The dynamics of protein noise can distinguish between alternate sources of gene-expression variability

Supplementary information

Contents

List of Supplementary Figures

Appendix A: Steady-state moment analysis of stochastic gene expression model

Consider a gene expression model where each expression event creates a burst of B mRNA molecules, where B is an arbitrary random variable with Probability$\{B = i\} = \alpha_i$ for $i \in \{1, 2, 3, \ldots\}$. The average number of mRNAs produced per expression event is then given by $\langle B \rangle := \sum_{i=1}^{\infty} i\alpha_i$. We denote by $\langle B^2 \rangle := \sum_{i=1}^{\infty} i^2\alpha_i$ the second order statistical moment of the burst size B. A special case of the model would be constitutive gene expression where mRNAs are made one at time and corresponds to $B = 1$ with probability one. Let $m(t)$ denote the number of mRNA molecules at time t. We treat mRNA expression and degradation events as stochastic events with probabilities of occurring in an infinitesimal time interval $(t, t + dt]$ given by

$$\text{Probability}\{m(t + dt) = m + i \mid m(t) = m\} = k_m \alpha_i dt, \quad \forall i \geq 1 \tag{1a}$$

$$\text{Probability}\{m(t + dt) = m - 1 \mid m(t) = m\} = \gamma_m m dt, \tag{1b}$$

respectively, where k_m is the rate of expression events and γ_m is the mRNA degradation rate. Based on the above formulation, the probability

$$P_j(t) := \text{Probability}\{m(t) = j\}, \ j \in \{0, 1, 2, 3, \ldots\} \tag{2}$$

evolves according to the Chemical Master Equation given by Eq. 1 in the paper. As the protein population count is typically large, the dynamics of

protein levels is modeled deterministically through an ordinary differential equation (Eq. 2 in the paper).

We next derive differential equations that describe the time evolution of the statistical moments of $m(t)$ and $p(t)$. For the above gene expression model we have

$$\begin{aligned}\frac{d\langle\psi(m,p)\rangle}{dt} &= \langle k_m\psi(m+B,p) + \gamma_m m\psi(m-1,p) - \psi(m,p)(k_m+\gamma_m m)\rangle \\ &\quad + \left\langle \frac{\partial\psi(m,p)}{\partial p}(k_p m - \gamma_p p)\right\rangle \end{aligned} \tag{3}$$

where $\psi(m,p)$ is any continuously differentiable function and $\langle . \rangle$ denotes the expected value of the corresponding quantity. Taking $\psi(m,p) = m^i p^j$ in (3) for appropriate integers i and j, the time evolution of the first and second order statistical moments of the population count are given by

$$\frac{d\langle m\rangle}{dt} = k_m\langle B\rangle - \gamma_m\langle m\rangle \tag{4a}$$

$$\frac{d\langle p\rangle}{dt} = k_p\langle m\rangle - \gamma_p\langle p\rangle \tag{4b}$$

$$\frac{d\langle m^2\rangle}{dt} = k_m\langle B^2\rangle + \gamma_m\langle m\rangle + 2k_m\langle B\rangle\langle m\rangle - 2\gamma_m\langle m^2\rangle \tag{4c}$$

$$\frac{d\langle mp\rangle}{dt} = k_p\langle m^2\rangle + k_m\langle B\rangle\langle p\rangle - \gamma_m\langle mp\rangle - \gamma_p\langle mp\rangle \tag{4d}$$

$$\frac{d\langle p^2\rangle}{dt} = 2k_p\ \langle mp\rangle - 2\gamma_p\langle p^2\rangle. \tag{4e}$$

Setting the left-hand-side of equation (4) to zero, and solving for the steady-

state moments we obtain

$$\overline{\langle m\rangle} = \frac{k_m\langle B\rangle}{\gamma_m}, \quad \overline{\langle p\rangle} = \frac{k_p\overline{\langle m\rangle}}{\gamma_p} \tag{5a}$$

$$\overline{\langle m^2\rangle} = \overline{\langle m\rangle}^2 + \overline{\langle m\rangle}\frac{\langle B^2\rangle + \langle B\rangle}{2\langle B\rangle} \tag{5b}$$

$$\overline{\langle mp\rangle} = \overline{\langle m\rangle}\ \overline{\langle p\rangle} + \overline{\langle p\rangle}\frac{\langle B^2\rangle + \langle B\rangle}{2\langle B\rangle}\frac{\gamma_p}{\gamma_p+\gamma_m} \tag{5c}$$

$$\overline{\langle p^2\rangle} = \overline{\langle p\rangle}^2 + \frac{\overline{\langle p\rangle}^2}{\overline{\langle m\rangle}}\frac{\langle B^2\rangle + \langle B\rangle}{2\langle B\rangle}\frac{\gamma_p}{\gamma_p+\gamma_m} \tag{5d}$$

where a bar denotes the steady-state value of the corresponding moment.

We quantify the steady-state protein noise level through the coefficient of variation squared defined as

$$\overline{CV^2} := \frac{\overline{\langle p^2\rangle} - \overline{\langle p\rangle}^2}{\overline{\langle p\rangle}^2}. \tag{6}$$

From (5) we find that the steady-state protein noise level is given by

$$\overline{CV^2} = \frac{\eta_m\gamma_p}{\overline{\langle m\rangle}(\gamma_p+\gamma_m)} \tag{7}$$

where

$$\eta_m := \frac{\overline{\langle m^2\rangle} - \overline{\langle m\rangle}^2}{\overline{\langle m\rangle}} = \frac{\langle B^2\rangle + \langle B\rangle}{2\langle B\rangle}. \tag{8}$$

is the steady-state Fano factor of the mRNA population count. Note that when gene expression noise arises from mRNA birth/death fluctuations (i.e., gene expression is constitutive and $B = 1$ with probability one), then $\eta_m = 1$

and the protein noise level reduces to

$$\frac{\gamma_p}{\overline{\langle m \rangle}(\gamma_p + \gamma_m)}. \tag{9}$$

In light of equation (9), the protein noise levels can be decomposed into two components:

$$\overline{CV^2} = \frac{\gamma_p}{\overline{\langle m \rangle}(\gamma_p + \gamma_m)} + \frac{(\eta_m - 1)\gamma_p}{\overline{\langle m \rangle}(\gamma_p + \gamma_m)}, \tag{10}$$

where the first component represents noise from mRNA birth/death fluctuations. The second component, which is non-zero only when $\eta_m > 1$, represents expression noise arising from stochastic promoter fluctuations between different transcriptional states (transcriptional bursting).

Appendix B: Connection between mRNA Fano factor and mean transcriptional burst size

Consider a two-state promoter model where the promoter stochastically transitions between an inactive (G_{off}) and active state (G_{on}). The promoter is represented by the following set of chemical reactions:

$$G_{off} \xrightarrow{k_{on}} G_{on}, \quad G_{on} \xrightarrow{k_{off}} G_{off}$$
$$G_{on} \xrightarrow{k_m} G_{on} + mRNA, \quad mRNA \xrightarrow{\gamma_m} \emptyset$$

where k_m is the rate of transcription from the active state and k_{on}, k_{off} are rate of transitions between the states. Let $g(t)$ denote the state of the

promoter, with $g(t) = 1$ and $g(t) = 0$ denoting that the promoter is active or inactive, respectively. We recall from Appendix A that $m(t)$ represents the mRNA population count at time t. Then, the time evolution of statistical moments are given by the following set of differential equations

$$\frac{d\langle g\rangle}{dt} = k_{on} - (k_{off} + k_{on})\langle g\rangle, \quad \frac{d\langle m\rangle}{dt} = k_m\langle g\rangle - \gamma_m\langle m\rangle \tag{11a}$$

$$\frac{d\langle g^2\rangle}{dt} = k_{on} + (k_{off} + k_{on})\langle g\rangle - 2(k_{off} + k_{on})\langle g^2\rangle \tag{11b}$$

$$\frac{d\langle m^2\rangle}{dt} = k_m\langle g\rangle + \gamma_m\langle m\rangle + 2k_m\langle gm\rangle - 2\gamma_m\langle m^2\rangle \tag{11c}$$

$$\frac{d\langle gm\rangle}{dt} = k_m\langle g^2\rangle + k_{on}\langle m\rangle - (k_{off} + k_{on} + \gamma_m)\langle gm\rangle \tag{11d}$$

(see supplemental in Singh et al. Biophysical Journal 2010). Analysis of (11) shows that the steady-state mRNA Fano factor is given by

$$\eta_m := \frac{\overline{\langle m^2\rangle} - \overline{\langle m\rangle}^2}{\overline{\langle m\rangle}} = 1 + \frac{k_{off}k_m}{(k_{off} + k_{on})(k_{off} + k_{on} + \gamma_m)} \tag{12}$$

In the limit $k_{off} \to \infty$ (i.e., the active state is unstable and the promoter spends most of the time in the inactive state)

$$\eta_m = 1 + \frac{k_m}{k_{off}}, \tag{13}$$

where k_m/k_{off}, is the mean transcriptional burst size, i.e., the average number of mRNA transcripts produced in one cycle of promoter activation and inactivation.

Appendix C: Change in gene expression noise in response to blocking mRNA production

Blocking mRNA production can be captured by setting the rate of expression events $k_m = 0$ at $t = 0$. Our goal now is to predict how statistical moments change as a function of the time t since the transcriptional block occurred. With $k_m = 0$, we have from (4) that the statistical moments now evolve according to the following differential equations

$$\frac{d\langle m\rangle}{dt} = -\gamma_m\langle m\rangle, \quad \frac{d\langle p\rangle}{dt} = k_p\langle m\rangle - \gamma_p\langle p\rangle \tag{14a}$$

$$\frac{d\langle m^2\rangle}{dt} = \gamma_m\langle m\rangle - 2\gamma_m\langle m^2\rangle, \quad \frac{d\langle mp\rangle}{dt} = k_p\langle m^2\rangle - \gamma_m\langle mp\rangle - \gamma_p\langle mp\rangle \tag{14b}$$

$$\frac{d\langle p^2\rangle}{dt} = 2k_p\ \langle mp\rangle - 2\gamma_p\langle p^2\rangle. \tag{14c}$$

Assuming that just prior to the transcriptional block the statistical moments had reached steady-state, we solve equation (14a) using the moments calculated in (5) as initial conditions. From (14a) it is straightforward to show that the mean mRNA and protein levels exponentially decay to zero as follows

$$\langle m(t)\rangle = \overline{\langle m\rangle}\exp(-\gamma_m t), \tag{15a}$$

$$\langle p(t)\rangle = \overline{\langle p\rangle}\left(\frac{\gamma_p\exp(-\gamma_m t) - \gamma_m\exp(-\gamma_p t)}{\gamma_p - \gamma_m}\right). \tag{15b}$$

Solving the remaining equations in (14) using *Mathematica* we find that the protein noise level (as measured by the coefficient of variation squared)

monotonically increases with time as

$$\frac{CV^2(t)}{\overline{CV^2}} = f(\gamma_m, \gamma_p, \eta_m, t), \quad CV^2(t) := \frac{\langle p^2(t)\rangle - \langle p(t)\rangle^2}{\langle p(t)\rangle^2} \tag{16}$$

where $\overline{CV^2}$ given by (7) is the protein noise level at $t = 0$ and the function f has the following asymptote

$$\lim_{t\to\infty} f(\gamma_m, \gamma_p, \eta_m, t) = \begin{cases} \frac{(\gamma_p+\gamma_m)(\gamma_p+\eta_m(\gamma_m-2\gamma_p))}{\eta_m\gamma_m(\gamma_m-2\gamma_p)} > 1 & \text{if} \quad \gamma_m > 2\gamma_p \\ \infty & \text{if} \quad \gamma_m \leq 2\gamma_p. \end{cases} \tag{17}$$

For given protein and mRNA degradation rates, Figure 2 in the paper plots the function $f(\gamma_m, \gamma_p, \eta_m, t)$ as a function of time t for different η_m.

Appendix D: Incorporating noise at the translation step in the gene-expression model

In this section we investigate how predictions in Fig. 2 change if protein synthesis and decay is modeled stochastically. Let the probability that a protein molecule is created or degraded in the next infinitesimal time interval $(t, t + dt]$ be given by

$$\text{Probability}\{p(t + dt) = p + 1 |\ p(t) = p,\ m(t) = m\} = k_p m \tag{18a}$$

$$\text{Probability}\{p(t + dt) = p - 1 \mid p(t) = p,\ m(t) = m\} = \gamma_p p dt, \tag{18b}$$

where k_p is the mRNA translation rate and γ_p is the protein degradation rate. The above equation together with Eq. (1) in Appendix A define a stochastic gene-expression model where both mRNA and protein levels evolve through stochastic jumps. For this model the moment dynamics is given by

$$\frac{d\langle m\rangle}{dt} = k_m\langle B\rangle - \gamma_m\langle m\rangle, \quad \frac{d\langle p\rangle}{dt} = k_p\langle m\rangle - \gamma_p\langle p\rangle \tag{19a}$$

$$\frac{d\langle m^2\rangle}{dt} = k_m\langle B^2\rangle + \gamma_m\langle m\rangle + 2k_m\langle B\rangle\langle m\rangle - 2\gamma_m\langle m^2\rangle \tag{19b}$$

$$\frac{d\langle mp\rangle}{dt} = k_p\langle m^2\rangle + k_m\langle B\rangle\langle p\rangle - \gamma_m\langle mp\rangle - \gamma_p\langle mp\rangle \tag{19c}$$

$$\frac{d\langle p^2\rangle}{dt} = k_p\langle m\rangle + \gamma_p\langle p\rangle + 2k_p\langle mp\rangle - 2\gamma_p\langle p^2\rangle \tag{19d}$$

which yields the following steady-state statistical moments of the population count

$$\overline{\langle m\rangle} = \frac{k_m\langle B\rangle}{\gamma_m}, \quad \overline{\langle p\rangle} = \frac{k_p\overline{\langle m\rangle}}{\gamma_p}, \quad \overline{\langle m^2\rangle} = \overline{\langle m\rangle}^2 + \overline{\langle m\rangle}\frac{\langle B^2\rangle + \langle B\rangle}{2\langle B\rangle} \tag{20a}$$

$$\overline{\langle mp\rangle} = \overline{\langle m\rangle}\ \overline{\langle p\rangle} + \overline{\langle p\rangle}\frac{\langle B^2\rangle + \langle B\rangle}{2\langle B\rangle}\frac{\gamma_p}{\gamma_p + \gamma_m} \tag{20b}$$

$$\overline{\langle p^2\rangle} = \overline{\langle p\rangle}^2 + \overline{\langle p\rangle} + \frac{\overline{\langle p\rangle}^2}{\overline{\langle m\rangle}}\frac{\langle B^2\rangle + \langle B\rangle}{2\langle B\rangle}\frac{\gamma_p}{\gamma_p + \gamma_m}, \tag{20c}$$

where a bar denotes the steady-state value of the corresponding moment.

From (20) the steady-state protein noise level is given by

$$\overline{CV^2} := \frac{\overline{\langle p^2\rangle} - \overline{\langle p\rangle}^2}{\overline{\langle p\rangle}^2} = \frac{1}{\overline{\langle p\rangle}} + \frac{\eta_m\gamma_p}{\overline{\langle m\rangle}(\gamma_p + \gamma_m)}. \tag{21}$$

Note that this noise level is similar to (7) in Appendix A, except for the $1/\overline{\langle p\rangle}$

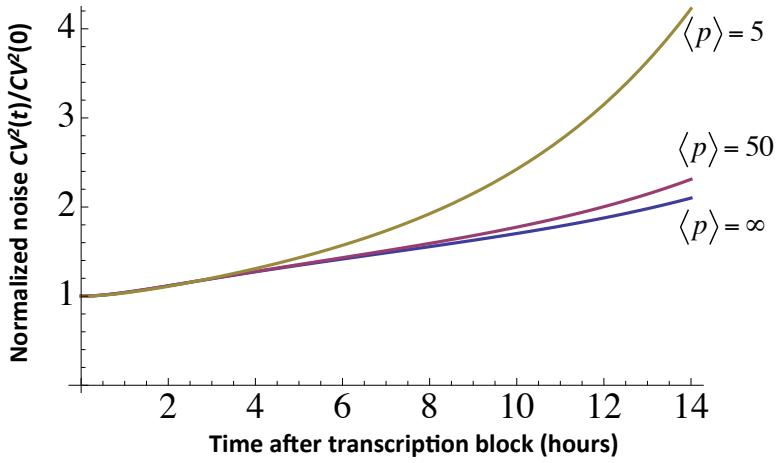

Figure S1: Predicted changes in protein expression noise after transcription is blocked for different mean protein copy number per cell ($\langle p \rangle$). $\langle p \rangle = \infty$ corresponds to deterministic protein birth-death dynamics. B was assumed to be geometrically distributed with a mean burst size of 10 transcripts and $\overline{CV^2} = 0.5$. mRNA and protein half-lives are taken as 3 and 2.5 hours, respectively. Noise levels are baselined by their corresponding values at time $t = 0$.

term which represents Poissonian noise arising from probabilistic birth and death of individual protein molecules. Transient changes in protein noise levels ($CV^2(t)$) are obtained by setting $k_m = 0$ at $t = 0$ and solving (19) using (20) as initial conditions. Fig. S1 plots protein noise levels after a transcriptional block for different steady-state mean protein levels per cell. As can be seen from the figure, as long as protein levels are larger than 50 copies per cell, $CV^2(t)$ is identical to noise levels obtained assuming deterministic protein dynamics ($\langle p \rangle = \infty$ line in Fig. S1). Typically, as long as

$$\overline{CV^2} \gg \frac{1}{\overline{\langle p \rangle}}, \tag{22}$$

modeling protein dynamics using mass-action kinetics is a good approximation. The clones considered in this paper have noise levels in the range of $\overline{CV^2} \approx 0.5 - 1$ and have on average over 10,000 copies of d2GFP protein molecules per cell (Singh et al. Biophysical Journal 2010). Thus we satisfy (22) by many orders of magnitude.

Appendix E: Actinomycin D does not change d2GFP half-life

To investigate the effects of Actinomycin D on d2GFP half-life, clone F32 was treated with Actinomycin D plus Cycloheximide or just Cycloheximide. Cells were collected at regular intervals after drug addition and the GFP expression was measured by flow cytometry. The kinetics of GFP decay is

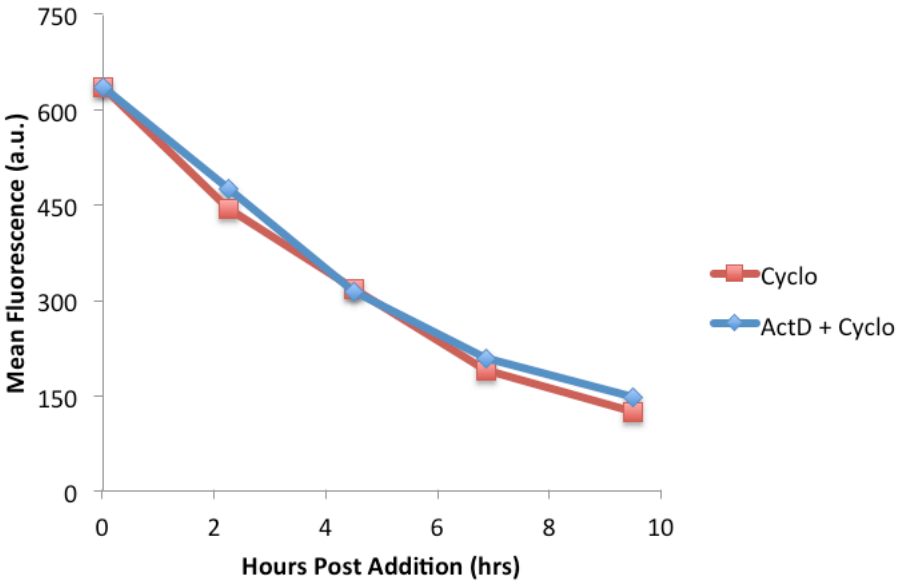

Figure S2: GFP decays at the same rate after translation is blocked using Cycloheximide, both in the presence and absence of Actinomycin D.

same both in the presence and absence of Actinomycin D (Fig. S2).

Appendix F: Transcriptional bursting is a significant source of noise in HIV-1 LTR gene expression

In the above section we showed that transient changes in $CV^2(t)$ after a transcriptional block can be used to quantify the relative contributions of different intrinsic noise mechanisms. Note that this theory was developed assuming a *complete-block* of transcription (i.e., k_m is set to zero). However, our experimental data shows that Actinomycin D only creates a *partial-block* of mRNA transcription (Fig. S3). More specifically, after treating clones with transcriptional inhibitors, mean reporter levels do not decay to

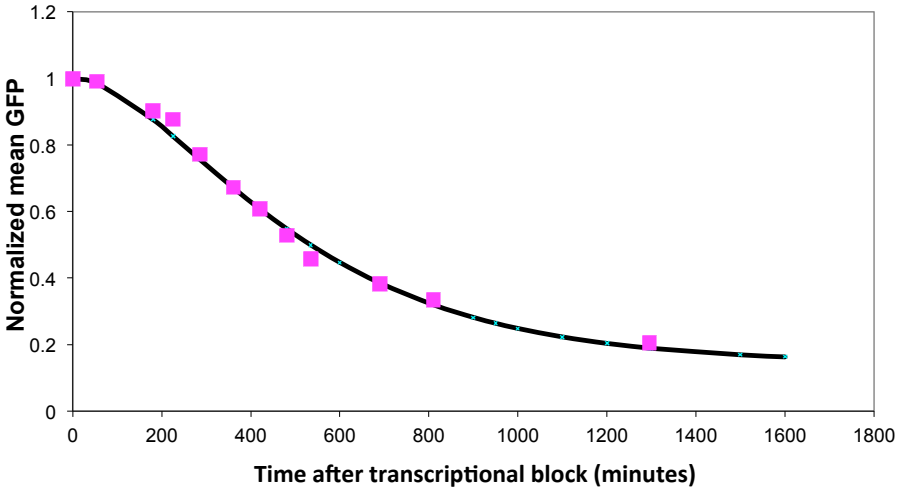

Figure S3: Transient changes in mean GFP levels after a transcriptional block. GFP levels are normalized by their corresponding value at $t = 0$. Clone F32 was treated with Actinomycin D at 10 μg/mL and mean expression levels were measure for 22 hours after drug addition (square data points). Black line represents the best fit of (23) to data, and it asymptotically approaches a value of $\approx 10\%$. The protein and mRNA half-lives are assumed to be 2.5 and 3 hours, respectively.

zero but converge to approximately 10% of its original value. Assuming, addition of transcription-blocking drug reduces k_m to δk_m where $0 < \delta < 1$, the transient change in mean GFP level is given by

$$\frac{\langle p(t)\rangle}{\overline{\langle p\rangle}} = \delta + (1-\delta)\left(\frac{\gamma_p \exp(-\gamma_m t) - \gamma_m \exp(-\gamma_p t)}{\gamma_p - \gamma_m}\right). \tag{23}$$

Note that in the constitutive promoter model the transcriptional burst size $B = 1$ with probability one by definition. So the transcription-blocking drug can only affect the burst frequency k_m. For comparison purpose, we also assume that the drug changes k_m to δk_m in the transcriptional bursting model.

We next predict transient changes in $CV^2(t)$ in response to a partial-block of transcription. This is done using the following steps:

1. Assume a certain distribution for the transcriptional burst size B.

2. Compute the mRNA translation rate k_p and frequency of transcriptional events k_m for the corresponding transcriptional burst size B using the following equations

$$\frac{\langle B\rangle k_p k_m}{\gamma_m \gamma_p} = \langle GFP\rangle \quad \frac{(\langle B^2\rangle + \langle B\rangle)\gamma_p \gamma_m}{2\langle B\rangle k_m(\gamma_p + \gamma_m)} = \overline{CV^2}, \tag{24}$$

   where $\langle GFP\rangle$ and $\overline{CV^2}$ are the experimentally measured mean reporter levels and steady-state coefficient of variation, respectively. Recall that both d2GFP protein and d2GFP mRNA half-life are known (i.e., γ_m and γ_p are known). Thus, k_p and k_m corresponding to the transcriptional burst size B chosen in step 1 can be computed from (24).

3. We assume that transcription-blocking drug reduces k_m to δk_m where $\delta = 0.1$. The transient change in gene-expression noise in response to a partial-block of transcription is computed by solving (4) using the moments calculated in (5) as initial conditions.

In Fig. 3, the black dashed line corresponds to performing the above steps for a transcriptional burst size $B = 1$ with probability one (i.e., $\eta_m = 1$). On the other hand, the red line corresponds to a geometrically distributed transcriptional burst size B with a mean of 15 mRNA transcripts (this

corresponds to $\eta_m = 15$ from (8)). Our analysis shows that the rise in gene-expression noise after a partial-block of transcription is inconsistent with a model where $\eta_m = 1$, but consistent with a model where expression noise primarily comes from transcriptional bursting.

In order to get a lower-bound on η_m, we note that the clones with the fastest rise in expression variability in Fig. 3 reach a value of ≈ 1.75 (after taking 95% error in $CV^2(t)$ into account) at the final time point ($t = 10$ hours). Thus, $CV^2(10) < 1.75$ across all clones. Recall from Fig. 2 that a lower value of η_m results in a faster increase in expression variability after transcription is blocked. Model predictions show that in order for the predicted protein expression variability to be less than 1.75 at ten hours post drug addition, η_m will have to be at least 10. This gives a lower bound on the mRNA Fano factor across integration sites. Given the error in our data, any value of η_m larger than 10 is consistent with the measured changes in expression variability. This shows an important limitation of our method: it is hard to estimate a precise value of η_m, especially when η_m is much larger than one.

In the above experiment, the GFP fluorescence level is always significantly higher than background florescence, even after a partial-block of transcription using Actinomycin D. However, this may not always be the case and if the measured reporter fluorescence is close to the background florescence, then the coefficient of variation will have to be corrected for it. More specifically, the measured reporter fluorescence x_M is given by

$$x_M = x_F + x_B, \tag{25}$$

where x_F denotes the actual reporter fluorescence and x_B is the background fluorescence. Assuming x_F and x_B to be independent of each other, the corrected coefficient of variation squared is given by

$$\frac{CV_M^2\langle x_M\rangle^2 - \sigma_B^2}{\left(\langle x_M\rangle - \langle x_B\rangle\right)^2}, \tag{26}$$

where $CV_M^2 := \frac{\langle x_M^2\rangle - \langle x_M\rangle^2}{\langle x_M\rangle^2}$ is the measured coefficient of variation squared; $\sigma_B^2 := \langle x_B^2\rangle - \langle x_B\rangle^2$ is the variance in background fluorescence; $\langle x_M\rangle$ and $\langle x_B\rangle$ denote the mean measured reporter fluorescence and mean background fluorescence, respectively.

Next, we estimate the change in η_m when Actinomycin D creates a partial-block of transcription. We assume that addition of Actinomycin D changes k_m to $\delta_1 k_m$ and η_m to $\delta_2\eta_m$ where $0 < \delta_1 < 1$ and $0 < \delta_2 < 1$. To get the best estimates of δ_1 and δ_2 we assume $\eta_m = 80$, the mRNA Fano factor determined from mRNA FISH (Appendix G). Note that $\delta_1\delta_2 = 0.1$ since the net transcription rate is reduced to 10% of its original value in the presence of Actinomycin D at 10 μg/mL (Fig. S3). Predicted change in protein expression variability for this partial transcriptional block is obtained by numerically solving (4) using initial conditions given by (5). Our analysis shows that $\delta_2 \approx 0.5 - 0.7$ provides the closest fit between the model prediction and the average transient expression variability across the four clones after adding Actinomycin D. This result shows that Actinomycin D blocks transcription by primarily affecting the burst frequency k_m ($\delta_1 \approx 0.2$) and there is a 2-fold reduction in η_m at the highest non-toxic dose of Actinomycin D.

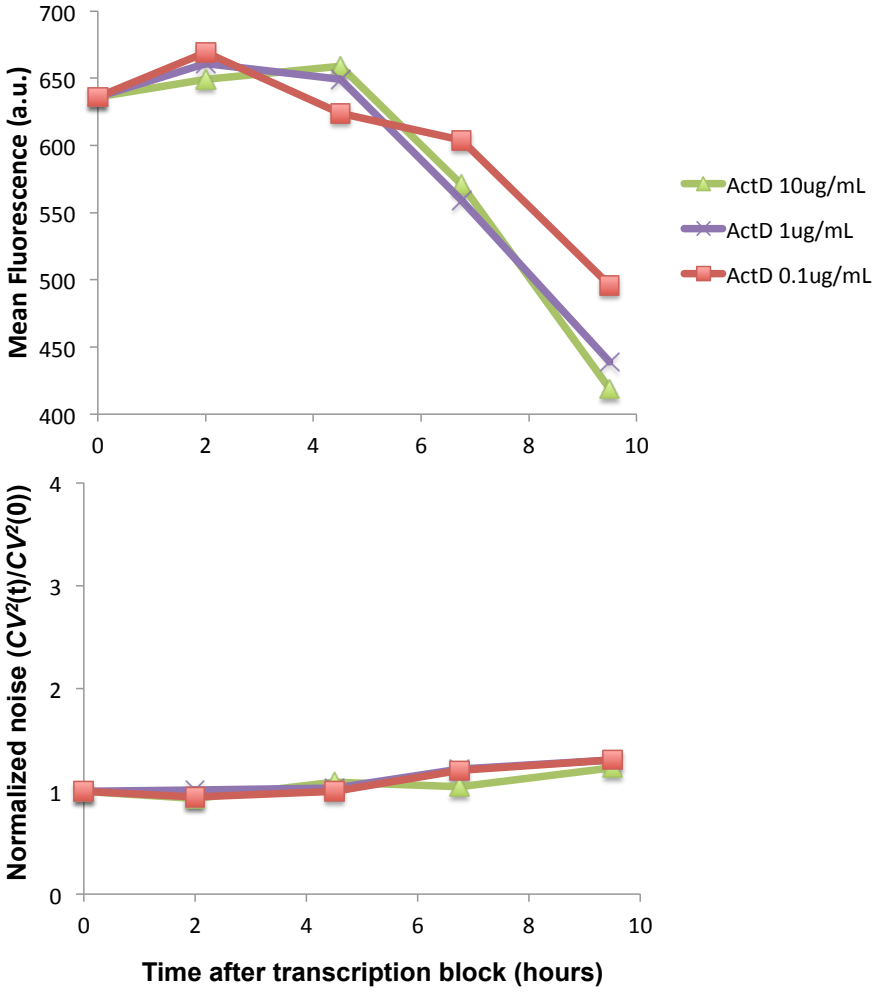

Figure S4: Actinomycin D titration changes the decrease in mean GFP over time (top graph), however, the effect on $CV^2(t)$ remains the same (bottom graph). Actinomycin D was either added at 0.1 (red line), 1 (purple line), or 10 µg/mL(green line) to clone F32 and mean GFP expression and CV were tracked using flow cytometry at 0, 2, 4.5, 7, and 9 hours after drug addition.

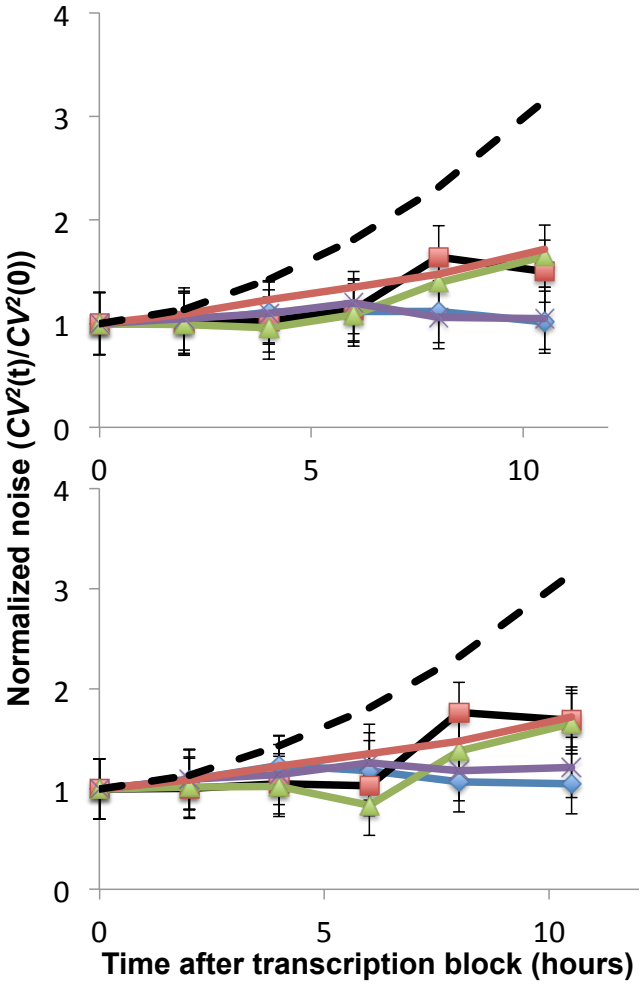

Figure S5: Transcriptional bursting is a significant source of variability in HIV-1 LTR gene-expression across different integration sites. Time courses of $CV^2(t)$ of GFP expression, as measured by flow cytometry, for four isoclonal Jurkat T lymphocyte populations: F32 (blue), G95 (black), LL44 (green), LL8 (purple) after perturbation with transcriptional-blocking drugs Actinomycin D (top graph) or Flavopiridol (bottom graph) at 10 µg/mL. Black dashed-line corresponds to $\eta_m = 1$ and the red line corresponds to $\eta_m = 15$.

To investigate how $CV^2(t)$ changes with the concentration of the transcription blocking drug, clone F32 was treated with different concentrations of ActinomycinD. Mean GFP intensity and coefficient of variation was measured at 0, 2, 4.5, 7, and 9 hours after drug addition. Changes in expression variability after a transcription block is identical for varying concentrations of the transcription-block drug (Fig. S4). Changes in d2GFP expression variability in different isoclonal populations after treatment with transcription-blocking drugs at 10 μg/mL is shown in Fig. S5.

Appendix G: Transcriptional bursting at the HIV-1 LTR revealed by mRNA FISH

mRNA FISH was performed on clone F32 and mRNA population counts were measured in a total of 192 cells (see Materials and Methods). Clone F32 has on average 100 d2GFP mRNA transcripts per cell (Fig. S6). The steady-state Fano factor, η_m, defined as

$$\eta_m := \frac{\overline{\langle m^2 \rangle} - \overline{\langle m \rangle}^2}{\overline{\langle m \rangle}} \tag{27}$$

was calculated from the histogram (Fig. S6) using bootstrapping. We obtain $\eta_m \approx 80$ with a 95% confidence interval of $(65, 95)$. This value of η_m is high but reasonable compared to published values. For example, Raj et al., 2006 reports mRNA Fano factors of over 100. It is important to point out that unlike flow cytometry data, where we remove extrinsic noise by appropriate gating, mRNA FISH data has both intrinsic and extrinsic noise. Thus this

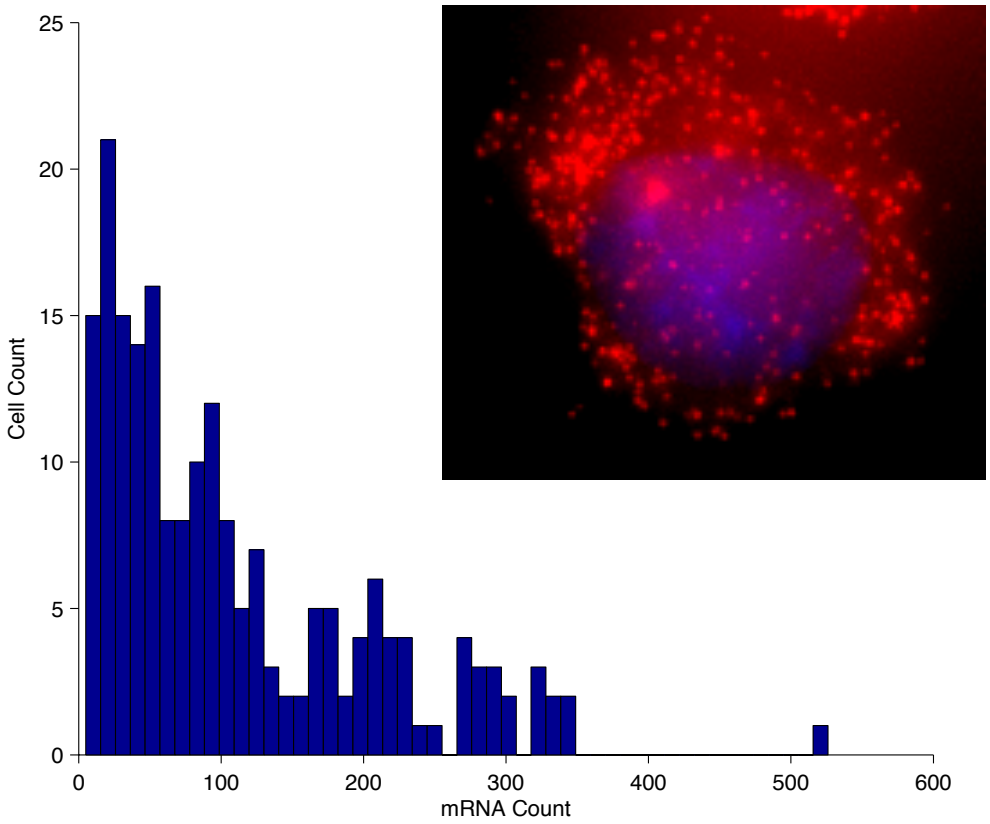

Figure S6: Histogram of mRNA populations counts across an isoclonal cell population determined using mRNA FISH. Inset: Example fluorescent image of DAPI stained nuclei (blue) and fluorscently labeled mRNA (red dots).

estimate of η_m is an upper bound since it is based on total expression noise rather than just intrinsic noise.

Appendix H: Quantifying extrinsic noise using a two-color reporter assay

To quantify the extent of extrinsic noise in HIV-1 gene expression, isoclonal populations containing a single integrated copy of both HIV-1 LTR driving d2GFP and a single integrated copy of HIV-1 LTR driving mCherry were constructed (see Materials and Methods). Populations gated around the forward (FSC) and side (SSC) scatter medians show little correlation between the GFP and Cherry signal (Fig. S7). This result shows that appropriate gating of cells removed most of the extrinsic noise.

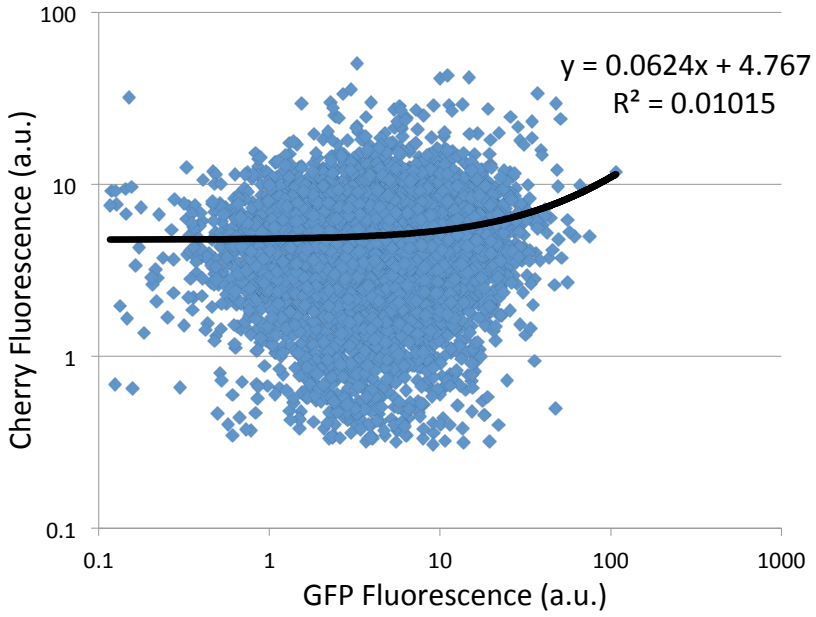

Figure S7: Scatter plot of single-cell intensities taken from flow-cytometry data for an isoclonal cell population shows little correlation between the GFP and Cherry signal. Populations were obtained by drawing a small gate around the forward (FSC) and side (SSC) scatter medians.