errors (Fig. 1), but how much noise comes from either source? Genome-scale studies that compare noise levels for different genes^{7,8} have supported the same overall conclusions as the many in-depth analyses that vary expression parameters^{9,10}: gene activation, transcription and translation are noisy processes, and the observed heterogeneity is well explained by models that disregard or idealize partitioning. Here we extend the mathematical models to separate noise that arises at cell division from noise that arises during the cell cycle, allowing us to examine how much of the observed noise attributed to gene expression instead may be explained by partitioning errors.

³Correspondence: Johan_Paulsson@hms.harvard.edu.

RESULTS

Noise from partitioning or gene expression depend indistinguishably on transcription and translation rates

In the most commonly used stochastic gene expression model, a component Y (e.g. an mRNA) is produced at a constant rate, a second component X (e.g. a protein) is produced at a constant rate per Y molecule, and each X and Y molecule has an independent and exponentially distributed lifetime (see Methods for details). For mathematical simplicity, the effects of cell growth and division are approximated by considering non-growing cells¹¹ where all components instead decay at a higher rate, essentially replacing dilution with additional degradation. The squared stationary coefficient of variation in X then follows: $CV^2 = \langle x \rangle^{-1} + \langle y \rangle^{-1} (1 + \tau_x / \tau_y)^{-1}$ where τ_x and τ_y are effective average lifetimes that account for both dilution and true degradation. This summarizes several of the main effects emphasized in the stochastic gene expression literature (where expression bursts correspond to special parameter combinations¹²) and can be extended to similarly account for other sources of upstream noise¹³.

Several models^{14–16} have also explicitly considered growth and division where molecules segregate binomially between the two daughters, typically following the approach introduced in two landmark 1970s papers by Otto Berg¹⁴ and David Rigney¹⁵. The normalized variance then increases by a term $Q_x^2 = 1 / \langle x \rangle_T$ upon cell division where Q_x is the partitioning error between daughter cells (Theory Box). However, cells could achieve more ordered segregation and reduce Q_x , while disorder in the segregation machinery, such as clustering or aggregation into vesicles and organelles, instead can increase Q_x greatly. For many types of mechanisms the partitioning error is expected to follow $Q_x = A / \langle x \rangle_T$, where A is a phenotypic proportionality constant that depends on the particular mechanism¹⁷. For example, molecules that form pairs that then are split up between the two daughters¹⁷ have $Q_x = A / \langle x \rangle_T$ with $A < 1$ where the exact value of A depends on the details of pair formation, while molecules that form clusters often have $Q_x = A / \langle x \rangle_T$ with $A > 1$. Effective clustering is common in eukaryotes and possibly also in bacteria where several of the main molecular machineries appear to be localized in the cell. By extending the model above to additionally account for growth and division with partitioning errors $Q_x^2 = A_x / \langle x \rangle_T$, $Q_y^2 = A_y / \langle y \rangle_T$ and $Q_{xy} = 0$, the noise at time t in the cell cycle follows (Methods):

$$CV_t^2 = \underbrace{\frac{S_{x,t}}{\langle x \rangle_t} + \frac{S_{y,t}}{\langle y \rangle_t}}_{\text{Noise from births and deaths}} + \underbrace{\frac{A_x U_{x,t}}{\langle x \rangle_t} + \frac{A_y U_{y,t}}{\langle y \rangle_t}}_{\text{Noise from partitioning and errors}}, \quad (1)$$

were the first two terms only reflect noise that arises from probabilistic births and deaths during the cell cycle, while the second two terms only reflect the noise from partitioning errors at cell division. Importantly, though S and U are complicated functions of the time in the cell cycle and the half-lives of the two components they are both independent of the synthesis rates (Methods). For binomial partitioning, $A_x = A_y = 1$, cancellations between terms in $S + U$ conceals whether the variation originates in partitioning errors or birth-and-death noise, and the models instead reveal how the total noise depends on the abundances of mRNAs and proteins.

Theory box

Stochastic changes in abundances in growing and dividing cells are described by probabilistic chemical reactions during the cell cycle combined with a statistical rule for

partitioning of molecules at cell division. For arbitrary nonlinear, multivariate and cell cycle dependent dynamics, the state vector of abundances $\mathbf{x}$ changes as $\mathbf{x} \xrightarrow{r_k(\mathbf{x},t)} \mathbf{x} + \mathbf{s}_k$ in reaction k where $\mathbf{s}$ is a vector of integers corresponding to the net change of the reactions and r_k is the rate. This results in a Chemical Master Equation where the corresponding averages $\langle x_i \rangle$ and covariances $\sigma_{ij} = \langle \Delta x_i \Delta x_j \rangle$, where $\Delta x_i = x_i - \langle x_i \rangle$, follow

$$\frac{d\langle x_i \rangle}{dt} = \langle \sum_{k=1}^N s_{ik} r_k \rangle \quad \text{and} \quad \frac{d\sigma_{ij}}{dt} = \langle \Delta x_i \sum_{k=1}^N s_{jk} r_k \rangle + \langle \Delta x_j \sum_{k=1}^N s_{ik} r_k \rangle + \langle \sum_{k=1}^N s_{ik} s_{jk} r_k \rangle. \quad (\text{B1})$$

The averages and covariances at the beginning of generation $g+1$ can then be calculated from the values at the end (time T in the cell cycle) of generation g :

$$\langle x_i \rangle_{t=0}^{g+1} = \langle x_i \rangle_{t=T}^g / 2 \quad \text{and} \quad \left. \frac{\sigma_{ij}^{g+1}}{\langle x_i \rangle \langle x_j \rangle} \right|_{t=0} = \left. \frac{\sigma_{ij}^g}{\langle x_i \rangle \langle x_j \rangle} \right|_{t=T} + Q_{ij}^g, \quad \text{where} \quad Q_{ij}^g = \frac{\langle (L_i - R_i)(L_j - R_j) \rangle}{\langle x_i \rangle \langle x_j \rangle} \bigg|_{t=T}, \quad (\text{B2})$$

where one daughter receives L_i and L_j molecules of the two components and the other cell receives R_i and R_j copies, respectively. If each X_i molecule has an independent 50% chance of being partitioned into either daughter cell, then $Q_{ii} = 1/\langle x_i \rangle_T$ while ordered and disordered mechanisms can create lower or higher partitioning errors respectively. The off-diagonal cross-partitioning errors Q_{ij} for $i \neq j$, are zero when X_i and X_j segregate independently of each other. After several generations of balanced cell growth, the process reaches a cyclostationary state where index g above can be dropped. The analytical results presented are all exact, combining Eqs. (B1-B2) in the cyclostationary state. For notational simplicity, the two-variable examples in the text drop subscripts and use Q_x for the square root of the diagonal elements for component X .

This simple result has remarkable consequences for interpreting the experimental noise literature. The most common strategy for analyzing the sources of gene expression noise is to modify the rates of gene activation, transcription or translation, and compare changes in the variances to predictions from stochastic models. For example, if most protein noise came from random transcription of a rare mRNA, then doubling the transcription rate would double the average number of mRNAs and halve the normalized protein variance. However, as seen in Eq. (1), if all noise instead came from segregation while transcription and all other aspects of gene expression were entirely noise-free, the exact same outcome is expected: doubling the transcription rate would still halve the normalized variances (Fig. 2a) because the partitioning error would be proportionally smaller. The same is true for the many studies that measure averages and variances for many different genes under several different conditions: the observed scaling that is typically attributed to stochastic gene expression follows naturally from random segregation as well. This is not an artifact of failing to synchronize the population with position in the cell cycle, but holds true within each age-class of cells. Measurements across imperfectly synchronized populations instead mix up effects of noise with systematic differences between cells of different ages. This creates non-zero variances regardless of abundances, and thus apparent ‘noise’ terms that are independent of synthesis or degradation rates, mimicking the effects of extrinsic noise sources^{10,12}. Some extrinsic noise sources have also been traced back^{10,12} to stable repressors present in low numbers, where partitioning errors again would produce the observed scaling even if its gene expression mechanisms were nearly deterministic. Even detailed experiments on stochastic gene expression that systematically modify gene

expression parameters and compare changes in variances to stochastic models thus do not separate between noise that arises in gene expression and the effect of partitioning errors.

Partitioning errors and stochastic gene expression produce abundance distributions of similar shapes

Variances measure the width of a distribution in a single number, while the full shape of the distribution in principle contains much more information about the underlying process. However, most experimental studies have been reluctant to use higher order statistics to test models, such as the skewness or the tails of distributions, because they are less robust to unknown measurement errors or slight mistakes in the model assumptions. For example, if cell age and size are imperfectly measured, or a slight cell cycle dependent synthesis rate is unaccounted for, this typically has a much greater effect on the skewness than the variance, because the former is a 3rd rather than 2nd order statistic. To test whether distributions could be used to infer stochastic mechanisms even if all such problems could be resolved, we consider the hypothetical extreme where mRNAs and proteins segregate binomially at cell division, but synthesis and degradation during the cell cycle are approximated as deterministic so that every cell produces and degrades an average amount conditional on the abundance in each newborn cell (Methods). The resulting distribution is then compared to predictions from models where transcription, translation and degradation are stochastic but that ignore partitioning errors. Surprisingly, the resulting histograms are practically indistinguishable from a negative binomial distribution (which can be approximated by a gamma distribution in certain continuity limits¹⁸) even when measuring levels across perfectly synchronized cells – the same distribution that has been predicted in several stochastic gene expression models¹⁹, observed in many experiments^{18,20–22} and used to infer the dynamic parameters of gene expression¹⁸ (Fig. 2b). The intuitive reason is that the shape of the distribution depends not only on low-number stochastic effects – which are not so different for partitioning errors and probabilistic births and deaths – but also on the rate functions, for example the fact that mRNAs make proteins and that proteins often decay more slowly than mRNAs. Allowing for non-binomial partitioning, more complicated expression mechanisms, or other sources of upstream noise, provide even more degrees of freedom and make it even more difficult to infer the sources of noise. Distributions thus cannot be used to reliably discriminate stochastic gene expression from partitioning errors.

Partitioning errors mimic time-series of stochastic gene expression

It is tempting to assume that time series data, e.g. tracking GFP abundances during growth and division, would directly test whether noise originates from partitioning or from synthesis and degradation. But even if the two daughters received equal amounts each division while levels change randomly during the cell cycle, all noise could still be *created* at cell division. The reason is that GFPs present in high enough abundances to be reliably tracked over several generations may still fluctuate greatly, but the noise then reflects the randomizing impact of upstream factors, not spontaneous partitioning errors or probabilistic births and deaths from the protein itself. Such transmitted noise always propagates gradually over time because the affected systems cannot respond immediately to parameter changes. Partitioning errors in an upstream component in the network can thus appear like synthesis or degradation noise in a downstream component. Since different types of noises spread with different time constants, and can affect downstream components in opposite directions, a seemingly meandering concentration trajectory during the cell cycle could be purely triggered by fluctuations arising at cell division (Fig. 2c).

We previously showed that systems with different nonlinear control mechanisms¹² or distributions of waiting times between events²³ can produce very different fluctuations, but still produce exactly the same experimentally observed features, and hence cannot be

separated by the most commonly used experiments. Here we show that it is not merely the detailed features of stochastic gene expression that cannot be determined in this way – there is in fact no way of using such observations to plausibly conclude that fluctuations originate in gene expression at all, despite excellent fits. The biological consequences of such noise – whether randomizing key processes or providing advantageous heterogeneity – may be similar regardless of where it originates, with at least one potentially important difference: partitioning errors typically introduce uncorrelated deviations in all cellular components at the same time, which potentially perturbs the dynamics more than many smaller deviations spread out in time.

Feedback control or rapid turn-over can increase the impact of partitioning errors

Cells could in principle correct partitioning errors after they arise, either using closed-loop feedback control that boosts synthesis in the daughter that receives too few copies, or open-loop mechanisms that simply set the half-lives such that deviations are short-lived.

But closed-loop noise suppression is highly non-trivial at the molecular level. Production and degradation processes inevitably create lags and delays along the loop, which can destabilize the dynamics and increase fluctuations instead. Most molecular control systems must also rely on intermediate species that are made in a small number of probabilistic reactions and therefore can randomize the response^{24,25}. Both types of challenges are particularly severe when suppressing partitioning errors: because new errors arise each cell cycle, they must be effectively corrected on a time scale of a fraction of a cell cycle to significantly reduce the heterogeneity. The relative durations of lags and delays are then long in comparison (destabilizing the dynamics), and the number of intermediate chemical events integrated over the effective time-scale of control is low (randomizing the response). Segregation also introduces large deviations to both higher and lower abundances. A system controlling synthesis rates may efficiently correct the down-fluctuations as well as prevent the synthesis machinery from creating up-fluctuations, but the up-fluctuations caused by partitioning errors would rely on dilution and degradation to be gradually eliminated. A system controlling degradation faces the opposite challenge, and cannot rapidly eliminate down-fluctuations generated by partitioning errors. Efficiently correcting partitioning errors thus requires both synthesis and degradation control, while noise arising from random births and deaths in principle can be efficiently suppressed by either one.

In any real system, a myriad of challenges combine to determine the exact effect of control, and in many cases negative feedback would *increase* rather than decrease fluctuations by randomizing or destabilizing the response. These principles are long established in control theory – Maxwell²⁶ mathematically analyzed how negative feedback can be destabilized by delays more than half a century before Cannon first coined the term ‘homeostasis’ in biology²⁷ – and cannot be done justice here, but a simple illustration of one such challenge is shown in Fig. 3a.

Open-loop control instead supposedly reduces the impact of partitioning errors by decreasing the half-life of the components, and thereby ensures that any post-segregation shortage or surplus is short-lived²⁸. But this apparently straightforward strategy also has serious caveats. First, shorter lifetimes require proportionally higher production rates to maintain the same abundance, and many macromolecules are too metabolically expensive to produce at high rates – at least at a genome-wide scale since a large fraction of the cell consists of RNAs and proteins. Second, and more counter-intuitively, decreasing half-lives will not necessarily reduce the total effect of partitioning errors, even when comparing for the same average abundance. Apart from being exposed to fluctuations in the degradation machinery, components with short half-lives merely adjust rapidly to the quasi-steady states set by the current parameters for synthesis and degradation. If these parameters fluctuate

over time, for example due to random segregation of upstream factors in the network, short-lived components are more susceptible to respond while long-lived components²⁰ can ‘time-average’ out some of the noise. For example consider cases where a long-lived transcription factor²⁹ segregates randomly, creating fluctuations in transcription rates from cell to cell. The abundance of a TF-regulated mRNA is then randomized both by the segregation of the mRNA itself and the segregation of the TF. A short mRNA lifetime reduces the impact of the first source of error, but instead amplifies the impact of the second, as shown in Fig. 3b. A short-lived TF would not have this effect, but would also fail to randomize gene expression because of time-averaging effects.

Reducing half-lives may thus increase the effect of segregation errors, increase the metabolic burden of cells, and could potentially affect a component’s dynamic response to stimuli in negative ways.

Experiments on partitioning errors versus gene expression noise

Is the strong focus on stochastic synthesis rather than on degradation or partitioning well motivated considering existing experiments? There are effectively three groups of studies:

- Segregation studies report a wide range of partitioning error for organelles, including independent partitioning of endosomes³⁰, lysosomes³⁰, and *Chlorella*³¹, and ordered partitioning of carboxysomes³², mitochondria^{33–35}, Golgi vesicles³⁶, and chloroplasts^{37,38}. Macromolecules have been less studied, but several reports suggest that mRNAs are partitioned binomially or in a mildly disordered manner²⁰, while the segregation of at least one fluorescent protein in *E. coli* shows³⁹ $Q_x = A / \langle x \rangle_T$
- In the stochastic gene expression literature, a few studies of simple model systems in *E. coli* have directly tracked the production of molecules with single-molecule accuracy^{20,40,41}, and shown exponential time intervals between birth events.
- Numerous in-depth studies measured the overall variation across the population^{7–10}. By testing how the width of the distributions change with the rates of gene activation, transcription and translation, these studies confirmed predictions from accompanying stochastic models. Yet other studies have used natural differences between different genes to survey how the variation correlates with average expression rates, covering much of the genome⁴² and measuring distributions under numerous different growth conditions⁴³.

Several studies have thus directly measured substantial partitioning errors, a few have directly measured birth-and-death noise, and the overwhelming majority have measured the total heterogeneity across cells and indirectly inferred the sources of noise by fitting models. This third group has almost indiscriminately been claimed in support of the stochastic gene expression hypothesis because the results fit gene expression models ignoring partitioning, while Eq. (1) and Figure 2 surprisingly show that they fit equally well to partitioning models that ignore gene expression noise. How can such issues be settled? Our results show that even a noisy time series during the cell cycle can be entirely caused by partitioning errors, because the effects of such errors are transmitted gradually over time, and that single molecule accuracy is required to identify the effect of probabilistic births and deaths. Dual reporter methods¹⁰, though highly useful for other purposes, are also unable to separate between these two types of noise: partitioning errors in the extrinsic factors contribute to the extrinsic noise, while partitioning errors in the intrinsic components contribute to the intrinsic noise. However, though measuring noise levels as a function of synthesis rates cannot distinguish *how* fluctuations arise, it still suggests *which* component introduces spontaneous low-copy fluctuations. Once such a component is identified, it could potentially

be labeled and monitored over time through cell division to determine the relative importance of births and deaths versus partitioning errors. For example, if the normalized variance in GFP abundance decreases proportionally with the rate of transcription but is largely unaffected by translation, the next step is to monitor the corresponding RNA in single cells. Fortunately, single cell methods are becoming increasingly quantitative^{10,20,21,40,41,44,45}, and it is only a matter of time before these matters can be routinely settled.

DISCUSSION

Numerous studies have sought to identify the molecular sources of non-genetic heterogeneity in cells, or to characterize the transmission, suppression or exploitation of fluctuations. Most such results have been interpreted in the light of mathematical models that ignore or idealize partitioning errors at cell division, and for good reason: the added algebraic complexity of accounting for non-trivial segregation in growing and dividing cells easily conceals other principles and prevents effective analysis. Most results and conclusions from those analyses will also hold regardless of the biological source of noise, and approximations are often necessary to make real progress. However, for one of the most central challenges – determining the mechanistic origin of fluctuations – such realism is key.

Our results show that noise arising from random segregation at cell division is difficult to suppress, and that it closely mimics gene expression noise. But what fraction of the fluctuations reflects either process? If molecules were made, degraded, and partitioned as independent units, the randomness introduced by degradation and segregation would be 33–75% of the total variance, depending on the time in the cell cycle (Methods). If the half-life is comparable to the cell generation time segregation and degradation contribute similarly, while for the many stable organelles, proteins, and even many RNAs, segregation dominates entirely.

A closer inspection of synthesis and degradation at first seems to upset this picture. Both processes can produce much more noise than expected from independent reactions, though degradation has been understudied, for example due to burst-like events or extrinsic noise sources. However, it is important to distinguish between processes that merely pass on fluctuations to downstream components, and the ‘prime movers’ that spontaneously generate noise due to the probabilistic nature of low-number processes⁴⁶. For example, burst-like translation is simply a special case of noise transmission where the prime movers are low-abundance mRNAs or active genes. The same is true for extrinsic noise sources, that often reflect spontaneously random events elsewhere in the cell. For each such spontaneously random upstream process, the noise can again be broken down into probabilistic synthesis, degradation, and segregation, much like the above. Though much noise is *transmitted* via synthesis and degradation, the same fraction may thus still *originate* in random segregation.

Interestingly, the reverse is typically not true: segregation in clusters or vesicles can greatly increase the partitioning error regardless of synthesis or degradation noise. A substantial part of the disorder in synthesis and degradation thus originates in random segregation, while disorder in segregation typically reflects spatial heterogeneity rather than randomness in synthesis and degradation. The asymmetry leads to a remarkable conclusion: from first principles we may well expect that random segregation, including its indirect effects, can be as great a source of noise as gene expression – possibly even much greater. This may seem unorthodox in the light of the extensive canon of literature on stochastic gene expression and cellular heterogeneity, but consider the foundations of that canon and our counter-arguments:

- *Numerous studies demonstrate that cellular noise depends on transcription and translation rates as expected from models of stochastic gene expression.* Here we show that these studies are equally consistent with all noise arising at segregation instead, because the changes in the kinetic parameters of gene expression have the same relative effects on the noise coming from partitioning errors.
- *Synthesis is often disordered or bursty, producing more noise than a simple Poisson process for independent events, while partitioning errors often seem to follow binomial statistics and independent segregation.* Here we show that the disorder transmitted via synthesis or degradation may originate in random segregation of upstream elements, while partitioning can be disordered by spatial effects without stochastic synthesis or degradation.
- *It seems cells could simply eliminate the effect of partitioning errors by correcting perturbations after they arise, or by decreasing the half-life of the molecules.* Here we show how surprisingly difficult such strategies are to implement effectively, and how easily they could have the opposite effect.

These results do not rely on controversial assumptions, but rather on eliminating mathematical approximations when evaluating the experimental literature on both gene expression and cell division. Testing the hypothesis that partitioning errors are behind much of the observed heterogeneity will be challenging – especially since many results that appear to favor one hypothesis in fact equally favor its opposite – but is also of fundamental importance to quantitatively understand life in single cells.

Methods

Many models in the main text are versions of the following birth and death process

$$\begin{array}{l} y \xrightarrow{\lambda_1} y+n_1 \quad x \xrightarrow{\lambda_2 y/n_2} x+n_2 \\ y \xrightarrow{\beta_1 y} y-1 \quad x \xrightarrow{\beta_2 x} x-1 \end{array}, \quad (M1)$$

for times $0 < t < T$ in the cell cycle, where T is the time of cell division when each chemical species segregates with arbitrary statistics (Theory Box). Here we present cyclostationary solutions and details for the examples.

Analytical expression for mRNA-protein model with arbitrary partitioning errors (Eq. (1))

When mRNAs (y) and their corresponding proteins (x) are made one at a time ($n_1=n_2=1$) the average dynamics follows $\partial_t \langle y \rangle = \lambda_1 - \beta_1 \langle y \rangle$ and $\partial_t \langle x \rangle = \lambda_2 \langle y \rangle - \beta_2 \langle x \rangle$, while the variances and covariance follow

$$\begin{aligned} \frac{d\sigma_{yy}}{dt} &= -2\beta_1\sigma_{yy} + \lambda_1 + \beta_1 \langle y \rangle, & \frac{d\sigma_{xy}}{dt} &= -(\beta_2 + \beta_1)\sigma_{xy} + \lambda_2\sigma_{yy} \\ \frac{d\sigma_{xx}}{dt} &= -2\beta_2\sigma_{xx} + 2\lambda_2\sigma_{xy} + \lambda_2 \langle y \rangle + \beta_2 \langle x \rangle \end{aligned}$$

with boundary conditions $4\sigma_{yy,0} = \sigma_{yy,T} + A_y \langle y \rangle_T$, $4\sigma_{xx,0} = \sigma_{xx,T} + A_x \langle x \rangle_T$ and $\sigma_{xy,0} = \sigma_{xy,T}$. For convenience let

$$R = \frac{\beta_2}{\beta_1}, r = \frac{3}{T} \left(\frac{\beta_2 - \beta_1}{\beta_1 \beta_2} \right), c_{i,t} = \frac{1}{\beta_i T} \left(1 - \frac{e^{-\beta_i t}}{2 - e^{-\beta_i T}} \right) \quad \text{and} \quad k_{i,j,t} = \frac{e^{-(\beta_i + \beta_j)t}}{4 - e^{-(\beta_i + \beta_j)T}}.$$

The averages then follow

$$\langle y \rangle_t = \lambda_1 T c_{1,t} \quad \text{and} \quad \langle x \rangle_t = \lambda_1 \lambda_2 T \frac{c_{1,t} - c_{2,t}}{\beta_2 - \beta_1} \quad (\text{M2})$$

and the normalized variances

$$\text{CV}_{y,t}^2 = \left[1 + k_{11,t} \frac{c_{1,t}}{c_{1,t}} (A_y - 1) \right] \frac{1}{\langle y \rangle_t} \quad \text{and} \quad \text{CV}_{x,t}^2 = (S_{x,t} + U_{x,t} A_x) \frac{1}{\langle x \rangle_t} + (S_{y,t} + U_{y,t} A_y) \frac{1}{\langle y \rangle_t} \quad (\text{M3})$$

where

$$\begin{aligned} S_{x,t} &= 1 - U_{x,t}, \quad U_{x,t} = k_{22,t} \frac{c_{1,t} - c_{2,t}}{c_{1,t} - c_{2,t}} \\ S_{y,t} &= \left[\frac{r + c_{1,t}}{1 - 2R} k_{22,t} + 2 \left(\frac{r}{R+1} + \frac{c_{1,t}}{R} \right) k_{12,t} - c_{1,t} k_{11,t} + \frac{R-1}{2R-1} \left(2 \frac{(R-1)}{R} c_{1,t} - \frac{r}{R+1} \right) \right] \frac{c_{1,t}}{(c_{1,t} - c_{2,t})^2} \\ U_{y,t} &= [k_{22,t} - 2k_{12,t} + k_{11,t}] \frac{c_{1,t} c_{1,t}}{(c_{1,t} - c_{2,t})^2}. \end{aligned}$$

This shows how terms S and U are independent of synthesis parameters λ , but also how algebraically complicated the solutions become even for simple cases.

Parameters used for numerical illustrations (Fig. 2)

In this model the mRNA (y) and protein (x) follow as above with $n_1 = 1$ and $\beta_2 = 0$ (no protein degradation), $\beta_1 = 0.07$, $\lambda_2 = 0.135$, $T = 50$, and λ_1 ranging from 12.6 to 84 (except for the example denoted by triangles: see below) which changes the average number of protein while keeping the time-constant and the mRNA-induced noise fixed. The values of y and x were sampled at $0.44T$ after the cyclostationary state was reached.

For proteins with bursty synthesis and independent partitioning (circle), n_2 is a random variable assumed to follow a geometric distribution¹⁹ with average of 4.4. The corresponding conventional degradation approximation (solid gray line) is plotted according to $\text{CV}_x^2 = (n_2 + 1) \langle x \rangle^{-1} + (1 + (\beta_1 + k)/(\beta_2/k))^{-1} \langle y \rangle^{-1}$, where $k = \ln(2)/T$. When mRNAs and proteins follow deterministic synthesis and degradation and partitioning errors are the only sources of noise (squares and triangles), the value of y and x in each cells during the cell growth are determined by numerical integration of the differential equations $dy/dt = \lambda_1 - \beta_1 y$ and $dx/dt = \lambda_2 y$. The values are then discretized. For the clustered partitioning example (squares), the number of vesicles (v) is sampled from a Poisson distribution with mean x/s' with $s' = 13$ so that $\langle v|x \rangle = x/s'$, then each protein is allocated into any of these vesicles with probability of $1/v$ (i.e., multinomial process), and vesicles are independently partitioned into each daughter cells. For the independent partitioning and deterministic synthesis and degradation (triangle), λ_1 is 60-fold reduced and λ_2 is 60-fold increased from the rate constants used in the rest of the simulations, and all the other rates are identical. For all three cases except for the degradation approximation, the remaining mRNAs at the end of the cell cycle were independently partitioned. The strategy used for generating the triangles of Fig.

2a was also used for the distribution result in Fig. 2b. with parameters $\lambda_1=0.03$, $\beta_1=0.004$, $\lambda_2=1.5$, and $T=50$, and the size of binning in the histogram is 22. For Fig. 2c we include four chemical species: the protein of interest (x_2), its protease (x_4), and their corresponding mRNAs (x_1 and x_3 , respectively):

$$\begin{aligned} dx_1/dt &= \lambda_1 - \beta_1 x_1 & dx_3/dt &= \lambda_3 - \beta_3 x_3 \\ dx_2/dt &= \lambda_2 x_1 - \beta_2 x_2 - \gamma x_4 x_2 & dx_4/dt &= \lambda_4 x_3 - \beta_4 x_4, \end{aligned}$$

with parameters $\lambda_1=0.25$, $\beta_1=0.01$, $\lambda_2=400$, $\beta_2=0.42$, $\gamma=0.01$, $\lambda_3=0.2$, $\beta_3=0.03$, $\lambda_4=3.0$, $\beta_4=0.5$, and $T=50$.

Partitioning errors are difficult to correct (Fig. 3)

For Fig. 3a we simulated three chemical species, all of which follow deterministic reactions during the cell cycle and independent partitioning at cell division. For the open loop system, we consider $dw/dt=\lambda_1-\beta_1 w$, $dx/dt=\lambda_2 w-\beta_2 x$, and $dy/dt=\lambda_3 x-\beta_3 y$, and for the negative feedback system λ_1 is substituted by $\lambda_0 K^h/(K^h+y^h)$ while the rest of the reactions are unchanged. The parameters are $\lambda_1=0.7$, $\beta_1=0.01$, $\lambda_2=0.032$, $\beta_2=0.005$, $\lambda_0=5.0$, $K=50$, $h=3$, $T=50$, and varying β_3 from 0.0281 to 20 and λ_3 from 0.0001 to 10 accordingly. The CVs are sampled at $t=T/2$, and plotted as a function of the average life time of y normalized over the cell cycle time, $\ln 2/\beta_3/T$. For the TF-mRNA model (Fig. 3b) the process for the stable TF (y) and its target mRNA (x) corresponds to $\beta_1=0$ and $n_1=n_2=1$ in (M1). The averages and variances can be obtained by taking the limit of $\beta_1 \rightarrow 0$ in Eqs. (M2) and (M3). In Fig. 3b, we plotted $U_x(t)A_x/\langle x \rangle_t$, $U_y(t)A_y/\langle y \rangle_t$, and their sum at $t=T/2$, as a function of the relative average mRNA life time, $\ln 2/\beta_2/T$, with $T=50$, $A_x=A_y=1$, $\lambda_1=0.1$, and varying λ_2 to ensure a constant average $\langle x \rangle_0=25$ despite of changing β_2 . With these parameters the noise coming from the TF is large enough that fast degradation of the mRNA can increase the total noise.

Fraction of noise coming from independent synthesis, degradation, and partitioning

To separate the sources of y noise in (M1), the degradation step-size is set to be n_3 instead of 1, and the process for y is solved for arbitrary n_1 , n_3 , and A_y . To compare the contribution of different origin of noise, we then return to set $n_1=n_3=A_y=1$, which gives

$$\langle y_t \rangle = \lambda_1 T c_{1,t} \quad \text{and} \quad CV_t^2 = \left[\underbrace{\frac{1-3k_{11,t}}{2\beta_1 T c_{1,t}}}_{\text{synthesis}} + \underbrace{\frac{\beta_1 T}{2} c_{1,t}}_{\text{degradation}} + \underbrace{\frac{c_{1,t} k_{11,t}}{c_{1,t}}}_{\text{Partitioning}} \right] \frac{1}{\langle y \rangle_t}.$$

Depending on parameters, the fraction of noise originating from synthesis is between 1/4 (fast degradation, at the beginning of the cell cycle) and 2/3 (no degradation, at the end of the cell cycle). Hence the total randomness introduced by degradation and segregation is always between 33–75% of the total variance. When the half-life equals the cell generation time ($\beta_1 T = \ln 2$), degradation and segregation contribute roughly equally when compared at time $t=(1/\ln 2-1)T$ in the cell cycle, corresponding to the average cell stage of the exponentially growing cell population.

References

1. Warren G. Membrane Partitioning During Cell Division. *Ann Rev Biochem.* 1993; 62:323–348. [PubMed: 8352593]

2. Marshall WF. Stability and robustness of an organelle number control system: modeling and measuring homeostatic regulation of centriole abundance. *Biophys J.* 2007; 93:1818–33. [PubMed: 17496020]
3. Birky CW Jr. The partitioning of cytoplasmic organelles at cell division. *Int Rev Cytol Suppl.* 1983; 15:49–89. [PubMed: 6343284]
4. Nordstroem K, Austin SJ. Mechanisms that contribute to the stable segregation of plasmids. *Ann Rev Genet.* 1989; 23:37–69. [PubMed: 2694936]
5. Mantzaris NV. From Single-Cell Genetic Architecture to Cell Population Dynamics: Quantitatively Decomposing the Effects of Different Population Heterogeneity Sources for a Genetic Network with Positive Feedback Architecture. *Biophys J.* 2007; 92:4271–4288. [PubMed: 17384073]
6. Cookson NA, Cookson SW, Tsimring LS, Hasty J. Cell cycle-dependent variations in protein concentration. *Nucleic Acids Research.* 2009; 38:2676–2681. [PubMed: 20019065]
7. Blake WJ, Kaern M, Cantor CR, Collins JJ. Noise in eukaryotic gene expression. *Nature.* 2003; 422:633–637. [PubMed: 12687005]
8. Raser JM, O'Shea EK. Control of Stochasticity in Eukaryotic Gene Expression. *Science.* 2004; 304:1811–1815. [PubMed: 15166317]
9. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A. Regulation of noise in the expression of a single gene. *Nature Genetics.* 2002; 31:69–73. [PubMed: 11967532]
10. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. *Science.* 2002; 297:1183–6. [PubMed: 12183631]
11. Paulsson J, Ehrenberg M. Noise in a minimal regulatory network: Plasmid copy number control. *Quarterly Reviews of Biophysics.* 2001; 34:1–59. [PubMed: 11388089]
12. Paulsson J. Summing up the noise in gene networks. *Nature.* 2004; 427:415–8. [PubMed: 14749823]
13. Paulsson J. Models of stochastic gene expression. *Phys Life Rev.* 2005; 2:157–175.
14. Berg OG. A model for the statistical fluctuations of protein numbers in a microbial population. *J Theor Biol.* 1978; 71:587–603. [PubMed: 96307]
15. Rigney DR. Stochastic model of constitutive protein levels in growing and dividing bacterial cells. *J Theor Biol.* 1979; 76:453–80. [PubMed: 439915]
16. Swain PS, Elowitz MB, Siggia ED. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci U S A.* 2002; 99:12795–800. [PubMed: 12237400]
17. Huh D, Paulsson J. in preparation.
18. Taniguchi Y, et al. Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science.* 2010; 329:533–538. [PubMed: 20671182]
19. Paulsson J, Berg OG, Ehrenberg M. Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation. *Proc Natl Acad Sci U S A.* 2000; 97:7148–53. [PubMed: 10852944]
20. Golding I, Paulsson J, Zawilski SM, Cox EC. Real-time kinetics of gene activity in individual bacteria. *Cell.* 2005; 123:1025–36. [PubMed: 16360033]
21. Zenklusen D, Larson DR, Singer RH. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat Struct Mol Biol.* 2008; 15:1263–1271. [PubMed: 19011635]
22. Chubb JR, Trcek T, Shenoy SM, Singer RH. Transcriptional Pulsing of a Developmental Gene. 2006; 16:1018–1025.
23. Pedraza JM, Paulsson J. Effects of Molecular Memory and Bursting on Fluctuations in Gene Expression. *Science.* 2008; 319:339–343. [PubMed: 18202292]
24. Paulsson J, Ehrenberg M. Random Signal Fluctuations Can Reduce Random Fluctuations in Regulated Components of Chemical Regulatory Networks. *Physical Review Letters.* 2000; 84:5447–5450. [PubMed: 10990965]
25. Lestas I, Vinnicombe G, Paulsson J. Fundamental limits on the suppression of molecular fluctuations. *Nature.* (In press).
26. Maxwell JC. On Governors. *Proceedings of the Royal Society.* 1867; 16:270–283.
27. Cannon WB. *The Wisdom of the Body.* 1932
28. Kar S, Baumann WT, Paul MR, Tyson JJ. Exploring the roles of noise in the eukaryotic cell cycle. *Proc Natl Acad Sci U S A.* 2009; 106:6471–6. [PubMed: 19246388]

29. Palsson BO, Lightfoot EN. Mathematical modelling of dynamics and control in metabolic networks. V. Static bifurcations in single biochemical control loops. *J Theor Biol.* 1985; 113:279–98. [PubMed: 3999779]
30. Bergeland T, Widerberg J, Bakke O, Nordeng TW. Mitotic partitioning of endosomes and lysosomes. *Curr Biol.* 2001; 11:644–51. [PubMed: 11369227]
31. McAuley PJ. Partitioning of Symbiotic *Chlorella* at Host Cell Telophase in the Green Hydra Symbiosis. *Philosophical Transactions: Biological Sciences.* 1990; 329:47–53.
32. Savage DF, Afonso B, Chen AH, Silver PA. Spatially Ordered Dynamics of the Bacterial Carbon Fixation Machinery. *Science.* 2010; 327:1258–1261. [PubMed: 20203050]
33. Wilson EB. The Distribution of the Chondriosomes to the Spermatozoa in Scorpions. *Proc Natl Acad Sci U S A.* 1916; 2:321–324. [PubMed: 16576154]
34. Wilson EB. The distribution of sperm-forming materials in scorpions. *J Morphol Physiol.* 1931; 52:429–483.
35. Hood RD, Watson OF, Deason TR, Benton CLB Jr. Ultrastructure of scorpion spermatozoa with atypical axonemes. *Cytobios.* 1972; 5:167–177. [PubMed: 5044627]
36. Shima DT, Haldar K, Pepperkok R, Watson R, Warren G. Partitioning of the Golgi apparatus during mitosis in living HeLa cells. *J Cell Biol.* 1997; 137:1211–28. [PubMed: 9182657]
37. Hennis AS, Birky CW. Stochastic partitioning of chloroplasts at cell division in the alga *Olithodiscus*, and compensating control of chloroplast replication. *J Cell Sci.* 1984; 70:1–15. [PubMed: 6389575]
38. Sheahan MB, Rose RJ, McCurdy DW. Organelle inheritance in plant cell division: the actin cytoskeleton is required for unbiased inheritance of chloroplasts, mitochondria and endoplasmic reticulum in dividing protoplasts. *Plant J.* 2004; 37:379–90. [PubMed: 14731258]
39. Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB. Gene regulation at the single-cell level. *Science.* 2005; 307:1962–5. [PubMed: 15790856]
40. Yu J, Xiao J, Ren X, Lao K, Xie XS. Probing gene expression in live cells, one protein molecule at a time. *Science.* 2006; 311:1600–3. [PubMed: 16543458]
41. Cai L, Friedman N, Xie XS. Stochastic protein expression in individual cells at the single molecule level. *Nature.* 2006; 440:358–362. [PubMed: 16541077]
42. Newman JR, et al. Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature.* 2006; 441:840–6. [PubMed: 16699522]
43. Bar-Even A, et al. Noise in protein expression scales with natural protein abundance. *Nature Genetics.* 2006; 38:636–643. [PubMed: 16715097]
44. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* 2006; 4:e309. [PubMed: 17048983]
45. Gruewald, D.; Singer, RH.; Czaplinski, K.; Lynne, EMaMK. *Methods in Enzymology.* Vol. 448. Academic Press; 2008. Chapter 27 Cell Biology of mRNA Decay; p. 553-577.
46. Paulsson J. Prime movers of noisy gene expression. *Nat Genet.* 2005; 37:925–926. [PubMed: 16132049]

Figure 1. Growing and dividing cells

(a) Cartoon of an individual cell line and segregating units (dots) followed through rounds of growth and division. (b) Sample time trace of copy number per cell (gray) and their average (black). Random changes are due to births and deaths during the cell cycle, and segregation at cell division (time T).

Figure 2. Partitioning errors mimic gene expression noise

(a) The CV in protein numbers per cell half-way through the cell cycle as a function of average protein level for different mechanisms of gene expression and segregation. The solid line (—) corresponds to the gene expression model above with probabilistic births and deaths but no cell cycle or division. For all other scenarios, mRNAs are made at constant intensities, degraded exponentially, and independently partitioned at cell division, while a stable protein is produced with constant intensity per transcript. Circles (○) correspond to an average burst size of 4.4 and independent protein segregation. Squares (□) correspond to clustered protein segregation with an average of 13 proteins per vesicle and a Poisson distributed number of vesicles with an average value proportional to the number of proteins, while all other reactions are modeled as deterministic. Triangles (Δ) correspond to independent partitioning of proteins and deterministic reactions during cell growth. (b) Bars correspond to a model in which mRNAs and proteins are made deterministically during the cell cycle, then discretized and independently partitioned at cell division. The distribution is fitted to a negative binomial (line) that is expected from commonly used stochastic gene expression models. (c) Protein time-series for a simplistic model including a protein, its unsaturated protease, and the two corresponding mRNAs. All synthesis and degradation reactions are approximated as deterministic and discretized, and heterogeneity is only created through independent partitioning errors (Methods), that eventually are transmitted to protein levels. The assumptions about deterministic effects in a–c are of course physically unrealistic and included only to demonstrate the fits even in extreme cases.

Figure 3. Partitioning errors are difficult to effectively correct during the cell cycle

(a) Negative feedback can exacerbate the effect of partitioning errors. Here we consider an open loop system where component w is synthesized at a constant rate, components x and y are synthesized at constant rates per w and x molecule respectively, and all three components are degraded exponentially. Synthesis and degradation reactions are modeled as deterministic and independent partitioning of each component is the only source of randomness. The negative feedback model system is identical except that the synthesis rate of w is down-regulated by y according to a negative Hill function with a Hill coefficient of three (Method). As the half-life of y relative to the cell cycle time increases, the closed loop system exhibits noisier behavior of x than the open loop. The average abundances of each species are kept the same for the open and the closed loop system, and for all half-lives, by changing the synthesis rate constant of w . (b) The total effect of random segregation can increase with shorter half-lives of the components. The mRNA noise originating from segregation, assuming a stable TF is made at a constant intensity and that the mRNA birth rate is proportional to the number of TF molecules, where both TF and mRNA molecules segregate independently. The CV is evaluated half-way through the cell cycle, with 5 TFs and 25 mRNAs per average newborn cell. As the mRNA half-life relative to the cell cycle time increases, the randomizing effect of its own segregation (green) increases, but the transmitted fluctuations from random TF segregation (red) instead goes through a maximum and then decreases. Shorter lifetimes can thus increase the contribution of random segregation to the population heterogeneity (purple). See Methods for derivations and details.