Title
The Role of Tat Positive Feedback in the Establishment of and Reactivation from HIV Latency

Author
Razooky, Brandon

Publication Date
2014

Peer reviewed|Thesis/dissertation

The Role of Tat Positive Feedback the Establishment of and Reactivation from HIV Latency

by

Brandon S Razooky

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biophysics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Dedication

This thesis is dedicated to my family and friends that have kept me sane throughout this process. I would like to thank my mother Rawa for her unconditional support. My father Samir for always pushing me no matter how far I think I've gone. I would like to thank my sisters, Chantel and Riva, as well as her husband Andre, for constantly reminding me of my shortcomings, thereby keeping me grounded. This thesis is also dedicated to my niece, London, who with the right guidance will receive the Nobel Prize in 2060. I would like to thank my cousin Rawnak—who I consider a brother—for being supportive and always making time for me, despite his own significant obstacles. My friends Samer and Anthony for truly helping me find the right work/life balance. I would also like to thank all of my friends in San Francisco who have made this city feel like home and have tolerated my unabashed spontaneity.

Acknowledgements

First, I want to thank my advisor Leor Weinberger for all of his help, guidance, and support throughout my PhD. It has been a humbling experience to work with someone who thinks so quickly and has the ability, scientifically, to see what very few can. His support has been paramount in my development as a scientist and in providing the opportunities available to me as I pursue my future endeavors.

I want to thank many of the Weinberger Lab mates, former and current, for the discussions, entertainment, and challenges during my PhD. Specifically, Tim Notton, Anand Pai, Seung-Young, Kay Aull, Igor Rouzine, Luke Rast, Elena Ingerman, Jonathan Klein, Jac Luna, Renee Ram, Grayson Kochi, Marleen Kawahara, Vince Metzger, Jeff Sasaki, Josep Sardenyer, Cynthia Bolovan-Fritts and Kim Osborn have all made this an enjoyable experience. I would like to especially thank Melissa "Bojing" Teng for all of her guidance and support throughout my PhD. I have asked Melissa for more career advice than anyone else I know. Winnie Yi Wen for working with me on many projects, and for her curt truths. Roy Dar for his unwavering and contagious excitement for science.

I want to acknowledge our collaborators from UCSD CFAR, Valeri Terry, Celsa Spina, and Matt Strain for giving me all of the necessary tools to work with HIV. Michael Simpson from Oak Ridge for acting as my co-advisor on many of the projects published within this thesis and giving me insight into truly quantitative biology. Alex Groisman and Edgar Guiterrez for teaching me everything I know about microfluidics. Jared Toettcher for advice, reagents, and helpful scientific discussions.

I would also like to thank Melanie Ott, Eric Verdin, Marielle Cavrois and Marianne Gesner for all of the help at the Gladstone Institute. I would like to especially thank JJ Miranda for discussions, both scientific and nonscientific and his unbridled support in my quest to find future research directions and positions.

The text of Chapter 3 in this thesis is a reprint of the material as it appears in Razooky, B.S., and Weinberger, L.S. (2011). Mapping the architecture of the HIV-1 Tat circuit: A decision-making circuit that lacks bistability and exploits stochastic noise. Methods *53*, 68-77. The co-author Leor Weinberger in this publication directed and supervised the research that forms the basis for Chapter 3 in the thesis.

The text of Chapter 5 in this thesis is a reprint of the material as it appears in Singh, A., Razooky, B., Cox, C.D., Simpson, M.L., and Weinberger, L.S. (2010b). Transcriptional Bursting from the HIV-1 Promoter Is a Significant Source of Stochastic Noise in HIV-1 Gene Expression. Biophys J *98*, L32-L34. The co-author Leor Weinberger in this publication directed and supervised the research that forms the basis for Chapter 5 in the thesis. Abhyudai Singh, Chris Cox, and Michael Simpson developed the theoretical model in the paper. Abhyudai Singh collected the used in the paper.

The text of Chapter 6 in this thesis is a reprint of the material as it appears in Singh, A., Razooky, B.S., Dar, R.D., and Weinberger, L.S. (2012). Dynamics of protein noise can distinguish between alternate sources of gene-expression variability. Mol Syst Biol *8*, 607. The co-author Leor Weinberger in this publication directed and supervised the research that forms the basis for Chapter 6 in the thesis. Abhyudai Singh developed the theoretical model in the paper. Roy Dar aided in acquiring the data for the paper.

The text of Chapter 7 in this thesis is a reprint of the material as it appears in Dar, R.D., Razooky, B.S., Singh, A., Trimeloni, T.V., McCollum, J.M., Cox, C.D., Simpson, M.L., and Weinberger, L.S. (2012). Transcriptional burst frequency and burst size are equally modulated across the human genome. Proc Natl Acad Sci U S A *109*, 17454-17459. The co-authors Leor Weinberger and Michael Simpson in this publication directed and supervised the research that forms the basis for Chapter 7 in the thesis. Roy Dar, Tim Trimeloni, and Abhyudai Singh analyzed the data for the paper. Abhyudai Singh, Chris Cox, and J.M. McCollum developed the theoretical model in the paper.

The text of Chapter 8 in this thesis is a reprint of the material as it appears in Razooky, B.S., Gutierrez, E., Terry, V.H., Spina, C.A., Groisman, A., and Weinberger, L.S. (2012). Microwell devices with finger-like channels for long-term imaging of HIV-1 expression kinetics in primary human lymphocytes. Lab Chip *12*, 4305-4312. The co-authors Leor Weinberger and Alex Groisman in this publication directed and supervised the research that forms the basis for Chapter 8 in the thesis. Edgar Gutierrez aided in the development of the microfluidic device. Celsa Spina and Valeri Terry provided reagents and advice for primary cell infections and culture.

Abstract

Proviral latency is the main persistence mechanism that precludes eradication of human immunodeficiency virus, HIV, from infected patients. Although latency is a viral phenotype, current theories posit that HIV latency is not 'hardwired' into viral circuitry but directly under cell-state control. Therefore, latency is thought to be a deterministic epiphenomenon with no role in the natural history of the virus. Here, we synthetically reengineer HIV regulatory circuits to define the role of viral gene circuitry and cellular state in regulating latency. The reengineered circuits demonstrate that latency is largely autonomous to cellular state in both minimal circuits and full-length replicating viruses. Strikingly, in primary cells, cell-driven silencing of viral transcription—the prevailing hypothesis for latency establishment—is overcome by tuning viral feedback strength. The reengineered minimal circuits also show that, through a combination of mathematical modeling and noise measurements, HIV transcription occurs through episodic bursts generating large stochastic fluctuations in HIV gene-expression. It is unclear if these stochastic fluctuations influence HIV Tat positive feedback, the decision-making circuit encoded by the virus. Surprisingly, upon stimulation of HIV transcription, Tat positive feedback immediately saturates, buffering against stochastic fluctuations once a fate decision has been made. This feedback saturation leads to three striking properties; (*i*) transcriptional kinetics and (*ii*) steady-state output from the LTR are insensitive to variable Tat inputs, and (*iii*) the LTR converts unimodal, graded Tat inputs, into bimodal expression patterns. Thus, stochastic fluctuations in Tat levels will have the most profound affect early in viral infection because, upon commitment to active replication, HIV circuitry displays robustness to noise. Overall these results argue that HIV latency

is an intrinsic, stochastic feature of the virus that most likely occurs early in the viral lifecycle. Since HIV circuitry can act autonomous to cell-state, latency seems to be 'hardwired' into viral circuitry and not simply an epiphenomenon stemming from host-cell factors. Given the rapid mutation rate of HIV, selection for and conservation of the latency phenotype suggests it has a fitness role in the natural history of the virus.

Table of Contents

Table of Figures

Chapter 1: Hypothesis

1. HIV latency is 'hardwired' into viral circuitry

Upon infection of a $CD4^+$ T lymphocyte, HIV undergoes a developmental bifurcation where the virus can either (*i*) enter a state of active replication, which creates viral progeny and destroys the host cell, or (*ii*) enter a quiescent, metabolically inert state, termed proviral latency (Fig. 1) (Chun et al., 1995). Latently infected cells are not detected by the immune system, and cannot be targeted by current standard-of-care therapies for HIV infected patients (Richman et al., 2009). Hence, latency remains one of the largest obstacles thwarting completing eradication of HIV from infected individuals due to the virus' ability to rebound from the latent state upon interruption of antiretroviral therapy (ART) (Richman et al., 2009). Although a barrier to a cure, the molecular mechanisms regulating latency are poorly understood. However, current dogma posits that HIV latency is a deterministic epiphenomenon dictated by infected host-cell state. The notion that HIV exhibits no control over latency has also lead to the hypothesis that latency plays no role in the natural history of the virus. ***Here, we hypothesize that HIV latency is a stochastic, natural viral phenomenon 'hardwired' into viral circuitry.*** This hypothesis is supported by published work (Dar et al., 2012; Razooky et al., 2012; Razooky and Weinberger, 2011; Singh et al., 2010a; Singh et al., 2012) highlighted in Chapters 3, 5, 6, 7, and 8 and unpublished work presented in Chapters 4 and 9.

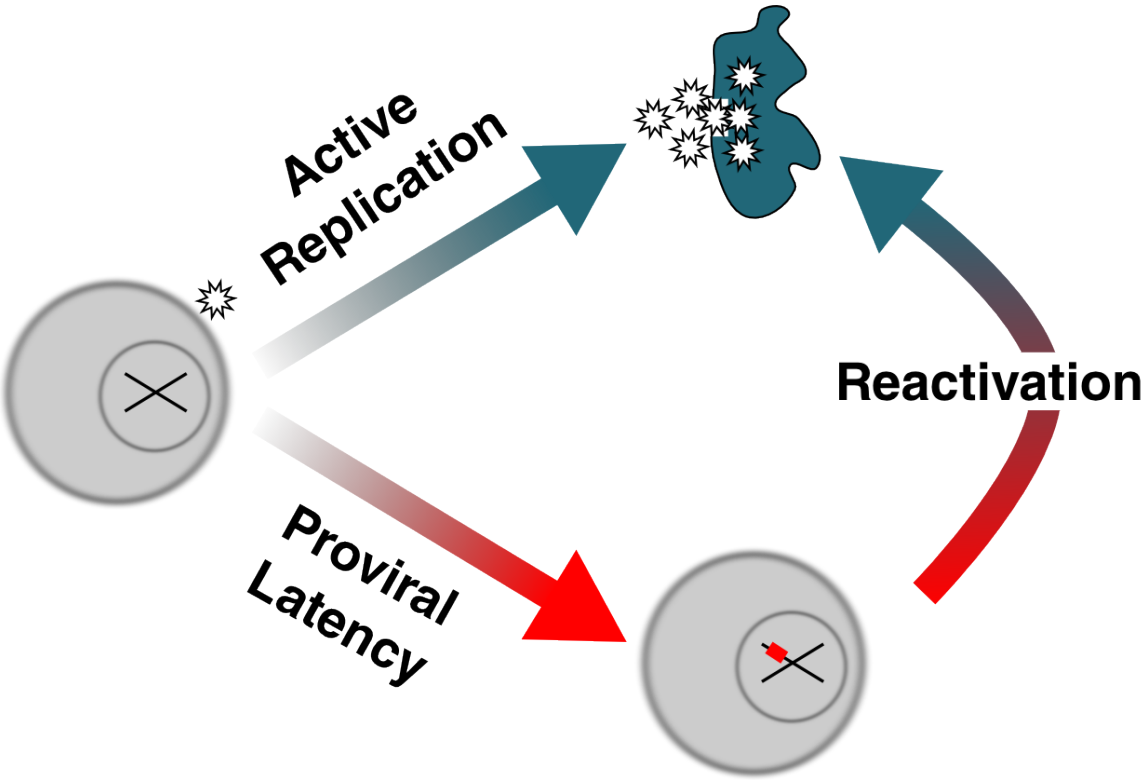

Figure 1. Schematic of the HIV developmental bifurcation.

Upon infection of a CD4+ T lymphocyte, HIV can actively replicate (blue arrow) and destroy the host cell (blue cell), or enter a state of silenced gene-expression (red arrow) after integrating into the host cell DNA (red square in bottom cell). Latent provirus can be rescued and enter the active replication fate (red to blue arrow).

Chapter 2: Background on HIV decision-making

1. Logic in looking for cellular determinants of HIV latency

The HIV field seems to be biased towards finding cellular determinants of HIV fate. This may be due to the lack of a dedicated gene-expression program for maintaining HIV latency (Siliciano and Greene, 2011). This is atypical for an organism that can enter a latent state. For instance, *Enterobacteria phage λ* is a phage that infects *Escherichia Coli* and can enter an actively replicating state termed lysis, or a dormant state termed

lysogeny (Ptashne et al., 1982). The fate of the phage is controlled by a genetic toggle switch where two competing negative feedback loops 'lock' the phage into either the lytic or lysogenic state (Ackers et al., 1982). Interestingly, each state is actively maintained through a separate dedicated gene-expression circuit and this feature of a dedicated latency program is found in many other viruses, including herpesviruses (Farrell et al., 1991). However, HIV, despite comprehensive identification of the open reading frames and characterization of associated protein functions, lacks an obvious gene-expression program for maintaining a latent state (Siliciano and Greene, 2011). There are no known virally encoded proteins that inhibit viral gene expression to maintain the latent state, but there are viral activators of transcription that maintain the actively replicating state. The inability to identify an HIV encoded gene-expression program for latency has led to a shift in focus away from viral mechanisms towards host-cell mechanisms that push and maintain HIV in a latent state.

2. General cellular restriction mechanisms

Upon infection of a target cell, typically activated $CD4^+$ T lymphocytes, the cell employs a few ubiquitous defense strategies against the invading virus; (i) cellular suicide, an altruistic action benefiting the overall host or cell population that ensures viral eradication along with the infected cell (Barber, 2001), and (ii) restriction of viral replication by targeting an aspect of the viral lifecycle (Duggal and Emerman, 2012). We will review a subset of strategies for cellular restriction of the viral lifecycle, specifically discussing the restriction of viral gene-expression. While there are numerous cellular-based hypotheses

for cellular control of HIV latency, we will focus on a select few that have heavily influenced the direction of the field.

2a. Chromatin state can influence viral fate

A difficult challenge in uncovering the molecular mechanisms controlling HIV latency is to recreate the phenomenon in a reproducible, physiologically relevant system, in the lab. One of the first reproducible model systems was developed by utilizing a full-length, replication-incompetent HIV virus encoding green fluorescent protein (GFP) (Jordan et al., 2003). In this model system, Jurkat T lymphocytes were infected with the replication-incompetent virus, and the population of cells that did not have active virus was sorted (the GFP negative population). To isolate cells with latent provirus, the authors took advantage of the NF-kB binding sites within the LTR, which are responsible for the responsiveness of the LTR to tumor necrosis factor alpha (TNF-α) since TNF-α activates p50/RelA. Upon activation of the LTR in the GFP negative population, single GFP positive cells were sorted into individual wells and grown into isoclonal populations. Surprisingly, in the J-lat system, ~10% of infections resulted in latency (Jordan et al., 2003), a dramatically larger fraction than suggested by studies of patient-derived cells where only one in a million cells harbor latent provirus (Chun et al., 1995). Analysis of the chromatin environment in the latently infected J-lat clones showed a significant correlation between entrance into the latent state and integration into heterochromatin environments (Jordan et al., 2003). Heterochromatin environments are typified by restrictive marks on chromatin that silence expression from the promoter by occluding transcription factors involved in either initiation or elongation of transcription (Fig. 2, top

left). Structural remodeling of the chromatin environment, towards an open euchromatin state (Fig. 2, top right), was necessary for reactivation of these latent clones.

Although integration of the virus is semi-random in the genome (Schroder et al., 2002), the above-described study established heterochromatin as a possible ubiquitous mechanism towards entrance into the latent state (Williams et al., 2006a). The possibility of heterochromatin restriction as a latency mechanism is supported by the findings that the HIV 5' long terminal repeat LTR promoter, LTR, has two characteristic nucleosomes that are dynamically modulated through acetylation and methylation marks that determine the activity of the promoter (Bednarik et al., 1990; Kauder et al., 2009; Pearson et al., 2008). The nucleosomes, termed nuc-0 and nuc-1 interfere with transcription factor binding and cause elongation stalls on initiated RNA polymerases, respectively (Van Lint et al., 1996). Nuc-1 has been subject to heavy investigation because this nucleosome causes RNA polymerase to stall at +60 from the HIV transcription start site (TSS) (Kao et al., 1987). The stalling mechanism causes the majority of HIV transcripts to become short abortive transcripts. In further support of a cell-based mechanism for J-lat viral latency was the observation that exogenous purified Tat protein—the viral transactivator of transcription, Tat, alleviates elongation stalls on the HIV long terminal repeat promoter, LTR (Laspia et al., 1989)— addition yielded modest effects on J-lat proviral reactivation (Jordan et al., 2003). In contrast, trichostatin A (TSA), a global histone deacetylase inhibitor (HDACi) (Yoshida et al., 1990), or tumor necrosis factor alpha (TNF-α), a global transcriptional inducer that works through the NF-κB signaling pathway (Osborn et al., 1989), were sufficient to reactivate the majority of the J-late

clones (Pearson et al., 2008; Williams et al., 2007b). Subsequent studies on heterochromatin and latency led to the observation that latent provirus could be reactivated by modulating heterochromatin states with other HDACi's (Reuse et al., 2009). These studies led to a therapeutic strategy for eradicating latent populations by administrating an HDACi, vorinostat (Archin et al., 2012). The administered HDACi would purge latent provirus by reactivating it from the latent state.

One pressing question remains from this model system: is integration into heterochromatin the mechanism of latency establishment in patient-derived primary lymphocytes? Recent evidence suggests that integration into heterochromatin environments does not sufficiently explain latency in resting lymphocytes isolated from HIV patients (Ho et al., 2013), hence further studies will have to establish the physiological determinant of HIV latency.

2b. Hypermethylation of the HIV 5' LTR restricts viral gene-expression

DNA methylation is an epigenetic marker that alters the expression of genes and the methylation tag is placed by DNA methyl transferases (Dnmt), which methylates DNA at cytosine-phosphate-guanine (CpG) dinucleotides (Okano et al., 1999). CpG islands are more abundant at the promoters of most genes and the methylation state of the island influences the level of expression from that promoter. The HIV 5' LTR contains CpG islands that can be methylated to restrict HIV gene-expression (Bednarik et al., 1990). However, whether methylation of the HIV promoter is sufficient to completely suppress HIV gene expression and force HIV into latency remains unclear.

In one model, a full-length HIV clone was reduced down to the regulatory elements necessary for HIV gene-expression regulation (Pearson et al., 2008). After infection of Jurkat cell lines with the HIV model virus, GFP positive cells were isolated and cultured over many months. Cells that were once GFP positive began to enter a GFP negative state, and immunoprecipitation of the HIV LTR showed methylation marks indicative of a silenced promoter state accumulating as these transitions occur (Fig. 2). Upon stimulation of latently infected populations with global cell regulators, such as TNF-α, the methylation marks on the LTR were lost as the cells were reactivated (Fig. 2). Once TNF-α was removed the cells entered into the latent state more rapidly than upon first infection, accumulating many of the same silencing methylation marks (Pearson et al., 2008). The studies were expanded into a long-term primary lymphocyte model of HIV latency. The authors were able to sustain primary $CD4^+$ T lymphocyte viability on a feeder line, infect the lymphocytes with the HIV model virus described above, and check for entrance into the latent state (Tyagi et al., 2010). The studies from this model yielded similar results to the Jurkat system, and hence, the authors concluded that methylation marks accumulating onto the HIV LTR are the cause of latency. Other studies have also found that CpG methylation can control reactivation of latent HIV (Bednarik et al., 1990; Blazkova et al., 2009; Kauder et al., 2009). However, patient-isolated latently infected resting lymphocytes lack the same methylation markers found in these latency models (Blazkova et al., 2012; Ho et al., 2013). Therefore methylation does not seem to be the dominant mechanism for entrance into the latent state.

2c. Polymerase collisions lead to quiescence

HIV integrates into a semi-random position in the host genome (Schroder et al., 2002), and so the virus must ensure preferential expression of the viral genome over surrounding host genes. Transcription of the randomly placed viral DNA leads to complications where transcription from neighboring genes affects the integrated provirus (Fig. 2). Specifically, if the host gene is transcribed frequently in the opposing direction, then the polymerase from the host gene can collide with the polymerase transcribing the viral genome and inhibit transcription of viral genes (Fig. 2) (Greger et al., 1998; Lenasi et al., 2008). Transcriptional collisions are a well-documented phenomenon and one method to overcome this is for the virus to increase the rate of transcription from the promoter and ensure transcription of its own genes. Another mechanism for transcriptional interference occurs when viral integration is oriented in the same direction as the surrounding host gene. This leads to promoter occlusion (Fig. 2), whereby the polymerase from the host gene reads-through the integrated provirus (Greger et al., 1998; Lenasi et al., 2008). Similarly to the convergent transcription hypothesis, one way to ensure viral transcription is to increase the rate of transcription from the viral promoter by bringing transcription factors at a more frequent rate than the surrounding gene (Shan et al., 2011). Perhaps HIV evolved Tat positive feedback to overcome transcriptional collisions and promoter occlusion to ensure robust HIV gene expression.

2d. Resting lymphocyte environment promotes restriction of HIV gene-expression

Whether a lymphocyte is in the resting or activated state drastically changes the global gene-expression landscape and overall intracellular environment. The intracellular state

in resting cells creates an environment that promotes the silencing of HIV transcription, however, resting lymphocytes are difficult to infect and HIV preferentially infects activated lymphocytes. These observations have lead to the hypothesis that latency establishment occurs when a cell transitioning from an activated to resting state is infected by HIV, silences HIV transcription during this transition, leading to a latent infection (Siliciano and Greene, 2011). Whether, and how, resting lymphocytes promote the restriction of viral replication remains subject to intense research. The global changes that result from lymphocyte cellular state transitions, can also affect the localization and activity of many of the proteins HIV relies on for gene expression (Ruelas and Greene, 2013). For instance, expression of polypyrimide tract-binding protein, PTB, is severely reduced in resting lymphocytes (Lassen et al., 2006). PTB is responsible for export of mRNA from the nucleus and is crucial for export of the class of multiply spliced HIV mRNAs that encode for Tat, and other HIV regulatory proteins (Lassen et al., 2006). The lack of expression of PTB in resting lymphocytes causes the multiply spliced transcripts to be retained in the nucleus, inhibiting the production of Tat. The lack of translation of Tat mRNA breaks HIV Tat positive feedback and may be responsible for entrance into the latent state. These restriction mechanisms can be overcome through either overexpression of HIV Tat, thereby reestablishing Tat positive feedback or overexpression of PTB (Lassen et al., 2006). Reactivation of the resting cell can also rescue the latent provirus, which allows for export of viral mRNA and reestablishment of Tat positive feedback. In addition to limited PTB expression, resting lymphocytes sequesters other transcription factors HIV relies on, such as NFAT, NF-κB, and pTEFb (Ruelas and Greene, 2013). These global transcription factors are crucial for robust HIV

transcription, and their absence can push HIV into the latent state. Strikingly, all of these regulators can be induced as the resting lymphocyte transitions to the activated state (Fig. 2) (Ruelas and Greene, 2013). Induction of these gene-expression regulators creates an environment conducive to HIV replication leading to activation of latent provirus. However, despite cell-state activation with the most potent activators only a fraction of latent provirus is reactivated (Ho et al., 2013) suggesting that cell-state modification is not sufficient to 'reawaken' 100% of latent viruses.

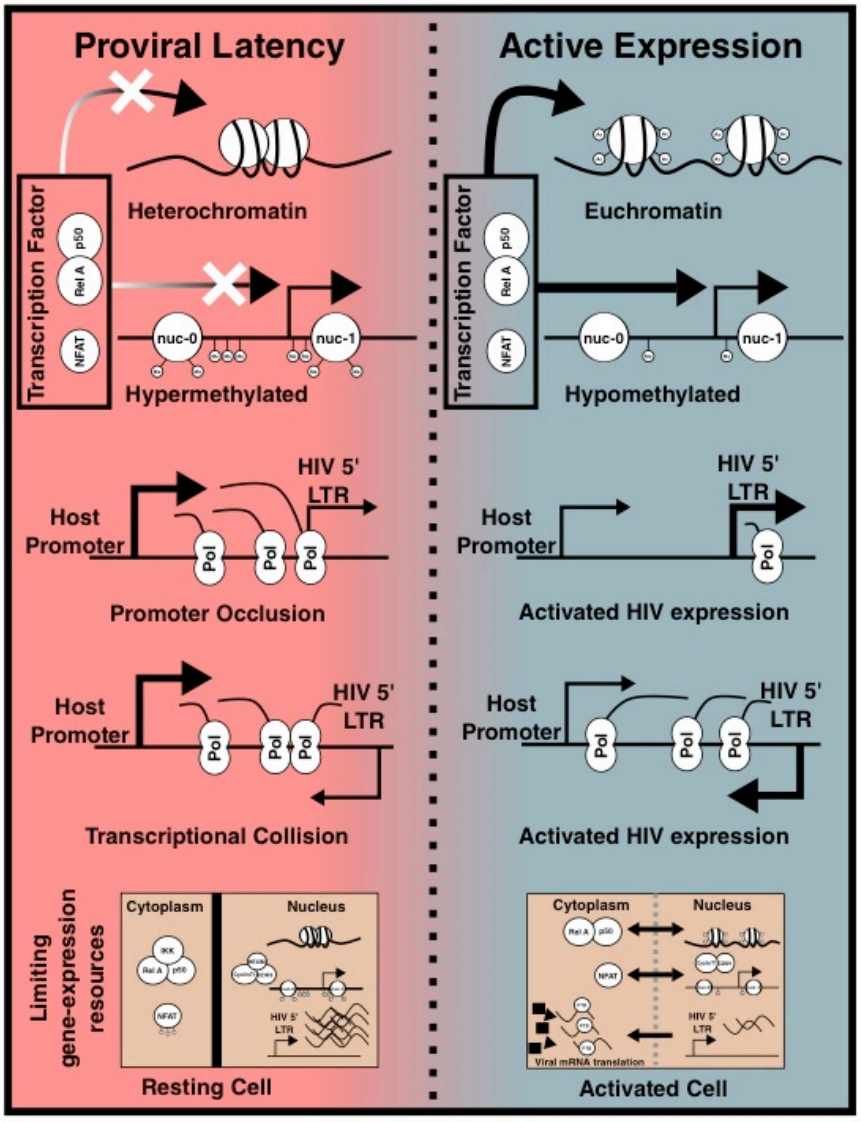

Figure 2. Potential cellular determinants of HIV latency and active replication.

The cell can restrict or promote viral gene-expression through chromatin and methylation modulation of the HIV LTR. The orientation of the HIV provirus relative to the surrounding gene and the relative strength of expression of said host gene to the HIV LTR will dictate if the virus can robustly transcribe its own genome. The state of the infected cell, whether resting or activated, will control whether necessary transcription and mRNA export factors are available for HIV gene-expression.

3. Viral control of fate

Although limited, there is some circumstantial evidence for viral control of HIV fate. Most studies where viral where the virus exhibits control of fate perturbed HIV circuitry or observed naturally occurring mutations that attenuated circuitry in some way.

3a. Tat over-expression suppresses establishment of HIV latency

Tat's role in controlling HIV latency has been reinforced many times through various studies. One recent study examined how constitutive Tat expression affects the establishment of a latent population across multiple viral replication cycles (Donahue et al., 2012). In this longitudinal study, a Jurkat line stably transduced with a lentiviral vector that expresses Tat protein from a constitutive promoter was infected with a replication-competent HIV virus. Overexpression of Tat reduced the size of the latent population compared to the control cell that did not constitutively express Tat (Donahue et al., 2012). These findings indicate a role of Tat in HIV latency.

3b. Manipulation of HIV gene-expression noise

Altering the level of noise in a system can increase or decrease the probability of the system entering into a certain state (Balazsi et al., 2011). Since stochastic fluctuations may influence HIV circuitry (Weinberger et al., 2005), manipulation of circuit strength and noise could potentially influence decision-making in HIV. One method to manipulate gene-expression noise is to alter the *cis* regulatory elements on the HIV LTR

promoter. The HIV LTR contains binding sites for ubiquitous cellular transcription factors such as NFAT, Sp1, NF-κB, AP-1, and TBP. The potential regulatory interactions within the LTR are complex (Ott et al., 2011). While studies have mutated these regions to quantify the influences on HIV decision-making, the affects of mutating *cis* regulatory elements on HIV gene-expression noise remained untested. To address this, a set of studies individually, and in combination, mutated and deleted the Sp1, NF-κB, and TATA binding elements in the minimal LGIT (Burnett et al., 2009). The data shows that mutating both NF-κB binding sites and the first Sp1 site had the most drastic effects on HIV transcription (Burnett et al., 2009). Subsequent analysis of the affects of these mutations revealed increased switching frequencies between the activated and resting state in the LGIT system most likely caused by increased levels of noise from the LTR (Burnett et al., 2009; Miller-Jensen et al., 2013).

Noise in HIV gene-expression can also be modulated through the addition of exogenous regulators such as TSA, TNF-α, and JQ1 (Dar et al., 2012). TSA and JQ1 are of special interest due to the possibility of using these compounds, or ones that work through similar mechanisms, to reactivate latent provirus in patients (Ruelas and Greene, 2013). TSA and TNF-α modulate noise by altering bursting dynamics from the HIV LTR (Dar et al., 2012; Singh et al., 2010a). The interest in JQ1 stems from recent studies that the bromodomain and extraterminal proteins (BET), specifically BRD2 and BRD4, stabilize the latent state (Banerjee et al., 2012; Bartholomeeusen et al., 2012). Knockdown of BRD2 or inhibition through the addition of a small molecule, JQ1, can modestly reactivate latent provirus highlighting the potential role of targeting these

proteins. To test the mechanism of action of JQ1, LTR-GFP cells were exposed to the compound and gene expression was tracked in single cells by quantitative time-lapse fluorescence microscopy. The study found that HIV expression noise significantly changes upon BRD2 and BRD4 addition by modulating transcriptional bursting dynamics (Boehm et al., 2013). Unlike TNF-α and TSA, JQ1 does not increase expression levels from the HIV LTR. It would be interesting to further study if the increase in gene-expression noise significantly contributes HIV decision-making by finding other modulators of HIV gene-expression noise that retain similar expression levels to more closely study the relationship between HIV noise and decision-making.

3c. Tat positive feedback can exhibit control of HIV fate

The viral transactivator of transcription, Tat, is obligate for HIV replication. As previously mentioned, the vast majority of transcription initiation events from the HIV LTR promoter lead to abortive transcripts, however, in the presence of Tat, the majority of transcription events read-through the viral genome (Kao et al., 1987). While Tat's interactions with cellular factors is complicated (Ott et al., 2011), in the simplest model, Tat functions by binding onto an RNA stem-loop structure known as the transactivating response element (TAR) hanging from stalled RNA polymerases on the HIV LTR, and promotes the interaction of positive elongation factor, pTEFb, with the stalled RNA polymerase. pTEFb then hyper-phosphorylates the C-terminal domain of the stalled RNA polymerase leading to transcription of the viral genome. This can in turn lead to the expression of more Tat, increasing the probability of another cycle of transcription,

and hence creating a positive feedback loop in HIV gene-expression (Pessler and Cron, 2004; Razooky and Weinberger, 2011; Weinberger et al., 2005).

There are a few studies that have performed an in-depth analysis of Tat positive feedback to understand its influence on HIV decision-making. In a set of studies testing for minimal sufficiency for control of HIV latency, lentiviruses encoding the HIV LTR promoter driving the expression of GFP alone (LTR-GFP) or Tat and GFP (LTR-GFP-IRES-Tat) were used to create isoclonal cell lines harboring a single copy of either integrated provirus (Weinberger et al., 2005). Strikingly, while the isoclonal LTR-GFP populations were unimodal in GFP expression, the LTR-GFP-IRES-Tat isoclonal populations displayed significant heterogeneity in GFP expression and bimodal distributions. In these model systems, an expressive 'ON' minimal circuit provirus represents an actively replicating virus, while a quiescent 'OFF' provirus represents the latent state.

A developed mathematical model (mathematical modeling of HIV circuitry is discussed in detail in Chapter 3) of HIV Tat positive feedback showed that stochastic fluctuations in HIV gene-expression would be sufficient to create a developmental bifurcation in HIV without the need for introducing cellular elements (Weinberger et al., 2005; Weinberger and Shenk, 2007). In the mathematical model, a low basal rate of HIV transcription, coupled with large stochastic fluctuations in the level of Tat could create a developmental bifurcation, despite the system starting from the same initial conditions. It is important to note that this system relies on no other elements (i.e. cooperativity) besides stochastic fluctuations in positive feedback to establish this bimodality in HIV

gene-expression. This system, unlike a myriad of other studied positive feedback loops, lacks the essential indicators of cooperativity or any obvious notions of bistability, and the mathematical model suggests that Tat positive feedback is monostable in the OFF state (Weinberger et al., 2005; Weinberger et al., 2008a; Weinberger and Shenk, 2007). Additional models show that Tat transactivation simply provides a transient pulse of HIV gene-expression and the duration of the pulse in transcription could be modulated by changing Tat positive feedback. One example of modulating Tat positive feedback included targeting a futile cycle of acetylation by p300 and deacetylation by SirT1 of Tat (Pagans et al., 2005). Since Tat is only functionally active when acetylated on certain residues, overexpression of SirT1 in the model system was sufficient to shift the futile cycle, attenuate Tat positive feedback, and decrease the transient pulse of expression leading to an increase in the number of latently infected cells (Weinberger et al., 2008a). Tat activity is directly correlated with entrance into the latent state in primary cells as well. Two separate studies characterized HIV latent provirus from patient-derived cells and found that latently infected lymphocytes typically carried mutations in Tat that would attenuate positive feedback (Emiliani et al., 1998; Yukl et al., 2009). The observation that latently infected cells carry debilitating mutations in Tat, along with the observation that Tat is necessary for HIV replication, reinforces the role of Tat in HIV decision-making, and more importantly for viral control over its own fate. The sufficiency of modulating Tat to alter HIV phenotype suggests that viral fate is not purely controlled by host-cell mechanisms. The amplification of stochastic fluctuations that stem from HIV LTR transcriptional bursting by Tat positive feedback suggests that HIV circuitry has evolved into this noisy phenotype as a method to control decision-making. Importantly,

this study establishes a ubiquitous mechanism for control of HIV fate and shows HIV latency decision-making is not deterministic but probabilistic.

Chapter 3: Modeling HIV circuitry

Chapter 3 was published in Razooky, B.S., and Weinberger, L.S. (2011). Mapping the architecture of the HIV-1 Tat circuit: A decision-making circuit that lacks bistability and exploits stochastic noise. Methods *53*, 68-77. The article can be found on Pubmed: http://www.ncbi.nlm.nih.gov/pubmed/21167940

1. Background

1a. Background on quantitative modeling and the need for kinetic data

From chemical engineering to meteorology, ecology and infectious-disease epidemiology, mathematical modeling has long been a vital and accepted tool for interpreting data, deriving mechanism of action, and predicting the behavior of complex systems. Modeling even has a storied past in molecular biology—notably Watson and Crick's seminal 1952 paper on the structure of DNA. During the 1990s, mathematical modeling approaches adopted from ecology were pivotal in determining fundamental kinetic rates of HIV replication and turnover in patients and transformed our understanding of HIV pathogenesis and the evolution of drug resistance (Herz et al., 1996; Ho et al., 1995;

Nowak et al., 1996; Perelson et al., 1997b; Perelson et al., 1996; Wei et al., 1995), for a review see (Nowak and May, 2000). These models of HIV were successful largely because high quality time-lapse data was available to 'fit' the models.

In general, the availability of high-frequency time-resolved data is key for many types of mathematical modeling and this data is essential for the mathematical modeling we focus on here: kinetic modeling with Ordinary Differential Equations (ODEs). Although the term 'mathematical modeling' encompasses a wide range of computational approaches, we focus on ODEs (and the stochastic version of ODEs) since this approach has a strong track record of successfully generating accurate, predictive, and testable models of many cellular signaling networks (Alon, 2007; Edelstein-Keshet, 1988; Lauffenburger and Linderman, 1993; Murray, 2002; Savageau, 1976; Sible and Tyson, 2007; Singh et al., 2010a; Singh and Weinberger, 2009; Weinberger et al., 2009; Weinberger et al., 2005; Weinberger et al., 2008a; Weinberger et al., 2003; Weinberger and Shenk, 2007). Time-lapse data to fit ODE models has been used from a variety of different *in vitro* experimental modalities including Western Blot, gel-shift assay, RT-PCR, and others—as long as the experimental system can be plotted to generate a kinetic curve of quantity versus time. The data from time-lapse experiments can potentially lead to a predictive and testable ODE model of the system being studied. During the past 15 years technical advances, such as the invent of fluorescent proteins (e.g. GFP) have allowed very high frequency time-lapse data to be collected from live cells by flow cytometry and live-cell time-lapse microscopy. Here, we describe how a time-lapse imaging approach can be combined with modeling analysis to study the HIV Tat

positive-feedback circuit, which regulates active viral replication and plays a crucial role in regulating the establishment of proviral latency and viral reactivation. The HIV Tat system represents an illustrative example because the network is tractable, many predictions of the model have been validated experimentally, and this model can serve as a 'module' that can be adapted and expanded to more complex models of HIV regulation.

The mathematical modeling approach we present here utilizes the computer as a form of model system to run 'experiments'. The computer experiments make predictions that must then be validated in other experimental model systems such as cell-culture or animal models. This approach is not dissimilar to how other experimental model systems are used: tissue culture acts as a model system that provides data which must eventually be validated in other model systems such as murine models and the data from murine models must in turn be validated in another system. When developing murine models, those models that do not recapitulate physiological realism are usually discarded and more relevant models are developed. Mathematical models, much like *in vitro* or *in vivo* models, undergo a cycle of development, testing, and reformation; for example, models that do not recapitulate experimental data from tissue culture, are discarded and new models are developed.

Each experimental model system has inherent benefits and drawbacks with some systems having greater physiological relevance and others having greater resolving power to differentiate specific mechanisms. The tradeoff between physiological relevance and resolving power is always a consideration and in this regard mathematical modeling

represents a powerful reductionist assay system for resolving between competing models. Specifically, the key benefit of mathematical modeling is that upon 'discarding' the mathematical model, very specific mechanisms can be negated; each ODE model ideally represents a specific mechanism and when this model cannot fit experimental data from tissue culture, that specific mechanism is eliminated from consideration. Thus, the most informative models are often the models that do not fit the data! In this way, ODE modeling can provide mechanistic and even structural insight.

There is extensive literature on mathematical modeling of HIV-1 intracellular dynamics (Althaus et al., 2009; Palsson et al., 1990; Reddy and Yin, 1999; Srivastava et al., 2002), intercellular viral transmission (Nowak et al., 1996; Nowak and May, 2000; Perelson et al., 1996; Wolthers et al., 1999), and HIV-1 epidemiology (Wilson et al., 2008). Here, we argue that the coupling of mathematical modeling with time-lapse microscopy experiments is a powerful method to differentiate between alternate models and shows that the HIV Tat circuit does not encode a 'bistable' circuit architecture. The lack of bistability in the Tat circuit leads to the stochastic model of the HIV Tat circuit where the circuit acts as a monostable 'timer' switch which inevitably shuts off (Weinberger and Shenk, 2007). Importantly, the HIV Tat circuit was the first characterized decision-making circuit that lacks bistability and the coupled modeling + imaging method we describe provided a predictive model for the establishment of HIV proviral latency.

1b. The Problem at Hand: HIV-1 proviral latency (a decision-making circuit)

Many viruses appear to have the ability to undergo a developmental 'bifurcation' between two lifecycle states. The lysis-lysogeny decision in bacteriophage-λ is the prototypical example, for a review see (Singh and Weinberger, 2009). Mechanistically, bacteriophage-λ appears to achieve this developmental bifurcation, in part, by encoding bistability (the ability to stably 'rest' in two different states) within its master regulator circuit, the λ-operator. Bistability within the bacteriophage-λ circuitry appears to be achieved by means of two competing negative-feedback loops acting on the λ-operator (Arkin et al., 1998; Ptashne, 2004). Similar to bacteriophage-λ, HIV-1 can also enter one of two developmental fates: upon infecting a $CD4^+$ T lymphocyte, HIV-1 can either enter an active replication state (productive infection) or enter a post-integration/proviral latent state (an analog of phage lysogeny) (Fig. 3A). HIV's ability to enter a proviral latent state in resting CD4 T cells is considered the most significant obstacle thwarting HIV-1 eradication from a patient (Han et al., 2007; Richman et al., 2009) since latent cells can 'reactivate' during interruption of highly active anti-retroviral therapy (HAART) to generate rapid viral rebounds that re-establish pre-treatment HIV-1 levels (Finzi et al., 1999). A substantial body of evidence has confirmed that HIV-1 proviral latent cells are quiescent for viral production and that viral gene expression is shut off during viral latency (Lassen et al., 2004; Perelson et al., 1996; Seth et al., 2005). Entry into proviral latency appears to be multifactorial with many molecular processes controlling the decision to enter latency including the integration site of virus within repressed chromatin regions (Jordan et al., 2003), transcriptional blocking due to surrounding genes (Lenasi et al., 2008), epigenetic silencing of proviral DNA (Coull et

al., 2000; Pearson et al., 2008; Ylisastigui et al., 2004), a transition from active to memory state of CD4+ T cells during infection (Siciliano and Siciliano, 2004), or a function attenuating mutation in the necessary HIV-1 transactivator of transcription, Tat (Yukl et al., 2009). Previous work by our group has demonstrated that the transcriptional master circuit of HIV-1, the Tat positive-feedback loop, can control the latency decision and appears to be sufficient to drive a 'decision' between two states: *bright* and *off* (Fig. 3B) (Weinberger et al., 2005).

Below, we describe how mathematical modeling coupled with flow cytometry and single-cell time-lapse imaging can be used to probe whether the Tat circuit architecture is bistable and switch-like, as in bacteriophage-λ, or whether Tat encodes a different circuit architecture (Fig. 3 C-D). The imaging and modeling experiments below demonstrated that the Tat circuit lacks bistability and instead acts as monostable 'timer' switch, where the latent state appears to be the only true stable state (Weinberger and Shenk, 2007). Finally, we discuss how stochastic modeling approaches demonstrated that the Tat circuit's decision between an *on* and *off* state can be accounted for by incorporating noise (i.e. molecular fluctuations) into models of Tat gene expression and how noise appears sufficient to control fate determination in the Tat circuit (Singh and Weinberger, 2009; Weinberger et al., 2005; Weinberger et al., 2008a; Weinberger and Shenk, 2007).

2. Approach

We present the following scheme to map the architecture of the Tat positive feedback loop and test for bistablility: (i) we develop minimal mathematical models of HIV Tat positive feedback that predict specific kinetic behaviors, (ii) we construct simplified HIV-1 based lentiviral vectors that examine Tat positive feedback in isolation from all other viral elements, and (iii) we analyze the kinetic behavior of these vectors using time-lapse fluorescence microscopy to test the various mathematical models. After demonstrating that the experimental single-cell data does not support the bistability model in the HIV-1 Tat positive-feedback decision-making circuit, we discuss how stochastic models of a monostable Tat positive-feedback circuit can account for HIV-1's ability to decide between two alternate states.

2.1 Starting Considerations for Generating a Predictive Model

The modeling discipline faces the philosophical dilemma as to whether models should be complex and attempt to fit all known mechanisms or whether models should be simple and attempt to fit only the most essential phenomena. Clearly, what phenomena and characteristics qualify as 'essential' is subject to interpretation and debate. However, modeling every molecular detail frequently results in models that are difficult to interpret and have little predictive value, so here we will focus on constructing 'simple' minimal models that consider only a skeletal set of processes needed to quantitatively fit a specific set of experimental data. Our approach follows an underlying principle of model development: *A model should aim to be predictive rather than descriptive*. Modeling every molecular detail frequently results in models that are difficult to interpret and have little predictive value. This simplified model approach has strong precedent: models that simplified much of the known biological and molecular detail were essential for

elucidating key molecular mechanisms such as bi-stability in the λ-phage lysis/lysogeny decision (Arkin et al., 1998; Dodd et al., 2001; Hochschild et al., 1986; Hochschild and Ptashne, 1988; Johnson et al., 1981; Ptashne, 2004; Ptashne and Gann, 2002), multi-stability and plasticity in the *lac* operon (Mayo et al., 2006; Ozbudak et al., 2004), robustness in the E. coli chemotaxis network (Alon et al., 1999; Barkai and Leibler, 1997) as well as many other key molecular mechanisms. Simplified models have been particularly successful for HIV, correctly predicting HIV-1 population dynamics (Anderson et al., 1991a; Anderson and May, 1988, 1996; Anderson et al., 1991b; Anderson et al., 1988; May and Anderson, 1987; Nowak et al., 1996), HIV-1 viral kinetics *in vivo* (Ho et al., 1995; Perelson et al., 1997a; Perelson et al., 1996; Wei et al., 1995), and have also demonstrated how stochastic molecular fluctuations in HIV-1 Tat contribute to viral latency (Weinberger et al., 2005). Arguably the best example of the utility of simple models over complex models is in protein folding which appears to be "relatively insensitive to details of the interatomic interactions" where Baker and colleagues established that low-resolution models have far better predictive power than high-resolution thermodynamic models (for a review see (Alm and Baker, 1999)). Complex models that account for many of the molecular details are eventually developed but it is very rare for comprehensive models to be constructed *de novo*. Helpful and comprehensive models must be developed in a stepwise fashion, and we describe the development of the initial models for HIV in a stepwise fashion in sections 3.1-3.3.

2.2 Designing Lentiviral Vectors and Creating Cell Lines to Test the Different Models

To determine quantitative values for different components of a system and 'parameterize' an ODE model, minimal circuit constructs can be used. We present maps for a number of different HIV-1 circuit constructs of increasing circuit complexity in Fig. 4 since these maps will provide a helpful guide for the ODE models presented in Fig. 5. Each minimal circuit construct is important for differentiating between the different levels of control in HIV-1 gene expression. The LTR-GFP (LG) system contains the HIV-1 5' long terminal repeat promoter (5' LTR) driving expression of the green fluorescent protein (GFP) followed by the 3' LTR. Gene expression from this system depends purely on the integration site of that particular cell, which controls the basal rate of transcription from the 5' LTR (Jordan et al., 2001). The LG provides a means to estimate basal promoter strength of the LTR and fold activation in response to inducers in the absence of the Tat positive-feedback loop.

The LTR-GFP-IRES-Tat (LGIT) construct removes many of the complex regulatory processes present in full-length HIV-1, such as splicing or other feedback components, and just leaves the interactions between Tat and the LTR. The LGIT system has the 5' LTR driving a single mRNA that contains GFP, an internal ribosomal entry sequence (IRES) (Pelletier and Sonenberg, 1988), and the HIV-1 Trans-Activator of Transcription (Tat), all followed by the 3' LTR (Weinberger et al., 2005; Weinberger et al., 2008a; Weinberger and Shenk, 2007). The single mRNA species does not contain any splicing signals, however, the IRES sequence allows for GFP and Tat expression from the same mRNA in related and measureable amounts with the Tat protein being expressed at a 10-

100 fold lower level (Mizuguchi et al., 2000). The HIVΔEnv system is a 'full-length' derivative of the HIV-1 pNL 4-3 virus (Adachi et al., 1986) with a point mutation at the start codon of the *env* gene and GFP in place of *nef* (Jordan et al., 2003). Like LGIT, HIVΔEnv can also be used to examine Tat positive-feedback kinetics since Nef, Tat, and Rev act as alternative-splice variants of one another.

Below, we will focus on the LGIT circuit, since we address the architecture of the feedback circuitry. However, the LG construct is helpful to probe LTR regulation in the absence of feedback and any other HIV factors and the HIVΔEnv construct is helpful to verify the results of studies with LGIT and to probe more complex regulation in HIV-1.

To create cell lines expressing our construct of interest, we use standard lentiviral packaging systems (Dull et al., 1998) and tranduce Jurkat T lymphocytes with the packaged lentivirus to create stable cell lines that expresses our genes of interest. We typically infect at low multiplicity of infection (MOI ≈ 0.01-0.1) to ensure that infected cells contain only a single integration. Enrichment of infected-cell populations is then achieved by fluorescence activated cell sorting (FACS) for fluorescent proteins such as GFP. For a complete discussion on the use of lentiviral vectors see (Franz et al., 2010).

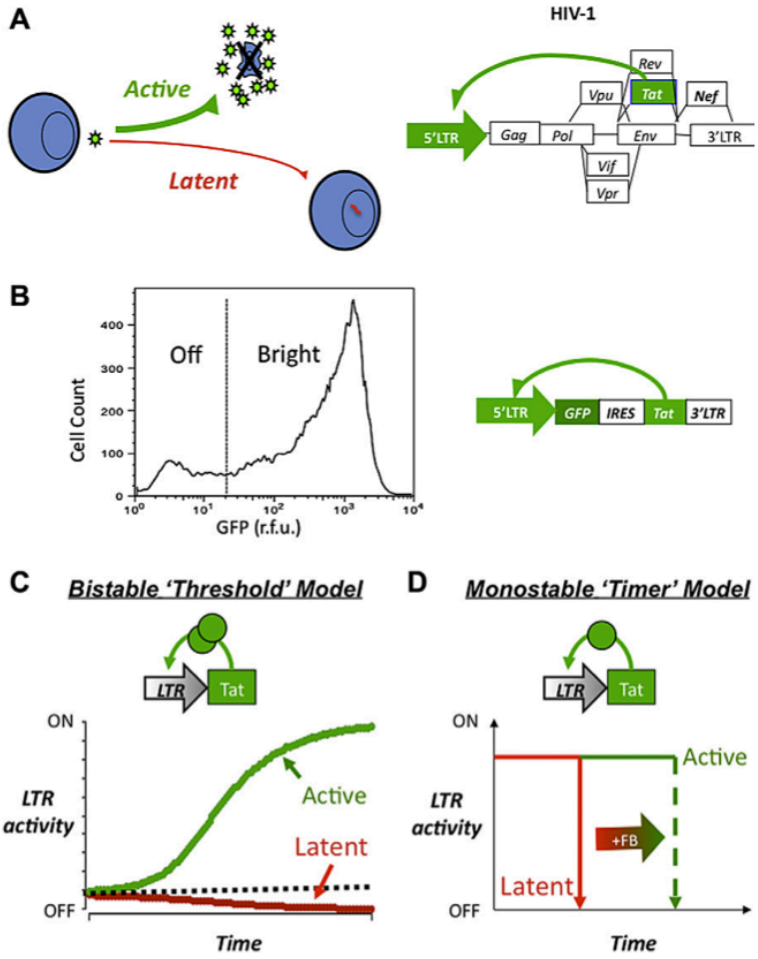

Figure 3. The HIV-1 proviral latency decision and two potential decision-making mechanisms.

(A) Schematic of HIV-1 infecting a CD4+ T cell and 'choosing' between active replication and proviral latency (left) and schematic of the HIV-1 proviral genome where the HIV-1 Tat positive feedback is essential for active replication and entry into proviral latency (right). (B) A minimal Tat circuit LTR-GFP-IRES-Tat (LGIT) can generate a bifurcation in GFP between two states: *Off* and *Bright* (Weinberger et al., 2005). Flow cytometry histogram of a Jurkat cell clonal population expressing LGIT from a single locus and exhibiting a developmental bifurcation. The bifurcation is not consistent with chromatin silencing or position effect variation but is consistent with stochastic fluctuations in Tat (Weinberger et al., 2005). (C) The bistability model for developmental bifurcation. If a transcriptional positive-feedback circuit encodes a self-cooperative threshold (e.g. homodimerization of the transactivator or multiple DNA binding sites that must all be bound by the transactivator), the circuit can exhibit bistability (the ability to stably rest in two alternate states). (D) An alternate model of kinetic partitioning (or a 'timer' switch) for developmental bifurcation. In this model, the positive-feedback loop is not required to encode a self-cooperativity threshold and all trajectories eventually fall to an off state (red) and the strength of positive feedback

determines the duration of time that a circuit resides in the on state (green). Figure originally published as Figure 1 in Razooky and Weinberger, Methods, 2011.

2.3 Choosing Between Different Methods for Obtaining Time Lapse-Data

Data from various experimental methods can be used for ODE modeling including data from qPCR, RT-PCR, Western blot, Western immunoprecipitation (IP), chromatin IP (ChIP), Northern blot, or other methods. If one has the luxury of choosing which experimental method to use to obtain time-lapse data for ODE models, a number of considerations come into play. Most of the biochemical assays mentioned above provide only data on the mean of a large population of cells in the culture (hundreds of thousands to millions of cells), which can obscure or complicate analysis if cell behaviors are not well synchronized. Alternative methods include flow cytometery, which presents the advantage of allowing one to assay both population means and distributions, or fluorescence microscopy that provides the unique advantage of tracking individual cells over time. The automated fluorescence microscope is an ideal tool to collect time-lapse data since data can be captured for thousands of cells in an automated fashion at an exceptionally high frequency (up to once every second) for hours to days. In this way, a vast amount of data is collected and this method has been used to probe dynamics of gene expression (Elowitz and Leibler, 2000; Elowitz et al., 2002; Weinberger et al., 2008a; Weinberger and Shenk, 2007), network cascades (Ting et al., 2001), and many other spatiotemporally regulated systems of the cell (Kagan et al., 2008; Maeda and Sano, 2006). Here, we focus primarily on the technique of fluorescence time-lapse microscopy to track gene expression of HIV-1 in single mammalian cells over time. We show how single-cell fluorescence trajectories can be interpreted by mathematical modeling to

provide insight into the architecture of the specific gene circuit and how HIV-1 may enter a latent state.

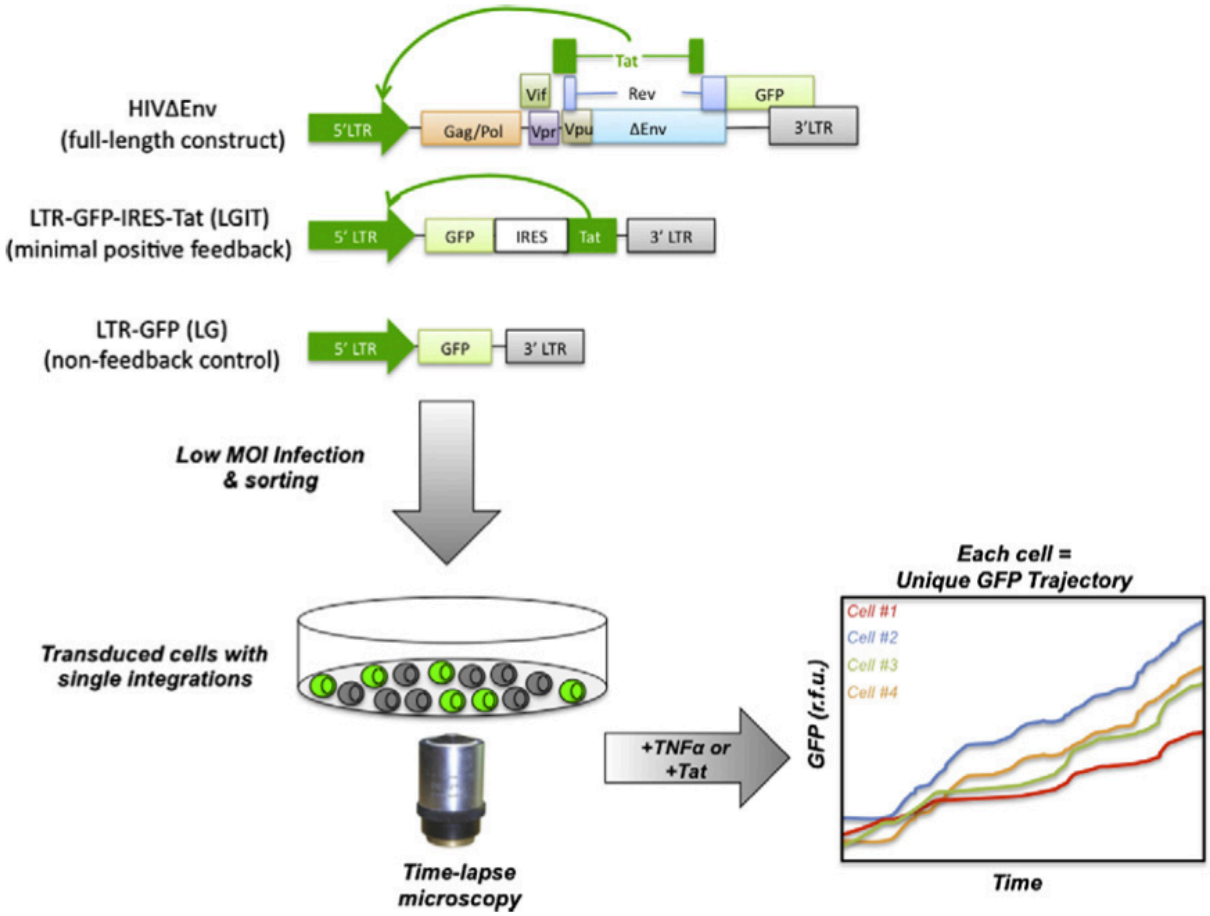

Figure 4. Experimental approach to map HIV-1 Tat circuit architecture using single-cell time-lapse imaging.

Various lentiviral vector constructs can be used for analyzing Tat circuitry. The HIVΔEnv construct contains the full-length HIV-1 genome, with GFP in place of the *nef* reading frame, and has a start codon mutation in the *env* so it does not produce infectious virus (Jordan et al., 2003). Nef and Tat are alternatively spliced, so GFP acts as an ectopic reporter for Tat kinetics. The LTR-GFP-IRES-Tat (LGIT) construct expresses a single bicistronic mRNA that codes only for GFP (the first cistron) and for Tat (the second cistron) which is translated from an internal ribosomal entry sequence (IRES). Tat positively feeds back onto the 5' LTR to transactivate it (black arrows). The LTR-GFP (LG) construct codes only for GFP driven by the HIV-1 5' long terminal repeat (LTR) promoter. The LG construct is useful as a non-feedback control. Each lentiviral construct can be packaged using standard approaches (Dull et al., 1998) and Jurkat T cells can be infected at low multiplicity of infection (MOI), FACS sorted to generate isoclonal (or polyclonal) populations, and then imaged on a live-cell fluorescence microscope system (Weinberger and Shenk, 2007). Single-cell trajectories are extracted from time-lapse imaging movies and can be analyzed by mathematical models. Figure originally published as Figure 2 in Razooky and Weinberger, Methods, 2011.

3. Methods

3.1 Developing and Analyzing ODE Models

3.1a ODE Model of Tat expression without feedback

To begin, we normally draw a schematic or cartoon of the system that describes the key interactions to be assessed (Sible and Tyson, 2007). We consider a set of alternate models for possible architectures of the Tat feedback circuit (Fig. 5) and we predict the output of each model in terms of HIV-1 gene expression. First, in Figure 5A we consider a model without any feedback and we draw the schematic for this model (Fig. 5A, left). Then, we deduce the ODEs that describe this schematic (Fig. 5A, middle), and we simulate the kinetics of Tat and GFP expression (Fig. 5A, right). The differential equations that describe this system are:

$$\frac{d}{dt}(mRNA) = \quad \alpha \quad - \quad \delta_{mRNA} * mRNA \qquad \textit{(Eq. 1)}$$

$$\frac{d}{dt}(Tat) = \lambda_{Tat} * Tat - \delta_{Tat} * Tat \qquad \textit{(Eq. 2)}$$

where α is the basal rate of mRNA expression, δ_{mRNA} is the per-capita decay rate of the expressed mRNA, λ_{Tat} is the translation rate of Tat from mRNA, and δ_{Tat} is the per-capita decay rate of Tat protein. Equations 1 and 2 can be lumped into a simplified version by making the quasi steady-state assumption of mRNA expression off of the LTR. We can solve for the steady-state mRNA levels by setting the left-hand side of Equation 1 to zero. We then obtain $\widehat{mRNA} = \alpha/\delta_{mRNA}$, and this can be plugged into Eq. 2 to yield:

$$\frac{d}{dt}(Tat) = \quad \beta_{Tat} \quad - \quad \delta_{Tat} * Tat \qquad \textit{(Eq. 3)}$$

where $\beta_{Tat} = (\alpha * \lambda_{Tat})/\delta_{mRNA}$. The numerical solution of Eq. 3 can then be plotted on a log scale using any number of software programs (e.g Matlab™ or Mathematica™) to predict/observe the kinetics of Tat expression in the absence of any feedback. Eq 3 predicts that Tat is expressed in a linear fashion since it only depends on basal rate of expression from the promoter and Tat reaches a steady state that is equal to β_{Tat}/δ_{Tat} (Fig. 5A, right). This steady state occurs when the rate of Tat production matches the rate of Tat decay; as the decay rate begins to approach the production rate we can see the plot approaching the asymptote of β_{Tat}/δ_{Tat}.

Next, in Fig. 5B, we consider a construct expressing both a GFP reporter and Tat driven by an internal ribosomal entry sequence (IRES) from the same mRNA as GFP. We assume this construct encodes no feedback (i.e. the LG construct). The equations that describe the dynamics of GFP expression were formulated in the same fashion as those for Tat from Fig. 5A. We assume the same α for Tat and GFP since they are derived from the same mRNA species. We present this model to illustrate that both Fig. 5A and Fig. 5B have very similar dynamics because the IRES allows for expression of both GFP and Tat in stoichiometric amounts off of the same mRNA species (Pelletier and Sonenberg, 1988). The log plot shows the same characteristic asymptote for GFP and Tat.

3.1b ODE Model of Tat Positive Feedback loop without potential for bistability (i.e. H=1, no Tat self-cooperativity)

Tat is known to establish a positive-feedback loop (D'Orso and Frankel, 2009; Feinberg et al., 1991; Frankel, 1992; Kwon et al., 2008; Weinberger et al., 2005) via binding to an RNA stem loop within the 5'LTR, termed the TAR loop (TransActivation Responsive

loop that extends from -18 to -70 in the 5'LTR) and relieving an elongation stall in RNA Polymerase II (RNAPII) (Fujinaga et al., 1998; Gatignol et al., 1991; Wei et al., 1998). There is a rich literature detailing the complex array of molecular interactions involved in Tat transactivation (D'Orso and Frankel, 2009; Feinberg et al., 1991; Frankel, 1992; Kwon et al., 2008) but we follow the minimalist philosophy above and by assuming that these molecular processes are non-limiting we 'lump' many of these processes into two parameters to generate a minimal model of HIV-1 Tat transactivation (Fig. 5C). The resulting minimal model can be described by the following set of ODEs:

$$\frac{d}{dt}(Tat) = \beta_{Tat} + \frac{x * Tat}{k_M + T_t} - \delta_{Tat} * Tat \qquad (Eq.\ 4)$$

We use the same terms found in Eqs. 1-3, however the middle term represents a saturable positive feedback loop, where x represents the positive-feedback strength, and k_M is the saturation constant of the system. The strength of the Tat activation term is dependent upon the amount of Tat present, so as the Tat concentration rises, the rate of activation also increases (i.e. the amount of Tat produced per unit time increases). This increasing rate manifests as an exponential rise, which on a log scale appears as a linear increase (Fig. 5C) and thus Eq. 4 generates behavior that differs significantly from the linearly increasing rate generated by Eqs. 1-3.

In this system with positive feedback, if the system operates far from saturation (i.e. $k_M >>$ Tat(t)), the middle term is approximately (x/k_M)*Tat(t) which gives an exponential increase. As the system approaches saturation (i.e. the amount of Tat(t) >> k_M) then the positive-feedback terms collapse to x, since Tat(t) + $k_M \approx$ Tat(t). This equation becomes very similar to the system without any feedback (Fig. 5A). Overall, the system displays exponential increase at early times and asymptotes at later times (i.e.

linear on a log scale at early times and asymptotes at later times). Below, we will show that the HIV-1 Tat feedback system exhibits this type of exponential increase over time.

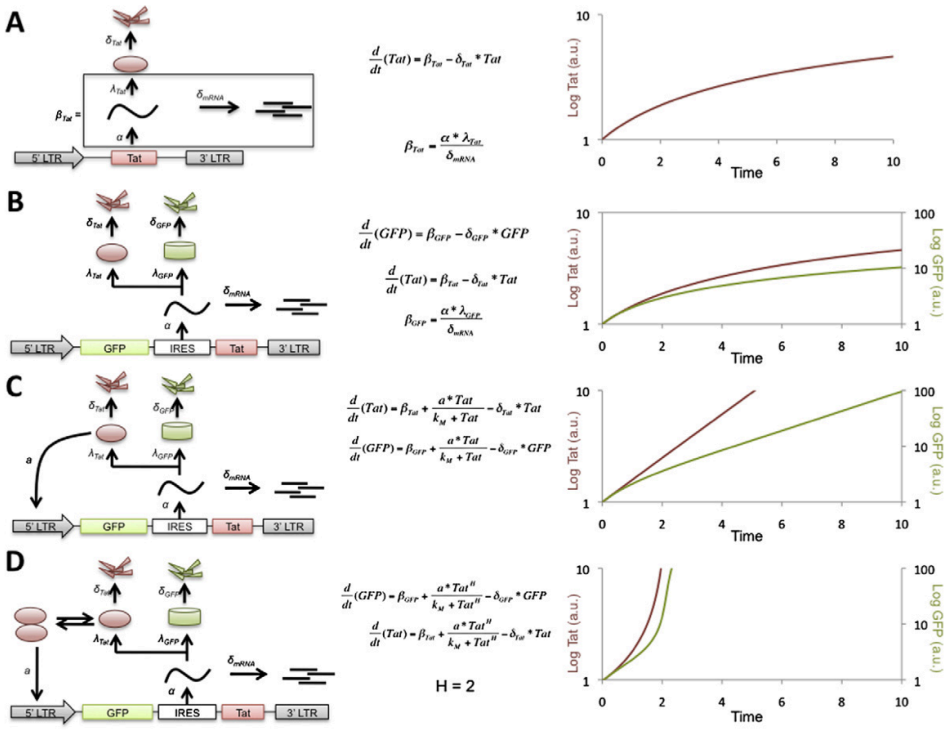

Figure 5. Mathematical models that predict the behavior of potential feedback architectures underlying the HIV-1 Tat circuit.

(A) Schematic, ODE model, and numerical solution of a hypothetical LTR-Tat circuit without any feedback; the basal rate of mRNA transcription, α, the degradation of mRNA, δ_{mRNA}, translation rate of Tat, λ_{Tat}, and the degradation rate of Tat, δ_{Tat}, are considered in the model. The basal rate, β_{Tat}, of Tat expression can be approximated by lumping the terms (bottom equation) within the black-framed box. The top equation is an ordinary differential equation (ODE) that describes the dynamics of Tat in this circuit. A plot of Tat versus time (with the log of Tat plotted versus time) shows that Tat approaches a steady state at a linear rate of increase (i.e. a sub-linear rate of increase on the log-linear plot). (B) Schematic, ODE model, and numerical solution of the LGIT system without any positive feedback. The production of GFP, β_{GFP}, was calculated using the approach as used above for Tat. The first (top) equation describes the GFP dynamics, the second (middle) equation is the same as that in panel A. The plot on the right shows that GFP and Tat dynamics mirror each other and are qualitatively very similar. (C) Schematic, ODE model, and numerical solution of the LGIT system positive feedback added. Tat and GFP expression both depend on a positive-feedback term with a Hill coefficient (H) = 1. This model generates an exponential rise in Tat and GFP levels (i.e. a linear rise on the log scale). For this simulation a=1, k_M=1, and all other parameters are unchanged from above. (D) Schematic, ODE model, and numerical solution of the LGIT system encoding a positive-feedback loop with nonlinear self-cooperativity (i.e. $H > 1$). $H = 2$ can drastically changes the shape of both GFP and Tat dynamics. All parameters are the same as in panel c except that H = 2 in this simulation. All simulations where performed in Mathematica™. Figure originally published as Figure 3 in Razooky and Weinberger, Methods, 2011.

3.1c ODE Model of Tat Positive Feedback loop with potential for bistability (i.e. H>1, Tat is self-cooperative)

Here we model the possibility that Tat positive-feedback is nonlinear or operates in a self-cooperative manner. Self-cooperativity in positive-feedback loops can generate multistability (or the ability to rest in multiple states) and provide a mechanism for choosing between alternate fates (Ferrell, 2002, 2008; Ozbudak et al., 2004). Self-cooperativity can be modeled by adding a Hill coefficient (*H)* to the positive feedback term in Eq. 4 to generate:

$$\frac{d}{dt}(Tat) = \beta_{Tat} \quad + \quad \frac{x*Tat^H}{k_M+Tat^H} \quad - \quad \delta_{Tat} * Tat \quad \textit{(Eq. 5)}$$

Self-cooperativity of *H*=2 or *H*=3 could have a number of molecular interpretations including (i) Tat forming a dimer or trimer to transactivate the LTR, (ii) that there are two or three Tat binding sites on the LTR that must be bound for transactivation to occur, or (iii) that Tat is multiply phosphorylated in a cooperative manner. The kinetics of GFP and Tat expression in a self-cooperative system (Fig. 5D) are qualitatively distinct from the kinetics of GFP expression in a system without cooperativity (i.e. Eq. 4 and Fig. 5C). Self-cooperative functions grow at rates that exhibit far greater curvature than exponential growth (*H*=1) or linear growth (*H*=0). The increased curvature is due to rates of increase being relatively low below the self-cooperative threshold (because feedback is not active), and the rate being relatively high and increasing quickly once the self-cooperative threshold is reached.

Notably, the model in Eq. 5 collapses into equation 4 when $H=1$ (and to Eq *3* when $H=0$) and we will utilize this fact to differentiate between models. The model described by Eq. 5 allows for the possibility that Tat could act as a multistable switch and that HIV-1 may be able to stably rest in multiple states (e.g. active replication or proviral latency). The model described by Eq. 4 does not allow for bistability or multistability in Tat positive feedback. Below, we differentiate between the models in Eqs. 4 and 5 by fitting the data from the single-cell microscopy experiments to our ODE models to determine the value of H.

3.2 Time-lapse Microscopy

3.2a Preparing Cells for Imaging

Once an appropriate fluorescently-labeled cell line expressing the constructs has been created, the next step in the process is imaging. Jurkat T lymphocytes are non-adherent cells and in order to perform time-lapse microscopy the cells must be immobilized. Non-adherent cells can be immobilized in a confluent monolayer by trapping them within microfluidic or small-chambered devices (Groisman et al., 2005), trapping within a polymer matrix (Upton et al., 2007), or using a 'sticky' film coating on a glass-bottom imaging dish (www.glassbottomdishes.com). The type of sticky substance used can be fibronectin, poly-L-lysine, or other substrates. We have successfully used BD Cell-Tak™ (www.BDBiosciences.com) a formulation of proteins isolated from marine mussels. A thin film of this substance allows for uninterrupted imaging of individual Jurkat cells for up to 30 hours. We use 35mm glass-bottom dishes or glass-bottom 96-

well plates on which we 'stick down' Jurkat cells as a monolayer of cells. Once cells are stuck down, the imaging process can begin.

3.2b Imaging Conditions and Image Acquisition

For time-lapse imaging of live cells, an inverted fluorescent microscope, with a motorized stage, and environmental incubation chamber (5% CO_2, 70-90% humidity, 37°C) is necessary. Software to control the microscope and automate the image acquisition process is essential (an excellent open-source microscope controller software is μManager which is available at: http://valelab.ucsf.edu/~nico/MMweb/overview.php) and microscope hardware and software that can minimize focal drift and maintain the focal plane of the microscope over the course of a multi-hour imaging experiment is also extremely helpful. To ensure cell viability during the course of the experiment, we usually monitor and carefully control excitation power, exposure time, humidity, CO_2, and temperature. Typically, 20-30 distinct X-Y positions (or nodes) on a glass-bottom dish are chosen, with a typical exposure time per node of 300-500ms, and one image captured every 5-10 minutes. We have determined that this exposure time minimizes photobleaching and phototoxicity in Jurkat cells, while still maintaining a high signal-to-noise ratio. It is important that the camera setting be tuned so that the dynamic range of fluorescence increase or decrease over the course of the experiment does not produce saturation or allow the signal to drop into a regime of poor signal-to-noise. The type of objective, gain, and offset settings can also be changed to ensure proper image quality (a number of web-based resources, e.g. Nikon www.microscopyu.com, provide helpful information on microscopy conditions).

After acquiring the time-lapse data series, analysis programs are needed to 'segment' the image. There are many options for segmentation programs capable of tracking single cells, (one commonly used program is CellProfiler (Jones et al., 2008)) and for an excellent review of single-cell imaging and segmentation see (Locke and Elowitz, 2009). Many labs, including us, utilize custom-written Matlab™ programs for automated cell tracking and to quantify GFP fluorescence intensity in individual cells over time (Weinberger et al., 2008a).

3.2c Sample Trajectories from an Experiment

Once the individual cell trajectories have been acquired by 'image segmentation' these trajectories must be processed before comparison to mathematical models and fitting. We follow a scheme of background subtraction of the segmented trajectories followed by conversion to a log scale (Fig. 6). In this experiment, Jurkat T Lymphocytes with a stable integration of the LGIT plasmid are stimulated by tumor necrosis factor alpha (TNFα) and our automated in-house Matlab™ software extracts and segments raw trajectories for 25 individual cells. Background subtraction is then performed so that the fold-increase in GFP expression over background can be analyzed – we have found that background subtracted data exhibits a fold-increase that agrees with flow cytometery and Western Blot data (Weinberger and Shenk, 2007). The individual cell trajectories are extracted from the single-cell movie (Fig. 6A) and the mean is calculated (Fig. 6B), and then converted to log scale (Fig. 6B, inset). Conversion to log scale is a typical approach used to fit data since log conversion minimizes the contribution of outlier data points. Although, the exact molecular interactions of Tat and the LTR cannot be deduced from the trajectories in Fig. 6C, the GFP expression kinetics do not appear to increase at a rate

greater than exponential (i.e. linear on a log scale)(Fig. 6B). The following sections will focus on fitting this time-lapse microscopy data to Eqs. 4 and 5 to determine the value of *H*.

3.2d Fitting the Single-cell Data to a model

To fit our microscopy data with the ODE model in Eq. 5, any number of different ODE solver software packages can be used (e.g. Matlab™ or Mathematica™). The mean of the 25 'segmented' trajectories is imported into the ODE-solver software package for fitting to Eq. 5. Ideally, parameter values for the basal rate of GFP and Tat expression – as well as the decay rates of GFP and Tat proteins – are determined from parallel experiments (e.g. with the LG construct or similar) but the values of these parameters should not effect the shape of the GFP increase. The only parameter that is essential to fit is the Hill coefficient, *H*. Fig. 6C shows that $H = 1$ provides the best fit to the single-cell microscopy data for the LGIT construct. This analysis of the single-cell time-lapse data argues strongly that the Tat positive-feedback loop does not encode self-cooperativity and that Tat positive feedback lacks the architecture required for bistability. In support of this finding, parallel dose-response experiments, sorting experiments, and FRET-based analysis confirm that the Tat positive-feedback loop is not bistable (Weinberger and Shenk, 2007).

The lack of bistability in Tat positive-feedback raises the question of how this circuit is able to mediate a decision between two states (on vs off) in the absence of bistability. Below, we show how simplified models of Tat positive feedback that consider stochastic

fluctuations are sufficient to generate a decision between two different states (on vs off) without requiring $H > 1$ or bistability.

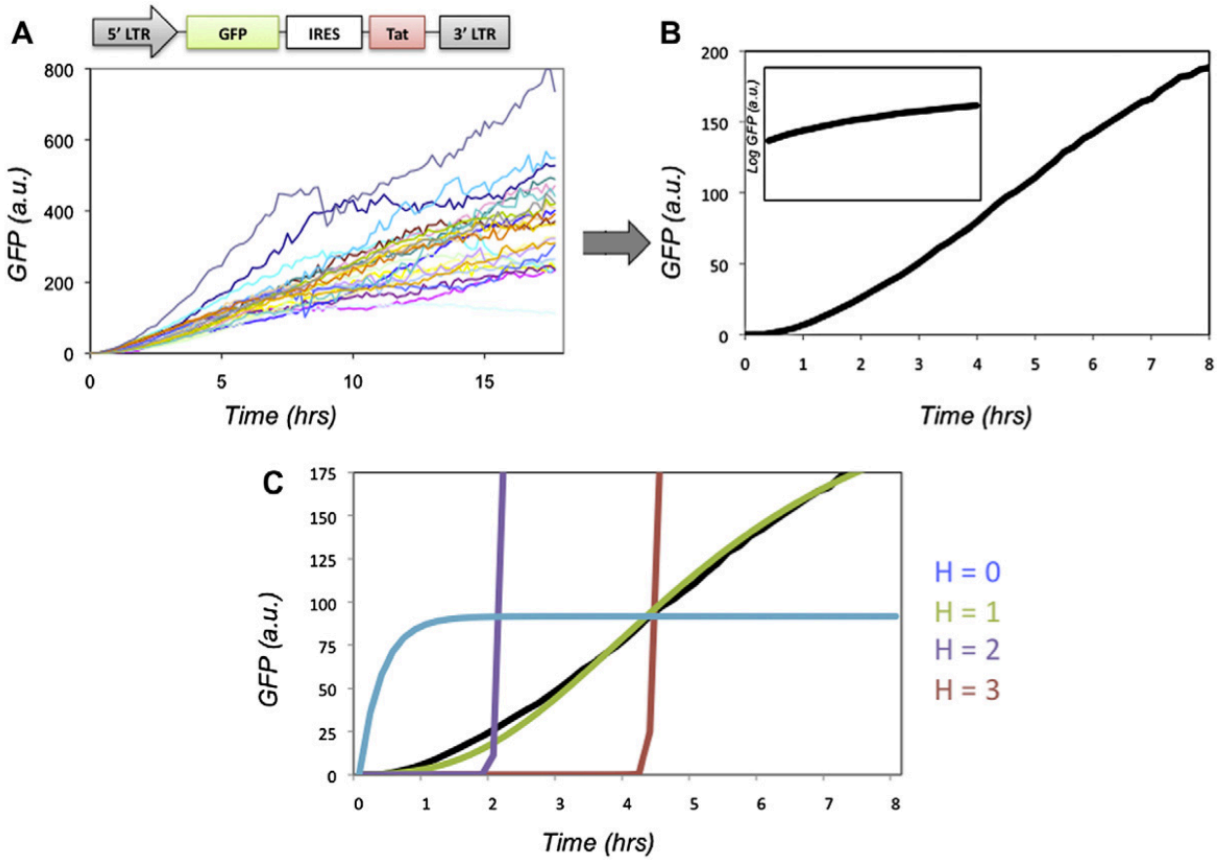

Figure 6. Time-lapse GFP trajectories from individual LGIT-infected cells show that *H* = 1 for the Tat positive-feedback circuit.

(A) Plot of raw GFP trajectories versus time obtained after image segmentation of a time-lapse confocal microscopy movie of LGIT Jurkat cells immobilized on a glass slide and imaged live for 18 hours after addition of TNFα. Each trajectory represents an individual cell and each trajectory has been pre-processed by background fluorescence subtraction. (B) Plot of the calculated mean of the individual cell trajectories (black line). Inset: Log-linear plot of the mean that is used for nonlinear least squares regression fitting to mathematical models. (C) Nonlinear least-squares regression fitting of the single-cell data to the mathematical model in Eq. 5. All parameters in Eq. 5 except *H* were fit to a parallel time-lapse movie of LG + TNFα (data not shown, see (Weinberger and Shenk, 2007)) and *H* was allowed to vary. The green line represents the best fit obtained, which is $H \approx 1$. Simulations were also performed by fixing the value of *H* to 0 (blue), 2 (purple), or 3 (red); none of these simulations generated a trajectory that could fit the data nearly as well as $H = 1$. (fits were performed on log-converted data to minimize the influence of outlier data points, fits trajectories were then back converted to linear scale). Figure originally published as Figure 4 in Razooky and Weinberger, Methods, 2011.

3.3 Stochastic Models of the Tat positive-feedback loop

3.3a Background on Stochastic Fluctuations and Stochastic vs ODE Modeling

Stochastic 'noise' arises from random thermal fluctuations in the concentration of protein, RNA or other molecules within the cell and is an unavoidable aspect of life at the single-cell level. Even cells in a clonal population (i.e. isogenic background) exhibit considerable cell-to-cell variation in the level of any specific gene product due to stochastic noise (Blake et al., 2006; Kaern et al., 2005; Raj and van Oudenaarden, 2008). The origin of this noise is biochemical: it arises from intracellular processes that are driven by reactant molecules randomly diffusing and colliding within the cell. Noise in gene expression can arise from the random timing in individual reactions associated with promoter remodeling, transcription and translation (Elowitz et al., 2002; Raser and O'Shea, 2004; Swain et al., 2002) and intercellular differences in the *amount* of cellular components (for example, RNA polymerase, transcription factors and ribosomes) also cause variations in expression levels. Measurements in live, single cells have shown that gene expression noise can lead to large statistical fluctuations in protein and mRNA levels in both prokaryotes and eukaryotes (Bar-Even et al., 2006a; Golding et al., 2005; Newman et al., 2006; Raj et al., 2006). These fluctuations (i.e. noise) can have significant effects on biological function and phenotype.

The ODE models we have considered until now are continuous approximations (e.g. ODE models consider concentration, a quantity that varies smoothly even when

describing the transition between a single molecule and zero molecules). ODE models describe the mean of a population, do not typically consider these molecular fluctuations, and are essentially an approximation for systems where a large number of molecules are present such that molecular fluctuations cancel out (Fig. 7A).

Unlike ODE models, stochastic models describe the state of the system in terms of numbers of molecules and they model discrete numbers of molecules for each species (not continuous values such as concentration). Stochastic models are implemented by writing down a reaction scheme (Fig. 7A) where the probability of any reaction going forward is modeled in a 'Monte-Carlo' fashion by choosing random numbers from distribution that describe the rate of random collisions between reactant molecules. Thus, the integer numbers of reactants and products fluctuates randomly between any two given simulation runs. In the regime of very large numbers of molecules, the fluctuations begin to overlap and cancel out such that stochastic models collapse to the ODE model (Gillespie, 2009). A clear difference between ODE models and stochastic Monte-Carlo models is that in Monte-Carlo models, each simulation trajectory can and will vary from every other simulated trajectory. Thus, by running many simulations we can generate a histogram of the trajectories for a given point in time, compare the simulated histogram to flow cytomtery histrograms (Weinberger et al., 2005), and even use the variation around the mean to quantitatively analyze promoter architecture (Singh et al., 2010a). For a thorough review on analyzing noise in gene circuits see (Kaern et al., 2005). Below, we present a stochastic model of HIV Tat transactivation, and demonstrate how this non-bistable model is sufficient to reproduce the HIV-1 decision-making phenotype.

3.3b Moving from an ODE to a Stochastic Model

Much like the setup of the wiring diagram in ODE models, a wiring diagram or 'cartoon' of the molecular reactions is helpful in generating a stochastic model. Based on our single-cell analysis (Weinberger and Shenk, 2007) and literature studies (Pagans et al., 2005), we can propose the wiring diagram and corresponding set of chemical reactions (Fig. 7B). The reaction scheme in Fig. 7B is by no means comprehensive and is intended to describe a minimal set of reactions that are sufficient to generate a 'decision' (or bifurcation) in a positive-feedback loop without self-cooperativity (Weinberger et al., 2005). Each arrow indicates the direction of the reaction and the speed (or probability) of each reaction is indicated above (or below) each arrow by a parameter constant. As described in the next section, these reactions can be coded into freely available simulation programs such as BioNetS (Adalsteinsson et al., 2004) and simulations can then be run to analyze the model of interest.

3.3c Running Stochastic Simulations

A variety of programs allow for stochastic modeling by running Monte-Carlo simulation using an algorithm now referred to as the 'Gillespie' algorithm (Gillespie, 1976, 1977). Chemical reaction schemes can be coded for simulation using the Gillespie algorithm in a programming language (e.g. FORTRAN or C++), Matlab™, or a web-based freeware graphical user-interface (GUI) software tool such as BIONETS (Adalsteinsson et al., 2004). In Fig. 7B, we show a schematic and reaction scheme for the LGIT circuit. The reaction scheme models a positive-feedback circuit lacking a bistable threshold

(Weinberger and Shenk, 2007) and we present a sample simulation of 10,000 individual trajectories (where each trajectory represents a single cell).

At time=0 (the start of the simulation) each cell contains identical initial conditions, but over time, each cell follows a different (random) path. At any point in time we can analyze all GFP values to generate a histogram. The simulated model of Tat positive feedback is sufficient to reproduce the GFP bifurcation from the flow-cytometry data in Fig. 1. From the histogram and simulation we can see that each individual cell has the potential of entering one of two states (Bright or Off). Importantly, in this model, every trajectory will eventually fall into the Off state (the off state is essentially a form of molecular extinction of Tat and acts as a trap from which the LGIT circuit cannot recover). This model coupled with experimental analysis (Weinberger et al., 2005; Weinberger and Shenk, 2007) was the first demonstration that a decision-making circuit, which lacked bistability, could generate a developmental bifurcation. Later work went on to demonstrate that the duration of the Tat transient in the *bright* state controlled entry and exit from HIV-1 proviral latency in full-length HIVΔEnv system (Weinberger et al., 2008a).

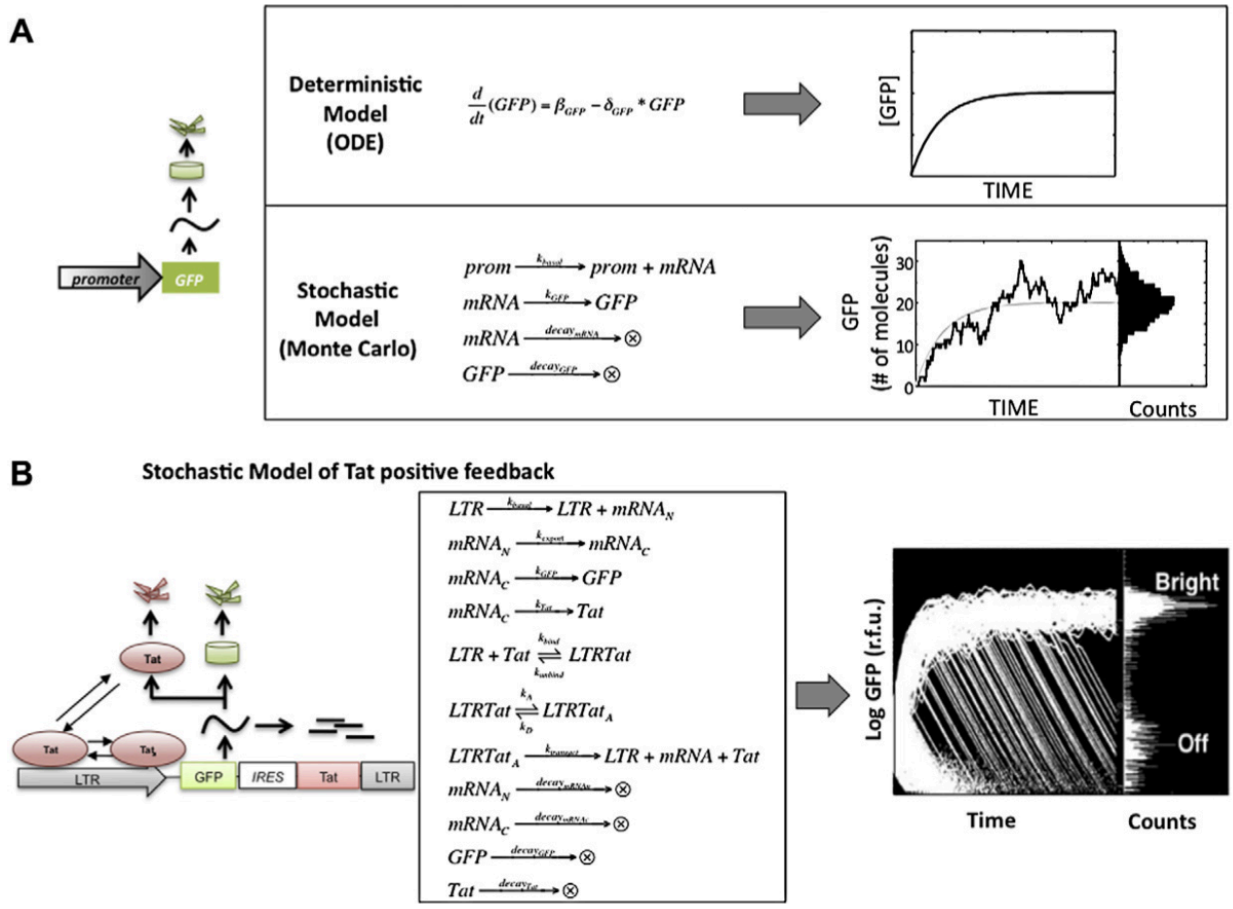

Figure 7. Stochastic fluctuations in the levels of Tat account for the developmental bifurcation of HIV-1.

(A) Deterministic (continuous) ODE models versus stochastic Monte-Carlo models for a simple gene circuit (left). The ODE model (top) generates a single smooth line that approximates the population average of a large population of GFP molecules. The stochastic model (bottom) considers a set of chemical reactions and generates a fluctuating trajectory that represents the number of molecules in the system (e.g. a single cell) over time. When simulated, each trajectory from such a stochastic model will vary and a histogram of GFP can be generated for any point in time. (B) Schematic, reaction scheme, and stochastic simulation results for the LGIT circuit (reproduced from (Weinberger et al., 2005)). The reaction scheme models a positive-feedback circuit lacking a bistable threshold (Weinberger and Shenk, 2007) and 10,000 individual trajectories (where each trajectory represents a single cell) are shown for simulated time-span of one week to generate the histogram of 10,000 cells at right. The simulated model of Tat positive feedback is sufficient to reproduce the GFP bifurcation from the flow-cytometry data in Figure 1. The initial conditions for this simulation are: all species begin with zero value except LTR (1 copy), Tat (1 - 5 molecules), and GFP (25,000 – 125,000 molecules); the parameter values are: $k_{basal} = 10^{-8}$/sec, $k_{bind} = 0.00015$/sec, $k_{unbind} = 0.017$/sec, $k_A = 0.001$/sec, $k_D = 0.13$/sec, $k_{transact} = 0.1$/sec, $decay_{mRNAn} = 0.000048$/sec, $k_{export} = 0.00072$/sec, $k_{GFP} = 0.5$/sec, $k_{Tat} = 0.00132$/sec, $decay_{mRNAc} = 0.000048$/sec,

$decay_{GFP}$ = 0.00000301/sec, $decay_{Tat}$ = 0.0000043/sec. Figure originally published as Figure 5 in Razooky and Weinberger, Methods, 2011.

4. Conclusions

Here we argue that a coupled single-cell imaging and mathematical modeling approach can differentiate between alternate models of the HIV Tat transcriptional circuit and enables mapping of the architecture of the HIV-1 Tat latency circuit. We demonstrate that the Tat circuit lacks bistability (the ability to stably rest in two alternate states) by measuring the Hill coefficient of Tat feedback in single cells. We also show that a monostable circuit architecture that exploits stochastic noise in gene expression can account for the Tat circuit's ability to 'choose' between two alternate states. Importantly, the Tat circuit represents the first example of a natural decision-making circuit shown to lack bistability and utilize stochastic noise to probabilistically 'choose' between two alternate states (Weinberger et al., 2005; Weinberger et al., 2008a; Weinberger and Shenk, 2007).

Chapter 4: HIV latency is controlled by viral genetic circuitry

1. Introduction

The role of intrinsic genetics and environmental factors in determining organism phenotype can be clear-cut. For instance, inheritance of X or Y chromosomes determines sex in mammals but the sex of sea turtles is determined by the temperature of the surrounding sand (Bull and Vogt, 1979). These straightforward examples are the exception rather than the rule, especially in cell-fate decisions, like in lambda phage, where it is unclear if intrinsic genetics (McAdams and Arkin, 1997; Ptashne et al., 1982) or environmental factors (St-Pierre and Endy, 2008) determines phenotype. Here, to

understand the contributions of intrinsic genetic circuitry and environmental factors in cell-fate decisions, we use the HIV decision-making process as a model system (Weinberger et al., 2005).

As previously mentioned, the accepted theory for latency regulation postulates that infected cell-state determines HIV's fate decision between active replication and proviral latency (Ruelas and Greene, 2013; Siliciano and Greene, 2011). Theories where cell-state controls latency postulate that activated lymphocytes are permissive to HIV replication while resting lymphocytes promote latency. In resting lymphocytes, HIV transcription is restricted due to sequestration of host transcription factors to the cytoplasm and accumulation of restrictive epigenetic markers on the HIV promoter (Pearson et al., 2008; Tyagi et al., 2010). Upon activation of the cell, these restriction factors are alleviated, allowing for HIV transcription and replication. Because host-cell state seems to completely determine viral phenotype, latency is thought to be an epiphenomenon with no role in the natural history of the virus (Coffin and Swanstrom, 2013; Eisele and Siliciano, 2012). If latency is simply an epiphenomenon stemming from host-cell state, then HIV should not be able to control its own phenotype. However, there is circumstantial evidence for viral control of fate, which confounds the deterministic hypotheses based on cell-state.

Evidence for viral control of HIV fate is based on the observation that modulation of HIV circuitry alters the fraction of active-to-latent infections. HIV encodes for a transcriptional positive feedback loop where the viral transactivator of transcription Tat,

amplifies expression from the HIV long terminal repeat promoter, LTR (Razooky and Weinberger, 2011). Attenuation of HIV Tat positive feedback strength can increase the number of latent infections (Weinberger et al., 2008a) and patient-derived latently infected cells are enriched for debilitating mutations in Tat's transactivation domain (Yukl et al., 2009). Strikingly, simple stochastic models of Tat positive feedback—that do not take cell-sate into account—can successfully recapitulate the developmental bifurcation in HIV (Weinberger et al., 2005). Although limited, this evidence suggests a role for viral circuit control of HIV fate. Consequently, since evidence exists for viral and cellular control of HIV phenotype, it is unclear if latency is simply an epiphenomenon or an intrinsic feature hardwired into viral circuitry.

The major obstacle in testing for viral circuit and cell-state control of HIV latency is that cell-state changes invariably affect viral circuitry, i.e. activation of cell-state recruits transcription factors to the LTR and initiates Tat positive feedback (Ott et al., 2011). To test if viral circuitry is sufficient to control HIV fate, it is necessary to independently manipulate viral circuitry and cell-state. If HIV fate is completely dependent on cell-state, then activation or relaxation of cell-state should determine HIV phenotype (Fig. 8A). However, if latency is hardwired into viral circuitry, tuning viral circuitry, independent of cell-state, should be sufficient to control HIV phenotype (Fig. 8B).

Here, to test if HIV latency is hardwired into viral circuitry we synthetically reengineer HIV to independently manipulate viral circuitry and cell-state. The data show that HIV Tat positive feedback can control HIV latency despite drastic changes in cellular state.

This suggests that HIV latency is an intrinsic virally encoded phenomenon that may play a role in the natural history of the virus.

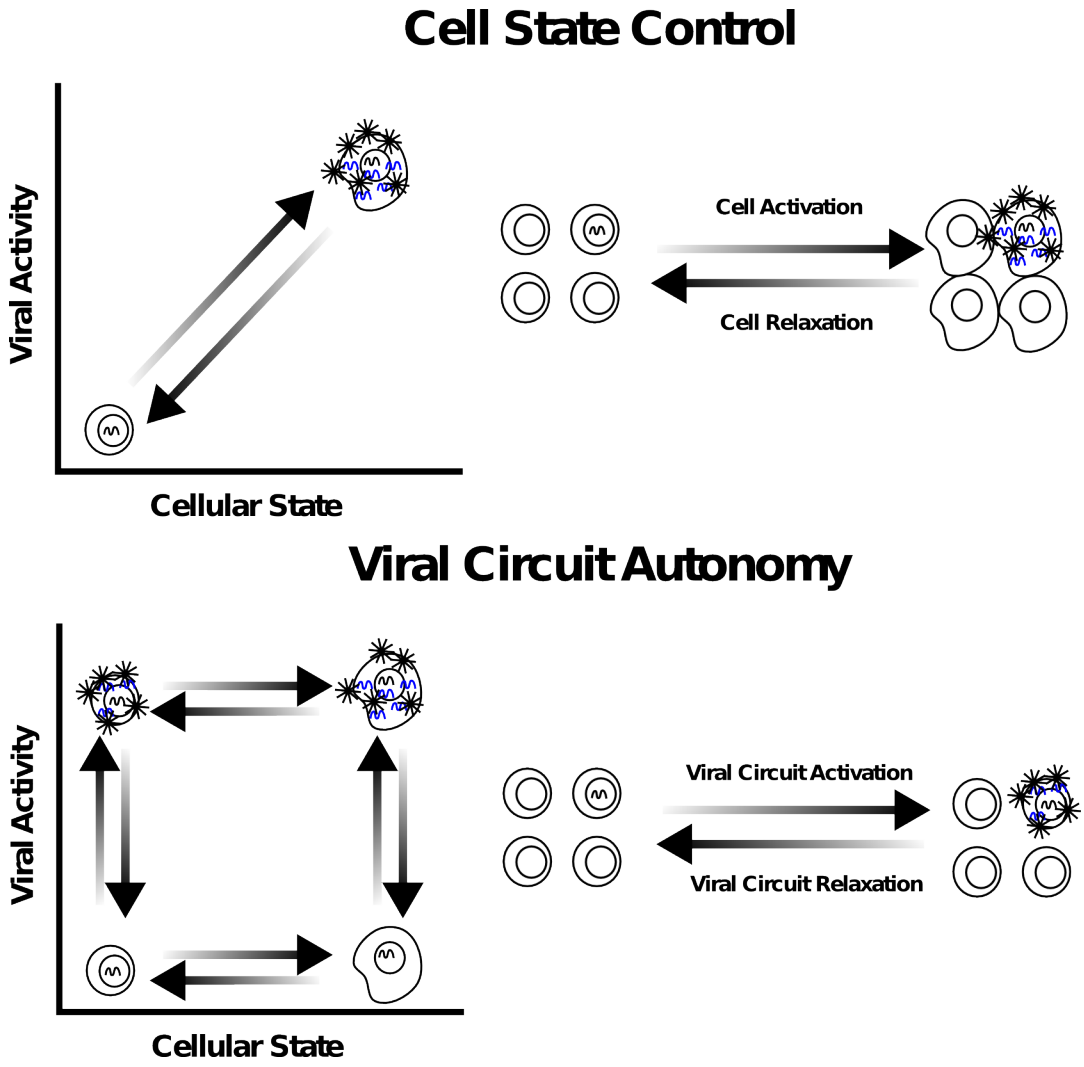

Figure 8. Schematic for activity of a viral circuit that is dependent on, or autonomous of, cellular state.

(A) Plot of viral circuit activity versus cellular state with the condition of cell state control over circuit activity (left). Cells in a resting state only harbor silenced, inactive virus, while cells in an active state only harbor activated, replicating virus. Transitioning between resting and activated will induce or silence all viruses in a population of cells (right). (B) If viral activity acts autonomously to cellular state (left), then circuit activity could act independently of cellular state and resting and activated cells could harbor latent or active virus. Additionally, circuit autonomy would allow for activation of silent, inactive virus without a need for changing cellular state (right).

2. Minimal Circuit Results

2a. Tuning viral circuit strength is sufficient to toggle HIV gene expression between active and latent

To determine if HIV gene-regulatory circuitry can autonomously control proviral latency without changes in cellular-activation state, we developed synthetic circuits where viral circuitry could be toggled *independent* of cell state (Fig. 10). The synthetic circuits are based upon a minimal model of the HIV latency circuit (Jordan et al., 2001; Weinberger et al., 2005) and encode for a transcriptional positive-feedback loop where the viral transactivator of transcription, Tat, amplifies expression from the HIV long terminal repeat promoter, LTR. The minimal LTR-Tat circuit is sufficient to recapitulate latent gene expression and stimulation with cell-state modifiers will reactivate proviral expression from an inexpressive 'OFF' state to expressive 'ON' state (Weinberger et al., 2005). Thus, the system provides a minimal-circuit model to test if Tat circuitry is sufficient to control latency.

The minimalist synthetic toggle circuit encodes Tat fused to a controllable-proteolysis tag, FKBP (Banaszynski et al., 2006) under the control of the HIV LTR (Fig. 10). FKBP degradation is reversibly inhibited by a small molecule, Shield-1, allowing Tat half-life to be rapidly tuned. The Tat-FKBP fusion was tagged with a photo-switchable fluorescent protein, Dendra-2 (Gurskaya et al., 2006), which allowed for light-based pulse-chase experiments (Zhang et al., 2007) to measure Tat half-life destabilization in single cells (Fig. 11). In this minimal LTR-Tat-Dendra-FKBP viral vector, Tat half-life is reduced to

2.5 hours, a ~3.3 fold reduction from the native 8 hour Tat half-life, in the absence of Shield-1 but returns to its native half-life (Weinberger and Shenk, 2007) in the presence of 1μM Shield-1 (Fig. 12). Stochastic theoretical models of the Tat positive feedback loop (Weinberger et al., 2005) predict that this change in Tat half-life would be sufficient to toggle the HIV positive feedback between 'ON' and 'OFF' at a majority of viral integration sites (Fig. 9).

As predicted by simulation (Fig. 9), altering Tat half-life alone, by addition or removal of Shield-1, is sufficient to toggle between latent and active expression in the isoclonal populations (Fig. 10). The observed reactivation is not due to pleiotropic effects of Shield-1 since, (*i*) Tat-Dendra fusion proteins lacking FKBP are insensitive to Shield-1 (Fig. 13), and (*ii*) activation is not simply due to an increase in the half-life of Dendra-2 (Fig. 14 and 15). To distinguish between Tat half-life changes from LTR activity, the half-life of the fluorescent reporter and half-life of Tat were decoupled (Fig. 16). The decoupled system corroborates the finding that Tat positive feedback is sufficient to control viral switching from an inexpressive 'OFF' to expressive 'ON' state (Fig. 17). These data show that Tat positive feedback is sufficient to toggle HIV gene expression between a quiescent state and an actively expressing state.

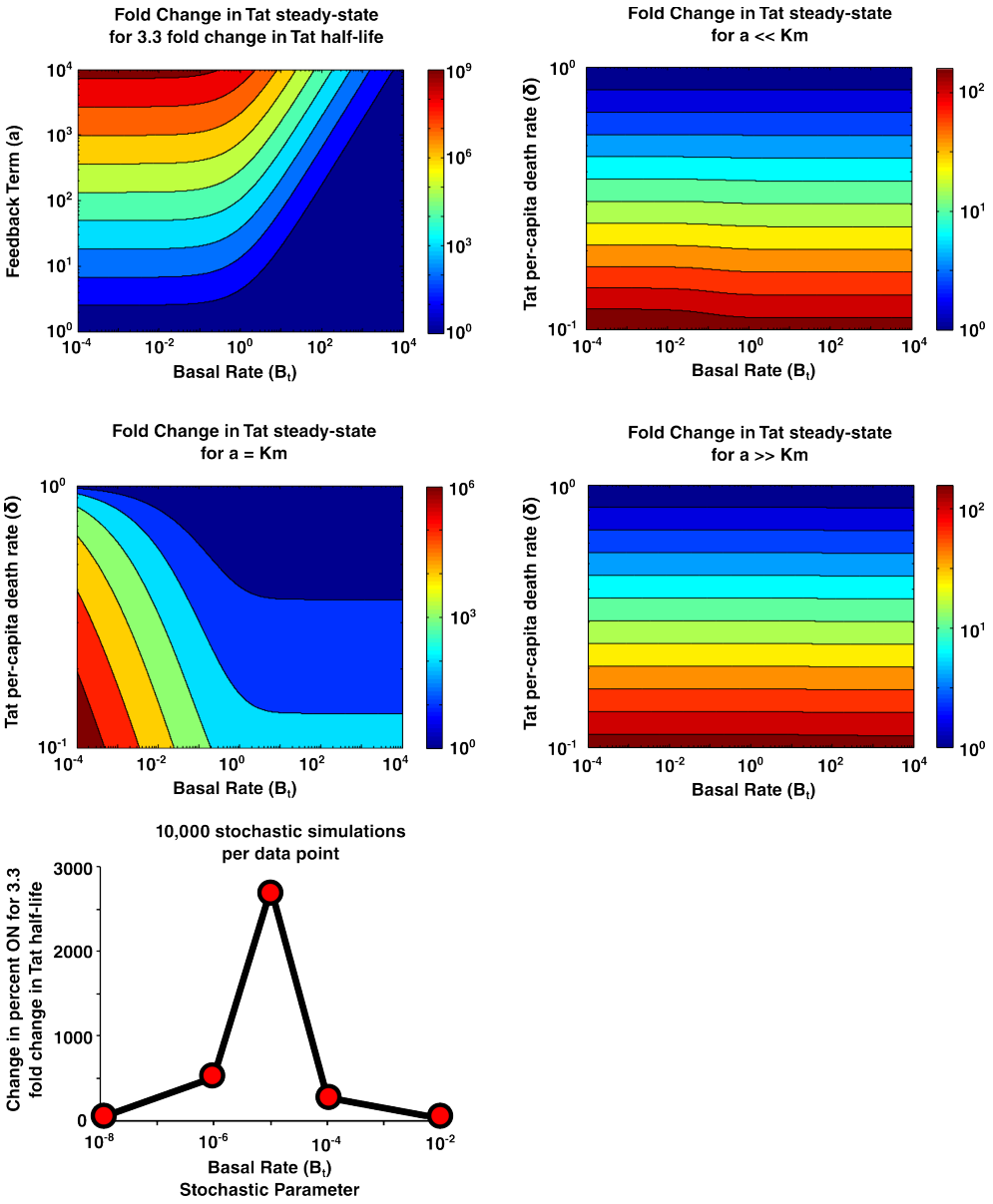

Figure 9. ODE and stochastic models of HIV positive feedback predict that changing Tat half-life 3.3 fold is sufficient to induce positive feedback.

The top left graph is an isocline plot showing the relationship between positive feedback strength and the basal rate of transcription. The top right, middle left, and middle right graph are isocline graphs showing the relationship between the changes in Tat half-life and basal rate of transcription for very strong feedback (a>>km), 'average' feedback (a = km) and weak feedback (a<<km), respectively. The color bar of each graph represents the log fold change in Tat steady-state values. Tat positive feedback was modeled using ODE's as described in Chapter 3 and in Razooky and Weinberger, 2011. The bottom left graph plots the results of 100,000 simulations. 10,000 simulations were run of Tat positive feedback for a given basal rate and Tat half-live (holding all other parameters equal). The number of simulations that resulted with at least 1 Tat molecule present was

considered ON as a most conservative estimate. Next, 10,000 more simulations were run for the same parameters, only changing the Tat half-life 3.3 fold. The number of ON simulations was then subtracted from the number of ON simulations from the first 10,000 simulations to get each data point. Tat positive feedback was modeled using the stochastic model shown in Chapter 2, Razooky and Weinberger, 2011, and Weinberger et al., 2005.

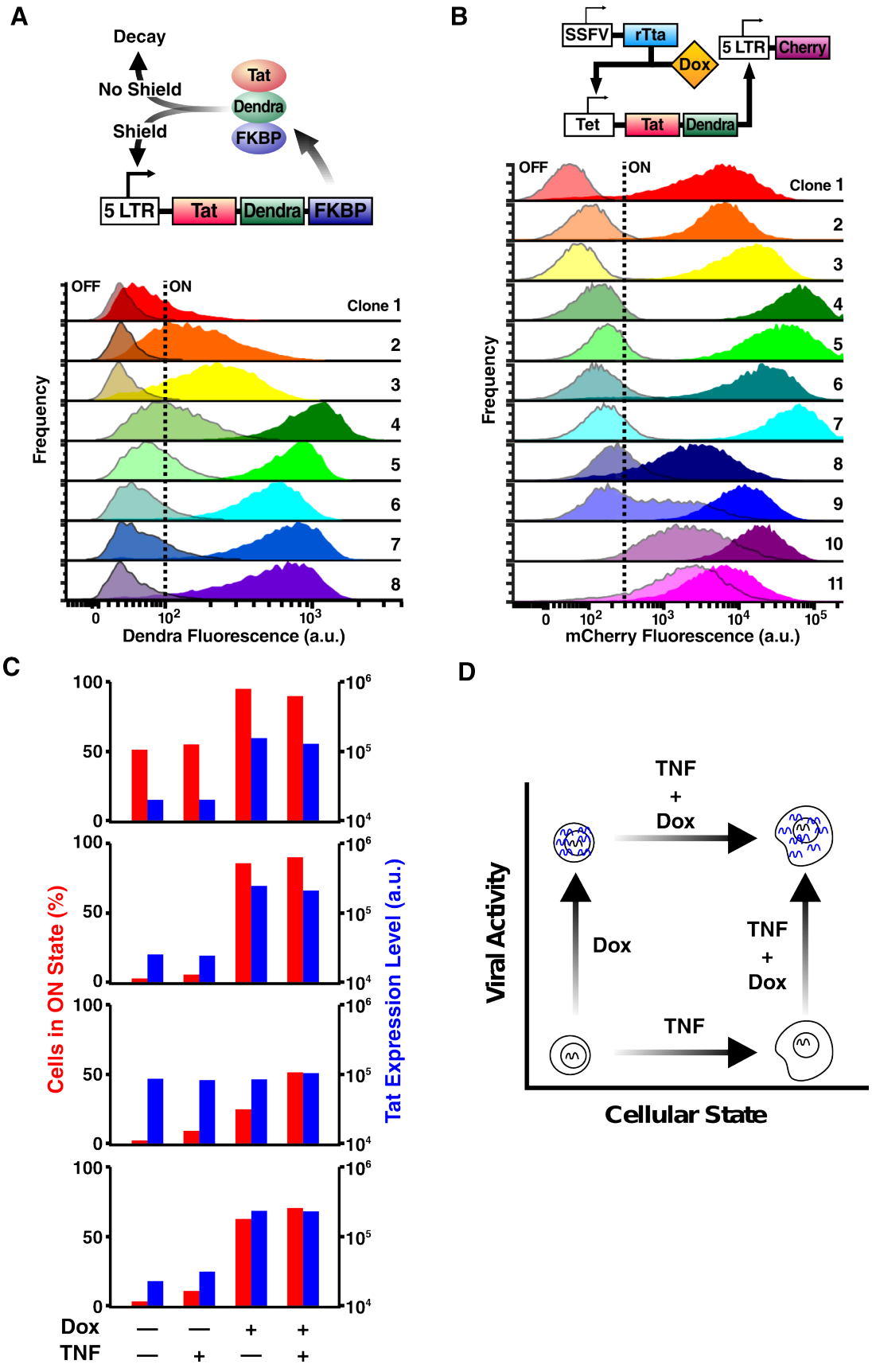
A
Decay
No Shield
Shield
Tat
Dendra
FKBP
5 LTR
Tat
Dendra
FKBP
OFF
ON
Clone 1
2
3
4
5
6
7
8
Frequency
0
10^2
10^3
Dendra Fluorescence (a.u.)
B
SSFV
rTta
Dox
5 LTR
Cherry
Tet
Tat
Dendra
OFF
ON
Clone 1
2
3
4
5
6
7
8
9
10
11
Frequency
0
10^2
10^3
10^4
10^5
mCherry Fluorescence (a.u.)
C
Cells in ON State (%)
Tat Expression Level (a.u.)
100
50
0
10^6
10^5
10^4
Dox — — + +
TNF — + — +
D
TNF
+
Dox
Dox
TNF
+
Dox
TNF
Viral Activity
Cellular State

Figure 10. Minimal HIV gene-regulatory circuitry is sufficient to control active-vs.-latent expression even in the absence of changes in intracellular state.

(A) LTR-Tat-Dendra-FKBP (LTDF) expresses Tat-Dendra-FKBP as a fusion protein. In the absence of Shield-1 (right), the fusion protein is rapidly degraded (red arrow), generating weak positive feedback (blue arrow). Shield-1 addition (left) blocks proteolysis (red arrow), allowing strong Tat positive feedback (blue arrow). Flow cytometry histograms of eight isoclonal LTR-Tat-Dendra-FKBP (LTDF) cell populations in the absence of Shield-1 (translucent histograms) or the presence of 1 μM Shield-1 (right histograms). 'OFF' and 'ON' thresholds were set based on cellular autofluorescence. (B) The SFFV promoter constitutively expresses rTta transcription factor, which binds to the Tet-ON promoter driving the expression of Tat-Dendra upon doxycycline addition. Tat-Dendra will then bind to the HIV LTR leading to the expression of mCherry. Eleven different isoclonal populations (isoclonal for LTR-mCherry) were exposed to 500ng/mL dox (opaque histograms) or no dox (transparent histograms). (C) 500ng/mL doxycycline, 1ug/mL TNFα, both, or neither were added to four isoclonal populations containing the same circuit as in (B) but with Tat-Dendra-FKBP instead of Tat-Dendra. Median Tat-Dendra expression (blue columns) and the percent of cells in the 'ON' state (red columns) for each population are shown for each condition. (D) Summary schematic of circuit activity versus cellular state. Starting from resting cell, 'OFF' populations (no doxycycline and no TNFα), resting cell, 'ON' (doxycycline alone), activated cell, 'OFF' (doxycycline and TNFα), and activated cell, 'ON' population (doxycycline and TNFα) phenotypes could be created.

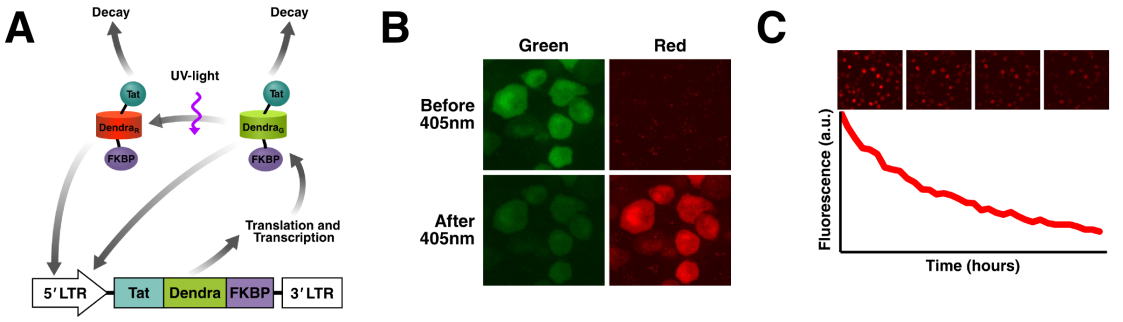

Figure 11. Schematic for measuring Tat half-life perturbations.

(A) The HIV 5' LTR drives the expression of a translational-fusion product of Tat, Dendra, and FKBP. Tat abundance is quantified by measuring $Dendra_G$ fluorescence in the green channel. However, upon exposure to UV-light, Dendra's spectral properties change and cause a fraction of $Dendra_G$ to fluoresce in the red channel ($Dendra_R$). $Dendra_G$ and $Dendra_R$ are degraded at the same rate so, by following $Dendra_R$ with time-lapse fluorescence microscopy, the half-life of the protein fusion product, and thus Tat, can be quantified. (B) Fluorescent images of Jurkat cells stably transduced with the LTDF vector before and after exposure to UV light. (C) After photoconversion, time-lapse fluorescence microscopy was performed to get the decay trajectories of the 'Red' signal over time. Using the mean of the single-cell trajectories, the half-life of Tat in different Shield concentrations was measured.

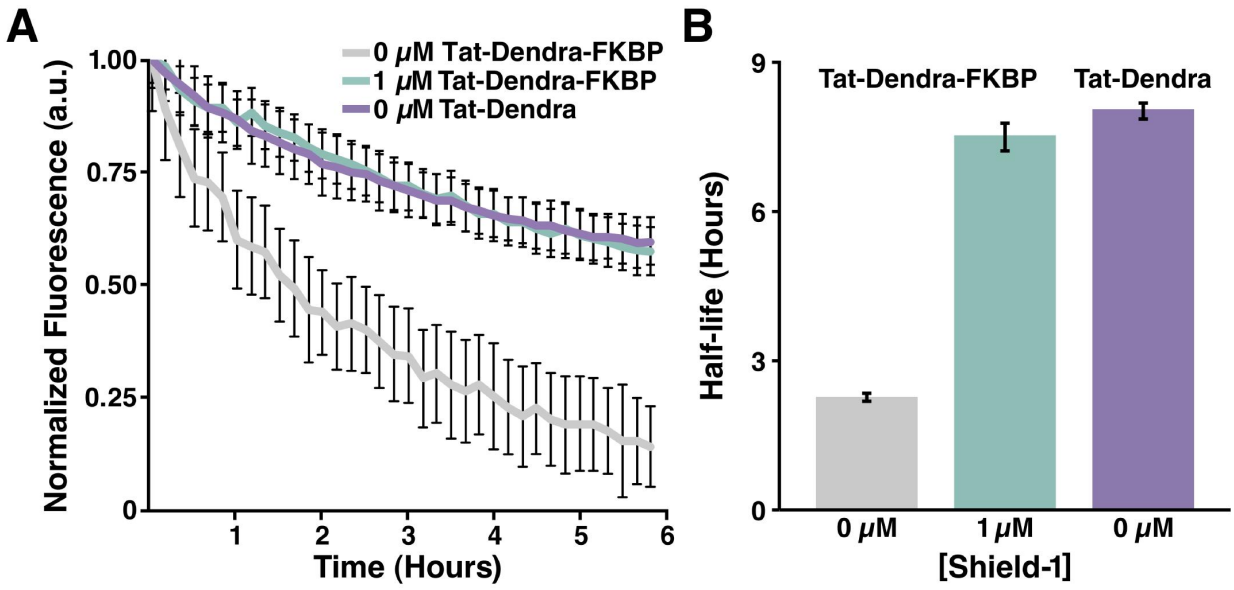

Figure 12. Tat half-life, as measured by single-cell time-lapse microscopy after photoconversion, can be tuned from 2.3 – 7.5 hours.

(A) Normalized mean fluorescence over time of Tat-Dendra-FKBP expressing cells (>50 cells per condition) after Dendra photoconversion, in the absence of Shield-1 (gray) or presence of 1uM Shield-1 (blue). The purple line is normalized mean fluorescence level of Tat-Dendra cells after photoconversion. Error bars = standard error of the mean. (B) Each curve from Fig. 1C were fit with an exponential function. In the absence of Shield-1, Tat has a half-life of 2.3 ± 0.1 hours (gray) and in the presence of 1 μM Shield-1, Tat's half-life is 7.5 ± 0.3 hours (blue). When Tat is fused to Dendra alone, Tat's half-life is 8.0 ± 0.2 hours (purple). Error bars represent the standard error of the mean from the single-cell trajectories for that condition.

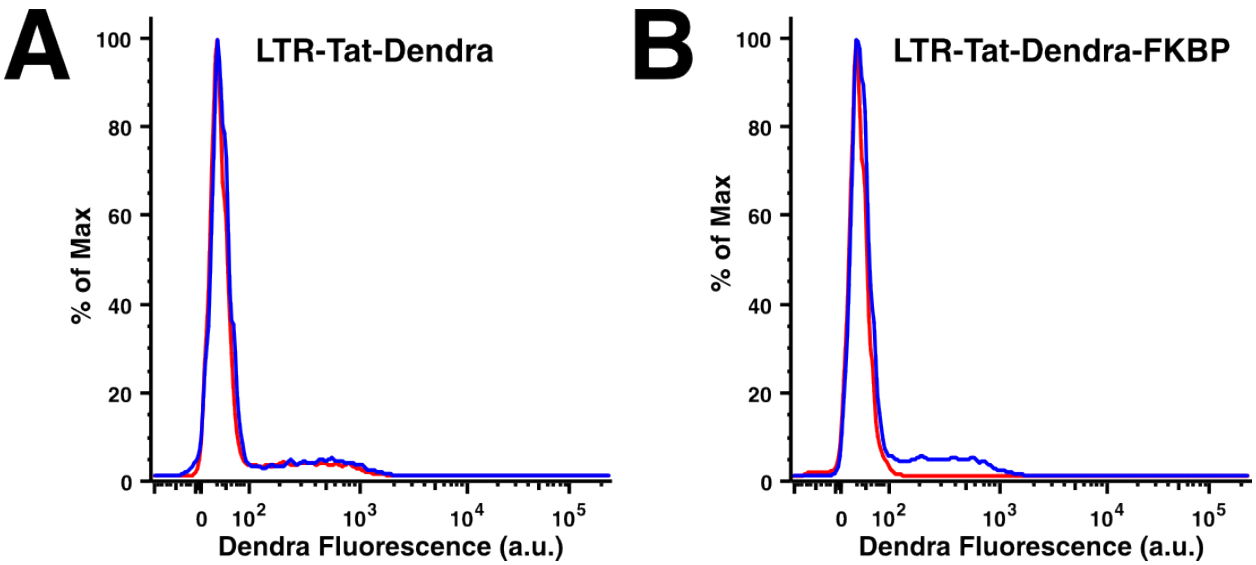

Figure 13. Shield-1 does not affect Tat positive feedback.

(A) An infected Jurkat LTR-Tat-Dendra population was cultured in the presence (blue) or absence (red) of 1 μM Shield-1. The level of fluorescence in the LTR-Tat-Dendra population does not change upon Shield-1 addition. (B) Incubating an infected Jurkat LTR-Tat-Dendra-FKBP population in the presence (blue) or absence (red) of 1 μM Shield-1 leads to large changes in fluorescence. Flow cytometry data were created as described in materials and methods.

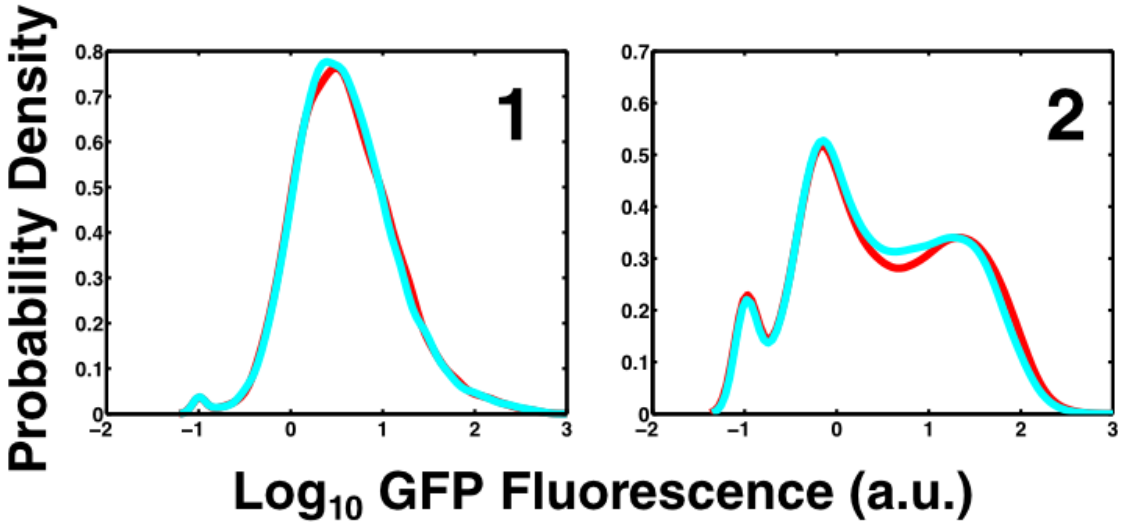

Figure 14. Shield-1 does not induce expression from the LTR.

Two isoclonal LTR-d2GFP-IRES-Tat populations—***without a FKBP domain***—were incubated in the presence (cyan) or absence (red) of 1 μM Shield-1. Shield-1 did not change expression from the HIV-1 LTR in either population. Thus, the pleiotropic effects of Shield-1 do not affect LTR gene-expression. This circuit is equivalent to the circuit in Figure 15 but mCherry fluorescent protein is replaced with GFP.

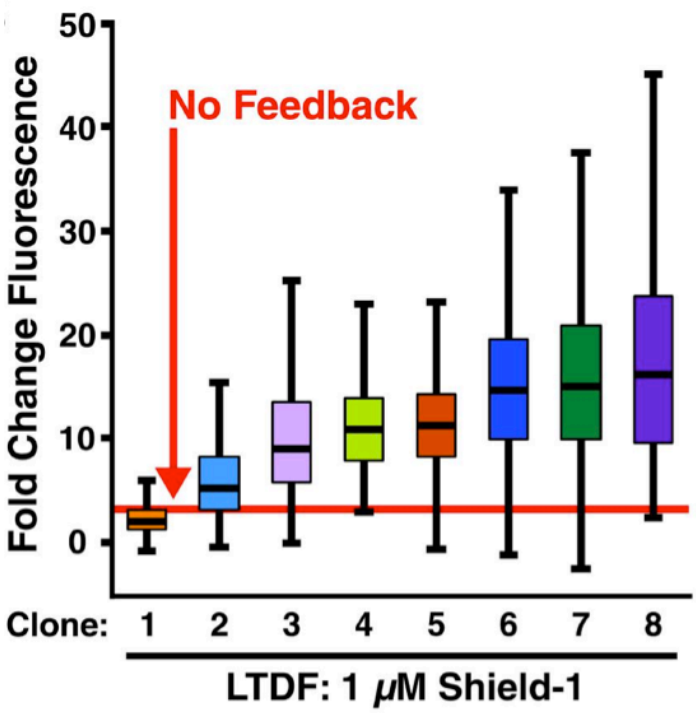

Figure 15. Stabilization of Tat leads to activation of positive feedback.

Boxplots showing the fold-change in median fluorescence after 1 μM Shield-1 addition for each isoclonal population from Figure 9A. The boxplots show the mean fold change in the median level (black horizontal line), the edge of the boxes represent the 25th to 75th percentile, and the whisker length represents extreme data points not considered outliers (MATLAB™ default setting). To generate the red 'No Feedback' line, we developed an ordinary differential equation model. Under the quasi steady-state assumption of mRNA expression, we lump transcription, mRNA decay and translation rate into a single basal term, β, and model protein kinetics by zero-order birth and first-order decay processes as follows:

$$\frac{d}{dt}(P) = \beta - \delta * P \qquad \text{(Eq. 6)}$$

where P is the protein abundance, β is the basal rate of expression, and δ is the per capita death rate. If we only change δ for different Shield-1 concentrations (i.e., β is not affected by Shield-1 (Fig. 12 and Fig. 13)), then the ratio of two different steady-state median protein abundances achieved at two different Shield-1 concentrations would be equal to the ratio of the two half-lives:

$$\frac{<P_1>}{<P_2>} = \frac{\tau_1}{\tau_2} \qquad \text{(Eq. 7)}$$

The half-life of Tat is determined for different Shield-1 concentrations (Fig. 11), so we can relate the fold change in abundance to the fold change in half-life in the absence of feedback.

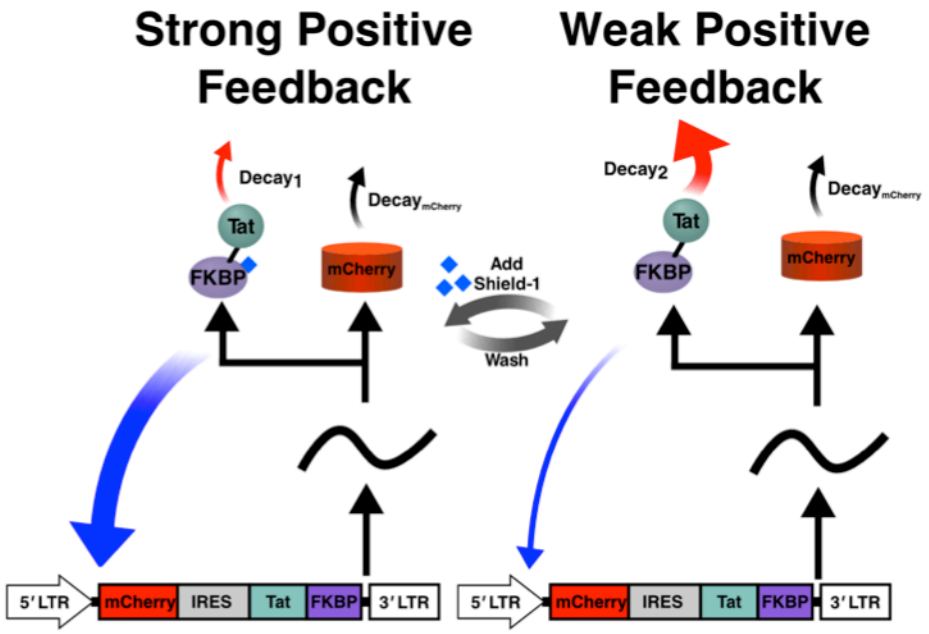

Figure 16. Schematic for a closed-loop feedback system where the perturbation to the trans-activator has been decoupled from expression levels.

In the LTR-mCherry-IRES-Tat-FKBP system, a single polycistronic mRNA is transcribed from the HIV-1 LTR that encodes for mCherry and Tat-FKBP in stoichiometric amounts. Unlike the LTDF system, this decouples the perturbation and transcriptional output so that Tat levels only affect fluorescence intensity through transcriptional mechanisms (i.e., not through half-life changes in the fluorescent reporter as in the LTDF system). In the absence of Shield-1, Tat-FKBP is rapidly degraded leading to weak-positive feedback, and lower mCherry levels. However, upon Shield-1 addition, Tat-FKBP is stabilized leading to more Tat and strong positive feedback. Subsequently, more Tat should lead to higher-level transcription from the LTR and mCherry expression.

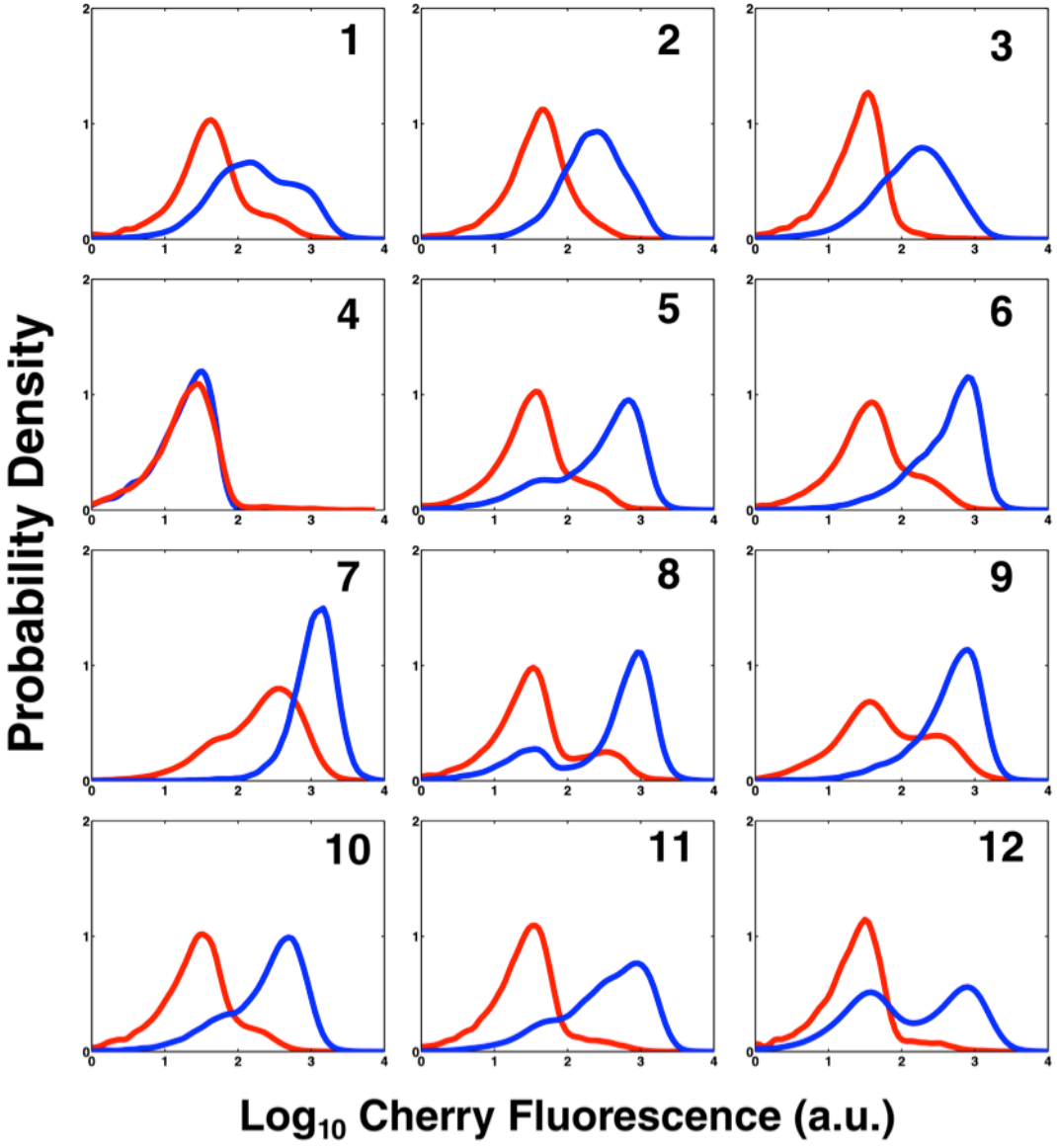

Figure 17. Tat-positive feedback controls gene-expression from the LTR (measured by mCherry).

Probability distributions of mCherry fluorescence for 12 LTR-Cherry-IRES-Tat-FKBP isoclonal populations with (blue histograms) or without (red histograms) Shield-1. Stabilization of Tat half-life leads to significant changes in Cherry expression in 11 of the 12 isoclonal populations tested. Probability density plots were created using the default ksdensity function in Matlab®.

2b. Tuning viral transcriptional strength stimulates HIV gene expression more effectively than perturbing cellular-activation state

One caveat of using tunable-proteolysis systems to toggle a circuit is that a minimal level of protein (i.e. Tat) must be present in the off state, since modulating protein half-life when protein concentration is zero would have no affect. Thus, the Tat-FKBP approach cannot test if Tat can reactivate latent cells that are fully silenced. To circumvent this obstacle so that we may test if Tat induction is sufficient to reactivate completely silent LTR's, we developed a set of open-loop circuits, based on the Tet-On system (Gossen and Bujard, 1992), to induce Tat expression *de novo*. These systems allow very tight induction of Tat expression upon doxycycline addition. To further reduce off-state Tat levels to extremely low levels, we developed a Tet-On system driving expression of Tat-Dendra-FKBP that requires both doxycycline and Shield-1 for activation (Almogy and Nolan, 2009). To test the effects of Tat induction on HIV gene-expression, these circuits were incorporated into cells that encoded an HIV LTR promoter driving mCherry fluorescent reporter (Fig. 10 and Fig. 18).

In agreement with the findings from the LTR-Tat-Dendra-FKBP circuits, the Tet-On circuits show that Tat alone is sufficient to control 'OFF'-versus-'ON' LTR expression state and LTR expression levels (Fig. 10, Fig. 19). Importantly, a number of clones ('Clones 1-3') exhibit no detectable LTR expression at all—the conventional threshold for latency (Jordan et al., 2003; Jordan et al., 2001; Pearson et al., 2008; Spina et al., 2013; Tyagi et al., 2010; Williams et al., 2006a; Williams et al., 2007b)—in the absence

of Tat induction, but inducing Tat expression is sufficient to fully reactivate these clones without the need for any cell-state activation signals.

Next, to test the effects of cell-state activation, isoclonal Tat-inducible populations were exposed to combinations of doxycycline and a cell-state activator (Fig. 10). TNFα potently activates expression from promoters containing NF-κB binding sites by stimulating recruitment of the p50-RelA heterodimer (Pazin et al., 1996). In the Tat-inducible populations, cell-state modifications alone only slightly increase expression and the percentage of cells in the 'ON' state (Fig. 10). In contrast, modification of cell-state and induction of Tat drastically increases expression from the LTR, but only slightly more than Tat induction alone (Fig. 10). Collectively, these data suggest that modulating cell-state alone is not sufficient to control HIV transcription (Fig. 10, Fig. 19). In contrast, the most potent reactivation was observed upon modification of cell-state and viral circuitry.

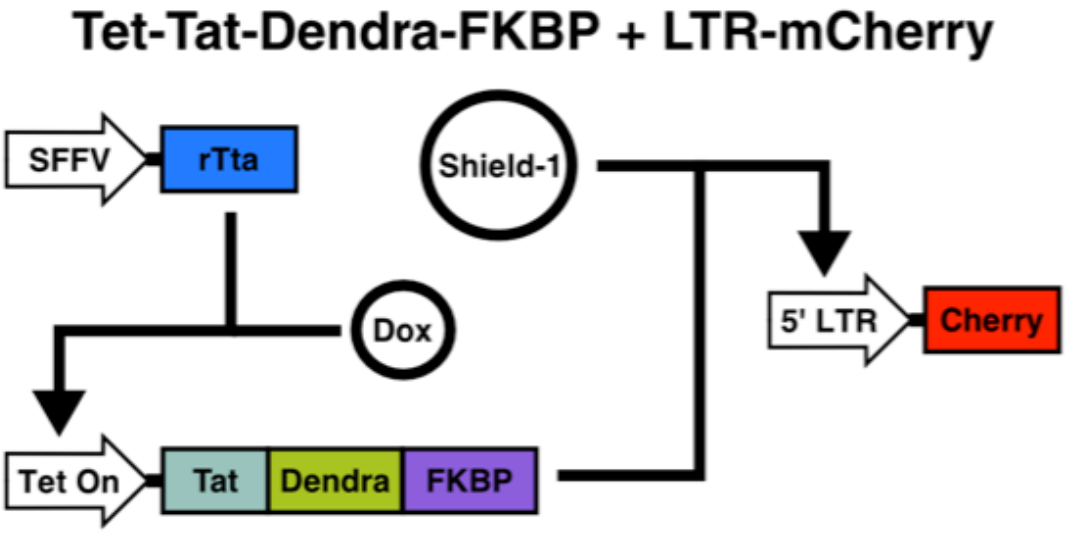

Figure 18. Schematic of Tet-Tat-Dendra-FKBP + LTR-mCherry.

The SFFV promoter constitutively expresses the rTta activator, which has the potential to transactivate the Tet-On promoter. Doxycycline (Dox) binds to the rTta transcription factor, which then binds to the Tet-On promoter and drives the expression of Tat-Dendra-FKBP. The junction point between the 'Dox' and rTta arrow represents the necessity for rTta to bind Dox for Tet-On transactivation. The level of Tat-Dendra-FKBP expression is titratable and dependent on the amount of Dox added to the system. Once Tat-Dendra-FKBP is expressed, Tat can bind to the HIV-1 LTR and trans-activate expression, thereby leading to significant increases in LTR expression in a dose-dependent manner. Tat-Dendra-FKBP stability can also be modulated through Shield-1 addition.

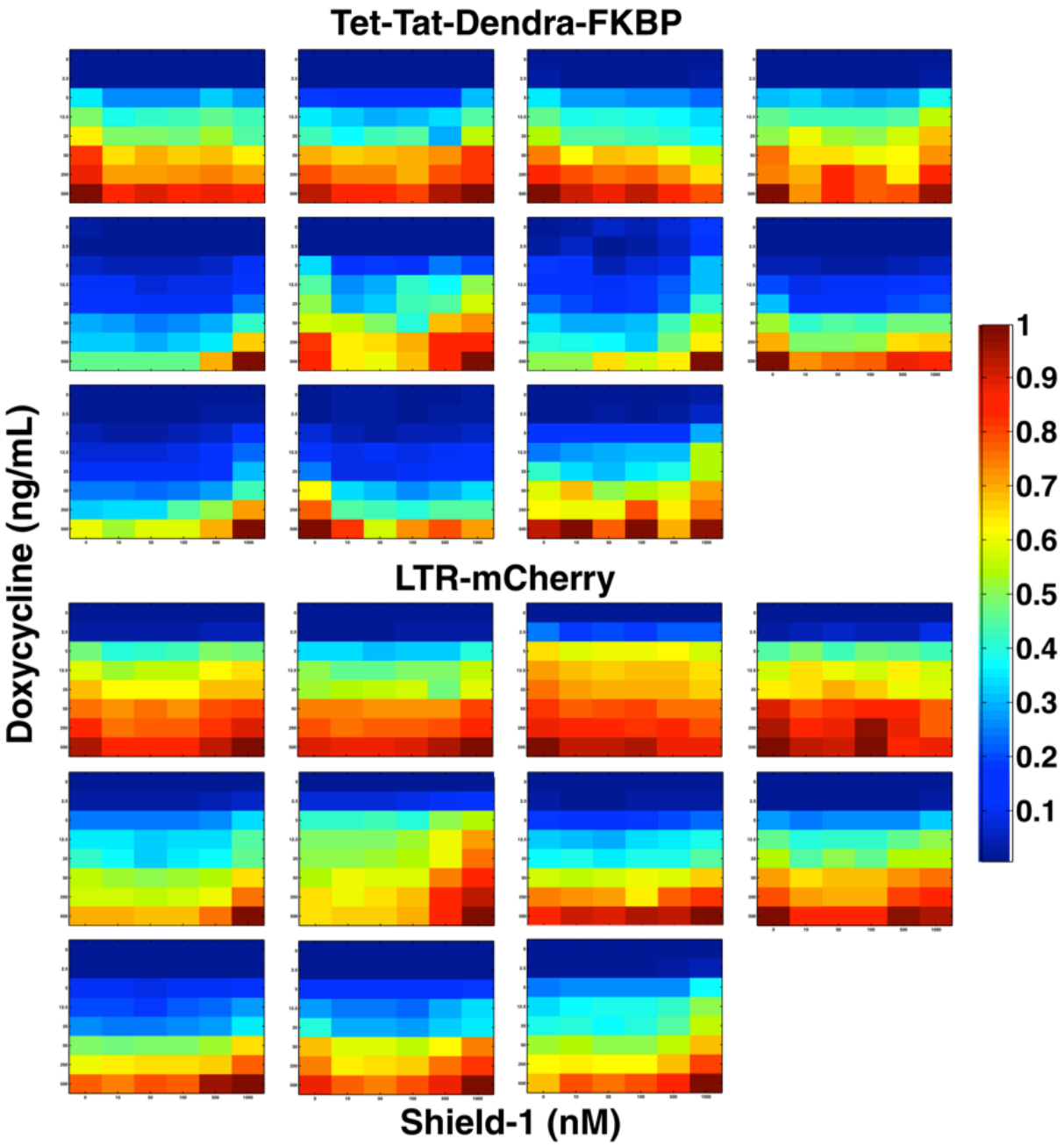

Figure 19. Providing Tat *in trans* controls expression from the LTR.

Each heat map consists of 48 different conditions for 11 isoclonal Tet-Tat-Dendra-FKBP + LTR-mCherry populations. The y-axis represents eight doxycycline concentrations (0 ng/mL, top; 500 ng/mL, bottom), and the x-axis represents six Shield-1 concentrations (0 nM, left; 1000 nM, right). The color in each graph represents the intensity of either Dendra or Cherry fluorescence for that particular heat map and fluorescence levels were normalized against the maximum intensity bin for that specific heat map. Increasing Dox or Shield-1 concentration leads to more Tat-Dendra-FKBP expression (top 11 'Dendra' heat maps). As the availability of Tat-Dendra-FKBP increases, Cherry expression from the HIV-1 LTR also significantly increases (bottom 11 'Cherry' heat maps). In most

cases the maximum Dendra and Cherry fluorescence occurs when the most Dox and Shield-1 are present (bottom right hand corner of heat maps). Note that the Tet-Tat-Dendra-FKBP + LTR-mCherry system expression is more dependent on Dox (vertical) than Shield-1 (horizontal), because the Shield-1 perturbation is only applicable post-transcriptionally, and the Tet-On system relies on varied levels of transcriptional activation for tuning levels of gene-expression.

3. Full-length Virus Results

3a. Establishment of HIV latency in full-length replicating virus is contingent on viral circuit strength

While the above experiments show Tat is sufficient to control gene-expression state in minimal circuits, we wanted to test if viral circuitry can control latency in full-length HIV. A full-length HIV molecular clone, pNL4-3, containing a Tat deletion, (termed ΔTat virus, (Huang et al., 1994)) was reengineered to encode either the Tat-Dendra-FKBP cassette (termed "Tunable-HIV", Fig. 20A, top), or a control Tat-Dendra cassette (termed "Control-HIV", fig. 21) in the *nef* reading frame, which is dispensable for replication in culture (Kim et al., 1989). In these *nef*-reporter viruses, actively replicating infections express reporter, while latent infections do not (Jordan et al., 2003).

Strikingly, modulating Tat positive-feedback strength with Shield-1 alters the percentage of actively infected cells by 141%, i.e. >2 fold (Fig. 20B). The reduction in actively infected cells is not due to reduced input virus since Shield-1-positive and Shield-1-negative cells are infected with equivalent titers of virus (i.e. MOIs). In striking contrast, Control-HIV infection efficiency in the presence and absence of Shield-1 has no measureable affect on active infection (Fig. 20B), demonstrating that Shield-1 is not inducing abortive infections and that any Shield-1 pleiotropic effects cannot explain the difference in active-versus-latent infection. In summary, modulating viral feedback strength is sufficient to control the establishment of active-versus-latent infection in full-length replicating virus.

3b. Full-length latent provirus can be reactivated independent of cellular-activation state

Given that viral regulatory circuitry can control *establishment* of latency in full-length virus, we next tested if viral circuitry was sufficient to control *reactivation* from latency in full-length virus. Unfortunately, as discussed above, the controllable-proteolysis systems create difficulties in interpreting reactivation because they require a minimal-threshold of Tat protein to toggle reactivation; completely silent latent infections, where no Tat is present, would likely not reactivate with Shield-1 thereby generating false negatives. Therefore, to test if Tat could reactivate full-length latent virus we developed a decoupled system where Tat expression is controlled by the cells (via Tet-On) and is induced (by doxycycline) completely independently of the virus. These Tet-Tat-Dendra cells provide *in trans* complementation for a reengineered Tat-deleted virus where mCherry fluorescent protein replaced *nef* (termed ΔTat Cherry virus) so that viral gene-expression can be toggled on from a starting-Tat level of zero (Fig. 20C). Inducing Tat expression in Tet-Tat-Dendra cells during infection with ΔTat Cherry virus shows a ~400% increase in active infection compared to non-induced ΔTat Cherry-infected cells (Fig. 20D), corroborating previous results that constitutive expression of Tat protein can inhibit the establishment of latency (Donahue et al., 2012). Infection of non-induced Tet-Tat-Dendra cells drives ΔTat-Cherry infections to enter latency and, most strikingly, subsequent induction of Tat expression by doxycycline (e.g. after >2 days), fully reactivates latent virus to levels observed in the initial infection with doxycycline (Fig.

20D). Hence, the ability to rescue latent provirus with Tat-induction and without altering cellular-activation state demonstrates that Tat is sufficient to control latent reactivation in HIV.

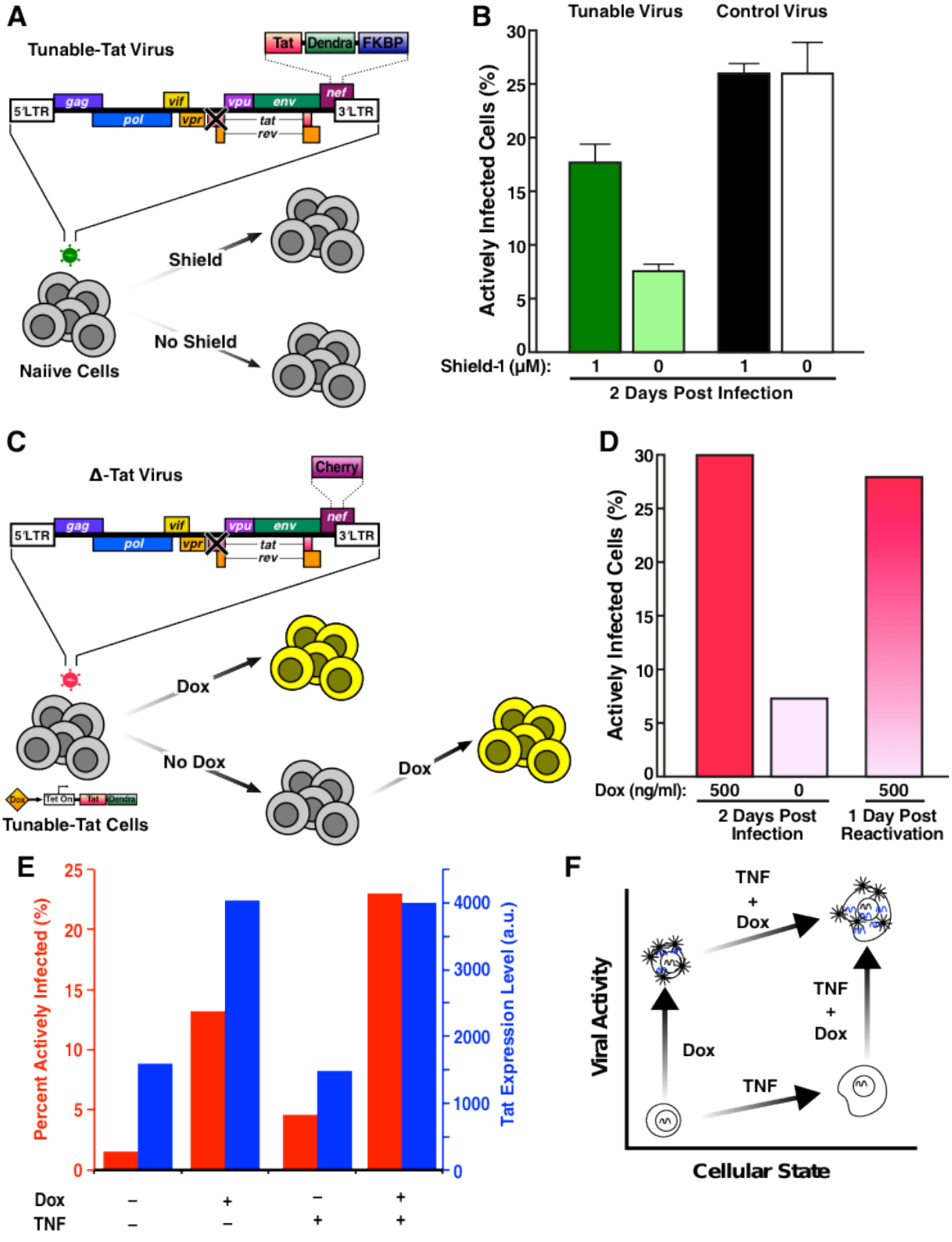
A
Tunable-Tat Virus
Tat
Dendra
FKBP
5'LTR
gag
pol
vif
vpr
vpu
env
nef
tat
rev
3'LTR
Shield
No Shield
Naiive Cells
B
Tunable Virus
Control Virus
Actively Infected Cells (%)
30
25
20
15
10
5
0
Shield-1 (μM):
1
0
1
0
2 Days Post Infection
C
Δ-Tat Virus
Cherry
Dox
No Dox
Dox
Tet On
Tunable-Tat Cells
D
Actively Infected Cells (%)
Dox (ng/ml):
500
0
500
2 Days Post Infection
1 Day Post Reactivation
E
Percent Actively Infected (%)
Tat Expression Level (a.u.)
Dox
TNF
F
TNF
+
Dox
TNF
+
Dox
Dox
TNF
Viral Activity
Cellular State

Figure 20. Viral gene-regulatory circuitry is sufficient to control latency in full-length replicating viruses in the absence of changes in cellular state.

(A) CEM cells were infected with full-length Tunable Tat virus, then split into cultures with (strong feedback arrow, top) or without 1μM Shield-1 (weak feedback arrow, bottom). (B) Two days post infection, the percentage of active infections could be changed from 17.5 ± 1.7% (green column, 1μM Shield-1) to 7.5 ± 1.0% (white column, 0uM Shield-1). The presence of Shield-1 did not change the infection rates for the Control-HIV infection (25.8 ± 1.0%, black column, 1μM Shield-1 and 26.0 ± 2.7%, white column, 0uM Shield-1). Infections were preformed in triplicate. Error bars = ±1 standard deviation. (C) Stable cell lines expressing Tat-Dendra from the doxycycline inducible Tet-ON promoter were infected with Δ-Tat mCherry virus, then split into cultures without doxycycline (weak promoter arrow, bottom) or with 500ng/mL doxycycline (strong promoter arrow, top). Two days post infection, cells from the weak promoter case (bottom) were incubated in 500ng/mL doxycycline. (D) Two days post infection, the presence of doxycycline changed the percentage of active infections from 8% (light red) to 30% (red). Doxycycline was added to the culture without doxycycline (light red), and 28% of cells expressed mCherry (light red to red gradient column) after 24hrs. (E) A plot of Percent Infected (red columns) versus Tat Dendra Fluorescence (blue columns) in Tunable Tat Cells from (c) infected with the Δ-Tat Virus in the presence of TNFα (4.6% actively infected), doxycycline (13.2%), neither (1.6%), or both (23%). (F) Summary schematic of viral phenotype versus cellular state. Starting from resting cell, latent populations (no doxycycline and no TNFα), resting cell, replicating (doxycycline alone), activated cell, latent (doxycycline and TNFα), and activated cell, replicating population (doxycycline and TNFα) phenotypes could be created.

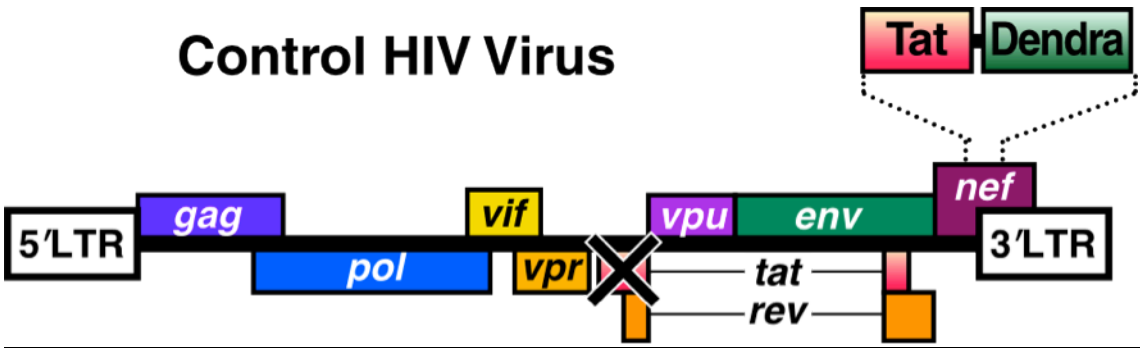

Figure 21. Control HIV was developed as a control virus.

Tat-Dendra was placed in the nef reading frame. Multiple premature stop codons were placed in the native Tat reading frame to prevent expression of Tat from the native locus. This virus is insensitive to Shield-1.

3b. Reactivating full-length latent provirus by tuning viral circuitry is 400% more effective than reactivating by perturbing cellular-activation state

To test the effects of altering cellular-activation state on viral latency, Tet-Tat-Dendra cells were infected with ΔTat virus in the presence of doxycycline or the cell-state activator TNFα (Fig. 20E). Modifying cell state with TNFα, in the absence of Tat induction, only leads to a 1.5-fold change in the percentage of active infections (Fig. 20E). In contrast, Tat induction alone, or in combination with TNFα, drastically increases the proportion of infections that are active by 400% and 700% more, respectively (Fig. 20E). In agreement with data from minimal-synthetic circuits (Fig. 10), perturbing viral circuitry (or both cell-state and viral circuitry) provides far more potent latent reactivation than targeting of cell-state alone.

4. Viral circuitry dominates the control of HIV latency despite changes in cellular-activation state in primary lymphocytes

Given that viral circuitry controls HIV latency in immortalized lymphocyte cells, we next tested if viral circuitry controls HIV latency in primary lymphocytes. Activated primary $CD4^+$ T lymphocytes, the main targets of HIV *in vivo*, undergo a dynamic transition from activated ($CD25^+$, $CD69^+$) to resting memory ($CD25^-$, $CD69^-$). The prevailing hypothesis for latency establishment is that a lymphocyte infected during this 'relaxation' from active to resting globally silences gene expression, including HIV gene expression, and generates a latently infected cell (Siliciano and Greene, 2011). This transitioning of

primary lymphocytes provides an ideal system to test between the cell-state and viral circuitry hypotheses; if viral circuitry is independent of (i.e. autonomous to) cellular state, global silencing of the lymphocyte gene expression during transition from active to resting-memory would not silence HIV gene-expression. To carry out this test, activated primary $CD4^+$ T lymphocytes were transduced with the LTR-Tat-Dendra-FKBP lentiviral construct and allowed to relax from an active to a resting-memory state while Tat feedback circuitry was reduced from wild-type strength to attenuated (Fig. 22A, Fig. 23).

If Tat positive feedback is attenuated (by absence of Shield-1) as lymphocytes relax from activated to memory, significant silencing of HIV gene expression occurs (Fig. 22B, red histograms). However, when Tat positive feedback is of wild-type strength (via Shield-1 addition), only a slight shift in HIV gene expression occurs as lymphocytes transition from active to memory (Fig. 22B, blue histograms). When cell-activation state is quantified by the change in CD25 and CD69 (two markers of T-cell activation state (Fig. 23)), a remarkable relationship emerges. If Tat feedback is attenuated, the cellular-activation state tightly controls entry to latency by significantly reducing the percentage of cells in active infection (Fig. 22C, red). However, when Tat feedback is active, the cellular activation state has no bearing on entrance into latency as the percentage of cells in active infection remains constant (Fig. 22C, blue)—i.e. the intact feedback circuit allows viral gene expression to act completely independent of cellular-activation state. Thus, active Tat feedback appears to buffer HIV from global gene silencing as primary lymphocytes transition from active to resting memory. Overall, these primary-cell studies show that only under attenuated Tat-feedback conditions is cellular-activation

state able to control HIV latency (Fig. 22D) and that wild-type Tat feedback circuitry appears to act autonomous of cellular state.

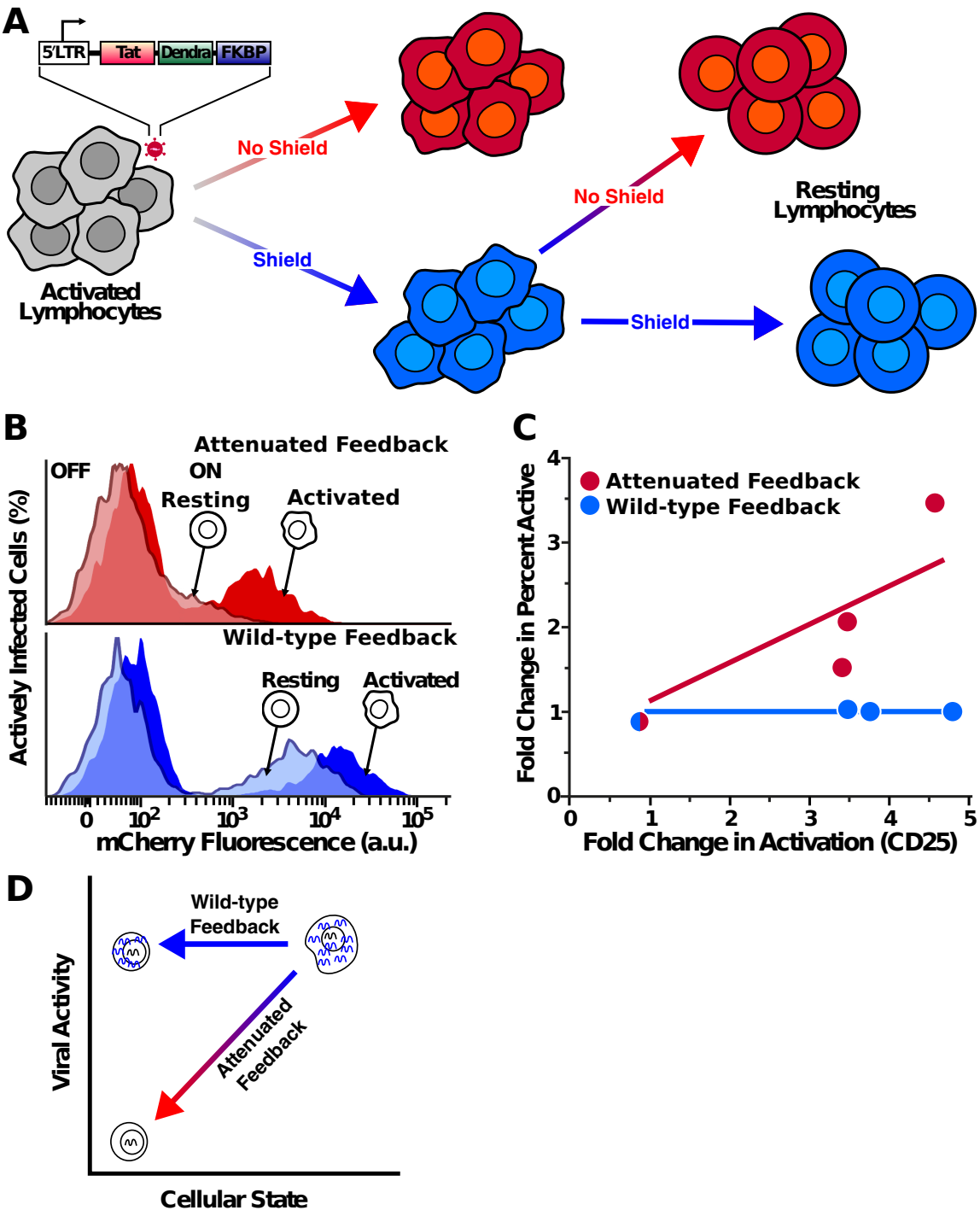
A
5'LTR
Tat
Dendra
FKBP
No Shield
Shield
Activated Lymphocytes
No Shield
Shield
Resting Lymphocytes
B
Attenuated Feedback
OFF
ON
Resting
Activated
Wild-type Feedback
Resting
Activated
Actively Infected Cells (%)
0
10^2
10^3
10^4
10^5
mCherry Fluorescence (a.u.)
C
Attenuated Feedback
Wild-type Feedback
Fold Change in Percent Active
0
1
2
3
4
5
Fold Change in Activation (CD25)
D
Wild-type Feedback
Attenuated Feedback
Viral Activity
Cellular State

Figure 22. HIV gene-regulatory circuitry dominates the regulation of active-vs.-latent expression even as cells relax from active to resting.

(A) Resting lymphocytes were activated then infected with LTR-Tat-Dendra-FKBP in the presence (strong feedback arrow, bottom) or absence (weak feedback arrow, top) of 1μM Shield-1. The 1μM Shield-1 activated lymphocyte culture was washed of activating agents and Shield-1 and split into a culture with and without 1μM Shield-1. (B) Flow cytometry of HIV LTR expression (Dendra fluorescence) in activated (opaque histograms) or resting (transparent histograms) cells in the presence of 1 μM Shield-1 (top panels) and absence of Shield-1 (bottom panels). (C) Plot of CD25 fold-change versus the percentage of actively expressing cells in the presence (blue) or absence (red) of 1μM Shield-1. (D) Summary schematic of circuit activity versus cellular state based on (A)-(C). Cells transitioning from an activated state can only silence HIV expression if positive feedback is attenuated. Otherwise, normal Tat feedback would cause cells to transition from 'ON' and activated to 'ON' and resting.

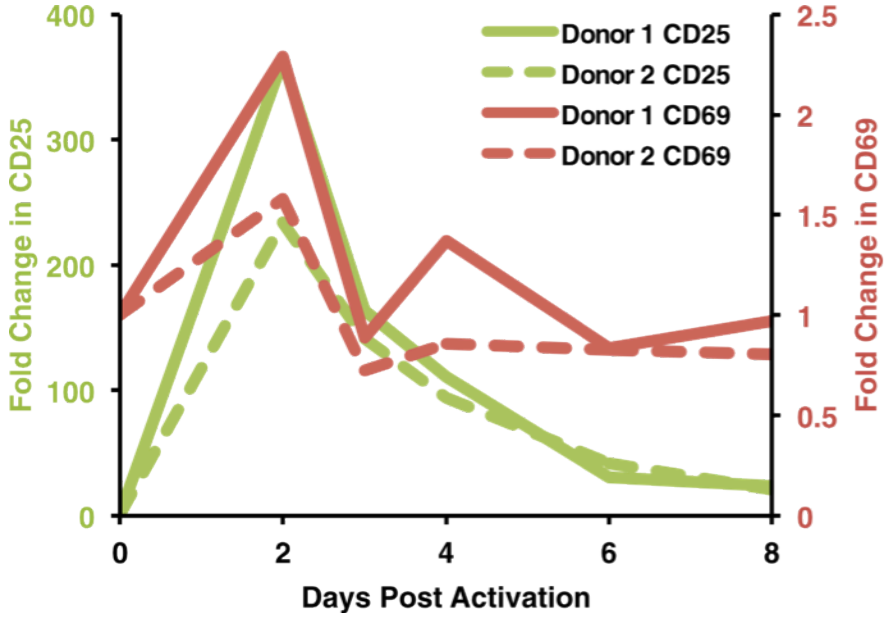

Figure 23. Activation and relaxation dynamics of primary cells.

Relaxation kinetics from two donors (Donor 1, solid lines and Donor 2, dashed lines) and the fold change in the levels of CD25 (green) and CD69 (red) are shown. Green lines correspond to CD25 y-axis on the left, and the red lines correspond to the CD69 y-axis on the right. Time points were taken at 0, 2, 3, 4, 6, and 8 days. At day 2, CD3/CD28 beads, IL-2, and virus were removed from the culture to allow the activated cells to relax back to the resting state. These data are related to Figure 21 of the main text.

5. Discussion

Here, synthetic engineering of minimal circuits and full-length HIV viruses allowed for decoupling of cell-state and viral circuit contributions to HIV decision-making. The minimal circuits establish the sufficiency of Tat to control HIV gene-expression state (Fig. 10). Engineering of full-length HIV showed that viral positive feedback strength alters the ratio of active to latent infections in the absence of cell-state modification (Fig. 20). Importantly, in primary cells, even in the face of drastic cell-state modifications, HIV circuitry dominated the decision between active expression and latency (Fig. 22).

Ideally a full-length virus would have been used in the primary-cell experiment (Fig. 22). However, actively replicating full-length viruses result in cell death after 1-2 days (Weinberger et al., 2008a), not allowing for long-term measurements of cell-state relaxation's effects on HIV transcription. Nevertheless, minimal circuit models are able to recapitulate all of the observed properties of latency (Jordan et al., 2001; Weinberger and Shenk, 2007) and silencing of HIV transcription is only dependent on cellular factors (Siliciano and Greene, 2011; Tyagi et al., 2010). Therefore, the result that wild-type feedback can sustain robust HIV transcription despite a changing cellular environment should translate to a full-length virus.

Why has the finding that viral circuitry can control latency remain unobserved? Mainly, HIV gene-regulatory circuitry doubles as an environmental sensor and the invariable coupling of cell-state changes to viral circuitry obfuscates the contributions of each (Ott

et al., 2011). Since cell-state was tuned in most studies (Siliciano and Greene, 2011; Tyagi et al., 2010; Williams et al., 2007b), reactivation of HIV from latency could only be contributed to the cell-state change. However, if viral circuit and cell-state changes are decoupled, a role for viral circuitry in HIV latency emerges (Jeeninga et al., 2008). Here, decoupling of viral circuitry and cell-state established the sufficiency of viral circuitry to regulate latency (Fig. 10, 20, 22).

The sufficiency of viral circuitry to control latency suggests it is an evolved phenotype with a role in the natural history of the virus. This is difficult to resolve with theories postulating latency is an evolutionary accident only advantageous in response to ART (Coffin and Swanstrom, 2013; Eisele and Siliciano, 2012). While latency may be an intrinsic feature of the virus, the natural fitness role remains unresolved.

In general, the tuning of feedback circuitry independent of cell-state changes provides a framework for delineating how intrinsic gene-regulatory circuitry and extrinsic environmental factors affect cell-fate determination. A diverse array of gene-regulatory circuits double as environmental sensors, so different circuits likely have diverse susceptibilities to environmental cues (Balaban et al., 2004; Balazsi et al., 2011; Lahav et al., 2004; Zeng et al., 2010). Thus, the ability to explore the effects of environmental cues independently of internal circuitry for diverse systems may isolate the dominant factors controlling cellular phenotype.

6. Materials and Methods

6a. Virus production, cell lines, and infections.

All lentivirus were generated in 293T cells and isolated as described(Weinberger et al., 2005). To generate the isoclonal populations, the LTDF lentivirus was added to Jurkat T Lymphocytes at a low MOI to ensure a single integrated copy of proviral DNA in infected cells. Cells were stimulated with tumor necrosis factor alpha (TNF-α) and Shield-1 for 18 hours before sorting for Dendra expressing cells. Isoclonal and polyclonal populations were created as described(Weinberger et al., 2005). Sorting and analysis of cells infected was performed on a FACSAria II. The same procedure was followed to create the LTD, LTR-mCherry-IRES-Tat-FKBP cell lines. The Tet-Tat-Dendra + LTR-mCherry populations were created by first infecting Jurkats with Tet-Tat-Dendra and SFFV-rTta lentivirus at a high MOI. Then LTR-mCherry lentivirus was added to the Jurkats at a low MOI. Before sorting for mCherry+ and Dendra+ cells, doxycycline was added at 500ng/mL for 24hrs, at which point isoclonal populations were isolated. The same procedure was followed for the Tet-Tat-Dendra-FKBP + LTR-mCherry populations, however, 24hrs before the sort 1uM Shield-1 and 500ng/mL doxycycline was added to the culture. All Shield-HIV or noShield-HIV infection experiments were performed by incubating 5 x 10^5 CEM cells in the same titer of Shield-HIV or the same titer of noShield-HIV in the presence or absence of Shield-1 and taking a flow cytometry time point after 48 hours.

6b. Primary CD4$^+$ T cell isolation and culture conditions for primary cells and cell lines.

Jurkat T Lymphocytes, CEMs, and Primary CD4$^+$ T lymphocytes were cultured in RPMI 1640 (supplemented with L-glutamine, 10% fetal bovine serum, and 1% penicillin-streptomycin) in a humidified environment at 37°C and 5% CO_2. Jurkats and CEM were maintained by passage between 2×10^5 and 2×10^6 cells/mL. Primary CD4$^+$ T cells were isolated from peripheral blood obtained from Stanford Blood Bank using RosetteSep^{TM} Human CD4$^+$ T Cell Enrichment Cocktail from STEMCELL^{TM} Technologies and Ficoll as described(Terry et al., 2009). Once isolated, cells were either cultured as described in Supplemental Figure 12 or frozen in 10% DMSO, 90% culture media at a density of 10^7 per mL. Cell activation was measured by flow cytometry with anti-CD25-PE-conjugated antibody and anti-CD69-APC-conjugated antibody from BD Biosciences^{TM}.

6c. Flow Cytometry and analysis.

Flow cytometry data was collected on a BD FACSCalibur DxP8, BD LSR II, or HTFC Intellicyt for stably transduced lines and BD FACSAria II for replication competent virus assays and sorting. All flow cytometry experiments on replication competent virus runs were performed in BSL3 conditions, and safety information will be provided upon request. Approximately 10,000 live cells were acquired for measurement and compared to an uninfected control to establish background fluorescence levels. Flow cytometry data was analyzed in FlowJo^{TM} (Treestar, Ashland, Oregon) and using customized MATLAB® code.

6c. Microscopy and analysis.

Microscopy experiments were performed on a Zeiss Axio Vert inverted fluorescence microscope equipped with a Yokagawa spinning disc, 405-, 488-, and 561-nm laser excitation light sources, CoolSNAP HQ^2 14-bit camera from Photometrics, computer controlled motorized stage, and environmental enclosure, maintaining a temperature of 37°C and a humidified atmosphere with 5% CO_2. Photoconversion was performed by exposing cells to: (*i*) 488 nm (10% laser power, 500 ms exposure time), (*ii*) 561 nm (50% laser power, 500 ms exposure time), (*iii*) 405 nm (100% laser power, 60 s exposure time), (*iv*) 488 nm (10% laser power, 500 ms exposure time), and last of all (*v*) 561 nm light (50% laser power, 500 ms exposure time). Time-lapse experiments immediately followed the photoconversion, and images were captured every 10 min, with a 40x oil, 1.3NA objective, 1 s exposure time, and 50% power on a 561-nm solid-state laser. At least 50 cells were collected for each experiment and analyzed as described(Weinberger et al., 2008a). Exponential fits for the Tat half-lives were performed on the first 6 hours of microscopy data collected.

6d. Cloning information

The HIV-1 LTR promoter driving a DNA fusion of Tat-Dendra (LTD) was generated using fusion PCR with Tat from the LGIT plasmid(Weinberger et al., 2005) and pDendra2-N from Evrogen. The primers were: (1) 5' Tat-linker-Dendra primer: TCC CGG GGT GTT ACT TCC TCC ACT TCC TCC CTT GTC ATC GTG GTC CTT GTA, (2) 3' Tat primer: GGG CCC GGA TCC ATG GAG CCA GTA GAT CCT AGA CTA, (3) 5' Dendra-linker-Tat: GAC GAT GAC AAG GGA GGA AGT GGA GGA AGT AAC

ACC CCG GGA ATT AAC CTG, (4) 3' Dendra primer: GGG CCC CTC GAG TTA CCA CAC CTG GCT GGG CAG GGG GCT. Primers 1 and 3 were added to the PCR reaction at a 50-fold lower concentration than 2 and 4. The GFP-IRES-Tat from LGIT was restriction digested with BamHI and XhoI, and the Tat-Dendra (TD) PCR product was ligated into the backbone of LGIT to generate LTD. The LTD plasmid was used to create the LTDF plasmid. To generate the LTDF plasmid, the FKBP #24 domain(Chu et al., 2008) from pBMN-HA-YFP-FKBP(E31G-R71G-K105E)-IRES-HcRed-Tandem was amplified using the following primers: (5) FKBP forward: GGA GTG CAG GTG GAA ACC ATC, and (6) FKBP reverse: TCA TTC CAG TTC TAG AAG CTC. The LTD backbone was then amplified using the following primers: (7) Tat-Dendra forward: TTC GAT GTG GAG CTT CTA GAA CTG GAA TGA CTC GAG ACC TGG AAA AAC ATG, and (8) Tat-Dendra reverse: TCC TGG GGA GAT GGT TTC CAC CTG CAC TCC CCA CAC CTG GCT GGG CAG GGG. The two PCR products were then incubated together in an isothermal assembly cocktail as described(Gibson et al., 2009) to generate the LTDF lentiviral vector. pNL4-3 TDF (Shield-HIV) was generated by amplifying TDF from LTDF with the following primers: (9) 5' Tat-Dendra-FKBP: GGA CCG CGG ATG GAG CCA GTA GAT CCT AGA and (10) 3' Tat-Dendra-FKBP: GCG TCT AGA TCA TTC CAG TTC GAG AAG CTC CAC ATC GAA GAC GAG AGT GGC ATG TGG. Primer 10 encoded a silent mutation where the third base pair of L105 in FKBP #24 was mutated from A to C to remove an XbaI site. The PCR product was cut with SacII and XbaI. The backbone of a double digest with SacII and XbaI was isolated from the pNL 4-3 ΔTat virus as described(Huang et al., 1994). The backbone and PCR product were then ligated to form pNL4-3 TDF. pNL4-3 TD (noShield-HIV in text) was

generated in the same way as Shield-HIV but by amplifying TD from LTD with the following primers: (1) 5' Tat-Dendra: GGA CCG CGG ATG GAG CCA GTA GAT CCT AGA and (2) 3' Tat-Dendra: GCG TCT AGA TTA CCA CAC CTG GCT GGG CAG GGG GC. Synthesizing d2GFP-IRES-Tat-FKBP, then swapping that into in the LTR-d2GFP-IRES-Tat construct using BamHI and XhoI restriction sites developed the LTR-d2GFP-IRES-Tat-FKBP plasmid. The sequence of Tat from recombinant clone pNL4-3, GenBank: AAA44985.1, M19921, was used. To clone the LTR-mCherry-IRES-Tat-FKBP construct d2GFP was swapped with mCherry using BamHI and EcoRI restriction sites. To clone the Tet-Tat-Dendra and Tet-Tat-Dendra-FKBP plasmids, Tat-Dendra or Tet-Tat-Dendra was swapped with YFP-Pif from the pHR-TREp-YFP-Pif plasmid (a gift from Wendell Lim's Laboratory at UCSF) using BamHI and NotI restriction sites.

Chapter 5: Transcriptional bursting from the HIV LTR is a significant source of noise in gene-expression

Chapter 5 was published in Singh, A., Razooky, B., Cox, C.D., Simpson, M.L., and Weinberger, L.S. (2010b). Transcriptional Bursting from the HIV-1 Promoter Is a Significant Source of Stochastic Noise in HIV-1 Gene Expression. Biophys J *98*, L32-L34. The article can be found on Pubmed: http://www.ncbi.nlm.nih.gov/pubmed/20409455

1. Introduction

Clonal (or isogenic) cell populations can exhibit considerable cell-to-cell variation in protein levels due to the inherent stochastic nature of biochemical processes involved in gene expression (Kaern et al., 2005; Raj and van Oudenaarden, 2008). This variation, or expression noise, can have significant effects on biological function and can 'flip' genetic switches to drive probabilistic fate decisions in bacteria (Maamar et al., 2007), viruses (Arkin et al., 1998) and stem cells (Losick and Desplan, 2008). Our recent work has shown that stochastic expression of human immunodeficiency virus type 1 (HIV-1) proteins immediately after infection can critically influence the HIV-1 fate decision between active replication and post-integration latency in single cells (Weinberger et al., 2005; Weinberger et al., 2008a; Weinberger and Shenk, 2007). However, the source of this noise has remained elusive. To probe the potential sources of noise in HIV-1 gene expression, we systematically quantify stochastic variation in HIV-1 promoter activity across different HIV-1 integration sites in the human genome.

2. Results

2a. Measuring the noise in isoclonal populations.

HIV-1 encodes a single promoter that drives expression of all its viral gene products. To study viral gene-expression noise, we exploit HIV-1s known ability to integrate semi-randomly into sites across the human genome (Schroder et al., 2002). Differences in local chromatin microenvironment at each integration site generate vast differences in mean expression level of HIV-1 (Jordan et al., 2001) and this difference provides a natural method to study noise as a function of mean expression levels. Using a minimal reporter virus encoding the HIV-1 5' long terminal repeat (LTR) promoter driving a short-lived GFP (the vector is referred to as LTR-GFP), we isolated 30 different clonal populations each carrying a single integrated copy of LTR-GFP in each cell. Our previous integration-sites analyses show that LTR-GFP clones integrate in positions similar to full-length HIV-1 (Weinberger et al., 2005). Clonal populations are analyzed at the single-cell level by flow cytometry and to minimize cell-to-cell differences in reporter levels due to heterogeneity in cell size, cell shape, and cell-cycle state (i.e., extrinsic noise), we adopt a previously used approach (Newman et al., 2006) of gating the smallest possible forward- and side-scatter region that contains at least 30,000 cells. As expected from previous findings (Jordan et al., 2001), the resulting gated data displays a 40-fold difference in mean GFP levels between the dimmest clone and the brightest clone (Fig. 25A). Importantly, integration-site also appears to shape the stochastic variability in gene-expression: two clones with the same mean GFP intensity can display vastly

different variability profiles (Fig. 25A, inset), suggesting that extrinsic noise factors cannot explain the difference.

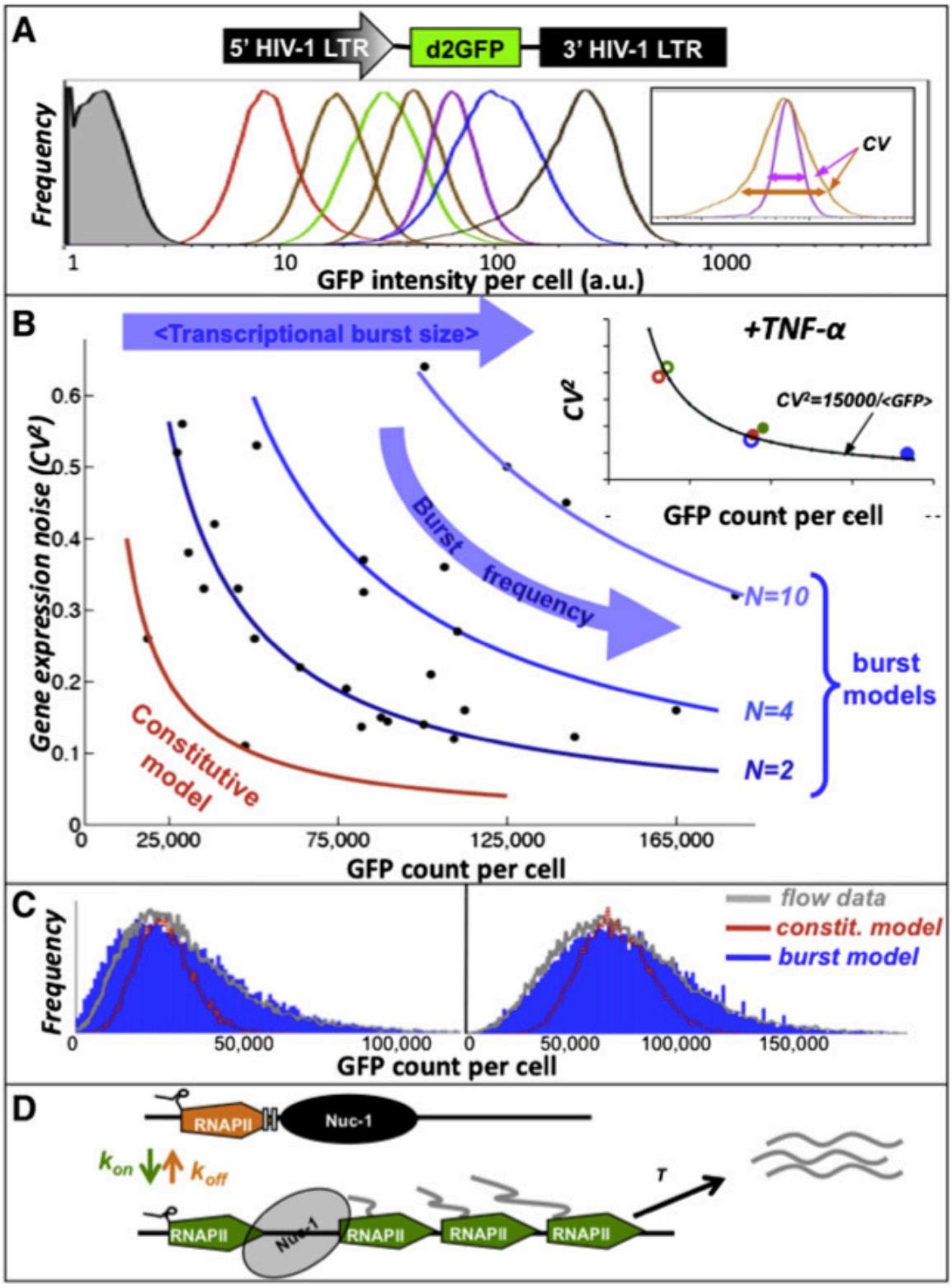
A
5' HIV-1 LTR
d2GFP
3' HIV-1 LTR
Frequency
CV
1
10
100
1000
GFP intensity per cell (a.u.)
B
<Transcriptional burst size>
+TNF-α
CV²
CV²=15000/<GFP>
GFP count per cell
Gene expression noise (CV²)
Burst frequency
N=10
N=4
N=2
burst models
Constitutive model
0
0.1
0.2
0.3
0.4
0.5
0.6
0
25,000
75,000
125,000
165,000
GFP count per cell
C
Frequency
flow data
constit. model
burst model
0
50,000
100,000
0
50,000
100,000
150,000
GFP count per cell
D
RNAPII
Nuc-1
k_on
k_off
RNAPII
Nuc-1
RNAPII
RNAPII
RNAPII
τ

Figure 24. Transcriptional bursting from the HIV-1 LTR.

(A) Schematic of the LTR-GFP lentivirus and flow cytometry histograms of seven representative Jurkat LTR-GFP clones (shaded histogram is uninfected control). Large differences in mean LTR expression are evident across clones and large differences in expression variability are present within each clonal population. Inset: two clones with same mean but different coefficient of variation (CV). (B) Plot of mean GFP abundance versus GFP noise level (measured by CV2) for 30 different clonal populations. Solid lines are predictions of noise scaling from a best-case, maximally conservative constitutive promoter model (red line) or a two-state bursty promoter model (blue lines) where average transcriptional burst size, N, is kept fixed and burst frequency is allowed to vary (N = 2 corresponds to Eq. 8; N = 4 corresponds to Eq. 9; N = 10 corresponds to 65,000/). Inset: Three representative clones (red, green, and blue) before induction with TNF-α (open circles) and after induction with TNF-α (solid circles). (C) Flow cytometry histograms of two representative clones (gray) along with predicted GFP histograms from a constitutive gene expression model (red line) and a two-state transcriptional burst model (blue). (D) Proposed schematic for the two-state transcriptional burst model: LTR promoter fluctuates between an inactive and active elongation state. Figure originally published as Figure 1 in Singh et al, Biophysical Journal, 2010. Abhyudai Singh performed the experiments and fit the data. Brandon Razooky performed the simulations and the schematic.

2b. Fitting noise to a model of transcriptional bursting.

To systematically quantify variability in GFP levels, fluorescence intensities were converted into GFP molecular abundance using EGFP Calibration Beads (BD Biosciences, Clontech, San Jose, CA). First subtracting the mean background Jurkat auto-fluorescence, and then multiplying the value by the measured MESF scaling-factor of 3000 molecules per unit fluorescence intensity calculated GFP abundance of a given clone. As in many studies (Bar-Even et al., 2006a; Newman et al., 2006), gene-expression noise is quantified using the coefficient of variation (*CV*) squared, defined as $CV^2=\sigma^2/\langle GFP\rangle^2$, where σ^2 is the variance in GFP abundance and $\langle GFP\rangle$ is the average number of GFP molecules/cell. For most clones, expression noise appears to decrease with increasing abundance (Fig. 25B) and fit the equation:

$$CV^2 = \frac{15{,}000}{\langle GFP \rangle}. \qquad (8)$$

Several clones exhibit *CV* values much larger than predicted by Eq. 8, but appear to match a second trend-line of

$$CV^2 = \frac{30{,}000}{\langle GFP \rangle}. \qquad (9)$$

The remaining clones can be fit by version of Eqs. 8 and 9 using different values in the numerator. Importantly, this inverse relationship between noise and mean protein levels cannot be explained by fluctuations in global or pathway-specific factors, as in that case *CV* would have shown no dependence on mean protein levels (Bar-Even et al., 2006a).

To explain this inverse relationship, we first explore constitutive models of gene expression that incorporate stochastic birth and death of individual mRNAs. These constitutive models assume that mRNAs are created one at a time in exponentially distributed time intervals and predict that

$$CV^2 = \frac{C}{\langle GFP \rangle}, \quad C = \frac{L}{dm + dp} \qquad (10)$$

where *L*, *dm,* and *dp* are mRNA translation rate, mRNA half-life and protein half-life, respectively. Qualitatively, the constitutive model gives a scaling of CV^2 vs. <GFP> similar to Fig. 25B. However, studies in eukaryotes report the proportionality factor C to be ~1,300 molecules (Bar-Even et al., 2006a), which is an order of magnitude smaller than the experimental proportionality factors of 15,000 and 30,000 in Eq. 8 and 9. To obtain an upper bound of what *C=L/(dm+dp)* could be for the GFP variant used in our study, we used a maximally conservative approach where *C* is set to the minimum value of CV^2 x <GFP> across all clonal populations. This maximally conservative estimate results in a proportionality factor *C* is approximately = *5000* molecules, which is still

many fold smaller than the proportionality factors experimentally observed for the HIV-1 LTR in Fig. 25B. Thus, even with the most conservative parameter estimates, a model where mRNAs are produced constitutively from the viral promoter cannot account for the high HIV-1 gene expression noise levels (Fig. 25B-C). This result contrasts with findings from yeast, where variation in the levels of many proteins results from thermal fluctuations in their corresponding mRNA counts (Bar-Even et al., 2006a; Newman et al., 2006; Zenklusen et al., 2008).

To explain the scaling of noise in Eqs. 8–9, we next consider the dynamics of the local chromatin environment at the HIV-1 LTR promoter. Transcriptional initiation from the LTR is efficient but the elongating RNA polymerase II (RNAPII) molecule is known to stall 50–70 nucleotides after initiating (Jordan et al., 2003). Stalling occurs just upstream of a nucleosome (termed nuc-1) until nuc-1 is remodeled by host factors such as SWI/SNF (Mizutani et al., 2009). Such blocks in transcriptional elongation have been reported across genomic loci (Guenther et al., 2007) and can create rate-limiting steps in mRNA production that lead to transcriptional bursting (Raser and O'Shea, 2004). Thus, we consider a model where the LTR promoter fluctuates between an inactive state (i.e., RNAPII stalled at nuc-1) and active state (nuc-1 remodeled and RNAPII unstalled) with rates kon, koff and transcriptional elongation only occurs from the active state at a rate T. In such two-state models (Kepler and Elston, 2001b), mRNAs are created in bursts during promoter transitions from inactive to active state, with kon and T/koff denoting the frequency and the average size of the transcriptional bursts, respectively. Solving the Chemical Master Equation corresponding to this two-state model yields:

$$CV^2 = \frac{C\left(1 + \frac{T}{koff}\right)}{\langle GFP \rangle}, \quad C = \frac{L}{dm + dp} \qquad (11)$$

if promoter transitions to the active state are infrequent. Equation 11 illustrates that changing the burst frequency kon for a fixed transcriptional burst size will result in a similar inverse scaling between noise and protein level as experimentally observed for the HIV-1 LTR in Fig. 25B. Moreover, by choosing an appropriate burst size one can match the high proportionality factors observed in Eq. 8 and Eq. 9. Using the maximally conservative estimate (*C* of ~5000), Eq. 11 predicts that the clones satisfying Eq. 8 have average transcriptional bursts of only two mRNA's, while clones satisfying Eq. 9, have bursts of four mRNA's. The clones that exhibit very high noise levels in Fig. 25B can have transcriptional burst sizes up to 10 mRNA transcripts. Thus, the two-state promoter model can explain the observed scaling of noise with protein levels if burst frequency and burst size vary across different integration sites.

The two-state model also provides insight into the mechanisms of HIV-1 LTR regulation by signaling factors such as Tumor Necrosis Factor-alpha (TNF-α). Experimentally, TNF-α induction raises mean GFP levels without changing the product CV^2 x <GFP> so that each clone appears to slide along the CV versus mean trend-line (Fig. 25B, inset). As Eq. 11 shows, an increase in the burst frequency (kon) will raise expression level but reduce noise such that CV^2 x <GFP> remains unchanged. Thus, Eq. 11 suggests that TNF-α enhances HIV-1 gene expression by primarily influencing the frequency of transcriptional bursts and not the size of the bursts. The two-state model may also explain recent reports on increases in LTR noise when SWI/SNF chromatin

remodeling complexes are removed (Mizutani et al., 2009) since removal of SWI/SNF reduces nuc-1 remodeling and would lower kon, thereby causing clones to slide up along the CV versus mean trend-line to higher noise levels.

3. Discussion

In summary, a two-state promoter model where the LTR infrequently transitions to an elongation-active state can explain the high stochastic variability in HIV-1 gene expression. These transitions cause mRNA's to be made in transcriptional bursts, with average burst sizes ranging from 2 to 10 mRNAs across integration sites. Our results indicate that the local chromatin environment of the HIV-1 promoter controls the extent of gene-expression noise by modulating the dynamics of transcriptional bursts, and integration sites with a low frequency of transcriptional bursts and/or high burst size will exhibit the broadest distributions in protein levels. Thus, viral integration site may play a critical role in biasing the viral fate-decision between active replication and proviral latency by influencing the stochasticity in the production of early viral proteins.

Chapter 6: Transcriptional bursting is the dominant source of noise in HIV gene-expression

This chapter was published in Singh, A., Razooky, B.S., Dar, R.D., and Weinberger, L.S. (2012). Dynamics of protein noise can distinguish between alternate sources of gene-expression variability. Mol Syst Biol *8*, 607. The article can be found on Pubmed; http://www.ncbi.nlm.nih.gov/pubmed/22929617

1. Introduction

Clonal cell populations exhibit considerable cell-to-cell variation in the levels of any specific protein (Blake et al., 2003; Elowitz et al., 2002; Kaern et al., 2005; Raj and van Oudenaarden, 2008; Raser and O'Shea, 2005; Sigal et al., 2006). This variation or expression noise is essential for diverse cellular processes, such as the regulation of probabilistic cell-fate decisions and the generation of phenotypic heterogeneity across isogenic cell lines (Chang et al., 2008; Losick and Desplan, 2008; Maamar et al., 2007; Munsky et al., 2012; Razooky and Weinberger, 2011; Singh and Weinberger, 2009; Suel et al., 2006; Sureka et al., 2008). Tight control of expression noise is also vital for optimal functioning of housekeeping proteins, and diverse diseased states have been attributed to an increase in expression noise in particular genes (Bahar et al., 2006; Cook et al., 1998; Fraser et al., 2004; Kemkemer et al., 2002). Collectively, these results suggest that gene-expression noise profoundly affects biological function and underscore the importance of developing methods that pinpoint the source of noise in a given gene circuit.

Gene-expression noise is often decomposed into intrinsic and extrinsic noise (Elowitz et al., 2002). While extrinsic noise arises from intercellular differences in the amounts of cellular components (e.g., RNA polymerases and ribosomes), intrinsic noise results from the random timing and discrete nature of biochemical reactions associated with promoter remodeling, transcription, translation, and degradation of mRNA and protein species. Although several techniques are available to separate extrinsic and intrinsic noise (for example, see ((Elowitz et al., 2002; Newman et al., 2006)), methods to discriminate between the different sources of intrinsic noise are far less developed. Experimental evidence suggests that intrinsic noise originates from two sources: (i) Poisson mRNA fluctuations arising from probabilistic synthesis and decay of individual mRNA transcripts (mRNA birth/death fluctuations) and (ii) promoter switching between different transcriptional states (promoter fluctuations). mRNA birth/death fluctuations constitute a major source of stochasticity in gene expression since many mRNA species are present at very low molecular counts within cells (Bar-Even et al., 2006a; Newman et al., 2006; Taniguchi et al., 2010). These mRNA fluctuations are transmitted downstream through translational bursting to generate intercellular variability in protein levels. Alternatively, protein variability results from promoter switching between different transcriptional states. An important consequence of promoter switching is transcriptional bursting, where multiple mRNAs are created per promoter-firing event (Golding et al., 2005; Muramoto et al., 2012; Raj et al., 2006; Raser and O'Shea, 2004; Suter et al., 2011a; Yunger et al., 2010). In this study, we sought to develop a method to determine the relative contributions of mRNA birth/death fluctuations and promoter fluctuations to

intrinsic gene-expression noise, assuming that protein levels are the only observable state of the system.

At a qualitative level, both mRNA birth/death and promoter fluctuations generate similar predictions for steady-state protein noise levels (Bar-Even et al., 2006a; Ingram et al., 2008). Thus, steady-state distributions of protein abundances across cells are insufficient to identify the source of intrinsic expression noise. One increasingly popular method to quantify the extent of stochastic promoter switching relies on counting individual mRNA transcripts in single-cells by single-molecule fluorescence *in situ* hybridization (smFISH) *(Raj et al., 2006; Tan and van Oudenaarden, 2010; Zenklusen et al., 2008)*. mRNA smFISH is a powerful and elegant method, but requires that individual mRNA molecules be long enough to accommodate the binding of at least 20 individual probes. This is needed to insure the spot is sufficiently bright to be distinguished in images and counted, and that the specific mRNA species being analyzed are in a concentration low enough that individual diffraction-limited spots do not spatially overlap.

Here, we report a method for discriminating mRNA birth/death and promoter fluctuations that can be used with highly expressed mRNAs and that is easily implemented across different cell types. More specifically, we show that the dynamical changes in protein noise levels, in response to perturbations, are sufficient to determine the source of intrinsic expression noise. These perturbations can be generated relatively easily with readily available small-molecule pharmaceutical agents that rapidly and

efficiently block transcription and translation. We illustrate the experimental utility of this method by perturbing gene-expression from the human immunodeficiency virus type 1 (HIV-1) long terminal repeat (LTR) promoter. Transient changes in reporter protein noise levels, in response to small-molecule drugs, show that LTR gene-expression noise results primarily (>90%) from promoter fluctuations.

2. Results

2a. Stochastic gene-expression model

To analyze the different sources of intrinsic noise in protein abundance, we considered a stochastic model of gene-expression that incorporates both low-copy mRNA fluctuations and transcriptional bursting (Fig. 26). In this model, mRNA transcription and degradation are stochastic events that occur at exponentially distributed time intervals. Each transcriptional event creates a burst of B mRNA molecules, where B is a discrete random variable with probability $\{B = i\} = \alpha_i$, $i \in \{1,2,3\ldots\}$ and mean $\langle B \rangle$, where the symbol $\langle\ \rangle$ represents the expected value. Note that $B=1$ with a probability of one corresponds to Poisson fluctuations in mRNA counts while a large average burst size $\langle B \rangle$ implies transcriptional bursting (i.e., promoter switching between active and inactive promoter states). If $m(t)$ denotes the mRNA population count at time t, then, the probability $P_j(t)$ of having j mRNA molecules at time t evolves according to the following chemical master equation:

$$\begin{aligned} \frac{dP_0(t)}{dt} &= \gamma_m P_1(t) - k_m P_0(t) \\ \frac{dP_j(t)}{dt} &= \sum_{i=1}^{j} k_m P_{j-i}(t)\alpha_i + \gamma_m (j+1) P_{j+1}(t) - (\gamma_m j + k_m) P_j(t), \qquad j \in \{1,2,\cdots\} \end{aligned} \tag{12}$$

where k_m is the frequency of transcription events and γ_m represents the mRNA degradation rate (Mcquarri, 1967). Fluctuations in mRNA counts are transmitted downstream to the protein level, which is assumed to be the only observable state of the system. As protein population counts are typically large, we can ignore Poisson noise arising from stochastic birth and death of individual protein molecules. Accordingly, protein dynamics are modeled deterministically as

$$\frac{dp(t)}{dt} = k_p m(t) - \gamma_p p(t) \qquad (13)$$

where k_p is the mRNA translation rate, γ_p is the protein degradation rate and $p(t)$ denotes the protein count at time t. Together, (12) and (13) constitute a stochastic hybrid gene-expression model in which mRNA time-evolution is discrete and stochastic, while protein levels evolve continuously and deterministically.

As in many studies, protein expression noise is quantified using the coefficient of variation squared, defined as $CV^2 = \sigma^2/\langle p\rangle^2$, where σ^2 is the variance in protein level and $\langle p\rangle$ denotes the average protein abundance (Paulsson, 2004, 2005; Pedraza and Paulsson, 2008). For the gene-expression model described above, the steady-state protein noise level is given by

$$\overline{CV^2} = \frac{\eta_m k_p}{(\gamma_m + \gamma_p)\overline{\langle p\rangle}} = \frac{k_p}{(\gamma_m + \gamma_p)\overline{\langle p\rangle}} + \frac{(\eta_m - 1)k_p}{(\gamma_m + \gamma_p)\overline{\langle p\rangle}} \qquad (14)$$

where

$$\overline{\langle p\rangle} = \frac{\langle B\rangle k_p k_m}{\gamma_p \gamma_m} \qquad (15)$$

denotes the steady-state mean protein abundance and

$$\eta_m = \frac{\langle B^2 \rangle + \langle B \rangle}{2\langle B \rangle} \qquad (16)$$

represents the steady-state Fano factor (also known as "noise strength") of the mRNA population count.

To understand our steady-state moment analysis of the 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*} = α_i for $i \in \{1, 2, 3, \ldots\}$. The average number of mRNAs produced per expression event is then given by $\langle B \rangle := \Sigma_{i=1}^{\infty} i\alpha_i$. We denote by $\langle B^2 \rangle := \Sigma_{i=1}^{\infty} i^2\alpha_i$ 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

$$\text{Probability}\{m(t+dt)=m+i|m(t)=m\}=k_m\alpha_i dt,\ \forall_i \geq 1 \qquad (17a)$$

$$\text{Probability}\{m(t+dt) = m-1 \mid m(t) = m\} = \gamma_m m dt, \qquad (17b)$$

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

$$Pj(t) := \text{Probability}\{m(t) = j\},\ j \in \{0,1,2,3,\ldots\} \qquad (18)$$

evolves according to the Chemical Master Equation given by Eq. 12. As the protein population count is typically large, the dynamics of protein levels is modeled deterministically through an ordinary differential equation (Eq. 13).

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

$$\frac{d<\psi(m,p)>}{dt} = \langle k_m \psi(m+B,p) + \gamma_m m\psi(m-1,p) - \psi(m,p)(k_m + \gamma_m m)\rangle + \langle \frac{\partial \psi(m,p)}{\partial p}\big(k_p m - \gamma_p p\big)\rangle \quad (19)$$

where $\psi(m, p)$ is any continuously differentiable function and < . > denotes the expected value of the corresponding quantity. Taking $\psi(m, p) = m^i p^j$ in (19) for appropriate integers *i* and *j*, the time evolution of the first and second order statistical moments of the population count are given by

$$\frac{d\langle m\rangle}{dt} = k_m\langle B\rangle - \gamma_m\langle m\rangle \quad (20a)$$

$$\frac{d\langle p\rangle}{dt} = k_p\langle m\rangle - \gamma_p\langle p\rangle \quad (20b)$$

$$\frac{d\langle m^2\rangle}{dt} = k_m\langle B^2\rangle + \gamma_m\langle m\rangle + 2k_m\langle B\rangle\langle m\rangle - 2\gamma_m\langle m^2\rangle \quad (20c)$$

$$\frac{d\langle mp\rangle}{dt} = k_p\langle m^2\rangle + k_m\langle B\rangle\langle p\rangle - \gamma_m\langle mp\rangle - \gamma_p\langle mp\rangle \quad (20d)$$

$$\frac{d\langle p^2\rangle}{dt} = 2k_p\langle mp\rangle - 2\gamma_p\langle p^2\rangle \quad (20e)$$

Setting the left-hand-side of equation (20) to zero, and solving for the steady-state moments we obtain

$$\overline{\langle m\rangle} = \frac{k_m\langle B\rangle}{\gamma_m},\ \overline{\langle p\rangle} = \frac{k_p\overline{\langle m\rangle}}{\gamma_p} \qquad (21a)$$

$$\overline{\langle m^2\rangle} = \overline{\langle m\rangle\langle m\rangle} + \overline{\langle m\rangle}\frac{\langle B^2\rangle+\langle B\rangle}{2\langle B\rangle} \qquad (21b)$$

$$\overline{\langle mp\rangle} = \overline{\langle m\rangle\langle p\rangle} + \overline{\langle p\rangle}\frac{\langle B^2\rangle+\langle B\rangle}{2\langle B\rangle}\frac{\gamma_p}{\gamma_p+\gamma_m} \qquad (21c)$$

$$\overline{\langle p^2\rangle} = \overline{\langle p\rangle\langle p\rangle} + \frac{\overline{\langle p\rangle\langle p\rangle}}{\langle\bar{m}\rangle}\frac{\langle B^2\rangle+\langle B\rangle}{2\langle B\rangle}\frac{\gamma_p}{\gamma_p+\gamma_m} \qquad (21d)$$

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

$$\overline{CV^2} := \frac{\overline{\langle p^2\rangle}+\overline{\langle p\rangle\langle p\rangle}}{\overline{\langle p\rangle\langle p\rangle}} \qquad (22)$$

From (22) we find that the steady-state protein noise level is given by

$$\overline{CV^2} = \frac{n_m\gamma_p}{\overline{\langle m\rangle}(\gamma_p+\gamma_m)} \qquad (23)$$

where

$$n_m := \frac{\overline{\langle m^2\rangle}+\overline{\langle m\rangle\langle m\rangle}}{\overline{\langle m\rangle}} = \frac{\langle B^2\rangle+\langle B\rangle}{2\langle B\rangle} \qquad (24)$$

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

$$\frac{\gamma_p}{\overline{\langle m \rangle}(\gamma_p+\gamma_m)} \qquad (25)$$

In light of Eq. 24, the protein noise levels can be decomposed into two components:

$$\overline{CV^2} = \frac{\gamma_p}{\overline{\langle m \rangle}(\gamma_p+\gamma_m)} + \frac{(n_m-1)\gamma_p}{\overline{\langle m \rangle}(\gamma_p+\gamma_m)} \qquad (26)$$

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).

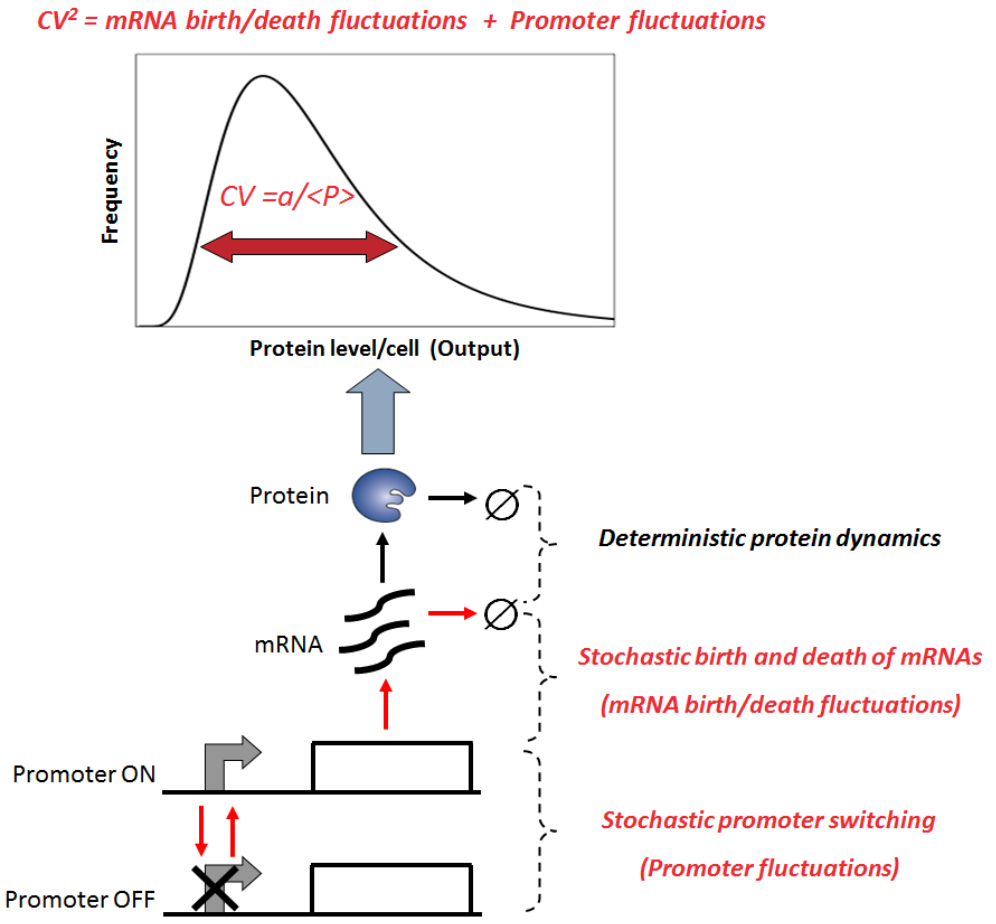

Figure 25. Stochastic hybrid model of gene-expression with two different sources of noise.

Schematic illustrating the two sources of noise: (i) Poisson mRNA fluctuations arising from stochastic production and degradation of individual mRNA molecules, and (ii) promoter fluctuations arising from slow promoter transitions (red arrows) between "Gene OFF" and "Gene ON" states. As protein population counts are often very large, stochasticity in protein dynamics is neglected. Originally published as Figure 1 in Singh et al., Nature MSB, 2012. Abhyudai Singh created and Brandon Razooky edited this figure.

Equation 14 represents the total intrinsic noise in gene-expression and can be decomposed into components representing expression variability originating from mRNA birth/death fluctuations and stochastic promoter switching. When $\eta_m = 1$, then mRNA population counts have Poisson statistics, and protein noise levels primarily arise from mRNA birth/death fluctuations. In contrast, when $\eta_m = 10$, then 90% of the protein noise is generated through promoter fluctuations and transcriptional bursting of mRNAs from the promoter. Finally, we point out that η_m is directly related to the mean transcriptional burst size $\langle B \rangle$ for many promoters. 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:

$$G_{off} \xrightarrow{k_{on}} G_{on} \; ; \; G_{on} \xrightarrow{k_{off}} G_{off}$$

$$G_{on} \xrightarrow{k_m} G_{on} + mRNA \; ; \; mRNA \xrightarrow{\gamma_m} \emptyset$$

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 Eqs. 17-26 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

$$\frac{d\langle g \rangle}{dt} = k_{on} - \big(k_{off} + k_{on)}\big)\langle g \rangle; \frac{d\langle m \rangle}{dt} = k_m \langle g \rangle - \gamma_m \langle m \rangle \qquad (27a)$$

$$\frac{d\langle g^2 \rangle}{dt} = k_{on} - \big(k_{off} + k_{on)}\big)\langle g \rangle - 2\big(k_{off} + k_{on)}\big)\langle g^2 \rangle \qquad (27b)$$

$$\frac{d\langle m^2 \rangle}{dt} = k_m \langle g \rangle + \gamma_m \langle m \rangle + 2k_m \langle gm \rangle - 2\gamma_m \langle m^2 \rangle \qquad (27c)$$

$$\frac{d\langle gm\rangle}{dt} = k_m\langle g^2\rangle + k_{on}\langle m\rangle + (k_{off} + k_{on} + \gamma_m)\langle gm\rangle \qquad (27d)$$

Analysis of (27) shows that the steady-state mRNA Fano factor is given by

$$n_m := \frac{\overline{\langle m^2\rangle} + \overline{\langle m\rangle\langle m\rangle}}{\overline{\langle m\rangle}} = 1 - \frac{k_{off}k_m}{(k_{off}+k_{on})(k_{off}+k_{on}+\gamma_m)} \qquad (28)$$

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)

$$n_m = 1 + \frac{k_m}{k_{off}} \qquad (29)$$

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.

In particular, analysis shows that for a two-state promoter model, where the ON state is unstable and the promoter spends most of the time in the OFF state, $\eta_m = 1 + \langle B \rangle$. Thus, η_m is representative of the mean burst size for many promoters including HIV LTR, which have been shown to reside mostly in the OFF state (Singh et al, 2010; Raj et al, 2006; Suter et al, 2011). We next explore methods to determine η_m for a specific promoter or gene.

Equations 14 & 15 show that increasing the frequency of transcription events, k_m, increases the steady-state mean protein abundance, but decreases the steady-state noise level such that the product $\overline{CV^2} \times \overline{\langle p \rangle}$ remains fixed. Thus, intrinsic noise scales inversely with mean protein abundance, consistent with experimental observations in both prokaryotes and eukaryotes (Bar-Even et al., 2006a; Kepler and Elston, 2001b; Newman et al., 2006; Ozbudak et al., 2002; Simpson et al., 2004a; Thattai and van

Oudenaarden, 2001). One method to determine η_m would be through the scaling of $\overline{CV^2}$ versus $\overline{\langle p \rangle}$, as the scaling factor is proportional to it (Eq. 14). However, the scaling factor also depends on the mRNA translation rate k_p and protein and mRNA half-lives, and thus, this method requires a priori knowledge of these parameters. Moreover, for this $\overline{CV^2}$ versus $\overline{\langle p \rangle}$ method, one needs to quantify protein abundances in absolute molecule counts rather than fluorescence intensities, as is typically the case. Thus, steady-state measurements of protein noise magnitude are insufficient to determine η_m and, hence, insufficient to discriminate between the different components of intrinsic noise. Below, we predict analytically and demonstrate experimentally that η_m can be inferred from transient changes of protein population mean and noise magnitude in response to transcriptional and translational perturbations to gene-expression.

2b. Analytical results: Dynamics of protein noise magnitude in response to perturbations

Small-molecule drugs provide a convenient method to perturb genetic circuits and realize transient changes in protein levels across different cell types. In this section, we theoretically analyze the effects of small-molecule drugs that specifically block transcription and translation. These drugs are routinely used to determine the stability of mRNAs and proteins (Hao and Baltimore, 2009; Yen et al., 2008; Zhou, 2004). Such experiments typically involve bulk assays that track transient changes in mean molecular abundances across a population of cells. 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 km = 0, the statistical moments evolve according to the Equations (20). Assuming that just prior to the transcriptional block the statistical moments had reached steady-state, we solve equation (20a) using the moments calculated in (21) as initial conditions. From (20a) it is straightforward to show that the mean mRNA and protein levels exponentially decay to zero as follows

$$\langle m(t)\rangle = \overline{\langle m\rangle}\, e^{-\gamma_m t} \qquad (30a)$$

$$\langle p(t)\rangle = \overline{\langle p\rangle}\left(\frac{\gamma_p e^{-\gamma_m t} - \gamma_m e^{-\gamma_p t}}{\gamma_p - \gamma_m}\right) \qquad (30b)$$

Solving the remaining equations in (20) using *Mathematica,* we find that the protein noise level (as measured by the coefficient of variation squared) (20c) monotonically increases with time as

$$\frac{CV^2(t)}{\overline{CV^2}} = f\left(\gamma_m, \gamma_p, \eta_m, t\right), \;\; CV^2(t) := \frac{\langle p^2(t)\rangle - \langle p(t)\rangle^2}{\langle p(t)\rangle^2} \qquad (31)$$

where CV^2 given by (23) is the protein noise level at $t = 0$ and the function f has the following asymptote

$$\lim_{t\to\infty} f\left(\gamma_m, \gamma_p, \eta_m, t\right) = \begin{cases} \dfrac{(\gamma_p+\gamma_m)\left(\gamma_p+\eta_m(\gamma_p-2\gamma_p)\right)}{\eta_m\gamma_m(\gamma_m-2\gamma_p)}, & \gamma_m > 2\gamma_p \\ \infty, & \gamma_m \le 2\gamma_p \end{cases} \qquad (32)$$

For given protein and mRNA degradation rates, Figure 25 plots the function $f(\gamma_m, \gamma_p, \eta_m, t)$ as a function of time t for different η_m.

Therefore, in response to a complete block of translation or transcription, the mean protein levels $\langle p(t)\rangle$ decay as,

$$\frac{\langle p(t)\rangle}{\langle p(0)\rangle} = \exp(-\gamma_m t) \quad \text{or} \quad \frac{\langle p(t)\rangle}{\langle p(0)\rangle} = \frac{\gamma_p \exp(-\gamma_m t) - \gamma_m \exp(-\gamma_p t)}{\gamma_p - \gamma_m}, \qquad (33)$$

respectively, where t represents the time since drug addition. Thus, both mRNA and protein half-life can be quantified by monitoring $\langle p(t)\rangle$ in response to blocks in transcription and translation. Next, we investigate how cell-to-cell variation in protein levels (coefficient of variation squared) changes in response to these small-molecule drugs and whether this change provides information about unknown system parameters, such as η_m.

We first consider a translational block to gene expression. Recall that, in this gene-expression model, protein production and degradation are modeled as differential equations. A translational block would result in a deterministic exponential decay in protein levels in each individual cell, which would shift the distribution of protein level across the cell population to lower levels without changing its coefficient of variation. Thus, in case of a translational block, this model predicts that no additional information can be gained from the higher order statistical moments of $p(t)$.

We next consider a transcriptional block to gene expression by setting the frequency of transcriptional events $k_m = 0$ at time $t = 0$. Assuming the system was at steady state before the block, the protein-noise level (measured by the coefficient of variation squared, $CV^2(t)$) changes as

$$\frac{CV^2(t)}{\overline{CV^2}} = f(\gamma_m, \gamma_p, \eta_m, t) \qquad (34)$$

where $\overline{CV^2}$ is the initial protein noise level given by (14), and function $f(\gamma_m, \gamma_p, \eta_m, t)$ is a monotonically increasing function of t. With transcription blocked, mRNA levels exponentially decay, leading to an increase in cell-to-cell variability, as would be expected from the inverse scaling of noise and mean levels. This increase in mRNA noise is transmitted to encoded proteins, causing $CV^2(t)$ to monotonically increase over time (Fig. 26). Interestingly, this analysis indicates that the rate of increase of $CV^2(t)$ is dependent on the source of intrinsic noise. This point is illustrated in Fig. 26, which plots the function $f(\gamma_m, \gamma_p, \eta_m, t)$ for different values of η_m, and hence different relative contributions of mRNA birth/death and promoter fluctuations to expression noise. More specifically, as we increase η_m, the initial rise in the protein-noise level becomes more and more gradual. Intuitively, this occurs because, when expression noise results primarily from promoter fluctuations, then mRNA counts are sufficiently high, and mRNA degradation is essentially a deterministic process. Thus, when transcription is blocked, mRNAs degrade approximately deterministically across single cells, which reduce protein levels and generate a slow increase in $CV^2(t)$. On the other hand, when expression noise originates from mRNA birth/death fluctuations, then a transcriptional block results in stochastic degradation of mRNAs across the cells, which rapidly increases the protein noise levels (Fig. 26). In summary, the increase in protein cell-to-cell variability after transcription is blocked is dependent on η_m, and the dynamics of $CV^2(t)$ can discriminate between alternative sources of intrinsic stochasticity in gene expression.

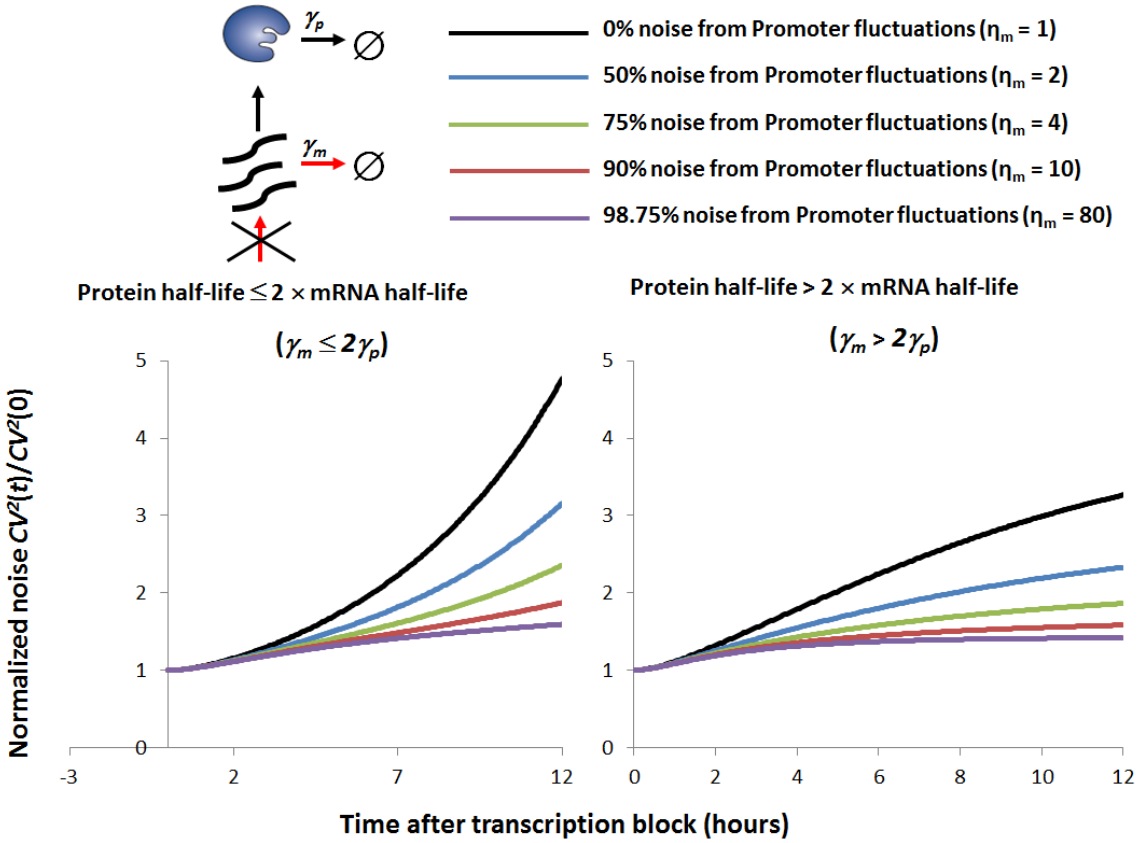

Figure 26. Transient changes in gene-expression noise can discriminate between mRNA birth/death and promoter fluctuations.

Predictions for changes in gene-expression noise after transcription is blocked when protein half-life ≤ 2 × mRNA half-life ($\gamma_m \leq 2\gamma_p$, left) and protein half-life > 2 × mRNA half-life ($\gamma_m > 2\gamma_p$, right), for different contributions of promoter fluctuations to gene-expression noise. If variability in protein levels mainly results from stochastic birth and death of individual mRNA molecules then $CV^2(t)$ is predicted to change much more rapidly than when protein variability results from promoter fluctuations. The y-axis is normalized by the $CV^2(t)$ at $t = 0$. These simulations were preformed assuming a 2.5-hour protein half-life with a 3-hour (left) and 1-hour (right) mRNA half-life. Originally published as Figure 2 in Singh et al., Nature MSB, 2012. Abhyudai Singh and Brandon Razooky created this figure.

To test how predictions in Fig. 26 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

$$\text{Probability}\{p(t+dt) = p+1 \mid p(t) = p,\ m(t) = m\} = k_p m \qquad (35a)$$

$$\text{Probability}\{p(t+dt) = p-1 \mid p(t) = p,\ m(t) = m\} = \gamma_p p dt, \qquad (35b)$$

where k_p is the mRNA translation rate and γ_p is the protein degradation rate. The above equation together with Eq. (17) 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

$$\frac{d\langle m\rangle}{dt} = k_m\langle B\rangle - \gamma_m\langle m\rangle, \qquad \frac{d\langle p\rangle}{dt} = k_p\langle m\rangle - \gamma_p\langle p\rangle \qquad (36a)$$

$$\frac{d\langle m^2\rangle}{dt} = k_m\langle B^2\rangle + \gamma_m\langle m\rangle + 2k_m\langle B\rangle\langle m\rangle - 2\gamma_m\langle m^2\rangle \qquad (36b)$$

$$\frac{d\langle mp\rangle}{dt} = k_p\langle m^2\rangle + k_m\langle B\rangle\langle p\rangle - \gamma_m\langle mp\rangle - \gamma_p\langle mp\rangle \qquad (36c)$$

$$\frac{d\langle p^2\rangle}{dt} = k_p\langle m\rangle + \gamma_p\langle p\rangle + 2k_p\langle mp\rangle - 2\gamma_p\langle p^2\rangle \qquad (36d)$$

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

$$\overline{\langle m\rangle} = \frac{k_m\langle B\rangle}{\gamma_m},\ \overline{\langle p\rangle} = \frac{k_p\overline{\langle m\rangle}}{\gamma_p} \qquad (37a)$$

$$\overline{\langle m^2\rangle} = \overline{\langle m\rangle\langle m\rangle} + \overline{\langle m\rangle}\frac{\langle B^2\rangle+\langle B\rangle}{2\langle B\rangle} \qquad (37b)$$

$$\overline{\langle mp\rangle} = \overline{\langle m\rangle\langle p\rangle} + \overline{\langle p\rangle}\frac{\langle B^2\rangle+\langle B\rangle}{2\langle B\rangle}\frac{\gamma_p}{\gamma_p+\gamma_m} \qquad (37c)$$

$$\overline{\langle p^2\rangle} = \overline{\langle p\rangle\langle p\rangle} + \langle p\rangle + \frac{\overline{\langle p\rangle\langle p\rangle}}{\langle\overline{m}\rangle}\frac{\langle B^2\rangle+\langle B\rangle}{2\langle B\rangle}\frac{\gamma_p}{\gamma_p+\gamma_m} \qquad (37d)$$

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

From (37) the steady-state protein noise level is given by

$$\overline{CV^2} = \frac{1}{\langle p \rangle} + \frac{n_m \gamma_p}{\overline{\langle m \rangle}(\gamma_p + \gamma_m)} \qquad (38)$$

note that this noise level is similar to (23), except for the 1/<p> 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 (36) using (37) as initial conditions. Fig. 27 plots protein noise levels after a transcriptional block for different steady-state mean protein levels per cell. As can be seen from Fig. 27, 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 (<p> = ∞ line in Fig. 27). Typically, as long as

$$\overline{CV^2} \gg \frac{1}{\langle p \rangle} \qquad (39)$$

modeling protein dynamics using mass-action kinetics is a good approximation. The clones considered in this paper have noise levels in the range of $CV^2 \approx 0.5 - 1$ and have on average over 10,000 copies of d2GFP protein molecules per cell (Singh et al. Biophysical Journal 2010). Thus we satisfy (39) by many orders of magnitude.

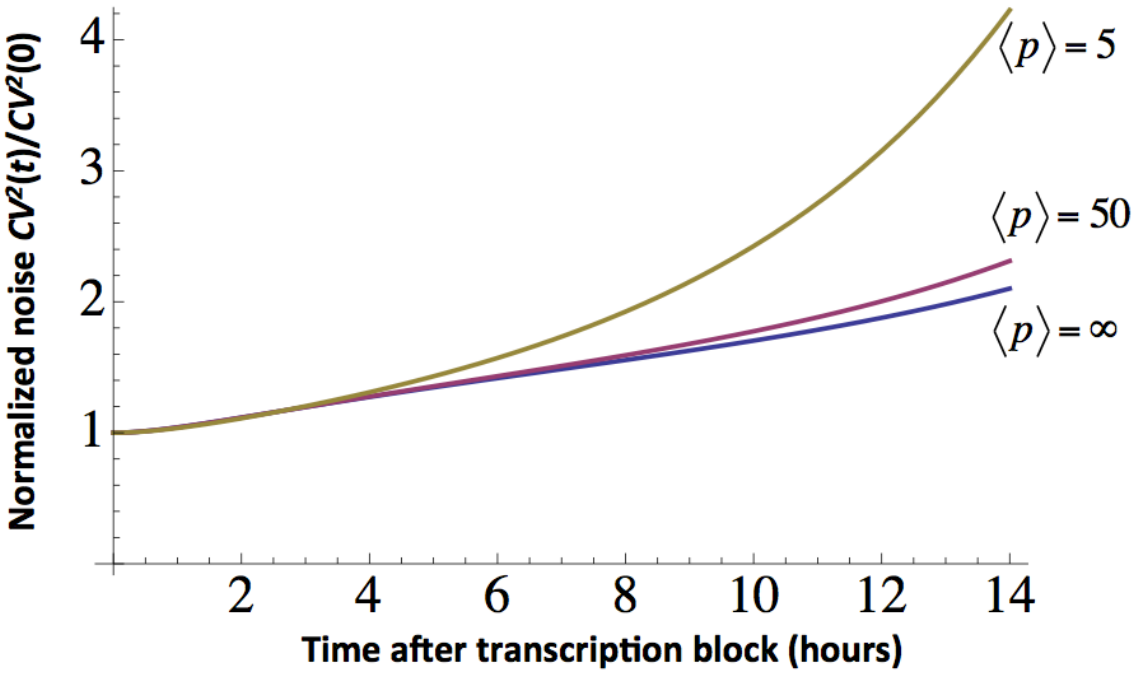

Figure 27. Predicted changes in protein expression noise after transcription is blocked for different mean protein copy number per cell.

$\langle \boldsymbol{p} \rangle = \infty$ corresponds to deterministic protein birth-death dynamics. B is assumed to be geometrically distributed with a mean burst size of 10 transcripts and $CV^2 = 0.5$. mRNA and protein half-lives are taken as 3 and 2.5 hours, respectively. Noise levels are baselined by their corresponding values at time $t = 0$. Originally published in supplemental information in Singh et al., Nature MSB, 2012. Abhyudai Singh created this figure.

Therefore, the above predictions in Fig. 26 do not alter significantly if protein synthesis and decay is stochastically modeled.

Eq. 34 shows that the rise in protein noise levels after a transcription block is dependent on η_m, γ_m and γ_p. Assuming protein and mRNA degradation rates are known or have been determined from the kinetics of protein decay (Eq. 33), η_m can be directly inferred from $CV^2(t)$. Since this procedure to quantify η_m depends on examining relative changes in $CV^2(t)$, it will work irrespective of whether protein levels are quantified in terms of fluorescence intensity or molecular counts. However, if data on absolute protein molecular count are available, then the frequency of transcriptional events k_m and the mRNA translation rate k_p can also be determined from the steady-state protein mean and noise level (Eqs. 14 & 15). Hence, by combining transient data on the statistical moments of $p(t)$ (Eqs. 33 & 34) together with their steady-state values (Eqs. 14 & 15), one can infer all the parameters of the gene-expression model.

One limitation of this method is that it may not be applicable in cases where the mRNA half-life is much shorter than the protein half-life. This limitation arises because when mRNA half-life is short, mRNA transcripts quickly decay to zero once transcription in blocked. With no available mRNA, protein decay will be dominated by protein degradation alone and will not contain any information about the underlying mRNA dynamics. For example, when γ_m is much larger than γ_p, then decay in mean protein levels is always given by $p(t)=\exp(-\gamma_p t)$, irrespective of whether transcription or

translation is blocked (Eq. 33), hence providing no information on the mRNA half-life. Thus, for this method to work, reporter systems need to be used where both mRNA and protein degradation reactions are first-order processes that occur with known comparable rates. Moreover, small-molecule drugs used for perturbing transcription/translation should not alter these decay rates.

Importantly, the change in protein expression variability after a transcription block becomes insensitive to η_m, when η_m is large. For example, model predictions for η_m equal to 1 and 10 are well separated in Fig. 26, but predictions for η_m equal to 10 and 80 are close to each other. Thus, it may not be possible to get precise estimates of η_m by this method, especially when η_m is large (i.e., high levels of transcriptional bursting form the promoter).

2c. Experimental Results: HIV-1 LTR gene-expression noise results primarily from transcriptional bursting

In this section, we illustrate the experimental utility of this method by quantifying the relative contributions of mRNA birth/death and promoter fluctuations to HIV-1 expression noise. Stochastic expression of viral proteins from the HIV-1 long terminal repeat (LTR) promoter critically influences the viral fate-decision between active replication and post-integration latency (a dormant state of the virus analogous to phage lysogeny) in single cells (Weinberger et al., 2005; Weinberger et al., 2008a; Weinberger and Shenk, 2007). Thus, this is an important system to identify the source of intrinsic noise in gene-expression.

To discriminate between the different sources of intrinsic noise using the dynamics of protein noise magnitude, it is essential to use a reporter system in which the mRNA half-life is not very short, compared to the protein half-life. Towards that end, we used flow cytometry measurements of a destabilized version of GFP (d2GFP) that exhibits a 2.5-hour half-life. Previous measurements quantified a 3-hour half-life for the d2GFP mRNA (Raj et al., 2006), so $\gamma_m \approx \gamma_p$ for d2GFP, making it an ideal reporter system for this study. To quantify HIV-1 LTR expression noise, we use our previously described library of isoclonal populations (Singh et al., 2010a), where each isoclonal population carries a single copy of a minimal vector encoding HIV-1 LTR driving d2GFP, integrated at unique location in the human genome. To focus on intrinsic noise and remove extrinsic contributions to noise measurements, gating of flow cytometry data was performed as described (Newman et al., 2006; Singh et al., 2010a). To filter extrinsic noise, a two-color fluorescent reporter system (Elowitz et al., 2002) was implemented where each cell contains a single integrated copy of the LTR driving d2GFP and a second copy of the LTR driving mCherry at a different integration site. These populations were gated around the forward (FSC) and side (SSC) scatter medians show little correlation between the GFP and Cherry signal (Fig. 28). This result shows that appropriate gating of cells removed most of the extrinsic noise. After filtering of extrinsic noise, isoclonal populations exhibit an inverse scaling of steady-state noise level ($\overline{CV^2}$) with mean protein abundance (Singh et al., 2010a), as predicted by the model (Eq. 14). This inverse scaling of $\overline{CV^2}$ versus $\overline{\langle p \rangle}$ is consistent with both mRNA birth/death and promoter

fluctuations (Eq. 14), and as mentioned above steady-state protein noise levels are insufficient to discriminate between the different sources of intrinsic noise.

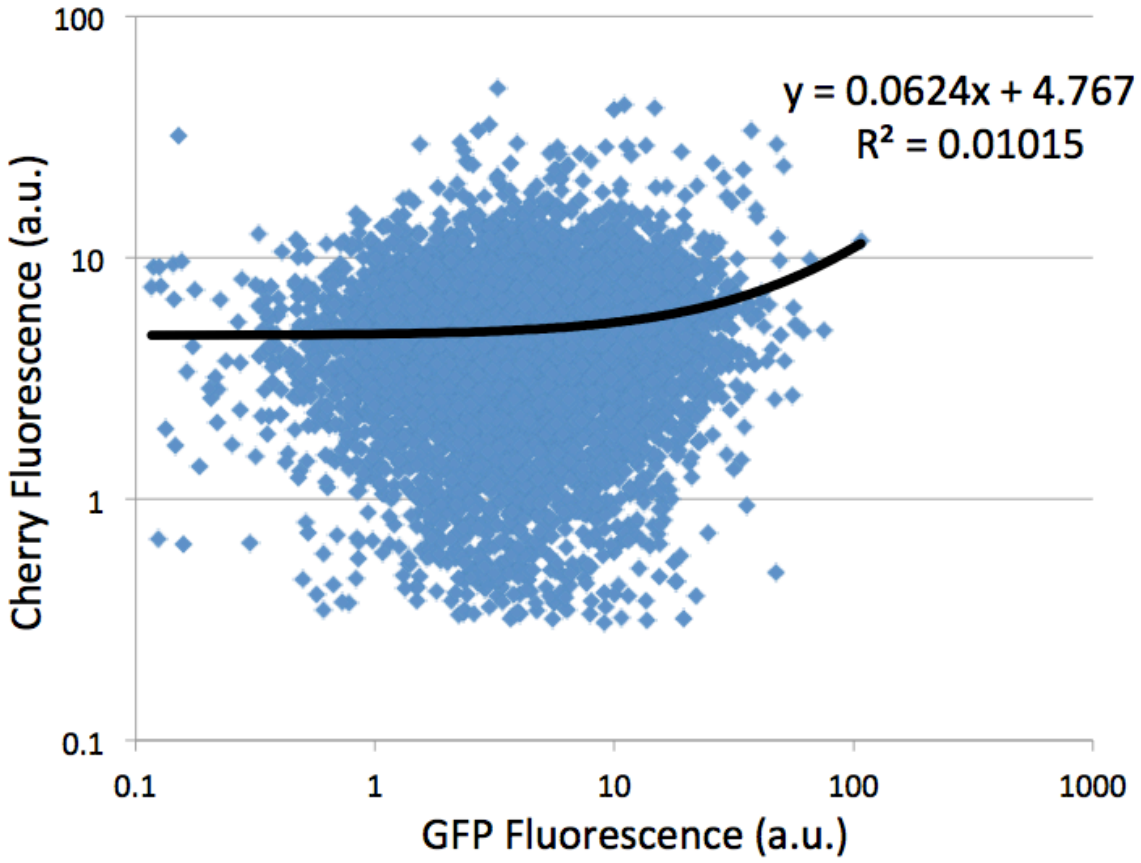

Figure 28. Scatter plot of single-cell intensities taken from flow-cytometry data for an isoclonal cell population shows little correlation between the GFP and Cherry signal.

Populations were obtained by drawing a small gate around the forward (FSC) and side (SSC) scatter medians. Originally published in supplemental information in Singh et al., Nature MSB, 2012. Brandon Razooky created this figure.

To discriminate between mRNA birth/death and promoter fluctuations, we analyzed dynamic changes in HIV-1 LTR expression noise after perturbation with two transcriptional inhibitors, actinomycin D, a small-molecule drug that rapidly and

efficiently blocks transcription by preventing RNA polymerase II (RNAp II) elongation (Sobell, 1985), and Flavorpiridol, a small-molecule drug that blocks transcription by inhibiting the interaction of P-TEFb with RNAp II (Chao and Price, 2001). Introducing either transcriptional inhibitor to any of the isoclonal populations resulted in a decrease in d2GFP fluorescence intensity levels that was consistent with the reported 2.5-hour d2GFP protein and 3-hour d2GFP mRNA half-lives (Fig. 29). The fact that the decay in d2GFP fluorescence after actinomycin D treatment is consistent with the reported 3-hour mRNA half-life (Raj et al., 2006) suggests that the dynamics of the reporter decay are not affected by the presence of the drug. The data also show that d2GFP fluorescence decays, in the presence of cycloheximide, at ~3.5hrs both in the presence and absence of actinomycin D (Fig. 30).

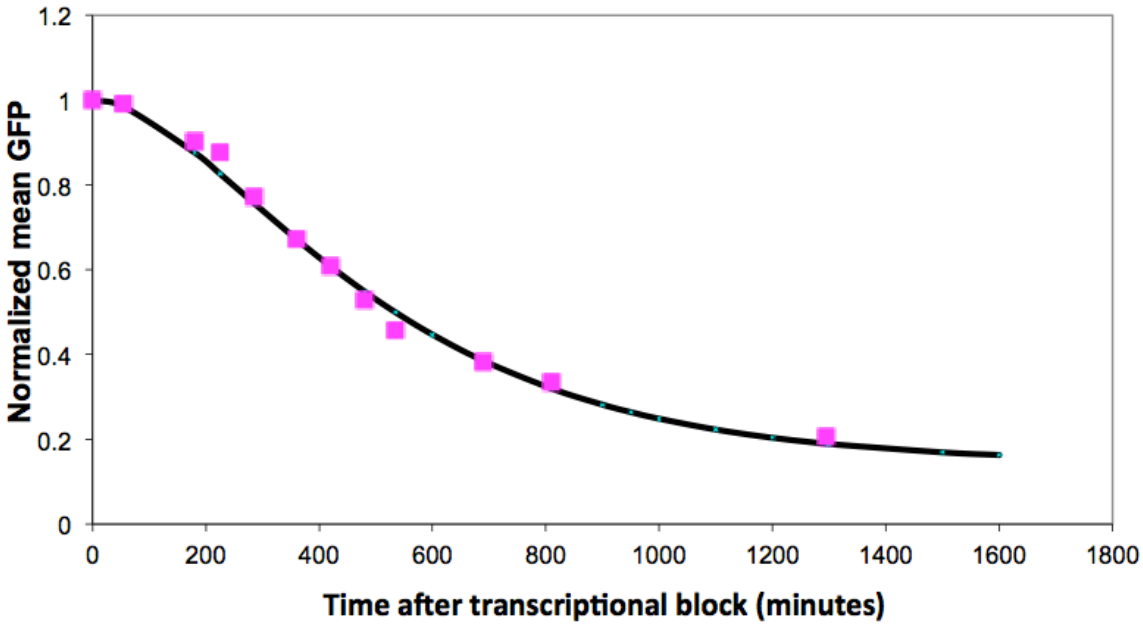

Figure 29. Transient changes in mean GFP levels after a transcriptional block. GFP levels are normalized by their corresponding value at $t = 0$.

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 (40) to data, and it asymptotically approaches a value of ≈ 10%. The protein and mRNA half-lives are assumed to be 2.5 and 3 hours, respectively. Originally published in supplemental information in Singh et al., Nature MSB, 2012. Abhyudai Singh created this figure.

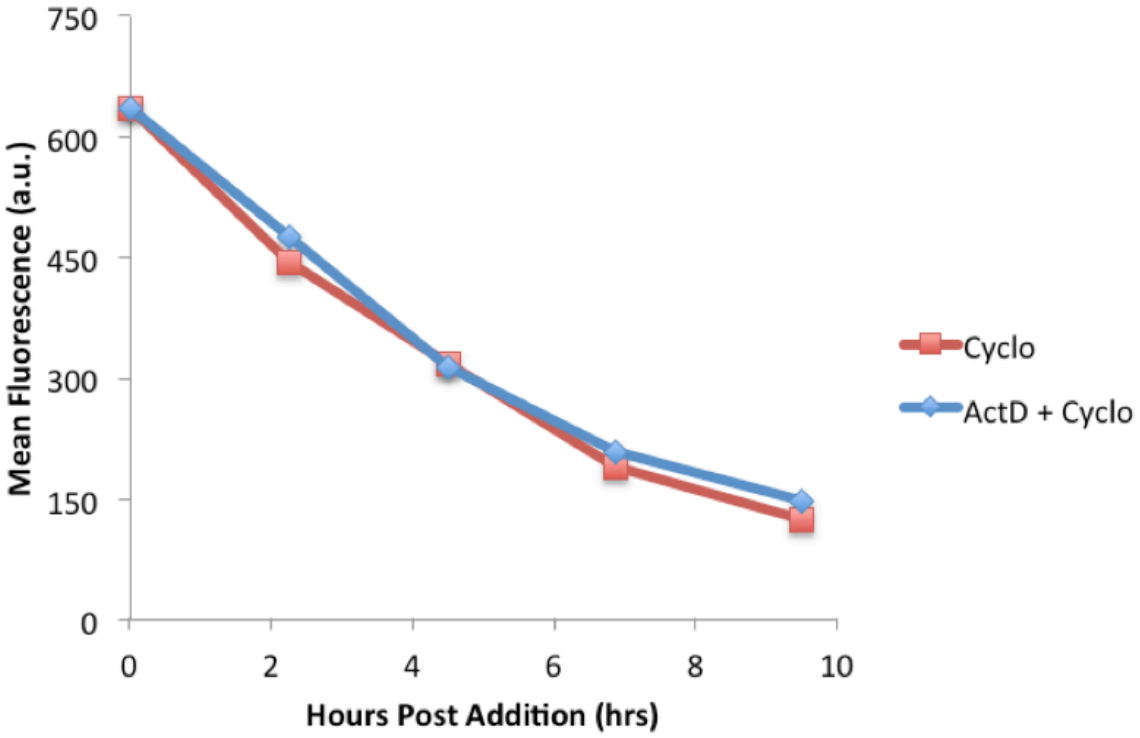

Figure 30. GFP decays at the same rate after translation is blocked using Cycloheximide, both in the presence and absence of Actinomycin D.

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 same both in the presence and absence of Actinomycin D. Originally published in supplemental information of Singh et al., Nature MSB, 2012. Roy Dar and Brandon Razooky created this figure.

In response to both transcriptional inhibitors, the cell-to-cell variability in fluorescence intensity levels across all the isoclonal populations ($CV^2(t)$) gradually increases over time (Fig. 31). Fitting the data to model predictions shows that this increase in $CV^2(t)$ is inconsistent, for all clones, with a model where intercellular variability in protein levels originates only from mRNA birth/death fluctuations (Fig. 31). Instead, the data indicate that promoter fluctuations act as the dominant source of gene-expression noise in the HIV-1 LTR promoter with a mRNA Fano factor (η_m) of at least 10 across four different integration sites. Direct measurements of mRNA population statistics by mRNA smFISH also show high levels of transcriptional bursting from the HIV-1 LTR ($\eta_m \approx 80$; Fig. 32) and provide an independent experimental verification of the proposed method.

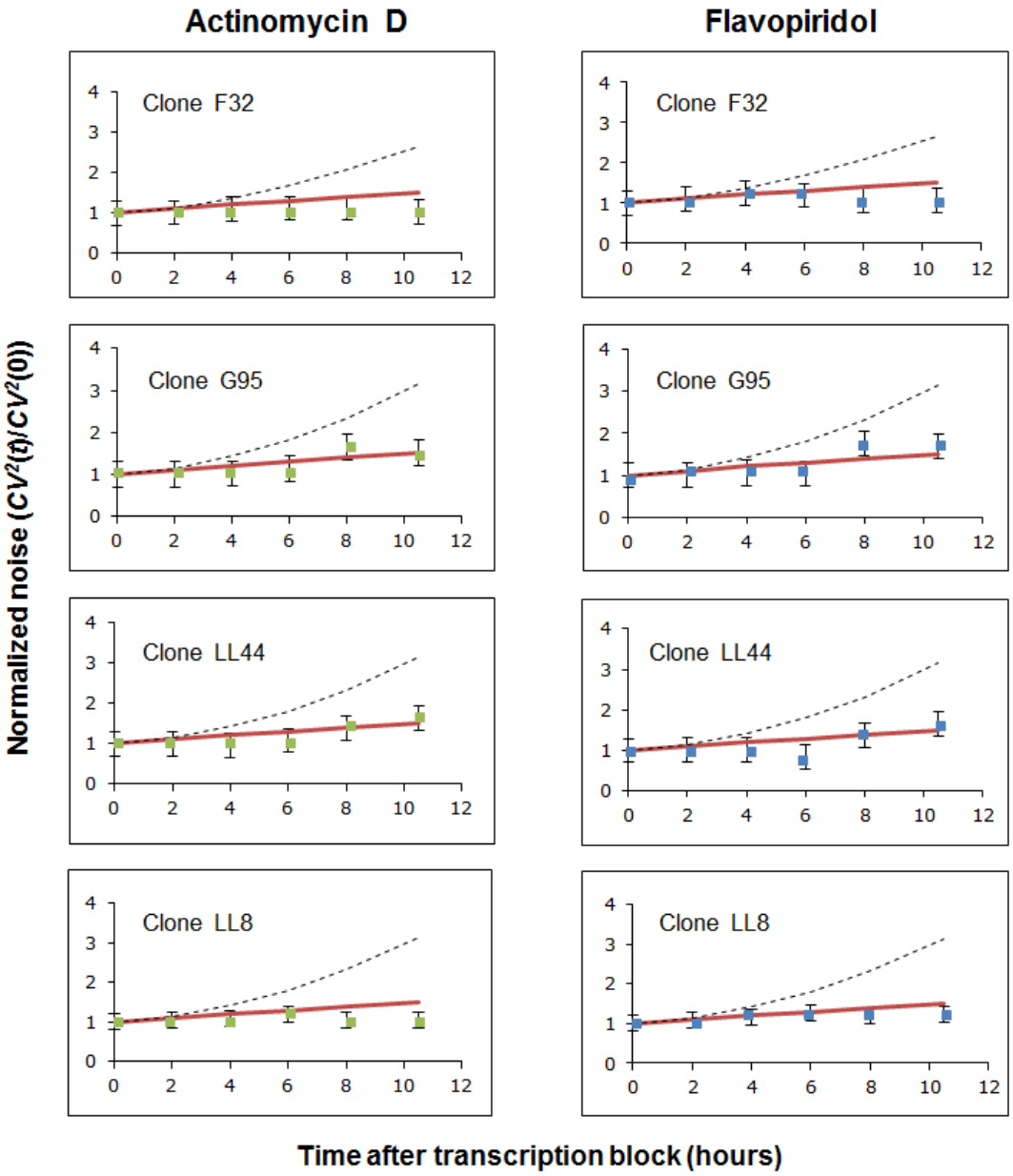

Figure 31. Transcriptional bursting is a significant source of variability in HIV-1 LTR gene-expression across different integration sites.

Time courses of $CV^2(t)$ of GFP expression, as measured by flow cytometry, for four isoclonal Jurkat T lymphocyte populations (F32, G95, LL44, and LL8) after perturbation with transcriptional blocking drugs actinomycin D (left column) or flavopiridol (right column). The transient increases in gene-expression noise after transcriptional blocking were consistent with a model in which HIV-1 LTR expression noise is primarily due to transcriptional bursting of mRNAs from the viral promoter (red lines; $\eta_m=15$). In contrast, a model in which mRNA birth/death fluctuations act as the dominant source of gene-expression noise (black dashed lines; $\eta_m=1$), overestimates the observed transient increase in $CV^2(t)$ for all clones. All quantities are normalized by their corresponding

values of $CV^2(t)$ at $t = 0$; error bars show 95% confidence interval for the $CV^2(t)$. Originally published in supplemental information in Singh et al., Nature MSB, 2012.

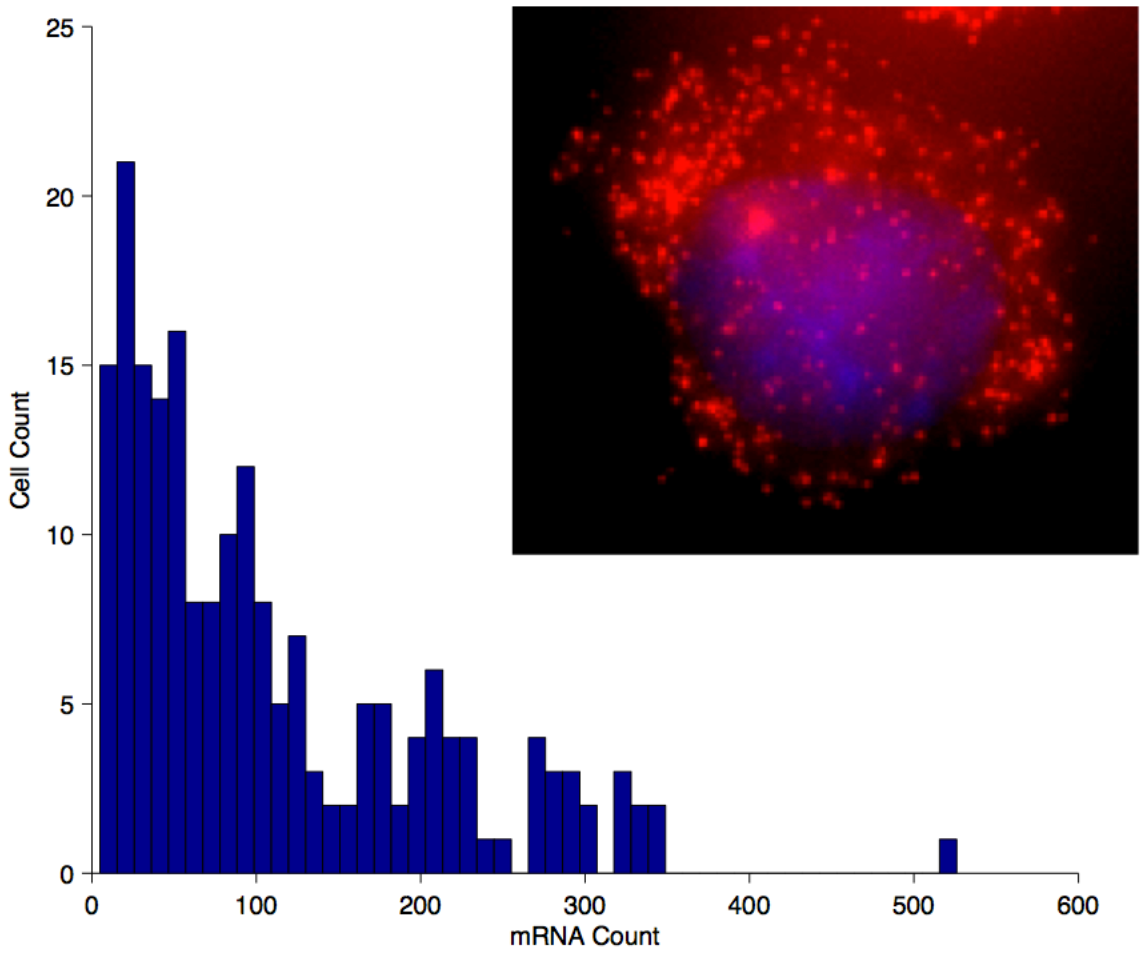

Figure 32. Histogram of mRNA populations counts across an isoclonal cell population determined using mRNA FISH.

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

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

was calculated from the histogram using bootstrapping. We obtain ηm ≈ 80 with a 95% confidence interval of (65, 95). This value of ηm 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

Inset: Example fluorescent image of DAPI stained nuclei (blue) and fluorescently labeled mRNA (red dots). Originally published in supplemental information in Singh et al., Nature MSB, 2012.

3. Conclusions

In summary, we present a simple method to discriminate between mRNA birth/death and promoter fluctuations by measuring the dynamics of protein-expression noise in response to perturbations. This perturbation approach is complementary to other methods, such as mRNA smFISH, as it only requires data at the protein level and can work at high mRNA-expression levels. With the large libraries of GFP-tagged proteins and reporters already exist for many systems, it may be possible to quantify the extent of transcriptional bursting across the genome without having to design individual mRNA smFISH probes for each gene.

One potential shortcoming of this approach is that changes in expression noise in response to perturbations can be sensitive to extrinsic noise in the system. For example, in the absence of extrinsic noise, different intrinsic-noise mechanisms predict very distinct changes in $CV^2(t)$ in response to perturbations (Fig. 2). However, a constitutive promoter with high levels of extrinsic noise in the transcription rate and a promoter with transcriptional bursting can have similar changes in $CV^2(t)$ after transcription is blocked. Thus, the presence of extrinsic noise hampers the ability to discriminate the different sources of intrinsic noise. These results emphasize that, to discriminate mRNA

birth/death and promoter fluctuations, extrinsic noise in gene expression must be filtered out by appropriate gating of cells or use of two-color reporter systems (as done in this paper; see Appendix H).

Recent theoretical observations suggest that, during cell division, stochastic partitioning of molecules results in the cell-to-cell variability observed in clonal populations (Huh and Paulsson, 2011). However, the present study focuses on transient changes in noise within a clonal population instead of steady-state measurements. The 12-hour duration of our experiments is also much shorter than the ~30-hour doubling time of Jurkat cells, therefore minimizing effects that may be due to stochastic partitioning during the course of the experiment. Most importantly, the perturbations performed in the present study utilize small molecules, actinomycin D and flavopiridol, that induce cell-cycle arrest (Kuerbitz et al., 1992; Nelson and Kastan, 1994) and flavopiridol (Schrump et al., 1998; Shapiro et al., 1999), therefore circumventing any added stochastic effects arising from partitioning.

Measuring transient changes in reporter expression noise in response to a transcriptional block showed that transcriptional bursting of mRNAs from the HIV-1 LTR is a significant source of noise in viral gene-expression. It is now well established that the HIV-1 LTR exhibits a block in transcriptional elongation: after transcriptional initiation from the LTR, RNAp II is poorly processive and stalls 50–70 nucleotides after initiation proximal to a nucleosome referred to as nuc-1 (Jordan et al., 2003; Kao et al., 1987). One attractive model is that the stochastic recruitment of elongation and chromatin

remodeling factors can enhance elongation processivity and remodel nuc-1, respectively, resulting in a stochastic burst of multiple stalled mRNAs (Singh et al., 2010a). Such transcriptional bursting can be an important regulatory mechanism to generate high cell-to-cell variability in the levels of expressed viral proteins for HIV-1 decision-making between active replication and latency in single cells. These results add to an increasing body of work (Austin et al., 2006; Cox et al., 2008a; Dunlop et al., 2008; Munsky et al., 2009; Sanchez et al., 2011; Warmflash and Dinner, 2008; Wong et al., 2011) showing that fluctuations in protein levels carry important information, and that measuring statistical properties of these fluctuations can be an important tool for characterizing the regulatory mechanisms of genetic circuits.

Chapter 7: Transcriptional bursting is the predominant mode of transcription for HIV throughout the genome

Chapter 7 was published in Dar, R.D., Razooky, B.S., Singh, A., Trimeloni, T.V., McCollum, J.M., Cox, C.D., Simpson, M.L., and Weinberger, L.S. (2012). Transcriptional burst frequency and burst size are equally modulated across the human genome. Proc Natl Acad Sci U S A *109*, 17454-17459. The article can be found on Pubmed; http://www.ncbi.nlm.nih.gov/pubmed/23064634

1. Introduction

The predominant mode of gene expression in prokaryotes and eukaryotes is unclear. The classical view of gene expression as a constitutive, Poisson-like accumulation of gene products (Fig. 33A) is supported by a comprehensive large-scale study in bacteria, demonstrating that >400 genes appear to follow constitutive (or Poisson-like) gene expression (Taniguchi et al., 2010). Constitutive expression has also been reported for subsets of human genes (Yunger et al., 2010). Conversely, several elegant studies showed that specific promoters in bacteria and yeast express gene products in an episodic process (Fig. 33B), characterized by pulsatile bursts in transcription (Bar-Even et al., 2006b; Blake et al., 2003; Cai et al., 2006; Golding et al., 2005; Newman et al., 2006; Pedraza and Paulsson, 2008; So et al., 2011). Is episodic bursting the predominant mode of gene expression across a genome or just a highlighted exception? If bursting is predominant, does it depend on genomic location?

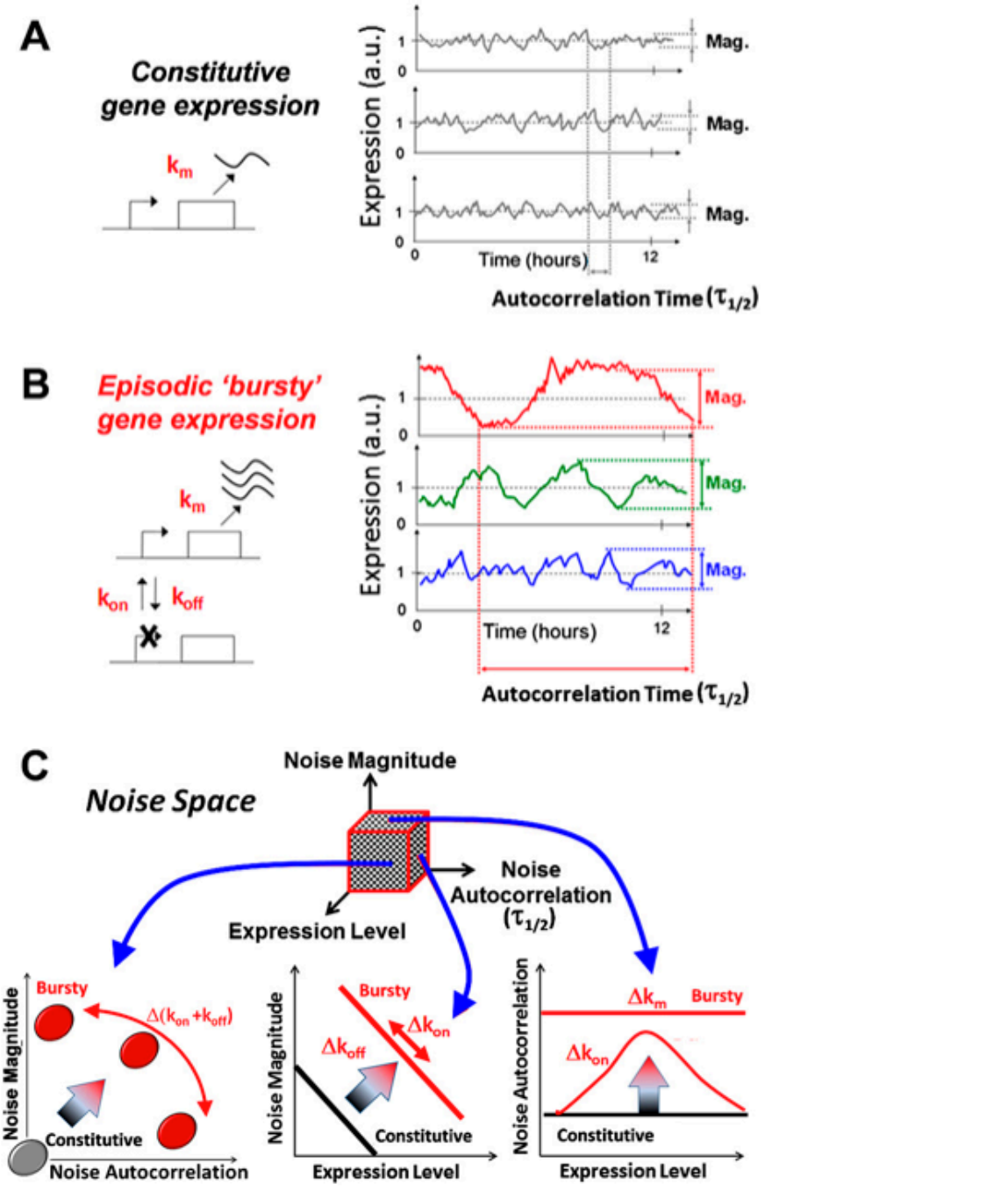

Figure 33. Fluctuations in gene expression to differentiate between alternate models of transcription across the genome.

(A and B) Schematics of the constitutive, Poisson-expression model and the episodic, bursty gene-expression model, together with three expression trajectories from hypothetical genomic loci. Sites that exhibit constitutive (i.e., Poisson) expression exhibit small and relatively fast fluctuations in gene products over time. Alternatively, loci that exhibit episodic expression bursts generate large, slow fluctuations in gene expression. (C) The principle of noise space. The three-dimensional noise space consists of noise magnitude, noise autocorrelation, and mean expression level. Small, fast fluctuations

have a small noise magnitude and short autocorrelation times and thus cluster (after normalization) at the origin of the noise magnitude-autocorrelation plane (gray region, Lower Left). Large, slow (i.e., bursty) fluctuations have expanded noise magnitude and extended autocorrelation times (red ovals). The three-dimensional space can be decomposed into two additional two-dimensional projections of noise magnitude and noise autocorrelation versus mean expression level (Lower Center and Lower Right). For episodic-bursty expression, a trajectory's noise-space coordinates are invariably shifted away from the constitutive model into the burst model space depending on changes to their transcriptional parameters. Originally published as Figure 1 in Dar et al., PNAS, 2012. Roy Dar created this figure.

To globally determine if constitutive "Poisson-like" expression or episodic "bursty" expression dominates throughout the human genome, we capitalize on a recently proposed theoretical framework (Cox et al., 2008b) for extracting the details of gene regulation from the time-resolved structure of fluctuations (i.e., "noise") in gene expression. This analysis quantifies time-lapse expression trajectories to obtain three orthogonal measures of expression: the average expression level, the magnitude of expression fluctuations (as measured by the coefficient of variation squared, CV^2), and the autocorrelation time of expression fluctuations (as measured by the noise autocorrelation time at half of its initial value, $\tau_{1/2,}$ (Austin et al., 2006; Weinberger et al., 2008b) (Fig. 33C). While this 3-D "noise space" is impractical to analyze directly, different two-dimensional projections of noise space allow the quantification of rate parameters in gene-regulatory models and provide a convenient method to differentiate between underlying gene-expression mechanisms, such as constitutive versus bursty transcription (Fig. 33C). For example, transcriptional bursting increases both noise magnitude and noise autocorrelation time (i.e., points in the CV^2-versus-$\tau_{1/2}$ plane shift to the upper-right quadrant relative to a constitutive expression model) (Fig. 33C, bottom left). Conversely, translational bursting shifts noise magnitude, but not the autocorrelation time (Cox et al., 2008b; Simpson et al., 2004b).

Importantly, analysis of the $\tau_{1/2}$ axis is critical to fully parameterize "two-state" transcriptional bursting models (Fig. 33C), which always include at least three unknown parameters: the rate of transition to a transcriptionally active state (k_{on}), the rate of transitioning to a transcriptionally inactive state (k_{off}), and the rate of transcription once in

the active state (k_m) (Kepler and Elston, 2001a; Simpson et al., 2004b). Analyses of a single 2-D plane (e.g., CV^2 versus expression level) cannot fully determine these three rate parameters. Conversely, analyses of CV^2 versus expression level and $\tau_{1/2}$ versus expression level allow the determination of these three parameters, and analysis of CV^2-versus-$\tau_{1/2}$ facilitates direct comparisons of data containing widely varying expression levels, since it removes the reciprocal dependence of noise magnitude on expression level (Cox et al., 2008b).

The ability to accurately quantify these transcriptional rate parameters is essential for answering basic questions about the mechanisms that regulate transcription. Previous studies elegantly applied flow cytometry (Bar-Even et al., 2006b; Newman et al., 2006) and time-lapse microscopy (Cohen et al., 2008; Sigal et al., 2006; Taniguchi et al., 2010) to analyze gene-expression noise in large subsets of genes; however, a tedious experimental bottleneck of sub-cloning and expansion of isogenic populations necessarily limits the throughput of these noise-analysis approaches. Here, we circumvent this sub-cloning requirement to globally apply the analytical framework of noise space across the human genome and quantify these transcriptional rate parameters. The analysis addresses two specific questions regarding transcriptional regulation in human cells. (i) Does constitutive (Poisson-like) expression or episodic (bursty) expression dominate throughout the human genome? (ii) Does genomic location influence either mode of expression? For example, if bursting is operant, does genomic location influence burst size or burst frequency, and which is predominantly influenced? Recent studies (Singh et al., 2010b; Skupsky et al., 2010; Suter et al., 2011a) have tackled these questions for specific genes, but no clear and broad consensus has yet emerged.

2. Results

2a. Measuring noise and correlation time using single-cell time-lapse microscopy

To globally apply the analytical framework of noise space to screen for constitutive versus bursty expression across the human genome, we capitalized on the semi-random pattern of integration exhibited by the HIV-1 lentivirus, where the majority of integrations (~69%) occur within transcriptionally active regions (Mitchell et al., 2004; Schroder et al., 2002). Jurkat T lymphocytes were infected with HIV-based lentiviral vectors encoding a short-lived, two-hour half-life, version of green fluorescent protein (referred to as d_2GFP), to generate a library of cells in which the vector is integrated at a distinct genomic position in each individual cell (i.e., a 'polyclonal' library) (Fig. 34A). To focus on measuring the intrinsic fluctuation dynamics of the genomic region surrounding the vector-integration site, we utilized a vector encoding the HIV-1 long terminal repeat (LTR) promoter, which is relatively weak and heavily influenced by the expression dynamics of the local chromatin environment (Jordan et al., 2001).

2b. Distinguishing between different sources of noise.

Initially, the analytical framework for distinguishing bursty expression was applied to five isoclonal populations: each population was grown from a single parent cell, and all daughter cells, therefore, share the same LTR genomic integration site (Fig. 34A-D). Cell fluorescence was then imaged for 18 h, and the resulting fluorescence intensity

trajectories were used to construct a three-dimensional noise space with three axes: noise magnitude as measured by the coefficient of variation squared (CV^2), noise autocorrelation represented by the half-autocorrelation time ($\tau_{1/2}$), and mean expression level (Fig. 34B-C). Analysis of these trajectories in the noise space allows comparison between isoclonal populations by extraction and identification of global expression characteristics, including differentiation between isoclones that exhibit constitutive transcription versus episodic bursts of transcription (Fig. 34D). To generate the baseline origin for noise-space analysis of single-cell data, we used a library of isoclonal populations (Singh et al., 2010b), each containing a single integration site, and identified isoclones that were the "most Poissonian" in their expression fluctuation profiles (Fig. 35). Two isoclonal populations exhibiting the fastest fluctuation autocorrelation decays (i.e., shortest autocorrelation times) were selected as the most-Poissonian and used to establish an origin of the CV^2-versus-$\tau_{1/2}$ noise map (Fig. 34D, clones 1 and 2).

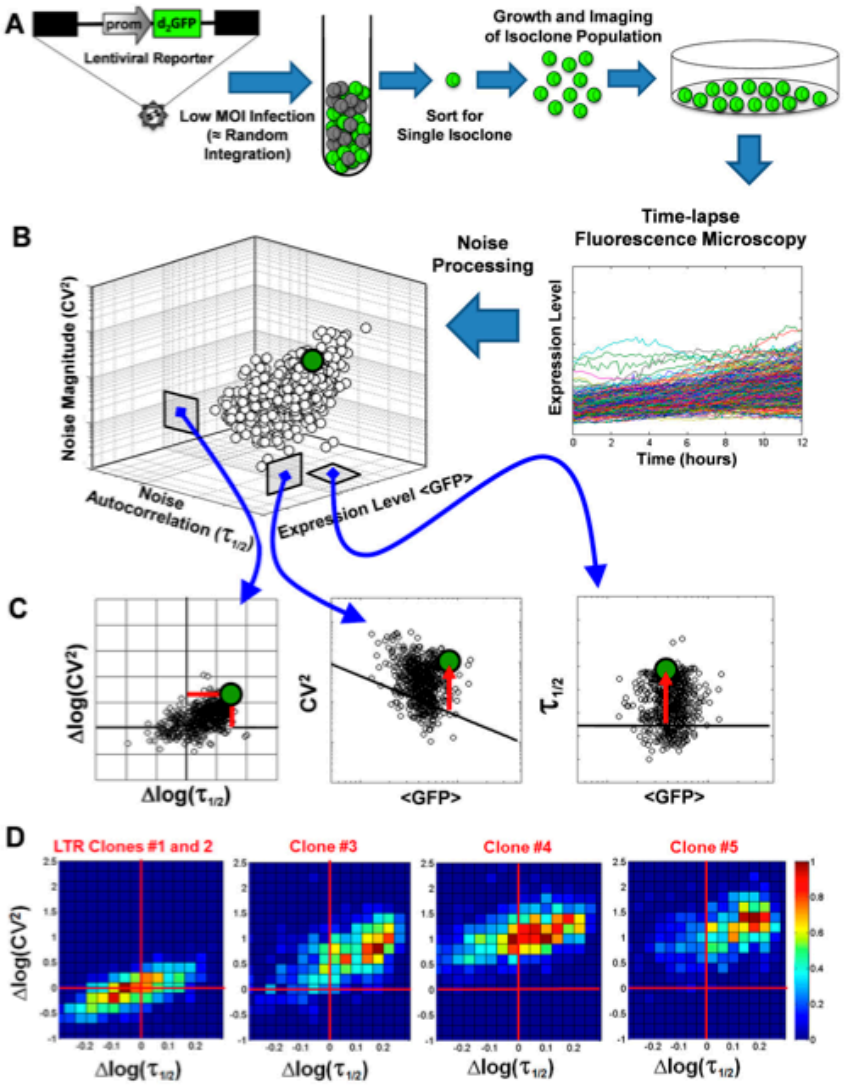

Figure 34. Extracting transcriptional parameters from the noise space. In individual isoclones, burst dynamics vary with genomic location.

(A) Cells are infected with a lentiviral vector expressing a 2-h half-life GFP reporter (d2GFP) at a low multiplicity of infection (moi) to ensure a single semirandom integration in each cell. Individual single cells are isolated, grown (creating isoclone populations), and imaged by time-lapse fluorescence microscopy. (B and C) Single-cells are tracked for 12–18 h, and an individual cell's mean expression level, variance (σ^2), and autocorrelation time ($\tau_{1/2}$) are extracted from the time trace (e.g., the green circle represents a single cell's noise space coordinate). A constitutive model of gene expression that displays abundance dependence (bold red arrows from black model lines) was used to normalize each cell's noise magnitude (CV^2) and autocorrelation ($\tau_{1/2}$). The normalized noise magnitudes and autocorrelations are plotted in a $\Delta\log CV^2$, $\Delta\log\tau_{1/2}$ noise map (Left). (D) Consistent shifts to the Upper Right quadrant in $\Delta\log CV^2$, $\Delta\log\tau_{1/2}$ space observed for three LTR isoclones (clones 3, 4, and 5), are indicative of transcriptional bursting relative to the least bursty isoclones (clones 1 and 2). Bursting dynamics varies between different clones as evidenced by shifts in both noise

autocorrelation and magnitude. The isoclonal signature is taken from 18 h trajectories of 400 cells. Originally published as Figure 2 in Dar et al., PNAS, 2012. Roy Dar created this figure. Brandon Razooky collected the data for this figure.

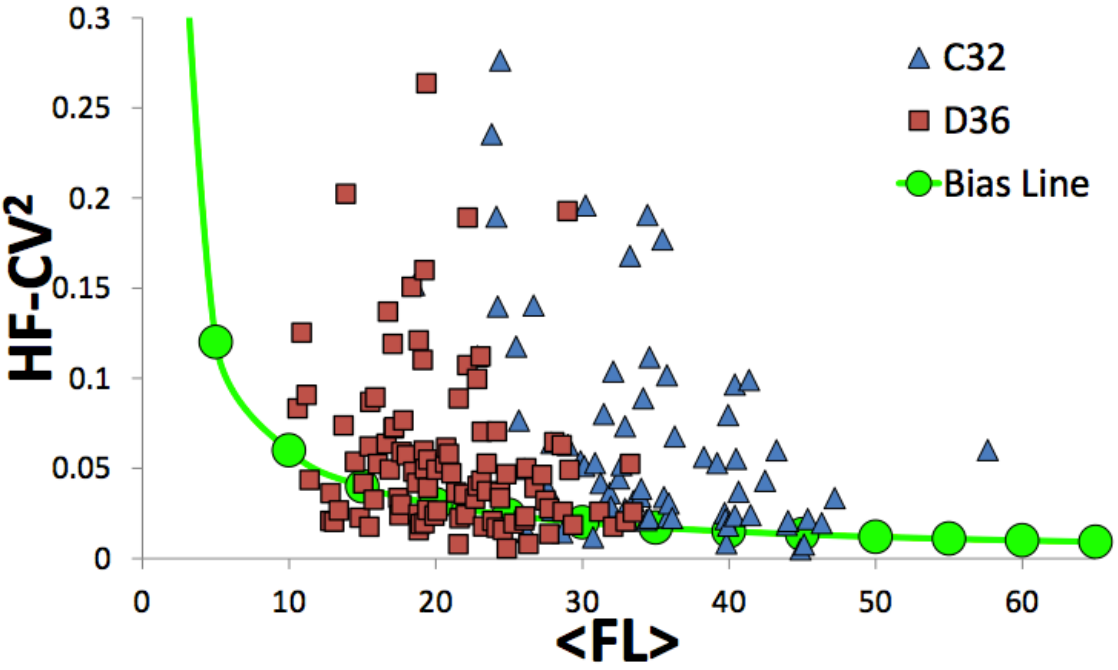

Figure 35. Least bursty isoclones to determine the bias vector.

HF-CV2 vs. average fluorescence level for clones C32 (58 cells) and D36 (87 cells). From these measurements the *CV*2 component of the bias vector (green line) was found as 0.6/<*fl*>. Originally published in supplemental information in Dar et al., PNAS, 2012. Roy Dar created this figure. Brandon Razooky collected the data for this figure.

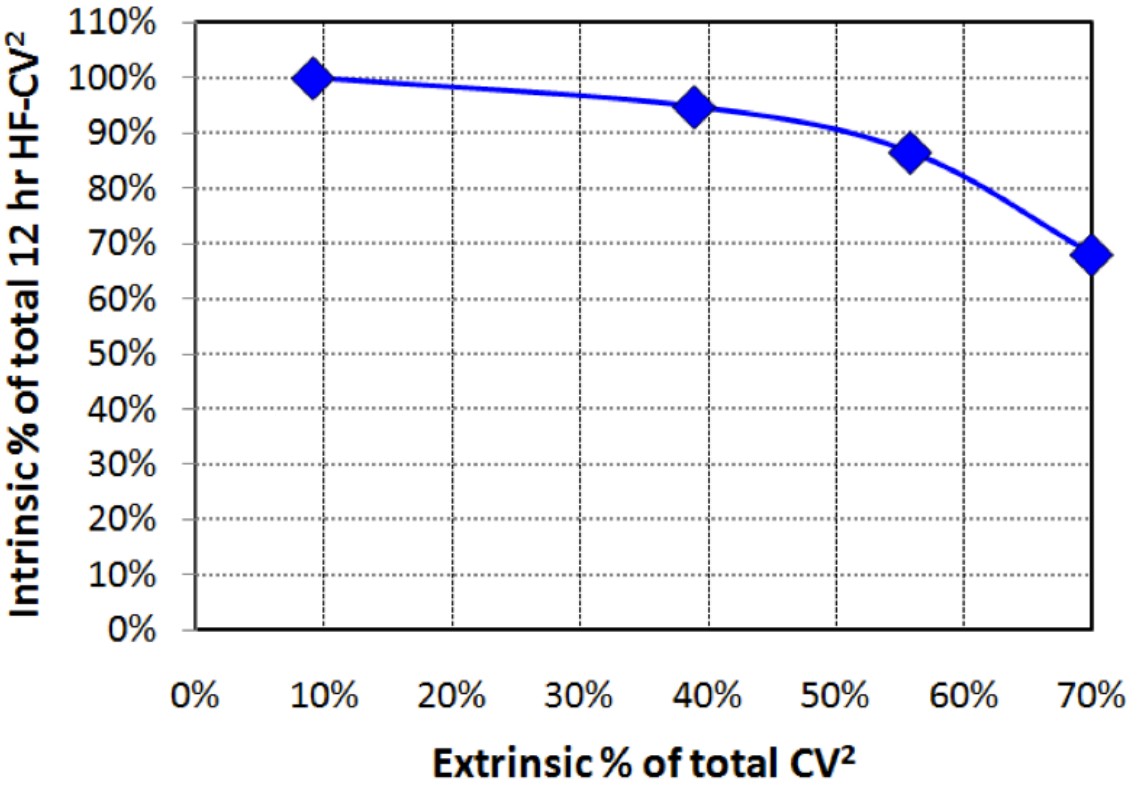

Figure 36. HF-processing focuses on intrinsic noise and filters out extrinsic noise.

Using the described intrinsic/extrinsic noise simulation, we can estimate how much noise magnitude is filtered out or emphasized with the 12-hHF-noise processing. For a large range of extrinsic noise contribution, intrinsic noise contribution is enhanced ~1.1-2.3 times of the total noise, while extrinsic noise is de-emphasized (filtered) down (e.g. 55% extrinsic of total noise filters down to ~15% of total noise, 40% of total to 5% of total, and so on). Originally published in supplemental information in Dar et al., PNAS, 2012. Roy Dar created this figure.

Importantly, non-transcriptional phenomena that affect noise behavior (e.g., protein and mRNA lifetimes, GFP maturation, extrinsic noise) were already embedded in the noise of these isoclonal populations, specifically the "most Poissonian" reference clones, thereby precluding the possibility that noise-map shifts were due to these non-transcriptional phenomena. In addition, the high-frequency noise-processing technique used minimizes extrinsic noise effects (Fig. 36 and (Weinberger et al., 2008b)). Since it is likely that these two isoclones were somewhat bursty in their expression (Fig. 37), comparisons to these isoclones represented a highly conservative assay for bursty expression. Nevertheless, the results showed that even in a small panel of five clones, there were marked changes in both noise magnitude and autocorrelation time of ~1.5-fold in normalized space (Fig. 34D). Clones #4 and #5 had significant changes in autocorrelation time with smaller differences in noise magnitude. These differences in noise magnitude were validated against conventional flow cytometry measurements (Fig. 38). The agreement between data from the 12–18 h microscopy experiments and flow cytometry demonstrates that this CV^2-versus-$\tau_{1/2}$ analysis has the fidelity to differentiate transcriptional dynamics between different isoclones. Importantly, when only noise magnitude of the isoclones is analyzed (noise magnitude is the only noise measure that can be obtained from flow cytometry or mRNA FISH) the clones are predicted to only have different burst frequencies. However, the differences in autocorrelation time and noise magnitude suggest that both the burst frequency and the burst size vary between the isoclonal populations (Fig. 34D).

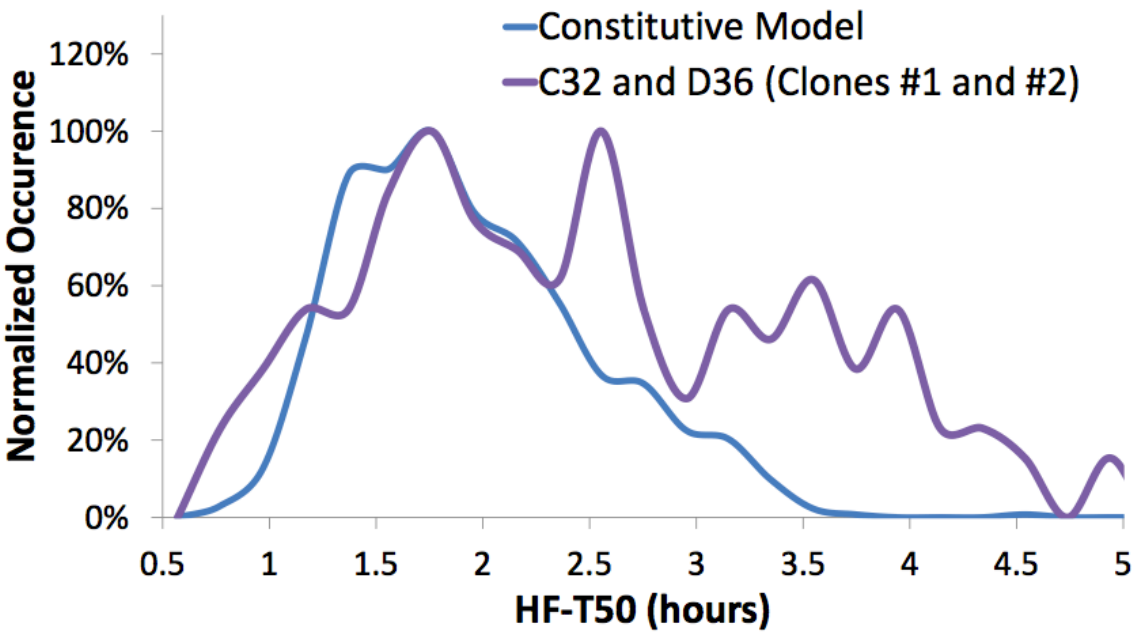

Figure 37. Isoclones with lowest correlation time to determine correlation bias vector.

The distribution of HF-T50s measured for clone C32 80%and D36 (purple). The bias vector value of HF-T50 was selected as 1.92 hours. The blue line shows the simulated HF-T50 distribution for constitutive expression and HF-T50 = 1.92 h, which is seen to fit well with the lower mode of the C32 and D36 HF-T50 distribution. The higher HF-T50 peaks in the isoclone distribution (purple) are indicative of some transcriptional bursting in these clones. Originally published in supplemental information in Dar et al., PNAS 2012. Roy Dar created this figure. Brandon Razooky collected the data for this figure.

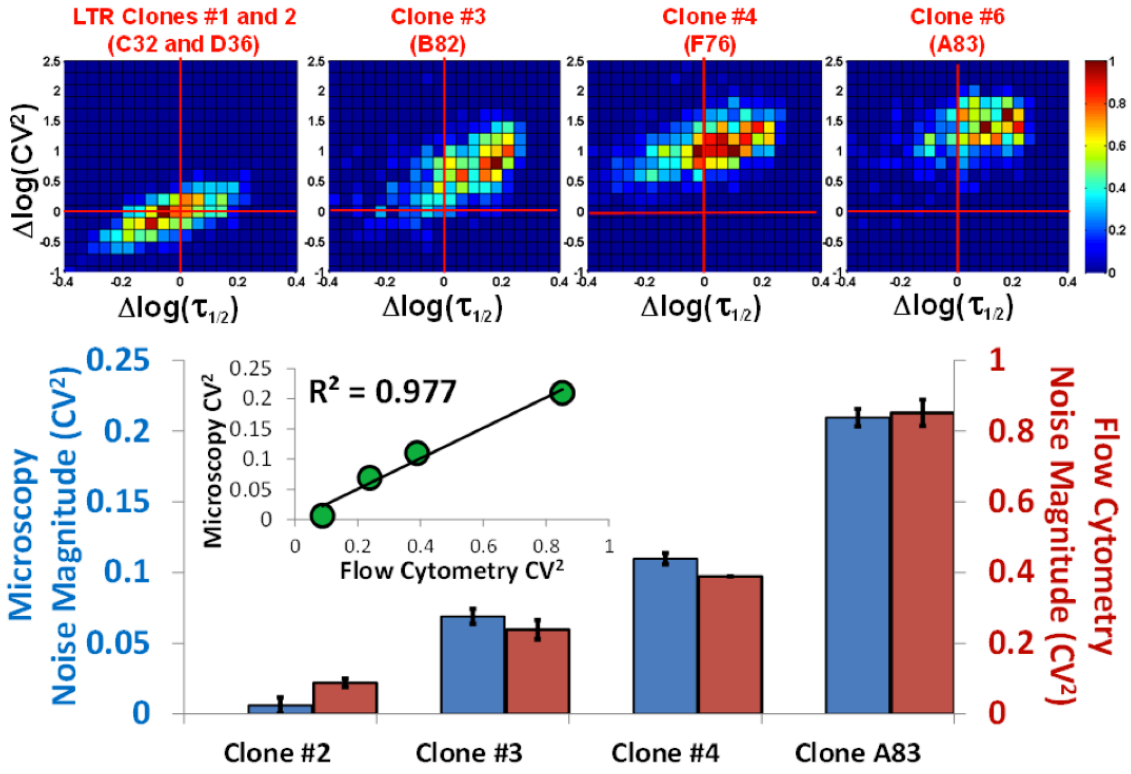

Figure 38. Benchmarking and validation of microscopy noise measurements with conventional flow cytometry.

Top: microscopy noise maps for four isoclones from 18-h microscopy experiments. Bottom: Comparison of CV^2 from microscopy (blue) to flow cytometry for the same clones (red). Inset: comparison of microscopy to flow cytometry CV's show measurements are consistent with one another to within a constant. High- frequency processing decreases noise magnitude from flow values consistently by ~0.25 for 18 h imaging durations. Originally published in supplemental information in Dar et al., PNAS, 2012. Roy Dar created this figure. Brandon Razooky collected the data for this figure.

2c. Measuring noise in thousands of integrations sites in a single experiment.

We next extended the analysis to image polyclonal populations—consisting of thousands of integration sites (Fig. 39A)—to globally apply the analytical framework of noise space to screen for constitutive versus bursty expression across the human genome. We analyzed more than 8,000 distinct genomic loci with three different promoters

integrated throughout the genome by imaging cells for 12–18 h (Fig. 3B). To control for LTR-specific or vector-specific artifacts, we also tested self-inactivating lentiviral vectors that encode promoters for either human elongation factor 1α (EF1A) or human ubiquitin C (UBC) that in turn drive d_2GFP. UBC and EF1A are essential cellular housekeeping genes: UBC promotes the ubiquitinization cascade by marking proteins for proteosomal degradation, and EF1A promotes the GTP-dependent binding of an aminoacyl-tRNA to ribosomes. UBC and EF1A are among the most abundant proteins in eukaryotic cells, and their promoters exhibit robust high-level expression across integration sites in different cell types (Kim et al., 1990; Ramezani et al., 2000). The constitutive expression origin of the CV^2-versus-$\tau_{1/2}$ noise map previously determined for the most-Poissonian isoclones (Fig. 2D, and SI text) was compared to the 8,000 loci. Virtually all examined genomic loci exhibit noise-map shifts to the upper right (Fig. 3B), indicating significant bursting in gene expression at virtually all genomic loci for the LTR and the strong human promoters. Similar widespread bursting was also observed for all three promoters in the THP-1 human monocyte cell line (Fig. S2). Both synchronized and unsynchronized cells exhibit similar shifts in noise space (Fig. S3), indicating that transcriptional bursts appear to be common throughout the cell cycle, as reported (Harper et al., 2011).

The UBC and EF1A promoters showed markedly lower CV^2s than the LTR promoter (Fig. 3B and Fig. S4), which was consistent with our previous study (Singh et al., 2010b) showing that the LTR promoter displays relatively higher levels of noise than other eukaryotic promoters in yeast (Bar-Even et al., 2006b; Newman et al., 2006). This shift in noise magnitude is consistent with the well-known transcriptional elongation "stall"

that characterizes LTR expression (Kao et al., 1987). This stall results in delayed switching to the transcriptional "ON" state in a two-state transcription model (Singh et al., 2010b) and predicts that noise frequency (not only noise magnitude) (Sakane et al., 2011) is modulated in different genomic or chromatin environments. In agreement with this prediction, the distribution of points in the LTR noise map indicates significant differences in kinetics from the UBC and EF1A noise maps, and an alternate representation with centroids of the noise-map distribution (with error bars) can be used to conveniently visualize these differences (Fig. 40) (Austin et al., 2006).

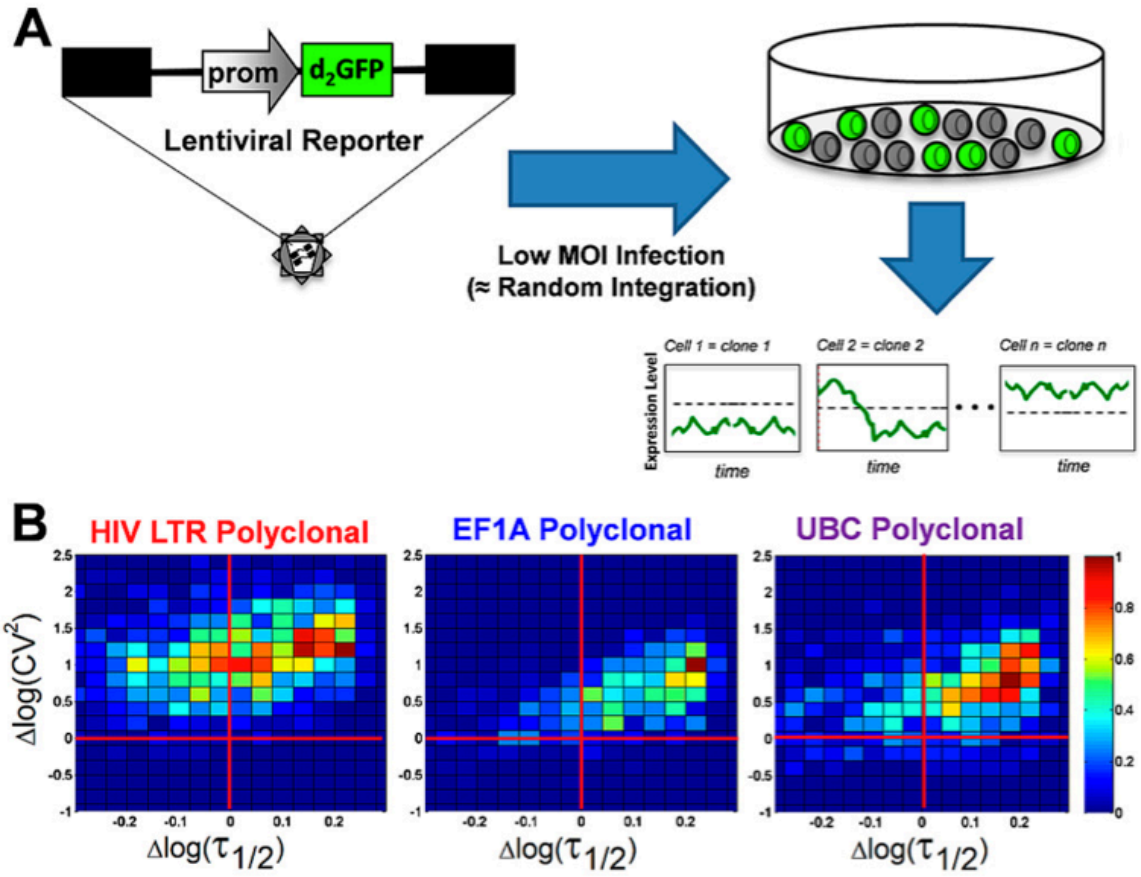

Figure 39. Episodic-bursty expression dominates across the human genome.

(A) To create the polyclonal population, cells are infected with a lentiviral vector expressing d2GFP so that each cell represents a unique clone harboring a single semirandom integration of reporter. (B) Resultant noise maps for over 8,000 individual cell trajectories for the HIV-1 LTR promoter, EF1A promoter, and UBC promoter. The

constitutive origin is derived from Fig. 34D (18 h). Originally published as Figure 3 in Dar et al., PNAS, 2012. Roy Dar created this figure. Brandon Razooky collected the data for this figure.

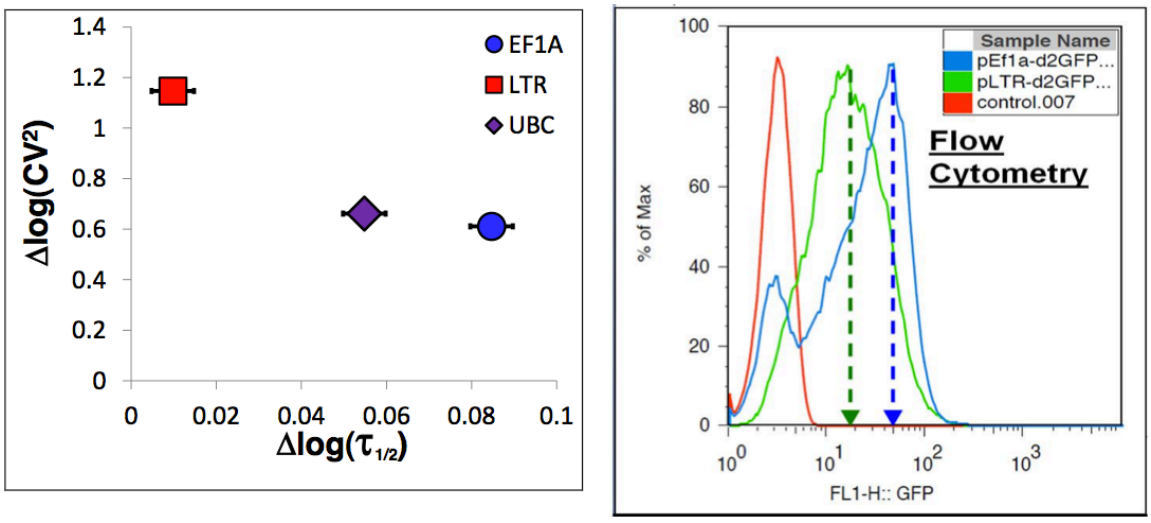

Figure 40. The HIV-1 LTR exhibits a greater noise-magnitude shift compared to EF-1α and UBC promoters.

(Left) Noise-map centroids for the polyclonal populations of LTR, EF-1α and UBC promoters determined from noise maps in Figure 3 of the main text. The LTR promoter has a higher noise magnitude and decreased correlation shift. The error bars show ± 1 sigma uncertainty in the centroid positions as determined by the number of cells in the sample. Noise-map centroids are determined from the base-line in Figure 2D (Clones 1 and 2), (Right) Flow cytometry verification showing distributions for EF1A-d2G (blue) and LTR-d2G (green). $CV_{LTR} > CV_{EF1A}$. The stronger EF1A promoter displays increased mean expression compared to the HIV-1 LTR. Originally published in supplemental information in Dar et al., PNAS, 2012. Brandon Razooky and Roy Dar created this figure.

2d. Transcriptional bursting is the dominant mode of transcription throughout the human genome.

Given conflicting reports on whether burst frequency varies with genomic location (Lo et al., 2012; Singh et al., 2010b; Skupsky et al., 2010; Suter et al., 2011a), we next determined if transcriptional burst size, burst frequency, or both changed across the genome (Fig. 41A). As mentioned above, transcriptional bursting can be quantified by a two-state model of transcription (Kepler and Elston, 2001a; Simpson et al., 2004b) in

which switching between the two states occurs at rates k_{on} and k_{off}, and transcription only occurs in the 'on' state with a rate k_m (Eqs. 1–3, Materials and Methods). The burst size, or number of mRNAs generated per activity pulse, is typically defined as k_m/k_{off} and, in the limit of $k_{off} >> k_{on}$, the burst frequency is defined as k_{on} (Fig. 41A) (Singh et al., 2010b). To directly test if transcriptional burst frequency changes across genome location, we analyzed the polyclonal 3-D noise-space data to fit values for k_m, k_{off}, and k_{on}. Polyclonal trajectories are sub-clustered into groupings of ~60 cells, so that each sub-cluster represents cells in a specific range of gene-expression levels (Fig. 42), and average noise autocorrelation is calculated for each sub-cluster by autocorrelation analysis (Weinberger et al., 2008b). To validate this sub-clustering approach, we compared the CV^2s from sub-clustered polyclonal trajectories to traditional flow cytometry data from isoclonal populations and found that the CV^2 patterns from sub-clustered data were coincident with isoclonal flow cytometry CV^2 patterns (Fig. 43). Thus, sub-clustering produces data that directly parallel the conventional noise processing of flow cytometry measurements for isoclonal populations, where each polyclonal sub-cluster corresponds to an isoclonal population in terms of average expression level (Fig. 43). Strikingly, the sub-cluster data demonstrate that autocorrelation time first increases with increasing expression and once an expression threshold is reached (gray line, Fig. 41B), autocorrelation time decreases as expression level increases (Fig. 41B). This pattern of concavity is inconsistent with constant burst frequency (i.e., constant k_{on}) across genomic locations and provides a direct measurement of k_{on} and k_{off} changes (independent of k_m (Cox et al., 2008b)) across genomic loci.

Conventional approaches to quantify transcriptional burst kinetics utilize the CV^2-versus-<GFP> plane of noise space (Bar-Even et al., 2006b; Newman et al., 2006), and the polyclonal data from Fig. 39 can also be analyzed on the CV^2-versus-<GFP> plane of noise space. This CV^2-versus-<GFP> analysis showed a strong initial decrease in CV^2 at low expression levels, then when an expression threshold is reached (gray line, Fig. 41C) the CV^2 levels off at higher expression levels (Fig. 41C). However, the CV^2-versus-<GFP> plane was insufficient to uniquely parameterize the two-state model; since burst size couples k_m and k_{off}, it was only through the $\tau_{1/2}$ measurement (which are not influenced by k_m) that the two parameters could be differentiated from one another. Note, that the gray line in Fig. 41B and Fig. 41C correspond to the same expression threshold.

2e. Burst size and frequency are equally modulated.

Fitting of the two-state model in the polyclonal 3-D noise map space showed a strong initial increase in burst frequency at low expression levels while burst size remained almost constant (Fig. 41D and 40E). Upon reaching a threshold expression level (gray vertical line in Fig. 41D), a switch in burst dynamics occurred, and burst size increased, while burst frequency remained constant (Fig. 41D and 40E). The fold change in transcriptional burst size and burst frequency values revealed that both vary equally across genomic loci (Figs. 40F and 42). In addition, the measured burst-size range predicted an average mRNA level of 110 molecules per cell, which is consistent with Singh *et al.* (Singh et al., In Review, 2012), and with previous measurements (Raj et al., 2006). The success of fitting the 3D noise space was reflected in the close agreement between a simulated autocorrelation curve and the experimental trend (Fig. 41B) with the

fit model parameters. This showed that our assumed two-state model is sufficient to describe the measured system. Much like the LTR, UBC and EF1A displayed similar fold changes in burst size and frequency and exhibited a similar pattern of increasing burst frequency, followed by increasing burst size (Fig. 41G). These data indicated that integration site influences burst kinetics, irrespective of promoter type (i.e. *cis* sequence). However, UBC and EF1A exhibit almost constant $\tau_{1/2}$ at the highest expression levels indicating increases in only k_m at these levels (Fig. 44). Interestingly, these two strong promoters individually span the range of burst frequencies recently reported for a variety of mammalian genes (Suter et al., 2011a) (Fig. 44), while the LTR functions at much lower burst frequencies (Fig. 41).

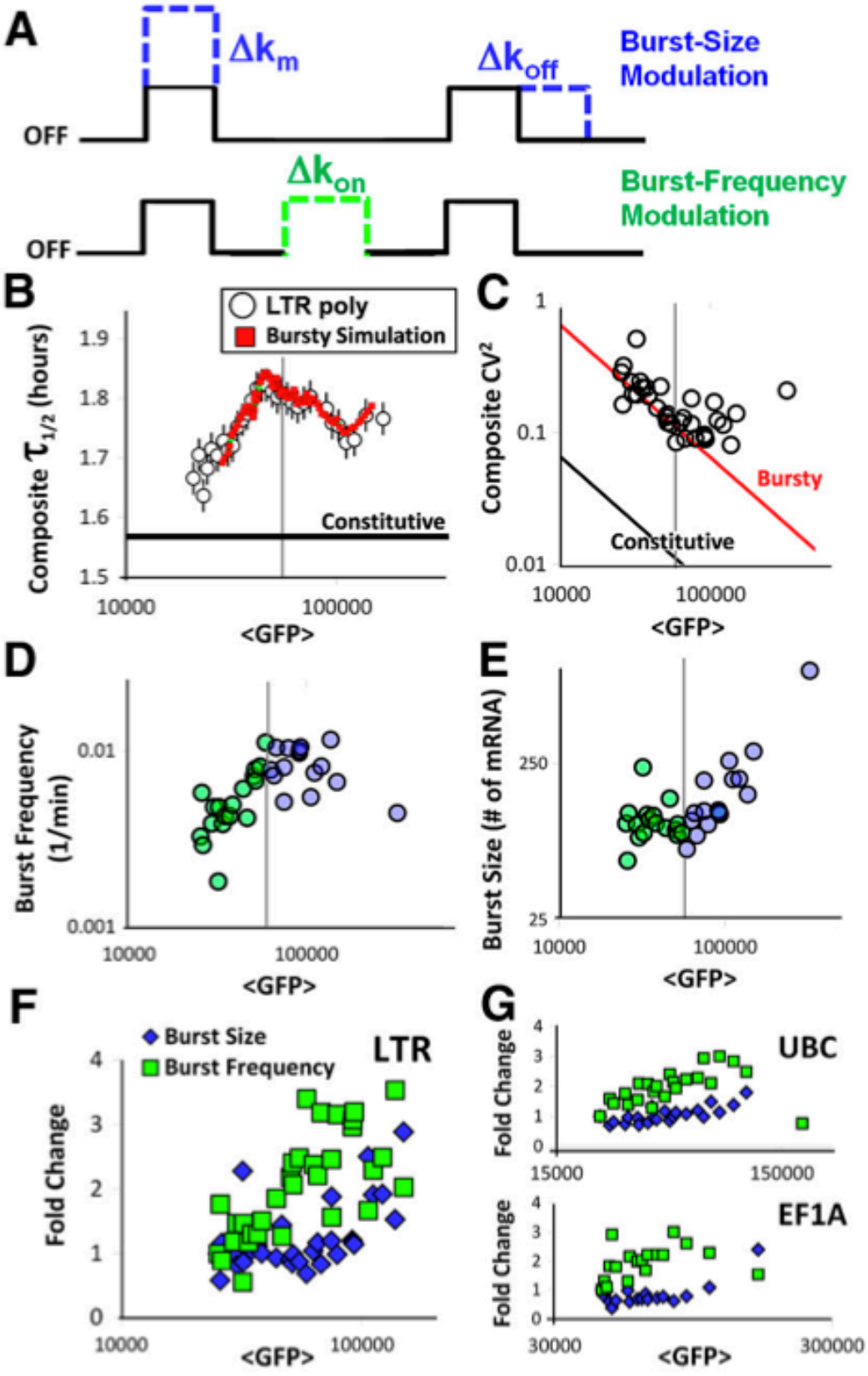
A
Δk_m
Δk_off
Burst-Size Modulation
OFF
Δk_on
Burst-Frequency Modulation
OFF
B
LTR poly
Bursty Simulation
Composite τ1/2 (hours)
1.9
1.8
1.7
1.6
1.5
Constitutive
10000
100000
<GFP>
C
Composite CV²
1
0.1
0.01
Bursty
Constitutive
10000
100000
<GFP>
D
Burst Frequency (1/min)
0.01
0.001
10000
100000
<GFP>
E
Burst Size (# of mRNA)
250
25
10000
100000
<GFP>
F
Burst Size
Burst Frequency
LTR
Fold Change
4
3
2
1
0
10000
100000
<GFP>
G
UBC
Fold Change
4
3
2
1
0
15000
150000
EF1A
Fold Change
4
3
2
1
0
30000
300000
<GFP>

Figure 41. Transcriptional burst frequency and burst size vary equally across the genome and are strongly dependent on expression level.

(A) Schematic of the two-state model of transcriptional bursting, where the promoter switches between ON and OFF states at rates kon and koff and transcribes at rate km in the ON state. Transcriptional dynamics are modulated through changes in burst size, burst frequency, or both. (B) Noise autocorrelation, noise magnitude (C), burst frequency (D), and burst size (E) versus abundance for polyclonal subclusters of 2,000 12-h Ld2G single-cell trajectories. Low and high abundance domains are separated by a solid gray threshold line, which indicates the changes in the trends of noise autocorrelation, noise magnitude, and hence burst size and burst frequency is observed. (F and G) As a function of <GFP>, fold changes in burst size and frequency are comparable, with an initial increase of frequency in all promoters investigated. Originally published as Figure 4 in Dar et al., PNAS, 2012. Roy Dar created this figure. Brandon Razooky collected the data for this figure.

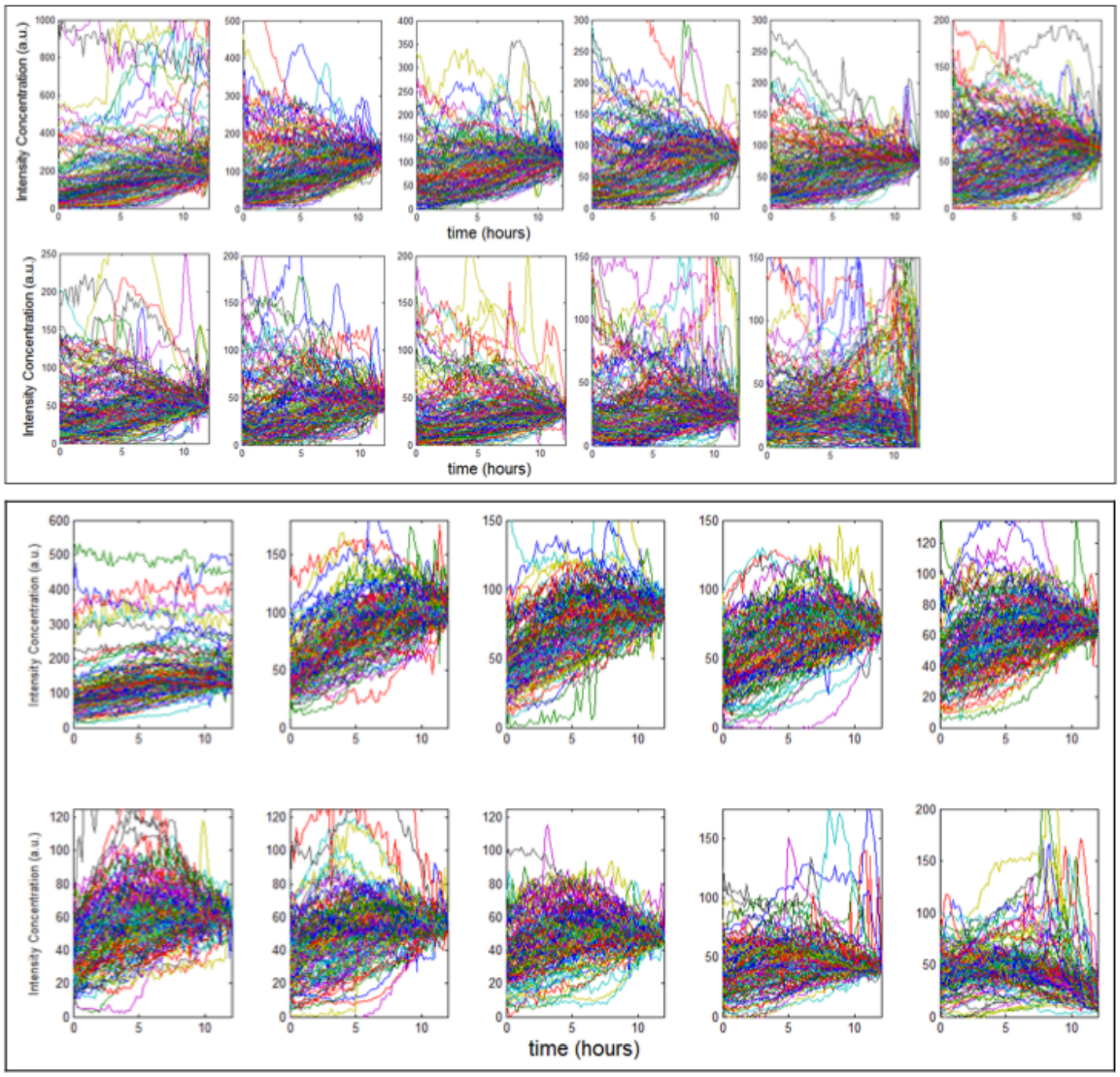

Figure 42. Sub-clustered GFP intensity trajectories of polyclonal populations.

(Upper) LTR polyclonal cells separated into 11 clusters of 170 cells by their last intensity values. Sub-clusters decrease in intensity from the upper-left to lower-right panels. (Lower) 10 sub-clusters of Ef1A d2GFP poly clustered in the same way as the LTR polyclonal cells. The sub-clustering of the UBC populations is not shown. Originally published in supplemental information in Dar et al., PNAS, 2012. Roy Dar created this figure. Brandon Razooky collected the data for this figure.

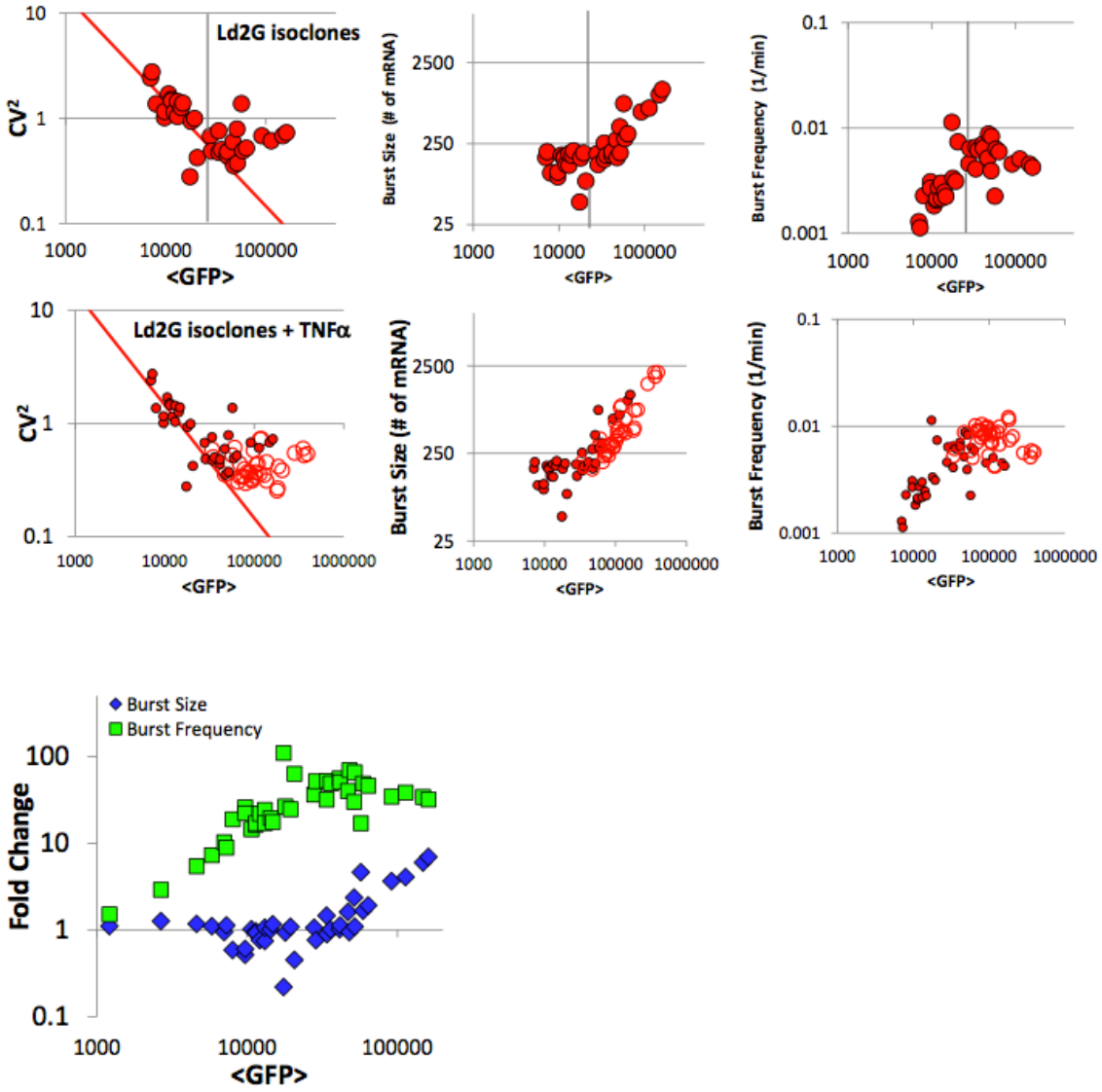

Figure 43. LTR isoclone yield similar noise and burst trends to polyclonal sub-cluster processing when measured using traditional flow cytometry methods.

35 Ld2G isoclones were measured for their fluorescence distributions by flow cytometry. The measurements of a range of LTR isoclones yielded noise and burst dynamic trends similar to the polyclonal microscopy data. The coefficient of variation squared and mean levels were used to quantify gene expression noise magnitude in the LTR isoclones, which generally land along a single burst model line inversely proportional to the fluorescence abundance. Here, deviations from the trend are only observed at high abundances. At low abundances, the burst size remains constant, while the burst frequency increases (left of vertical gray line). At higher abundances, the BF hits an upper bound and plateaus, and abundance increases through increase in burst size. Upon TNF-alpha addition, shifts along the noise and burst trends identical to the polyclonal sub-clusters are observed. Originally published in supplemental information in Dar et al., PNAS, 2012. Roy Dar created this figure. Brandon Razooky collected the data for this figure.

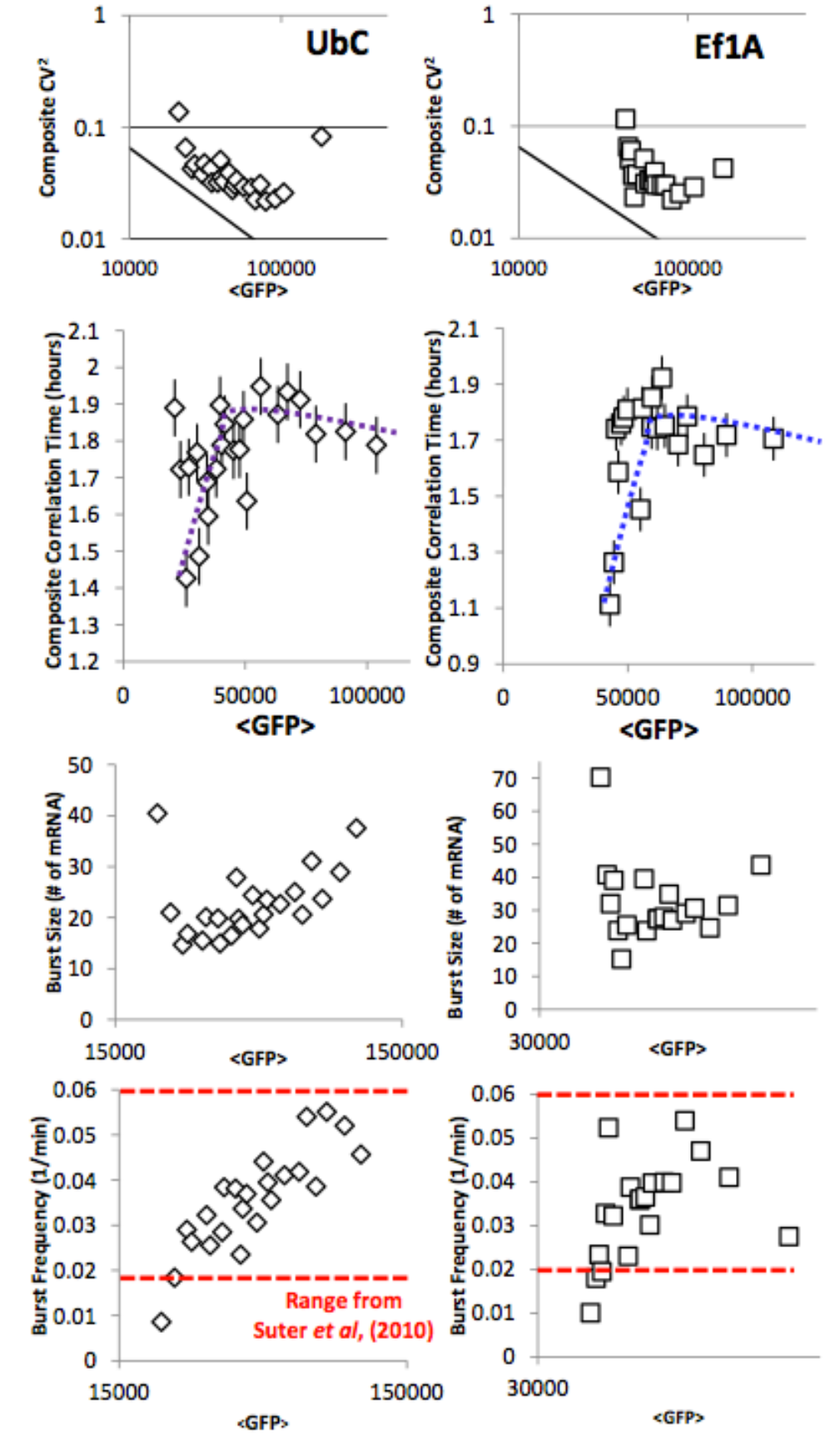
UbC
Ef1A
Composite CV²
1
0.1
0.01
10000
100000
<GFP>
Composite Correlation Time (hours)
2.1
2
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
0.9
0
50000
100000
Burst Size (# of mRNA)
50
40
30
20
10
0
70
60
15000
150000
30000
Burst Frequency (1/min)
0.06
0.05
0.04
0.03
0.02
0.01
Range from
Suter et al, (2010)

Figure 44. Two housekeeping promoters display increasing episodic expression with increasing intensity levels.

The 2-state model and burst size expressions predict burst size = 0 for constitutive for continuous gene expression. BothUBC and EF1A have markedly less BS than the LTR. Interestingly, the burst frequency range for these strong promoters are consistent with the range of burst frequencies observed by Suter et al., and further emphasizes that each individually span the kinetic range with varying integration site. Originally published in supplemental information in Dar et al., PNAS, 2012. Roy Dar created this figure. Brandon Razooky collected the data for this figure.

2f. Transcriptional activators alter bursting dynamics along the same burst trends.

To test whether exogenous stimuli explore the same burst trends (Fig. 4), transcription was perturbed with transcriptional activators, including the cell-signaling molecule tumor necrosis factor α (TNF). TNF enhances expression by stimulating recruitment of a p50-RelA heterodimer to nuclear factor κB (NFκB) binding sites (Vallabhapurapu and Karin, 2009), and the HIV-1 LTR encodes multiple NFκB binding sites and is potently activated by TNF (Jordan et al., 2001). We previously reported in a few isoclonal populations that TNF only changes burst frequency of the LTR while conserving burst size (Singh et al., 2010b), and were interested to see how widespread this phenomena was across the genome. The $\tau_{1/2}$-versus-<GFP> analysis of TNF stimulation showed a significant decrease in $\tau_{1/2}$ with increasing expression level. $\tau_{1/2}$ decrease with increasing abundance was a direct indication of kinetic changes and demonstrated that increasing expression level cannot be explained solely by modulations of k_m (Cox et al., 2008b; Simpson et al., 2004b).

Fitting of the 3-D noise map upon TNF induction demonstrated that both burst frequency and size significantly increase as expression levels increase with burst frequency increasing at low expression levels and burst size increasing at higher

expression levels (Fig. 45C-D). Interestingly, there appeared to be a threshold in expression level, above which k_{on} plateaus to values observed before adding TNF, and k_{off} appeared to decrease. The data and analysis agreed with results obtained from conventional flow cytometry measurements of 35 isoclonal populations (Fig. 43). Overall, these results suggested that TNF induces transcription from the LTR along the same burst trends (Fig. 45C-D), and the use of Trichostatin A (TSA), although acting through a different mechanism than TNF for induction (VanLint et al., 1996), corroborates this observation. This observed decrease in k_{off} with TNF induction, leading to extended duration of bursts, is consistent with the reported inhibition of p50-HDAC1 repressive-complex formation at LTR NFκB sites by p50/RelA heterodimers (Hayden and Ghosh, 2004b). The successful formation of HDAC1 leads to weakened recruitment of RNA polymerase II, transcriptional initiation(Williams et al., 2006b), and increases in k_{on} are expected by known recruitment of RNA polymerase to the LTR promoter by NFκB (Barboric et al., 2001; West et al., 2001). Fitted parameter estimates of LTR residency time in the presence of TNF were used to represent an average over the first 12 h of stimulation given the dynamic non-linear nature of the NFκB response (Hoffmann et al., 2002; Williams et al., 2007a). Collectively, these results enable estimation of LTR residency time in the transcriptional ON and OFF states and show that TNF extends duration in the ON state up to eightfold (Fig. 45E).

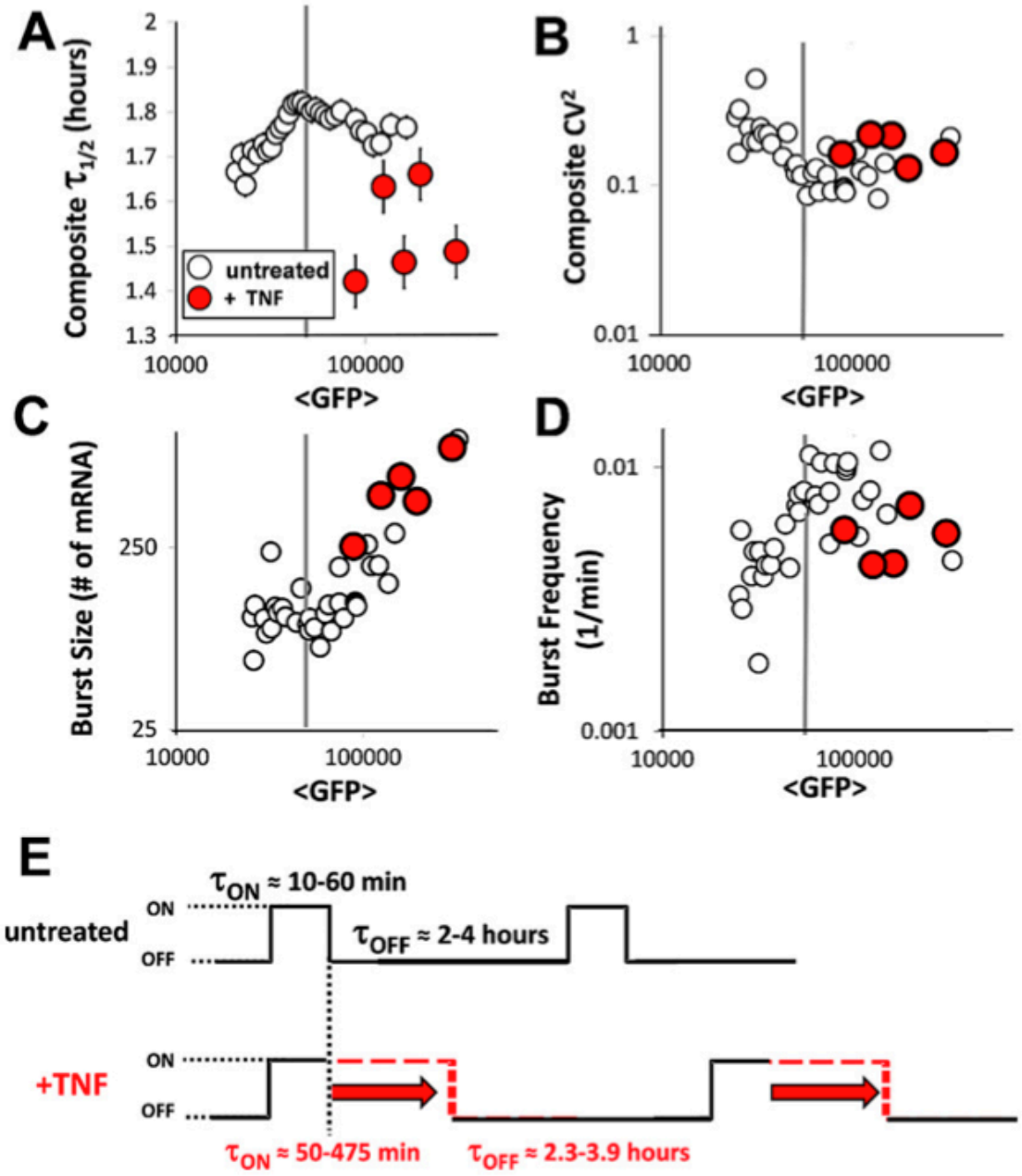

Figure 45. Transcriptional burst size and frequency are altered by transcriptional activators.

(A–D) TNF addition (filled red circles) shifts the measured integration sites to the higher abundance and burst dynamic domain along the nondrug curve (empty circles). Large autocorrelation shifts implicate changes in burst kinetics. (E) Estimated residence times in the active (ON) and inactive (OFF) states. Originally published as Figure 5 in Dar et al., PNAS, 2012. Roy Dar created this figure. Brandon Razooky collected the data for this figure.

3. Discussion

The analysis of noise space presented here provides a high-throughput method to dynamically profile gene-regulatory mechanisms and the effects of perturbations on gene expression. A significant methodological advantage of analyzing three dimensions of noise space is the ability to more accurately constrain two-state transcriptional burst models and the polyclonal nature of the approach enables 'shotgun' mapping of gene regulation dynamics on a genome-wide scale.

The resulting genome-wide data demonstrate that constitutive transcription is rare across the human genome. Instead, the overwhelming majority of human genomic loci appeared to stochastically fire in episodic bursts. Analysis of noise space demonstrated that both transcriptional burst frequency and burst size vary in roughly equal degree across the human genome (Fig. 41D-G). Intriguingly, there appeared to be a threshold expression level below which integrations modulate only transcriptional burst frequency and above which only burst size is modulated (Fig. 45B-G and 42-43). This transition could result from recently reported refractory periods inherent to bursting kinetics (Suter et al., 2011a; Suter et al., 2011b). Burst frequency can be increased at loci where transcriptional bursts are infrequent, but as frequency increases, the refractory period temporally precludes further increases in frequency. Therefore, once this frequency ceiling is reached, the only way to increase expression is to extend the duration of each burst.

As proposed (Cai et al., 2008), widespread episodic bursting may allow limited transcriptional resources within the cell to be efficiently allocated to achieve high-level transcription across large numbers of loci. This efficient allocation of resources may be

the biological analog of 'time-domain multiplexing' approaches used to efficiently transmit data in signal processing applications.

4. Materials and Methods

4a. Lentiviral Vectors

Lentiviral vectors were cloned as described (Weinberger et al., 2005) and used to infect 5×10^5 Jurkat cells at a multiplicity of infection < 0.1, resulting in 25,000–50,000 infected cells each with a unique integration site. Cells were then sorted by FACS and fluorescently imaged on glass-bottom dishes in RPMI 1640 with 10% fetal calf serum and 1% penicillin-streptomycin.

4b. Imaging

Imaging was performed in humidified conditions at 37°C and 5% CO_2 for 12–24 h with a 40X (1.2 NA) oil-immersion objective on an Olympus DSU microscope equipped with an automated linear-encoded X-Y stage, as described (Weinberger et al., 2008b; Weinberger and Shenk, 2007). Image processing and cell tracking were performed in Matlab with an in-house algorithm (Weinberger et al., 2008b) and a single 12 h experiment could generate up to 1000 trajectories for analysis.

4c. Calculations

For each trajectory, noise autocorrelation ($\tau_{1/2}$) and noise magnitude (CV^2) were calculated using an established noise-processing algorithm (Austin et al., 2006; Weinberger et al., 2008b). A reported theory (Cox et al., 2008b; Simpson et al., 2004b) of

the two-state transcriptional bursting model yields analytical expressions for both the autocorrelation of the noise, $\tau_{1/2}$, and the noise magnitude.

Transcriptional burst dynamics are quantified by deriving analytical expressions for burst size (BS) and burst frequency (BF) with formulations from previous analyses (Kepler and Elston, 2001a; Simpson et al., 2004b; Singh et al., 2010b) and low promoter activity assumptions where $k_{off} >> k_{on}$, $k_{off} >> k_m$, $k_{off} >> g_p$, and $k_m >> (g_m + g_p)$:

$$CV^2 = \frac{C_1(1+BS)}{\langle p \rangle}, \ \langle p \rangle = \frac{BF*BS*k_p}{\gamma_m*\gamma_p} \qquad (40)$$

$$C_1 = \frac{k_p}{(\gamma_m+\gamma_p)} \approx \frac{k_p}{\gamma_p} \ or \ b \qquad (41)$$

$$BS = \frac{CV^2*\langle p \rangle}{C_1} - 1 \qquad (42)$$

$$BF = \frac{\langle p \rangle}{BS*C_1*C_2}, \quad C_2 = \frac{(\gamma_m+\gamma_p)}{\gamma_m*\gamma_p} \approx \frac{1}{\gamma_p} \qquad (43)$$

where BS is the burst size, BF is the burst frequency, k_m is the transcription rate, k_p is the translation rate, g_m and g_p are the mRNA and protein decay rates, respectively, <P> is the mean protein abundance, and b is the translational burst rate. <P>, or the mean number of GFP molecules in the measurements, is assumed to be directly proportional to <FL>, the mean fluorescence intensity. Equations 42 and 43 reveal that measurements of CV^2 and <FL> are sufficient to quantify burst size and burst frequency within a constant, which is only dependent on the translation rate and decay rates of mRNA and protein. Assuming these remain constant, while varying integration site or promoter sequence, an abundance dependent burst size and frequency trend can be directly resolved.

Chapter 8: Microwell device for tracking HIV gene-expression kinetics in primary cells

This chapter was published in Razooky, B.S., Gutierrez, E., Terry, V.H., Spina, C.A., Groisman, A., and Weinberger, L.S. (2012). Microwell devices with finger-like channels for long-term imaging of HIV-1 expression kinetics in primary human lymphocytes. Lab Chip *12*, 4305-4312. The article can be found on Pubmed; http://www.ncbi.nlm.nih.gov/pubmed/22976503

1. Introduction

Time-lapse microscopy is a powerful quantitative technique in modern biomedical science and is increasingly used to probe dynamic processes in individual cells(Locke and Elowitz, 2009). Time-lapse fluorescence microscopy has been critical to the study of stochastic cell-fate decisions by enabling single-cell analysis and mapping of the gene-regulatory circuits that control bacterial sporulation(Eldar et al., 2009), bacteriophage-λ lysis-vs.-lysogeny(Zeng et al., 2010), and development of antibiotic persistence(Balaban, 2011).

A similar cell-fate decision occurs in human immunodeficiency virus type 1 (HIV-1), the causative agent of AIDS, as the virus undergoes a stochastic transition between active replication and proviral latency (a long-lived viral dormancy state)(Han et al., 2007; Singh and Weinberger, 2009; Weinberger et al., 2005). In seroconverted patients, HIV-1 establishes a small reservoir of these latently infected cells. This reservoir is considered the largest obstacle thwarting eradication of HIV-1 from infected patients since

interruption of highly active antiretroviral therapy (HAART) allows for viral rebound from these reservoirs. Hence, mapping the genetic circuitry regulating HIV-1 latency is crucial to developing new therapies(Richman et al., 2009). Time-lapse fluorescence microscopy led to progress in mapping the HIV-1 latency circuit, making it possible to quantify single-cell viral-expression kinetics in immortalized tissue culture cell lines, such as Jurkat cells(Razooky and Weinberger, 2011; Weinberger et al., 2008a; Weinberger and Shenk, 2007). However, a major remaining challenge is to map the latency circuitry within primary $CD4^+$ T lymphocytes isolated from patients, as it is these cells that are the physiologically relevant target for HIV-1 replication and latency *in vivo*(Siliciano and Greene, 2011).

Single-cell time-lapse assays require tracking of individual cells many hours or even days, and to achieve statistical significance, large numbers of cells must be tracked in parallel, necessitating some degree of cell immobilization. High experimental throughput is relatively straightforward to achieve with adherent cells, which move relatively slowly, when plated at the bottom of a cell culture dish. However, for non-adherent or weakly adherent cells, such as T lymphocytes, which can be displaced by uncontrolled flow of the culture medium caused by convection or motion of the microscope stage, time-lapse microscopy becomes significantly more difficult. Uncontrolled displacements of non-adherent and weakly adherent cells can be prevented by using substrates with microwells (Faley et al., 2008; Gong et al., 2010; Guldevall et al., 2010; Han et al., 2010; Khademhosseini et al., 2004; Li et al., 2004; Love et al., 2006; Ogunniyi et al., 2009), microfluidic devices with semi-permeable barriers (weirs), which also enable time-controlled exchange of the medium(Deutsch et al., 2006; Kobel et al., 2010;

Schiffenbauer et al., 2009; Yang et al., 2011), or adhesive coatings, which have been successfully applied to immobilize Jurkat T cells for up to 30 hours (Razooky and Weinberger, 2011; Weinberger et al., 2005; Weinberger et al., 2008a).

However, long-term tracking of primary human $CD4^+$ T lymphocytes presents a set of specific challenges. When resting, $CD4^+$ T lymphocytes are non-adherent suspension cells. Nevertheless, to support HIV-1 infection, viral replication, and reactivation from latency, primary $CD4^+$ T lymphocytes must be activated, which is typically achieved by stimulation with small molecules (e.g., cytokines or anti-CD3 receptor cross-linking)(Han et al., 2007). As a result of this activation, T lymphocytes become weakly adherent, spontaneously attaching to and detaching from the substrate, and gain a capacity to rapidly migrate along the substrate while attached(Ridley et al., 2003; Volkov et al., 1998). Importantly, the adhesion molecules, which reliably immobilize cultured T lymphocytes, prove to be inadequate for activated primary T lymphocytes. With all the adhesive coatings we tested (including collagen, fibronectin, and Cell-Tak™), primary human T lymphocytes could only be immobilized and attached to the substrate for 2 – 3 hours at best (unpublished results). Furthermore, unlike strongly adherent cells, activated T lymphocytes do not form confluent monolayers and do not acquire easily distinguishable shapes, making their motion difficult to track over a two-dimensional region of a substrate. Time-lapse microscopy of HIV-1 gene-expression in infected lymphocytes was demonstrated before in regular cover glass-bottom dishes(Saez-Cirion et al., 2006). However, very few cells were tracked and the duration of tracking was only ~8 hrs(Saez-Cirion et al., 2006), much shorter than the 40 hrs duration of the life cycle of HIV-1 in an infected cell(Weinberger et al., 2008a).

Several microwell and microfluidic techniques of cell confinement and immobilization have been applied to experiments on lymphocytes. Arrays of microwells engraved on the surface of polydimethylsiloxane (PDMS) elastomer and sealed by glass slides have been used to analyze molecules secreted by individual lymphocytes(Gong et al., 2010; Han et al., 2010; Love et al., 2006; Ogunniyi et al., 2009; Thorslund et al., 2008). Arrays of sealed microwells have also enabled tracking of lymphocytes in short-term experiments (up to several hours)(Guldevall et al., 2010; Lindstrom et al., 2009; Ozawa et al., 2009; Schiffenbauer et al., 2009; Tokimitsu et al., 2007; Yamamura et al., 2005). However, a sealed microwell limits the access of cells to fresh medium, and a low cell density in a microwell, while postponing depletion of the medium, limits cell-cell interactions that are physiologically important for T lymphocytes. So, when tracking of T lymphocytes in sealed microwells was extended to 2 days, it resulted in considerable cell loss (Guldevall et al., 2010). Microfluidic devices have been used to capture and count human T lymphocytes on substrates with various coatings(Thorslund et al., 2008). A microfluidic device with bucket-like (weir) structures has been used for capture and imaging of non-adherent T cells by exposing them to a steady hydrodynamic drag(Di Carlo et al., 2006a; Di Carlo et al., 2006b; Faley et al., 2008). However, maintaining good viability of primary T cells required careful adjustment of the flow rate(Faley et al., 2008; Kobel et al., 2010). At an optimal flow rate, the 24 hrs viability was ~68% (and much lower at two-times faster or slower flow)(Faley et al., 2008), whereas experiments on HIV-1 latency require time-lapse microscopy over intervals longer than the HIV-1 life cycle of 40 hrs. Furthermore, while *in vivo* lymphocytes are exposed to hydrodynamic stress when circulating in the blood plasma, they migrate to lymph nodes within 30

minutes(Chen et al., 2002; von Andrian and Mempel, 2003). Therefore, for multi-day time-lapse microscopy of primary T lymphocytes, a static environment may provide a better emulation of the *in vivo* conditions than continuous perfusion.

To circumvent these technical problems, we developed, built, and tested microfabricated devices, which enable easy loading of primary T lymphocytes into arrays of narrow and shallow microchannels that are dead-ended (Fig. 46). These finger-shaped microchannels restrict spontaneous migration of lymphocytes to a nearly one-dimensional region, greatly facilitating their tracking. A single device has either 4 or 9 separate round wells accessible to a pipette for cell loading, making it possible to test 4 to 9 different cell lines or medium conditions in parallel. The devices were tested with activated primary $CD4^+$ T lymphocytes, resting primary $CD4^+$ T lymphocytes, and THP-1 monocyte-macrophage cells, which all maintained viability over multiple days. As a further test of the devices, we performed long-term time-lapse microscopy of T lymphocytes that were infected with HIV-1 derived lentiviruses and expressed green fluorescent protein (GFP). We were able to track the GFP expression of a large number of individual T lymphocytes for ~60 hours, 20 hours longer than the HIV-1 life cycle.

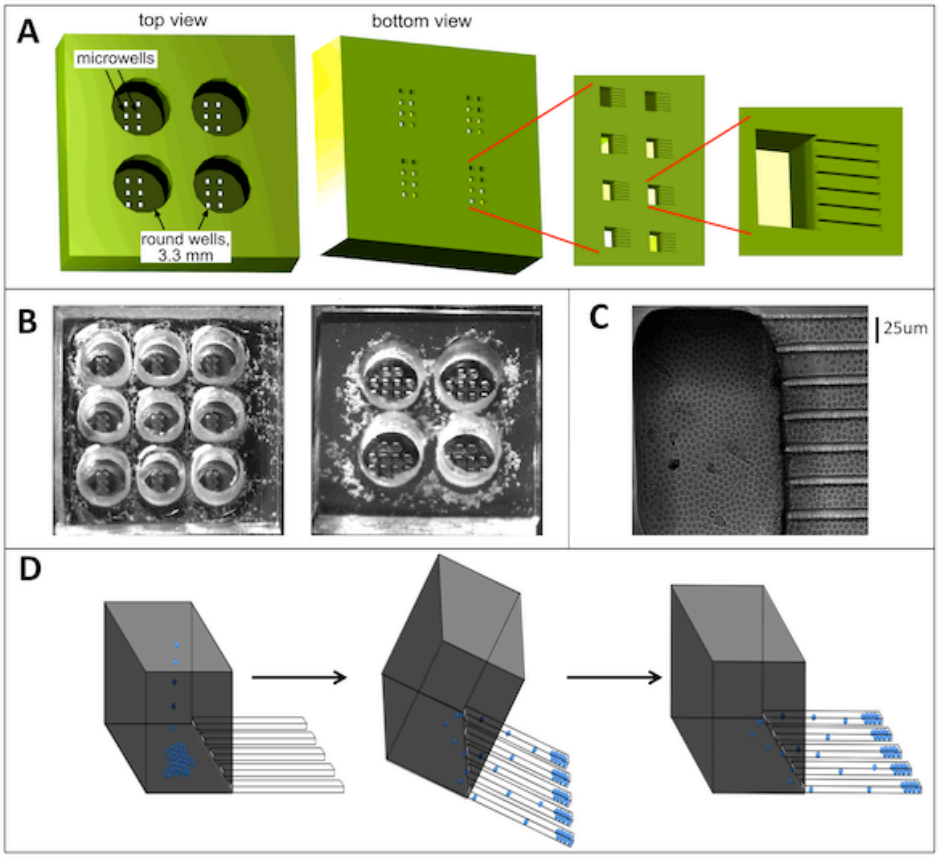

Figure 46. The microfabricated device.

(A) Schematic of a PDMS chip with a 2×2 array of 3.3 mm in diameter, 3 mm deep round wells. At the bottom of each round well there is a 2×4 array of 100×200 μm rectangular openings in a 100 μm thick layer of PDMS, forming microwells. At the bottom of each microwell there are finger-like horizontal extensions, 25×25 μm in cross-section and ~100-200 μm in length, forming cell-imaging channels. (B). Photographs of PDMS chips with 3×3 and 2×2 arrays of round wells. The microwells are visible at the bottom of the round wells. (C) Brightfield image of resting primary $CD4^+$ T lymphocytes loaded into a microwell and the adjacent finger-like channels. (D) Loading of cells into finger-like channels. First, cells are loaded into a round well and allowed to settle onto the bottoms of the microwells by gravity. Then the device is tilted by 45° with the finger-like channels pointing downward, making cells slide into the finger-like channels. Third, the normal orientation of the device is restored and it is carefully placed on the microscope stage. Originally published as Figure 1 in Razooky et al., Lab on a Chip, 2012. Brandon Razooky, Alex Groisman and Leor Weinberger created this Figure.

2. Experimental

2a. Design and operation of the devices

The finger-like microchannels, which are the structures that ultimately "collect" the T lymphocytes for long-term imaging, are 25×25 mm in cross-section and ~100-200 mm long. Seven parallel finger-like microchannels form a cluster connected to a single microwell. Each microwell is 200 mm in width, 100 mm in length, and 100 mm in depth (Fig. 46A). The bottom of the microchannels and microwell is formed by a regular #1.5 microscope cover glass, making cells in the microchannels accessible to high-resolution microscopy. The top of the microwell opens to the bottom of a round well, which is 3.3 mm in diameter and ~6 mm deep. The total volume of a round well is ~50 mL, and medium can be readily loaded into it and aspirated from it with a micropipette. A single round well has an array of 8 rectangular microwells at the bottom, and the microfabricated chips have round wells in either 3x3 (Fig. 46B, left) or 2x2 (Fig. 46B, right) square arrays, with each well connected to a total of 56 finger-like microchannels at the bottom.

The three-level hierarchal structure of the devices facilitates loading of T lymphocytes into the dead-ended finger-like channels, where the low ceiling and small width limit the motion of cells in the vertical and lateral directions, hindering their ability to crawl past each other (Fig. 46C). To load T lymphocytes in the finger-like channels, they are pipetted directly into a round well and allowed to settle by gravity, first reaching the bottom of the round well and then the bottoms of the microwells (Fig. 46D, left). At this point (usually after 10–15 minutes), cells at the bottom of the round well are removed by repeated aspiration and dispensing of the medium with a micropipette. Importantly,

the aspiration and dispensing do not agitate cells at the bottom of the microwells, since these cells are shielded from flow of the medium by the microwell walls(Carvalho et al., 2011). Next, the device is tilted at 45° in the direction of the finger-like channels (so that they point downward) and cells are allowed to slide into the finger-like channels by gravity (Fig. 46D, middle). Because of the small width of the microwells (100 mm), the sliding of cells from the bottom of the microwell into the finger-like channels only takes several seconds. The horizontal orientation of the device is then restored (Fig. 46D, right), leaving cells in the finger-like channels.

The footprint of the finger-like channels is nearly the same as the footprint of the microwell, making it simple to estimate the concentration of cells in the suspension loaded into the well that would result in an appropriate number of cells in the finger-like channels. Because the depth of the round wells is much greater that the diameter of a T lymphocyte (6 mm vs. 12 μm), a dense monolayer of cells at the microwell bottoms can be created using a relatively low number of cells in the suspension (~$1 \cdot 10^6$ cells per mL). Therefore complications that might arise from creating an excessively concentrated cellular suspension by centrifugation are avoided. Furthermore, loading of cells into the finger-like channels is nearly as simple as loading of cells into the regular multi-well plates and much simpler that loading of cells into microfluidic devices. Only ~50,000 cells (and ~50 μL of medium) are required to fill all finger-like channels within a single round well.

The volume of a round well is ~1000 times greater than the cumulative volume of all microwells (and finger-like channels) connected to it. Therefore, a round well represents a large reservoir of fresh medium for cells in microwells and finger-like channels.

Furthermore, due to the small depths of the microwells and small lengths of the finger-like channels, the molecular diffusion between a round well and finger-like channels leads to efficient equilibration of the medium content around cells with the medium in the round well. A characteristic time of the diffusion can be estimated as $\tau = L^2/(2D)$, where $L = 250$ μm is the cumulative distance from the bottom of a round well to the end of a finger-like channel, and D is the coefficient of diffusion. For small molecules, with $D = 5 \cdot 10^{-5}$ cm^2/s, the value of τ is ~60 s, indicating that, while the density of cells in the finger-like channels and at the bottom of the microwells may be very high (which is usually the case), the medium around cells is likely to be nearly identical to the fresh medium in the well (which does not have any cells in it). Moreover, the short medium equilibration time by diffusion implies that if the medium in the round well is exchanged, cells in the finger-like channels are rapidly exposed to the new medium, even without any active flow through the microwells and finger-like channels(Carvalho et al., 2011). Close proximity of the round wells (and the finger-like channels connected to them) limits the motion of the microscope stage to <12 mm along one dimension even with the 9-well device, making it possible to use high-resolution oil-immersion objectives and facilitating rapid scanning and high repeatability of the positioning.

2b. Fabrication and loading of the devices

Each microfluidic device consists of a microfabricated PDMS chip and a #1.5 microscope cover glass sealing the chip. The 4-well and 9-well PDMS chips have footprints of 14×14 and 19×19 mm, respectively. Each chip is assembled of two parts. The first part is 6 mm thick with 3.3 mm diameter through-holes in an either 2×2 or 3×3 array; the second part is 100 μm thick and has 200 μm long grooves with 25×25 μm

cross-section engraved on its surface (forming the finger-like channels) and 100x200 μm through-holes (forming the microwells; Fig. 46A). The 6 mm thick parts are cast of PDMS (Sylgard 184 by Dow Corning) using a master mold machined with a solid printer. To cast the 100 μm thick parts, a lithographically fabricated master mold is produced by spin-coating a silicon wafer with a 25 μm layer of a UV-curable epoxy (SU8-2015 by MicroChem), exposing it to UV-light through a photomask, spin-coating the wafer with a second layer of SU8 to a total thickness of 100 μm, exposing it to UV-light through another photomask, and developing the wafer.

The master mold is spin-coated with an ~100 μm thick layer of PDMS pre-polymer, so that the upper surfaces of the 100 μm relief features remain exposed (PDMS-free)(Kartalov et al., 2006), and the PDMS is cured by baking in an 80 °C oven. The 6 mm thick PDMS parts are bonded to the 100 μm thick PDMS layer on the mold by treating their surfaces with oxygen plasma. The 100 μm layer is cut around the perimeter of the 6 mm thick parts and the two-part monolith PDMS chips are separated from the mold. Each chip is reversibly bonded to a #1.5 microscope cover glass by overnight baking in an 80 °C oven forming a complete microfabricated device.

Before loading cells into a device, the microwells and finger-like channels are filled with a pH 7.5 PBS buffer. To this end, the device is treated with oxygen plasma to make its surface hydrophilic, the round wells are filled with the buffer using a micropipette, the device is placed into a plastic bottle with the buffer, and the bottle is pressurized to ~5 psi for 10 min. This procedure results in bubble-free filling of the finger-like channels, likely due to a combination of the buffer wicking into voids with hydrophilic walls and the

excessive pressure in the bottle pushing the residual bubbles from the microwells and finger-like channels into the bulk of the porous PDMS chip.

2c. Microscopy setup, cells and reagents

Experiments were performed on a Zeiss Axiovert inverted fluorescence microscope equipped with a Yokagawa spinning disc, 488nm laser excitation light source, a CoolSNAP HQ^2 14-bit camera from Photometrics, a computer controlled motorized stage, and environmental enclosure, maintaining a temperature of 37°C and a humidified atmosphere with 5% CO_2. In time-lapse experiments, images were captured every 10 minutes, using a 40x oil, 1.3NA objective, 300ms exposure time, 10% power on a 50mW 488nm solid-state laser, and analyzed as previously described(Weinberger et al., 2008a).

$CD4^+$ T Lymphocytes were isolated from patient blood using a negative selection method as previously described(Terry et al., 2009). Before infection experiments, cells were activated from the resting state for 48hrs using Dynabeads® Human T-Activator conjugated with anti-CD3/CD28 antibodies from Invitrogen. Primary $CD4^+$ T lymphocytes were cultured in RPMI 1640 supplemented with 5% Human Serum AB from GIBCO and 1% penicillin-streptomycin and kept at 1×10^6 cells/mL. THP-1 monocytes were cultured in RPMI 1640 with 10% fetal calf serum, 1% penicillin-streptomycin, and 0.05mM 2-mercaptoethanol and cultured at 2×10^5 - 8×10^5 cells/mL.

Cell viability was tested with Yo-Pro®-1 Iodide (Life Technologies™) as previously described(Faley et al., 2008). After isolation from peripheral blood (see above), primary cells were either placed in standard culture dishes or within the microwell device. At 24 hrs and 48 hrs time-points, ~100,000 cells from culture dishes were placed on a glass bottom slide and stained with Yo-Pro®-1. Within 10 min after

Yo-Pro®-1 addition, 4 images of cells in randomly selected areas were taken under brightfield and fluorescence illumination (with a FITC filter set), using a 10x 0.3NA objective. The total number of cells was 1800 at the 24 hr and 760 at the 48 hr time point. The percentage of dead cells was calculated by dividing the number of stained cells, as detected under the fluorescence illumination, by the total number of cells counted in the brightfield images. To evaluate the percentage of dead cells in the microwell device, Yo-Pro®-1 was applied to two different microwells at 24 and 48hrs and the numbers of dead cells and all cells were counted in finger-like channels connected to 4 different microwells at each time point. The total number of cells was 270 at the 24hrs and 609 at the 48 hr time point. All p-values were calculated using the Student's two-tailed t-test.

Four lentiviral constructs (LTR-GFP, LTR-GFP-Tat, HIVΔGag-GFP, and full-length HIVΔEnv-GFP)(Razooky and Weinberger, 2011) were packaged in 293FT cells as previously described(Razooky and Weinberger, 2011). Viral preps were clarified and ultra-centrifuged at 18,000rpm at 18°C for 1.5hrs, then re-suspended in 200 μL of medium together with primary $CD4^+$ T Lymphocytes. Infections were performed at a high multiplicity of infection (MOI) in a 96-well plate, and the suspension of cells and viruses was loaded into the microwell device. Activated primary cells cultured using standard techniques were infected with the LTR-GFP-Tat construct and GFP fluorescence of 10,000 live cells was measured at 48hrs post infection by flow cytometry on a FACSCalibur™ DxP8 instrument.

3. Results

3a. Primary cells loaded into the finger portion of the device are limited in mobility.

Loading of the microwell device normally resulted in ~25 cells in each finger-like channel and less than a monolayer of cells at the bottoms of rectangular wells (cf. Fig. 46C and 46C). To test for possible depletion of nutrients and accumulation of metabolites in the finger-like channels due to the high density of cells, we performed time-dependent numerical simulations in COMSOL (Fig. 48). The simulations indicated that medium conditions in the finger-like channels were substantially more favorable as compared to a monolayer cell culture in a dish (or multi-well plate). Because of relatively small total number of cells in the finger-like channels and large volume of medium in the circular wells, the depletion of nutrients (both low-molecular (Fig. 48A and 47C) and macromolecular (Fig. 48B and 47D)) in the finger-like channels was less severe than in the dish. Whereas, according to the simulations, concentrations of metabolites secreted by cells (especially those with low diffusivity) considerably varied along the finger-like channels, the average metabolite concentrations were several times lower than in the monolayer culture in a dish (Fig. 48, compare 47A to 47C, and 47B to 47D). Moreover, the oxygen permeability of PDMS is >5 times higher than water(Polinkovsky et al., 2009), so medium in the finger-like channels was expected to be better aerated than medium at the bottom of the dish.

The simulations suggested that cell viability would be maintained in the device since there would be no depletion of nutrients and sufficient oxygenation. We then experimentally checked the viability of resting primary $CD4^+$ T lymphocytes with Yo-Pro®-1 cell death stain. Resting primary $CD4^+$ T lymphocytes were either cultured using

standard techniques (Fig. 47A, blue columns) or placed within the finger-like channels of the device (Fig. 47A, red columns). After 24 and 48hrs in the device cells were just as healthy as cells cultured using standard techniques (Fig. 47A). To test if individual cells could be tracked over extended periods of time, we performed fluorescence time-lapse microscopy of $CD4^+$ T lymphocytes infected with an HIV-1 lentivirus expressing GFP. A minority (less than 30%) of cells initially infected within the finger-like channels crawled out (Fig. 47B) over the duration of the experiment. The cells that stayed in the finger-like channels of the device were readily tracked since the 2-dimensional movement is restricted in the finger-like channels (Fig. 47C, traces). In contrast, cells at the bottom of the microwells (Fig. 47C, right of dashed lines) freely moved in two dimensions, making them difficult to track over extended time intervals. The time of detection of fluorescence (indicating the onset of GFP expression) varied between cells (Fig. 47D), reflecting inherent variability in the timing of their infection. In agreement with previous reports in Jurkat cells(Weinberger and Shenk, 2007), the GFP expression of infected cells rapidly increased during the first 10–15 hours after the onset and then reached a plateau (Fig. 47D).

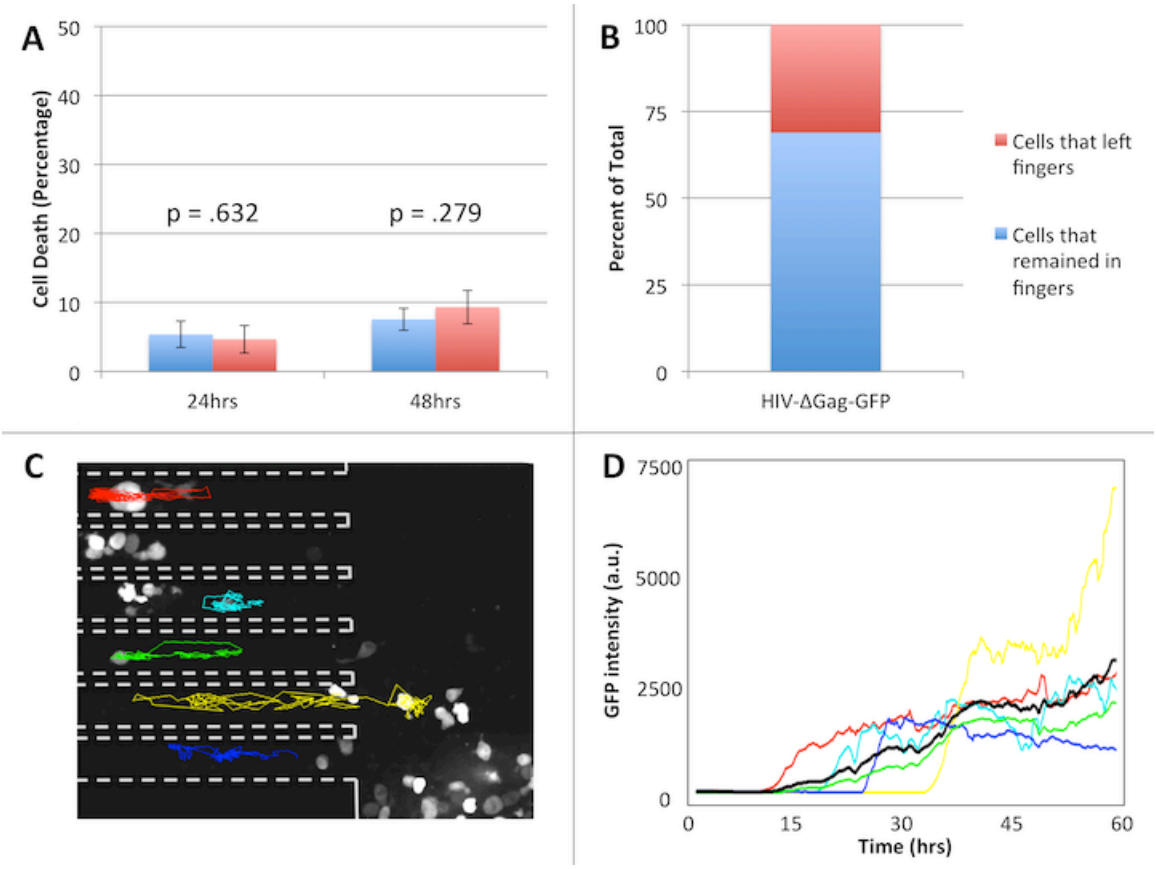

Figure 47. Primary human CD4$^+$ T lymphocytes in the microfabricated device maintain viability and can be reliably tracked.

(A) Cells placed within the finger-like channels of the device or in standard culture dishes were stained with Yo-Pro®-1 Iodide cell death stain after 24 and 48 hours. In the device, 4.66 ± 2.01 % of cells stained positive after 24 hours, a statistically insignificant difference from the 5.36 ± 1.91 % positively stained cells in culture (p = .632). There was also no statistically significant difference in the number of dead cells in the device, 9.32 ± 2.45, or in culture, 7.55 ± 1.60, after 48 hours (p = .279). Both p-values were calculated using the Student's two-tailed t-test. (B) The percent of infected cells that stayed within the finger-like channels of the device (69%, blue portion) compared to the number of infected cells that moved out of the finger-like channels (31%, red portion) over the duration of the 60-hour experiment. (C) Fluorescence micrograph of activated primary human CD4$^+$ T lymphocytes infected with a GFP-expressing HIV-1 virus at 26 hours post infection. Dashed outlines mark the sidewalls of the finger-like channels and colored tracks represent migration trajectories of individual cells over the course of a 60-hour imaging experiment. Random migration of cells in the finger-like channels is largely one-dimensional, making it possible to reliably track cells over the entire 60 hours of the time-lapse experiment. Cells at the bottom of the microwell (large area to the right of the dashed outlines) move about freely in two-dimensions and are difficult to track. (D) The intensity of fluorescence of GFP in a cell vs. time for the six infected cells tracked in panel B. The average fluorescence signal (black line) shows that cells start expressing

GFP at approximately 10 hours after infection. Originally published as Figure 2 in Razooky et al., Lab on a Chip, 2012.

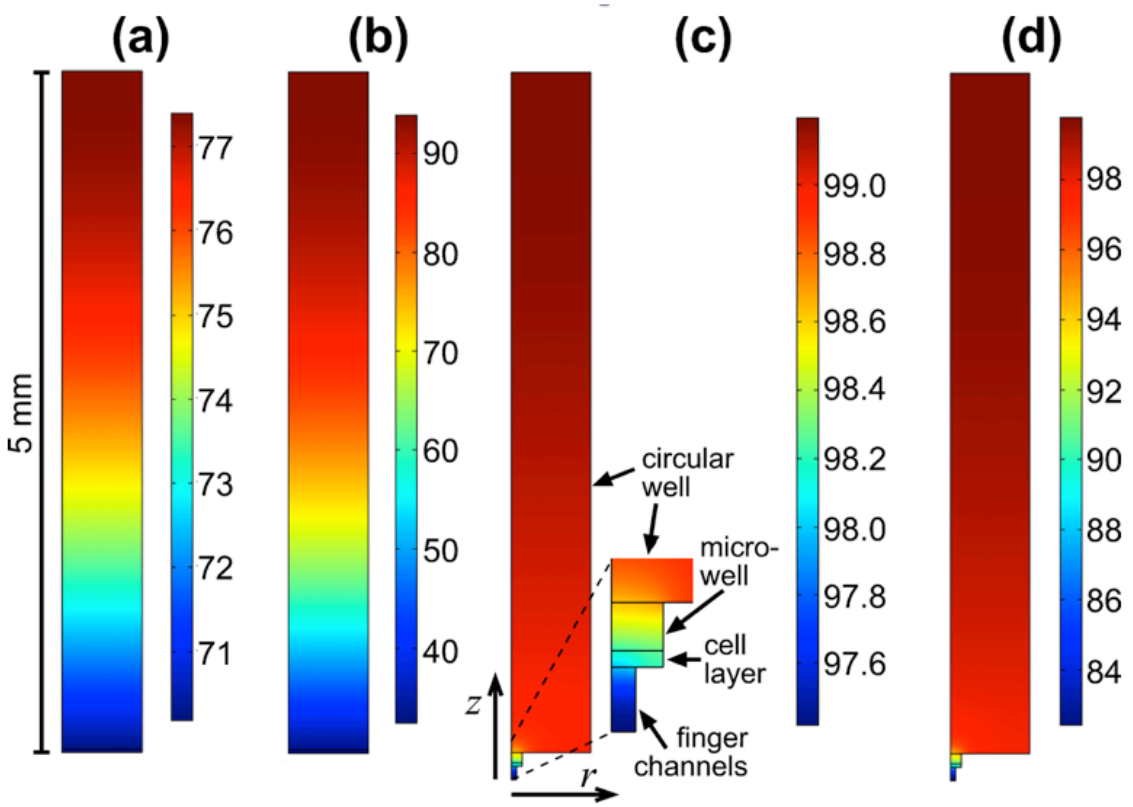

Figure 48. Results of numerical simulations on the diffusion and consumption of nutrients by cell culture at the bottom of a dish and in the microfabricated device.

The color-coded plots show nutrient distributions after 24 hrs of cell culture under a 5 mm deep layer of medium with an initial nutrient concentration of 100%. The legends, with concentrations in percents, are shown to the right of the corresponding distributions. (A) and (B) are two-dimensional simulations of cell culture at the bottom of a dish (~0.6 mm wide fragment is shown). Cell culture is represented by a 25 μm thick layer at the bottom of the computational domain with a uniform reaction rate of -0.0579%/sec, corresponding to consumption of 25% of the nutrient within 24 hrs. The coefficient of diffusion of the nutrient is D_1 = 500 μm^2/sec in (A) and D_2 = 50 μm^2/sec in (B). (C) and (D) are simulations of cell culture in the microfabricated device. A circular well of the device is represented by a 5 mm tall, 1.17 mm in diameter cylinder; a microwell is represented by a 100 μm tall, 160 μm in diameter cylinder; the finger-like channels are represented by a 100 μm tall, 75 μm in diameter cylinder. The three-dimensional configuration of the device is reduced to a two-dimensional axisymmetric computational domain, along the radial, r, and vertical, z, axes of the three-cylinder structure. Cells at the bottom of the microwell of the device are represented by a 25 μm thick layer with a uniform reaction rate of -0.0579%/sec at the bottom of the 160 μm diameter cylinder;

cells in the finger-like channels are represented by the same uniform reaction rate all over the 75 μm diameter cylinder, corresponding to 4 layers of cells with the same density as in (A) and (B). The coefficient of diffusion of the nutrient is 500 μm^2/sec in (C) and 50 μm^2/sec in (D). Originally published in supplemental information in Razooky et al., Lab on a Chip, 2012. Alex Groisman created this Figure.

3b. Numerical simulations of consumption and diffusion of nutrients in the microwell device

To test for variations in medium conditions in the microfabricated device, we numerically simulated the diffusion of nutrients in the device and their consumption by cultured cells. The goal of the simulation was to compare the medium conditions around cells in a finger-like channel of the device to cells at the bottom of a regular culture dish. The thickness of the medium layer was set at 5 mm for both the device and the dish. For the purpose of comparison, we chose a dish with a monolayer cell culture at the bottom, with a density of 1 cell per 64 μm^2 (8×8 μm), and set the nutrient consumption rate at such a level that the monolayer culture consumes 25% of the initial content of the 5 mm layer of the medium within 24 hours. The cell culture at the bottom of the dish was represented by a 25 μm thick layer (thicker than lymphocyte diameter to avoid unnecessarily high density of the simulation mesh) with a reaction (consumption) rate of -25%×5mm/25 μm/24 hrs or -0.0579%/sec. The initial concentration in the dish was set at 100%. The distributions of concentrations in a section of the dish after 24 hrs for a small-molecule nutrient, with a diffusion coefficient D_1 = 500 μm^2/sec (similar to that of glucose), and a macromolecule (protein), with a diffusion coefficient D_2 = 50 μm^2/sec (similar to that of albumin), are shown in Fig. 48A and 47B. As expected, the concentrations averaged over the bulk of the dish are 75% for both the small molecule and protein, but both concentration profiles are non-uniform and non-linear. Moreover,

the slower diffusion of the protein makes its distribution substantially less uniform, with concentration around the cells culture being as little as 32%, whereas it is ~70% for the small molecule nutrient.

The simulation of the lymphocyte culture in the microfabricated device was built to incorporate the three-dimensionality of the device architecture, comprising the finger-like channels, microwells, and circular wells, without making the simulation excessively complex or heavy. To this end, a rectangular microwell of the device (100×200 μm in the cross-section) was represented in the simulation by a cylindrical microwell, with the same cross-section area (160 μm in diameter) and the same depth of 100 μm. Because a single 3.3 mm in diameter circular well of the device has 8 microwells at the bottom, the circular well was represented in the simulation by a cylinder with 1/8 of the cross-section of the circular well (1.17 mm in diameter) and the same depth of 5 mm. Finally, all finger-like channels connected to the microwell, with an integral cross-section area of 4375 μm^2, were represented by a single cylindrical microwell, with the same cross-section area, corresponding to a 75 μm diameter, and the depth of 100 μm, equal to the length of the finger-like channels. The resulting simulation geometry, consisting of three co-axial cylinders, has the same connectivity and features with the same cross-section areas and lengths as their counterparts in the device, but is substantially easier to analyze and more simple to simulate, because of its axial symmetry. To represent cells remaining at the bottom of the microwell of the device after the finger-like channels are loaded, the 160 μm diameter microwell of the computational domain has a 25 μm tall layer at the bottom with a nutrient consumption (reaction) rate of -0.0579%/sec, the same as for the monolayer of cells at the bottom of the simulated dish. In our experiments, a typical

number of lymphocytes in a finger-like channel was ~25. Given the finger-like channel cross-section of 25×25 μm and the 8×8 μm footprint of a lymphocyte in the simulation of a monolayer in a dish, this number of lymphocytes per finger-like channel is equivalent to 2.6 layers of lymphocytes at the bottom of a dish. For our simulation, we set a uniform reaction rate of -0.0579%/sec in the 100 μm deep 75 μm diameter microwell (representing the finger-like channel), corresponding to 4 layers of lymphocytes or 39 lymphocytes per finger-like channel, that can be considered as a safe upper limit.

The results of the simulation of diffusion and consumption of a small molecule (D_1 = 500 μm^2/sec) in the model of the microfabricated device (Fig. 48C) indicate substantially more favorable cell culture conditions than in the model of a dish with the cell monolayer (Fig. 48A). Specifically, after 24 hours of culture, the nutrient concentration around cells in the finger-like channel is ~97% (Fig. 48C), as compared to ~70% in the culture dish model (Fig. 48A). The variation of concentration along the finger-like channel is only ~1.5%, suggesting nearly uniform concentration of small molecules. For a model protein (D_2 = 50 μm^2/sec), the concentration varies more appreciably, from 88% at the beginning of the model finger-like channel (top of the 75 μm diameter cylinder) to 82% at its end (bottom of the cylinder). Nevertheless, those variations are relatively minor, and the mean concentration (85%) still compares favorably with that in the monolayer culture in a dish (32%). The simulation can be corrected for partial blocking of a finger-like channel by cells in it, resulting in an effective reduction of the diffusion coefficient. Nevertheless, given that the total volume of 25 lymphocytes, ~6,400 μm^2, is ~10 times smaller than the volume of a finger-like channel (62,500 μm^2), the reduction of the

effective diffusion is expected to be ~10%, with only a minor effect on the profile of concentration in the finger-like channel.

Overall, the effects of depletion of both small-molecule and macromolecule nutrients are expected to be substantially weaker for cells in the finger-like channels (at 39 cells per channel) than for a cell monolayer at the bottom of a dish. According to our simulations (not shown), average concentration of small-molecule nutrient around cells in the finger-like channel at their typical population level is similar to that in a culture in a dish at a density of 1 cell per 640 μm^2, which is 1/10 of a monolayer. For macromolecules, the ratio is smaller, ~1/4, but still favorable for cells in the finger-like channels of the microwell device. The variation of concentration along the finger-like channel is negligible for small molecules and small for macromolecules. The variation would be larger for a more rapidly metabolized protein (or a lower initial absolute concentration), e.g. between 75 and 65% across a finger-like channel, if a monolayer culture in a dish reduces the concentration of the protein within 24 hrs by 50% (rather than 25%). However, such culture conditions would be somewhat extreme and not directly applicable to the lymphocyte cultures and the media used in our experiments.

The distributions of nutrients in Fig. 48 can be used to estimate distributions of molecules secreted (rather than consumed) by cells. If, in an analogy to the above discussion, we assume that for cells in a dish the average concentration of secreted molecules reaches 25% of some threshold value in 24 hrs, distributions of secreted molecules can be obtained from the distributions in Fig. 48 by subtracting these latter distributions from 100%. For small-molecules (metabolites and small signaling molecules) secreted by cells, this fact implies that their concentration around cells always

remains low (~2.5% after 24 hrs vs. ~30% for the dish culture; Fig. 48C vs. 47A). For secreted macromolecules with a representative diffusion coefficient of 50 μm^2/sec, the mean concentration around cells in a finger-like channel is still substantially lower than in a dish, ~15% vs ~68%, but the concentration varies along the finger-like channel by as much as a factor of 1.5 (from 18% to 12%). Variations of concentration in the finger-like channel would be even greater for secreted molecules or particles with lower diffusivity, such as HIV viruses, with high concentrations accumulating at the dead end.

If the length of the finger-like channels is increased to 200 μm, while the number of cells in them remains unchanged, the variations of concentrations of small-molecule nutrients along the finger-like channels still remain small, 98% at the beginning vs. 96.8% at the end of the finger-like channel after 24 hrs. Gradients of small-molecule metabolites become substantial, with an ~1.6-fold change from the beginning to the end, but their absolute concentrations are still very low, at 2.6% on average. Gradients of protein nutrients (D_2 = 50 μm^2/sec) are more significant, with ~88% concentration at the beginning and ~76.5 % at the end, but their average concentration still compares favorably to that in the dish culture (~30%). For macromolecules with the same diffusivity secreted by cells, the variation of concentration is stronger, from 12% at the entrance to 23.5% at the end, corresponding to an almost two-fold change. But again, the average concentration of macromolecular metabolites compares favorably to that in the dish culture (68%). Further increase of the finger-like channel length or the number of cells in them would lead to even stronger gradients of macromolecular compounds that can be detrimental for the functionality of the device.

3c. Expression kinetics for HIV are dependent on the regulatory elements.

We next compared the expression profiles for diverse HIV-1-derived lentiviral vectors encoding successively fewer HIV-1 regulatory and structural components and all encoding GFP as a reporter. These minimal synthetic lentiviral vectors are important controls. The comparison of their behavior with the behavior of the full-length virus reduces complexity to tractable levels and facilitates logical mapping of viral gene-regulation circuitry(Franz et al., 2011; Razooky and Weinberger, 2011; Weinberger et al., 2008a; Weinberger and Shenk, 2007). Four previously described viral constructs(Razooky and Weinberger, 2011) were used to infect activated primary $CD4^+$ T lymphocytes in a 9 well device. Each infection was performed in duplicate in adjacent wells. The single remaining well was retained as an uninfected control to check cell viability and autofluorescence over the course of the experiment (Fig. 49A). Cellular autofluorescence does not influence the GFP signal since autofluorescence is 30-fold lower than GFP fluorescence even for the very dimmest cell tracked (autofluorescence is 100-fold lower than the average GFP fluorescence from LTR-GFP-Tat expressing cells). A total of 59 cells were tracked for 60 hours, with at least 10 cells tracked for each lentiviral infection. Consistent with the results of the previous experiment (Fig. 48), cells within the finger-like channels could be reliably tracked, whereas tracking of cells at the bottom of the microwells was difficult due to their rapid two-dimensional migration. Cells that were infected within the finger-like channels tended to remain inside the finger-like channels for the duration of the experiment.

Cells infected with the LTR-GFP construct consistently exhibited the lowest fluorescence levels (i.e. expressed the lowest levels of GFP; Fig. 49B, top), which is

likely because this vector only encodes the relatively weak HIV-1 long-terminal repeat (LTR) promoter element and no other HIV-1 transactivating genes. Cells infected with the LTR-GFP-Tat had the strongest GFP expression (Fig. 49B, second from top), most likely due to the HIV-1 positive-feedback loop in which the transactivator of transcription (Tat) increases expression from the HIV-1 LTR(Razooky and Weinberger, 2011; Singh et al., 2010a; Weinberger et al., 2005; Weinberger et al., 2008a). When averaged over an interval from 40 to 50 hours post infection, in which the fluorescence was changing in time relatively slowly, the levels of fluorescence per cell for cells infected with LTR-GFP, LTR-GFP-Tat, HIV-ΔGag-GFP, and full length HIV-ΔEnv-GFP were, respectively, 542 ± 117, 1605 ± 532, 1062 ± 156, and 932 ± 207. These values indicated that the GFP expression from LTR-GFP was significantly lower than from the other three constructs, whereas the differences between LTR-GFP-Tat, HIV-ΔGag-GFP, and full length HIV-ΔEnv-GFP were statistically insignificant. The rate and level of GFP expression from the three lentiviruses that contain the Tat positive feedback loop (LTR-GFP-Tat, HIV-ΔGag-GFP, and full length HIVΔEnv-GFP) were significantly higher throughout the duration of the experiment (Fig. 49C), in agreement with previous measurements of these lentiviral vectors in the immortalized Jurkat T Lymphocyte cell line(Weinberger et al., 2008a; Weinberger and Shenk, 2007). To compare the behavior and function of the cells in the microwell device with standard bulk cell culture, cells in standard bulk culture were infected with the LTR-GFP-Tat lentiviral vector and 48hrs after infection fluorescence was measured by flow cytometry (Fig. 49D). The increase in GFP expression over 48hrs in bulk culture is ~130 fold over background autofluorescence which is equivalent to the

increase in GFP expression for cells residing in the microwell finger-like channels 48hrs post infection (Fig. 49D).

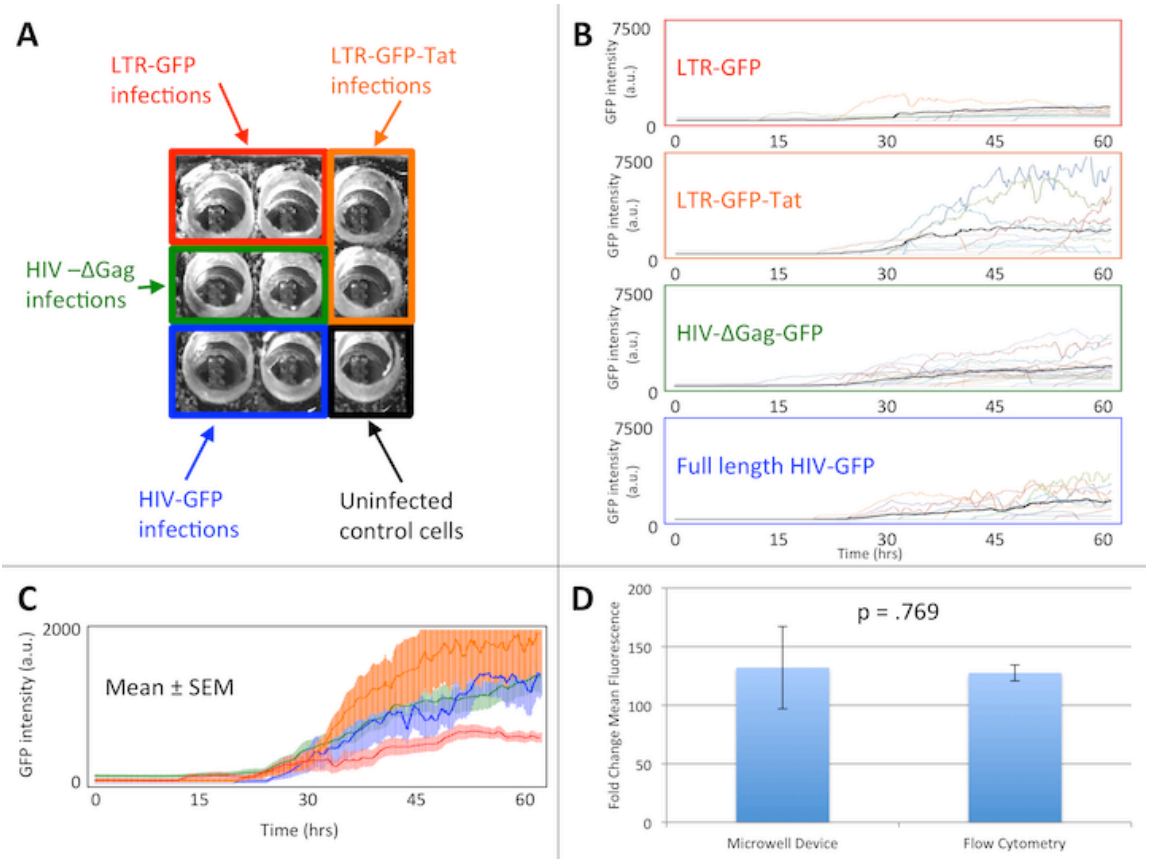

Figure 49. Single-cell gene-expression kinetics measured for primary human CD4$^+$ T lymphocytes infected with four different HIV-1 viral constructs in a single experiment.

(A) Experimental setup for the infection of activated primary CD4$^+$ T lymphocytes in a 3×3–well chip. Four previously described HIV-1 lentiviral constructs (LTR-GFP, LTR-GFP-Tat, HIV-ΔGag-GFP, and full-length construct HIV-ΔEnv-GFP(Razooky and Weinberger, 2011)) were used to infect activated primary human CD4$^+$ T lymphocytes. CD4$^+$ T lymphocytes infected with each of the four constructs were loaded into two nearby wells for time-lapse imaging. One well was left for uninfected control cells. (B) Single-cell GFP intensity (arbitrary units) time traces over a 60-hour period for the four lentiviral constructs. Fluorescence signal (reflecting the level of GFP expression) is shown for at least 10 individual cells for each construct. Different cells are represented by time traces of different colors. Mean fluorescence intensity per cell is shown as bold black line. (C) The average time-lapse trace for each construct with error bars representing the standard error of the mean. The LTR-GFP (red with red error bars)

average is significantly lower in rate and level of GFP expression than the LTR-GFP-Tat (orange with orange error bars), HIV-ΔGag-GFP (green with green error bars), and full-length HIV-ΔEnv-GFP (blue with blue error bars) average traces. (D) Activated lymphocytes were infected with the LTR-GFP-Tat virus and measurements were taken 48hrs post infection by flow cytometry. There is no statistically significant difference in the fold change in mean fluorescence for cells 48hrs post infection as measured by flow cytometry, 131 ± 7, or by microscopy in the device, 127 ± 35 ($p = .769$ as calculated by Student's two-tailed t-test). Error bars represent the standard error of the mean. Originally published as Figure 3 in Razooky et al., Lab on a Chip, 2012. Brandon Razooky and Leor Weinberger created this Figure.

4. Discussion and Conclusions

The results of our pilot experiments demonstrate the utility of the proposed microwell devices with finger-like imaging channels, microwells, and pipette-accessible wells. These devices have enabled tracking the fluorescence of a large number of individual primary human $CD4^+$ T lymphocytes over a time interval significantly exceeding the 40 hrs life-cycle of HIV-1, thus providing a substantial improvement upon what was previously shown in either culture dishes(Saez-Cirion et al., 2006) or continuously perfused microfluidic devices(Faley et al., 2008). The proposed devices combine an important benefit of multi-well plates, the ease of cell loading, with a benefit of microfluidic devices, the confinement of cells to microfabricated channels with a diameter comparable with that of the cells. $CD4^+$ T lymphocytes within the finger-like channels have access to a large reservoir of fresh medium in the round wells, and are in physical contact with each other, thereby allowing them to maintain their viability for multiple days. In addition, the device allows rapid and simple media exchange without displacement of cells. The results of the analysis of cell trajectories from the finger-like channels vs. bottoms of the microwells indicate that the confinement of activated T lymphocytes to a nearly one-dimensional region in a finger-like channel is sufficient for

reliable cell tracking. Moreover, the finger-like channels appear to provide a nearly optimal degree of cell confinement for long-term imaging, because more aggressive confinement and immobilization of T lymphocytes (e.g., by placing them in small closed wells or pushing them against a semi-permeable barrier) might create less physiological conditions or even compromise T lymphocyte viability.

The proposed devices have the potential to enable quantitative studies of kinetics of HIV-1 infection and reactivation from latency with single-cell resolution. The devices can be readily expanded to include a larger number of the round wells (each with microwells and finger-like channels at the bottom), making it possible to test a larger number of medium conditions and different cell lines in a single experiment. The proposed devices can also be modified and adapted to experiments on other non-adherent cells such as other leukocytes, hematopoietic stem cells, or yeast.

Chapter 9: Robustness in HIV decision-making

1. Introduction

When cells or viruses react to perturbation, the architecture of the responding signal transduction pathway or gene circuit shape the dynamic behavior and subsequent phenotype of the organism (Ferrell, 2002; Teng et al., 2012). One feature of many circuits controlling fate decisions is cooperative feedback (Ferrell, 2002), which marks a clear transition from one fate to another through concentration thresholds (Gardner et al., 2000). However, these fate-determining circuits are susceptible to stochastic fluctuations in gene-expression that dictate the overall state, for an organism such as lambda phage, stochastic fluctuations can influence if the phage enters a lytic or lysogenic state (Cagatay et al., 2009; McAdams and Arkin, 1997; Suel et al., 2006). In contrast to bistable circuits, some fate-determining systems, such as those found in HIV, and synthetic systems, lack bistability and utilize stochastic fluctuations to transition between fates (To and Maheshri, 2010; Weinberger et al., 2005). Intuitively, circuits relying on stochastic transitioning between fates would not be robustly commit fate-decisions. Here we test if circuits that stochastically transition between fates can robustly commit to one fate, and if so, how is robustness achieved.

To test if stochastic fate-determining circuits can robustly commit to fate decisions, we use the HIV-1 active-versus-latent fate decision as a model system to test if a previously described positive feedback motif contains sources of robustness (Razooky and Weinberger, 2011; Weinberger et al., 2005). Upon infection of a CD4+ T lymphocyte, HIV-1 can either actively express viral genes and create viral progeny that

destroy the host cell, or enter a quiescent state of silenced gene-expression termed proviral latency (Siliciano and Greene, 2011). The decision between active replication and latency is dependent on the availability of the viral transactivator of transcription, Tat, and stochastic fluctuations in the levels of Tat drive this decision making process (Weinberger et al., 2005). HIV Tat forms a positive feedback loop on the HIV long terminal repeat (LTR) promoter, and this positive feedback loop lacks cooperativity (Weinberger and Shenk, 2007). To test for robustness in HIV decision-making, HIV positive feedback was manipulated and (*i*) HIV fate, (*ii*) gene-expression kinetics, and (*iii*) steady-state output are followed.

2. Results

2a. Tuning Tat feedback to test for robustness in kinetics.

We first tune Tat availability by fusing Tat to a controllable proteolysis tag, FKBP, in a minimal circuit decision-making model. FKBP degradation is reversibly inhibited by a small molecule, Shield-1 (Banaszynski et al., 2006), allowing Tat half-life and positive-feedback strength to be rapidly tuned. The resulting Tat-FKBP translational-fusion cassette is encoded in a lentiviral vector driven by the HIV-1 long terminal repeat promoter (LTR), which also drives expression of a destabilized two hour half-life enhanced green fluorescent protein (GFP) (Li et al., 1998). In the LTR-GFP-IRES-Tat-FKBP (LGITF) viral vector, Tat-FKBP is actively degraded in the absence of Shield-1 leaving little Tat to activate the LTR promoter (Fig. 50A, left). However, upon Shield-1 addition, Tat-FKBP is stabilized, generating sufficient levels of Tat to feedback and transactivate the LTR (Fig. 50A, right). We first created and sorted a polyclonal

population (i.e. each cell contains a different integration site of the LGIT construct) of Jurkat cells stably transduced with this lentiviral construct (Jordan et al., 2001). The cells were then exposed to variable concentrations of Shield-1 in a fixed concentration of either tumor necrosis factor alpha (TNF-α) (Fig. 50B) or trichostatin A (TSA) (Fig. 50C) to stimulate the LTR transcription and activate Tat positive feedback. These molecules, although known potent inducers of HIV transcription (Dar et al., 2012), act through distinct mechanisms (Hayden and Ghosh, 2004a; Van Lint et al., 1996). Strikingly, we find that expression kinetics from the HIV LTR is insensitive to variable Tat levels (Fig. 50B and C).

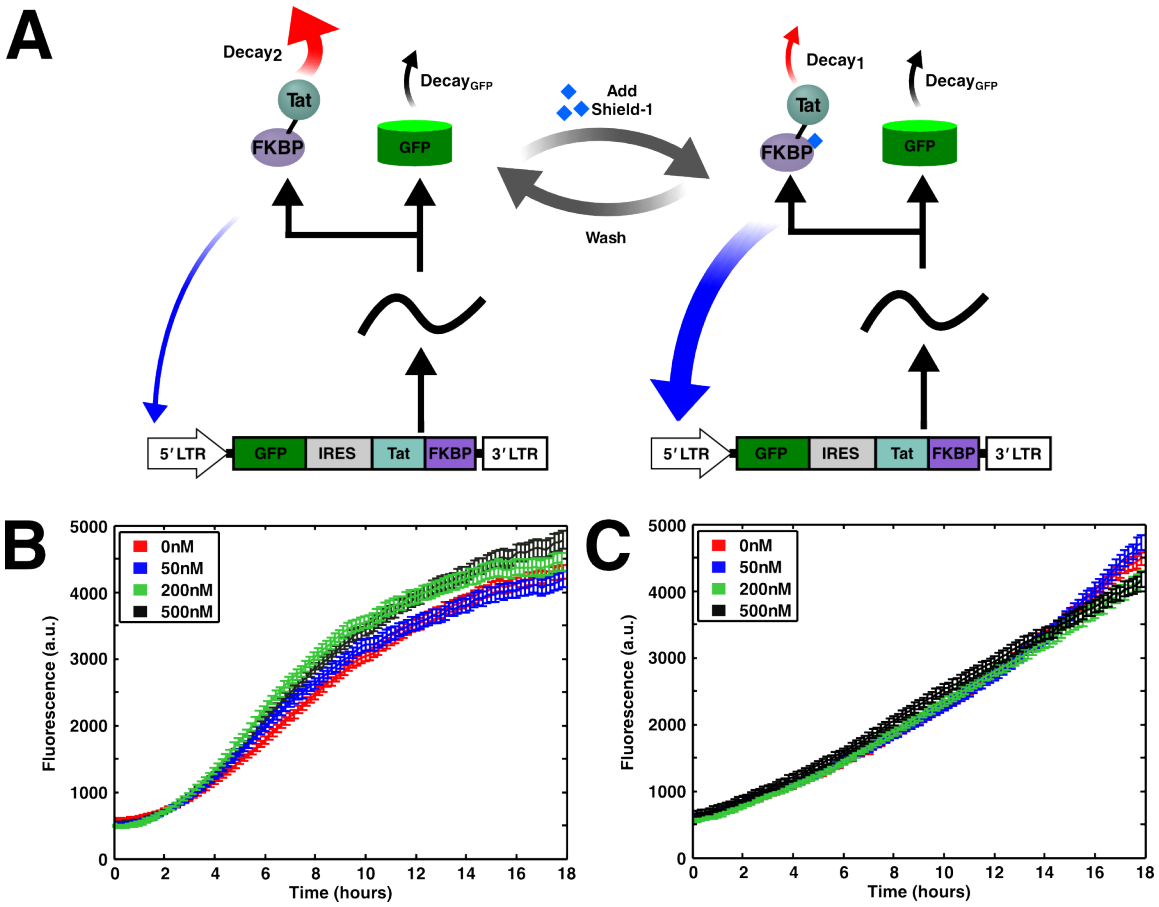

Figure 50. Modulation of Tat positive feedback does not alter gene-expression kinetics from the HIV LTR.

(A) The LGITF circuit expresses a single mRNA species, which is translated to produce GFP and Tat-FKBP. In the absence of Shield-1, Tat-FKBP rapidly decays, leading to weak positive feedback (left). Upon Shield-1 addition, FKBP mediated Tat degradation is blocked, and more Tat becomes available to feedback onto the LTR (right). (B) and (C) Single-cell trajectories of a polyclonal population of Jurkat cells stably transduced with LGITF stimulated with TNF-α (B) or TSA (C) in the presence of 0nM (red), 50nM (blue), 200nM (green), or 500nM (black) Shield-1. Error bars represent the standard error of the mean for at least 300 cells per condition.

2b. Model of HIV positive feedback predicts that it rapidly saturates.

To understand this behavior we adopted a previous mathematical model describing Tat positive feedback (Razooky and Weinberger, 2011). To simplify the

model we assume quasi-steady-state mRNA levels, and describe the LGITF circuit with two differential equations that describe Tat and GFP expression respectively:

$$\frac{d}{dt}(Tat) = \beta_{Tat} + \frac{\alpha+Tat}{K_M+Tat} - \delta_{Tat} * Tat \quad (44)$$

$$\frac{d}{dt}(GFP) = \beta_{GFP} + \frac{\alpha+Tat}{K_M+Tat} - \delta_{GFP} * GFP \quad (45)$$

where β_{Tat} is the basal rate of transcription and translation of Tat, α represents positive feedback strength, K_M is the saturation constant of the system, δ_{Tat} is the per capita death rate of Tat (which can be tuned), β_{GFP} is the basal rate of transcription and translation of GFP, and δ_{GFP} is the per capita death rate of GFP. Using Equations 44 and 45, we explored possible parameter regimes that could recapitulate the result of altering Tat half-life without affecting GFP expression kinetics (Fig. 51A). Two possibilities surface to explain the data. The first is that $K_M >> \alpha$ (Fig. 51A, bottom curves), which suggests positive feedback does not greatly contribute to expression level or kinetics expression and so expression is mostly dependent on the basal rates of transcription and translation, i.e. β_{Tat} and β_{GFP} (Fig. 51A, bottom curves). This regime is not physiologically relevant since Tat positive feedback is well-described (Weinberger et al., 2005) and can affect expression from the LTR up to 1000-fold (Jordan et al., 2001; Weinberger et al., 2005). The second possibility is that $K_M << Tat$, and the Michaelis-Menten positive feedback term rapidly saturates (Fig. 51A, top curves). Under the assumption of K_M being small the new equations become:

$$\frac{d}{dt}(Tat) = \beta_{Tat} + \alpha - \delta_{Tat} * Tat \quad (46)$$

$$\frac{d}{dt}(GFP) = \beta_{GFP} + \alpha - \delta_{GFP} * GFP \quad (47)$$

and GFP kinetics no longer depend on Tat concentration, much like the observation in Fig. 50. To verify that Tat positive feedback operates in the saturated regime, we performed additional experiments where Tat concentration, instead of being attenuated (Fig. 50), was increased significantly using a Tet-ON, doxycycline inducible system (Fig. 51B) (Gossen and Bujard, 1992). Cell lines were produced where Tat could be exogenously supplied upon doxycycline addition. The cell lines were then transduced with a lentiviral vector encoding the LTR driving expression of the fluorescent protein mCherry, an IRES, and Tat-FKBP protein fusion product. Therefore, the positive feedback loop could be exogenously perturbed upon doxycycline addition (Fig. 51B). The cells were exposed to TNF-α and varying concentrations of doxycycline, and single-cell time-lapse microscopy was performed to follow expression kinetics from the HIV LTR. Strikingly, despite large boluses of Tat, activation kinetics from the HIV LTR could not be modulated (Fig. 51C). These data agree with the previously described assumption of saturated Tat positive feedback and suggest that this saturated feedback motif provides robustness in HIV gene-expression kinetics.

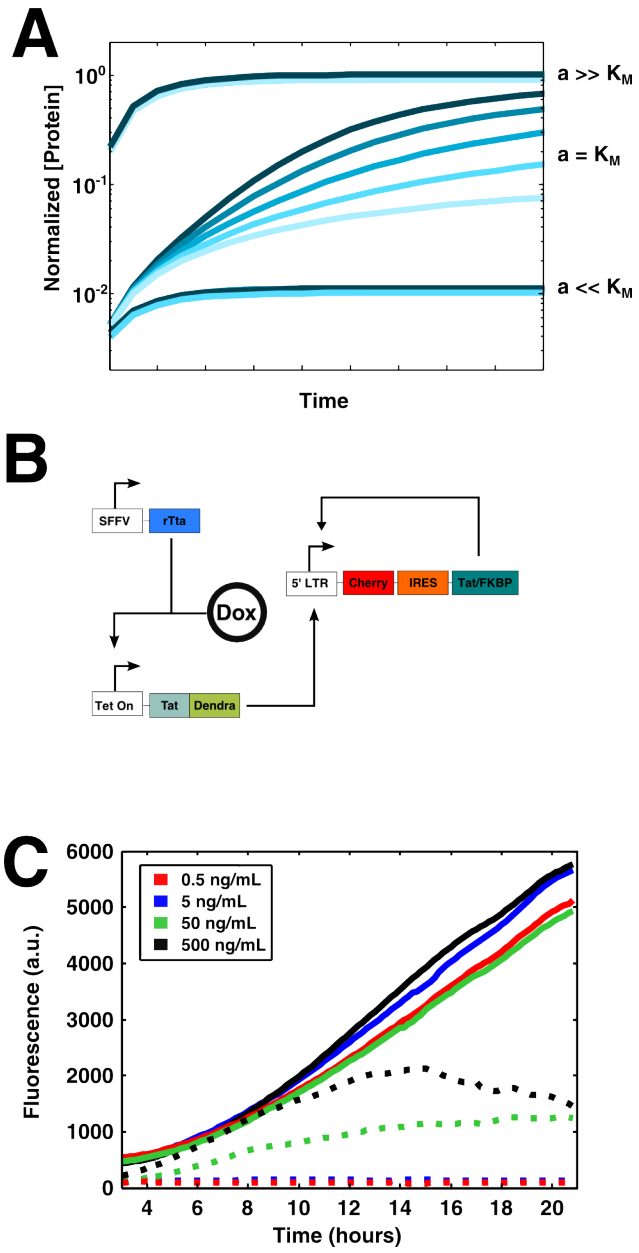

Figure 51. Tat positive feedback rapidly saturates upon induction and exogenous Tat addition does not alter expression kinetics.

(A) ODE simulations of expression kinetics from the LTR for various Tat half-lives. The half-life of Tat was varied 5-fold (light blue to dark blue). Simulations were performed in conditions ranging from readily saturable ($\boldsymbol{\alpha} >> \boldsymbol{K_M}$), intermediate ($\boldsymbol{\alpha} = \boldsymbol{K_M}$), and weak/non-existent ($\boldsymbol{K_M} >> \boldsymbol{\alpha}$) positive feedback. (B) The Tet-TD + LChITF circuit supplements LChITF positive feedback with exogenous Tat through doxycycline addition. LChITF is analogous to the LGITF circuit but GFP is replaced with mCherry fluorescent protein. (C) The mean of single-cell trajectories of Tet-TD + LChITF cells stimulated with TNF-α in the presence of 0.5 (red), 5 (blue), 50 (green), or 500 (black) ng/mL doxycycline. Error bar representations were left out for readability but are shown in the

corresponding supplemental figure XYZ. At least 800 single-cell trajectories were collected for each condition.

There are additional testable predictions that stem from the saturated feedback model. The isoclines from the analytical solutions to Equations 44 and 45 suggest that under situations where K_M is very small, GFP steady-state values, $< GFP >$, would be insensitive to Tat half-life (Fig. 52A). If we assume saturation, the analytical solution to Equation 47 simplifies to one that no longer depends on Tat concentration:

$$< GFP > = \frac{\beta_{GFP} + \alpha}{\delta_{GFP}} \qquad (48)$$

Additionally, the ratio of two GFP steady-state values for two different Tat half-lives should be equal to one:

$$\frac{<GFP>_{Tat(1)}}{<GFP>_{Tat(2)}} = 1 \qquad (49)$$

We tested this prediction in activated primary CD4+ T lymphocytes, which are the primary target cells for HIV infection. In agreement with the model, we find that stead-state GFP values did not change despite altered Tat half-lives (Fig. 52B).

Another testable prediction of the model is that in saturating feedback conditions, the ratio of two steady-state Tat concentrations should be equal to the ratio in the change of Tat-half-lives. Isoclines of Tat steady-state concentration show that in the regime where K_M is very small, the changes in Tat steady-state are linearly-dependent on Tat half-life (Fig. 52C). This becomes obvious when one solves for the analytical solution to Equation 46 for a given Tat half-life:

$$< Tat_{\delta_{Tat(1)}} >= \frac{\beta_{Tat}+\alpha}{\delta_{Tat(1)}} \quad (50)$$

where $< Tat_{\delta_{Tat(1)}} >$ is the steady state Tat concentration and $\delta_{Tat(1)}$ is the half-life of Tat for that given condition. The ratio of two-steady-state Tat concentrations for two different half-lives is given by:

$$\frac{<Tat_{\delta_{Tat(1)}}>}{<Tat_{\delta_{Tat(2)}}>} = \frac{\delta_{Tat(2)}}{\delta_{Tat(1)}} \quad (51)$$

$$\delta_{Tat(1)} = \frac{\ln 2}{\tau_1} \quad (52)$$

where τ_1 is the half-life, then:

$$\frac{<Tat_{\delta_{Tat(1)}}>}{<Tat_{\delta_{Tat(2)}}>} = \frac{\tau_1}{\tau_2} \quad (53)$$

To test this we transduced activated primary CD4+ lymphocytes with the LTR-Tat-Dendra-FKBP construct in the presence or absence of Shield-1. Previously, we quantified the change in Tat half-life with and without Shield-1 and found that we could modulate it ~3.3 fold (Fig. 11). Strikingly, the shift in expression-level between the Shield-1 and no Shield-1 cells in primary lymphocytes was only altered 3.6 ± 0.4 fold, showing that HIV positive feedback is saturated in activated lymphocytes (Fig. 52D).

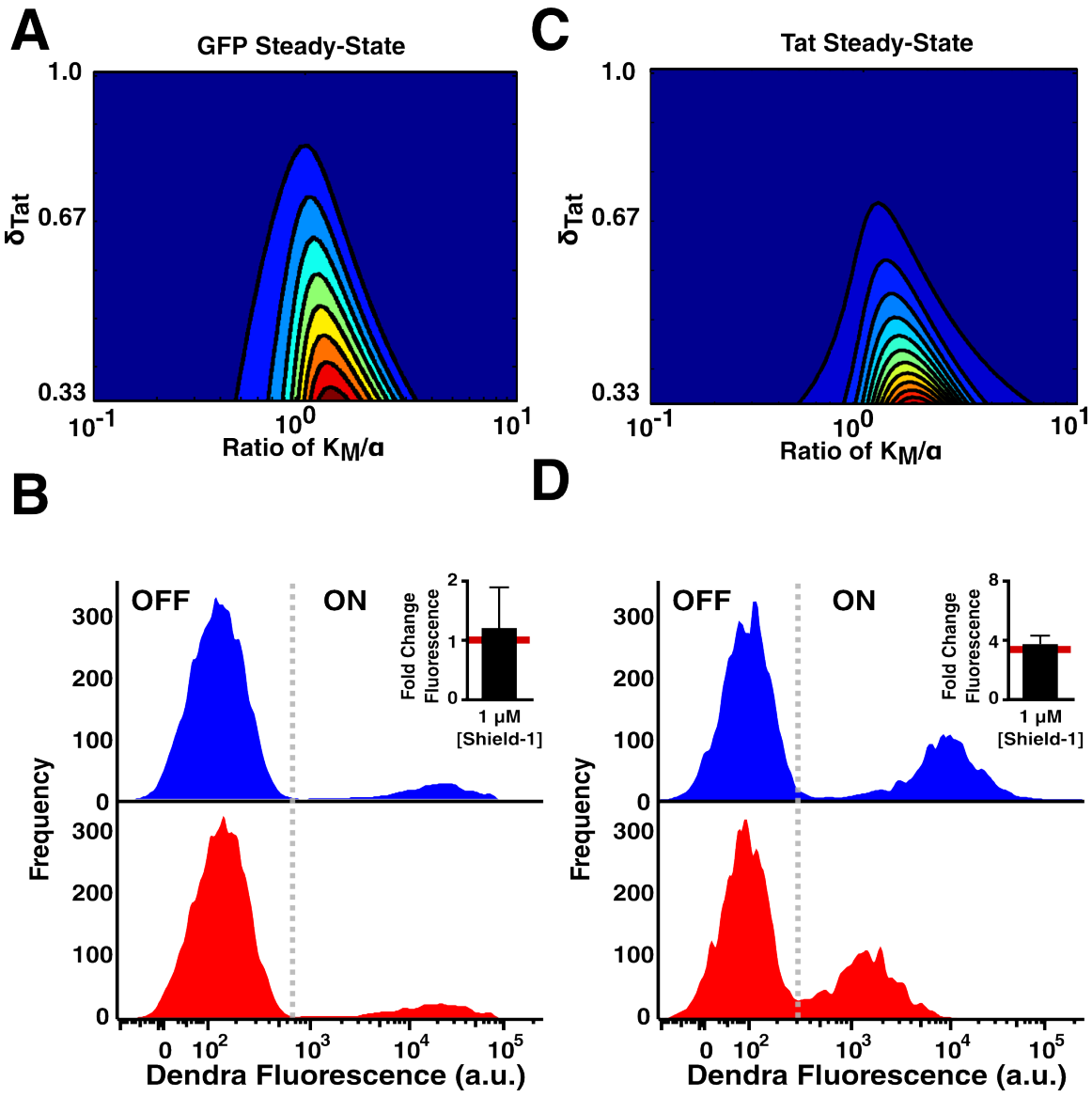

Figure 52. Tat positive feedback saturation leads to steady-state robustness in LTR output rendering LTR activity relatively insensitive to variable Tat concentrations.

(A) Isocline plot of normalized GFP steady-state values for different Tat half-lives ($\boldsymbol{\delta_{Tat}}$) versus the ratio of $\boldsymbol{K_M/\alpha}$. Outside of the bounds between ~0.5 and 5 of $\boldsymbol{K_M/\alpha}$, GFP steady-state values become relatively insensitive to Tat half-life changes. The plot was normalized by taking the minimum GFP value in each column of Km/alpha, which corresponds to the smallest $\boldsymbol{\delta_{Tat}}$, to express everything in terms of fold-changes. (B) Histogram of GFP fluorescence of activated primary CD4+ T lymphocytes infected with LGITF in the presence and absence of Shield-1. GFP median fluorescence value was taken from the ON population (inset). (C) Isocline plot of normalized Tat steady-state values for different Tat half-lives ($\boldsymbol{\delta_{Tat}}$) versus the ratio of $\boldsymbol{K_M/\alpha}$. Outside of the bounds between ~0.5 and 5 of $\boldsymbol{K_M/\alpha}$, Tat steady-state values begin to scale linearly with Tat half-life changes. The plot was normalized by taking the minimum Tat value in each

column of Km/alpha, which corresponds to the smallest $\boldsymbol{\delta_{Tat}}$, to express everything in terms of fold-changes. (D) Histogram of Dendra fluorescence for activated primary CD4+ T lymphocytes infected with LTDF in the presence and absence of Shield-1. Fold change in median Dendra fluorescence was quantified for the ON cells (inset).

2c. Transcriptional activators do not change the concentration at which Tat saturates.

Since altering Tat concentrations did not affect GFP steady-state output or expression kinetics (Fig. 50-51), we next wanted to test Tat dosing would affect expression levels from the LTR in the absence of positive feedback (Fig. 53A). Recall from Equation 47 that HIV expression is comprised of a Michaelis-Menten induction term ($\frac{\alpha+Tat}{K_m+Tat}$) and basal rate term ($\beta$). To understand how cells transition between a state dominated by β to a transactivated state dominated by $\frac{\alpha+Tat}{K_m+Tat}$, we developed cell lines where Tat transactivation was decoupled from basal LTR expression. Cell lines were stably transduced with Tat-Dendra-FKBP under the control of the Tet-ON promoter. In the same cells a single copy of LTR-mCherry was also transduced (Fig. 53A) (Weinberger et al., 2005). In the absence of doxycycline, expression is only dependent on β. However, upon induction with doxycycline, Tat can transactivate the LTR and expression will be equivalent to Equation 47. The cell lines were exposed to varying concentrations of doxycycline and Shield-1 in the absence or the same concentration of TNF-α—96 conditions total (Fig. 53B). If Tat reaches saturating concentrations, then there should be no observed fluorescence shift in the Tat Induced State (TIS) (Fig. 53B, right side of right two histograms). For very low Tat concentrations as observed by Dendra fluorescence (i.e. low doxycycline concentrations), expression in the TIS does indeed shift, suggesting that sub-saturating concentrations of Tat are achievable (Fig.

53C). However, these shifts only mark a small regime in the doxycycline-response curve, and Tat rapidly saturates upon further induction (Fig. 53C). These effects are further exaggerated in the presence of TNF-α (Fig. 53C). Importantly, although the TIS ceases to increase in fluorescence despite increasing Tat inputs, the percentage of cells entering the TIS is correlated with increasing Tat concentration (Fig. 53D). Strikingly, despite unimodal Tat inputs (Fig. 53B, top histograms), and no observable shift in the TIS (Fig. 53C), cells continue to enter the TIS state for increasing Tat concentrations (Fig. 53D). These results are analogous to reports of converting graded inputs into a binary response (Becskei et al., 2001).

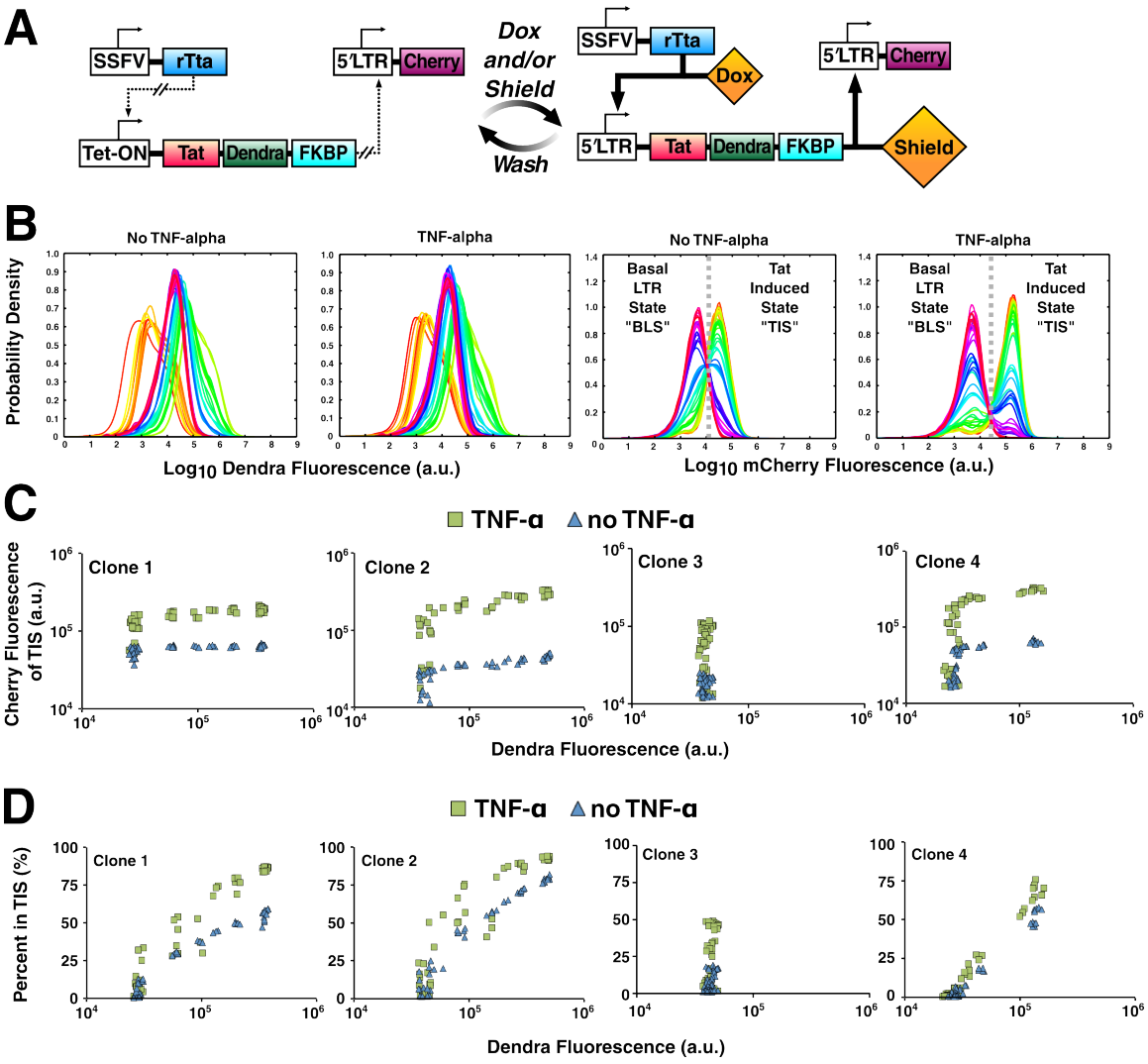

Figure 53. Tat transactivation converts graded inputs into binary responses by saturating the HIV LTR.

(A) Circuit design to exogenously supply Tat onto the HIV LTR. In the absence of doxycycline, the Tet-ON promoter is silent and Tat-Dendra-FKBP is unable to transactivate the LTR (left). Upon addition of doxycycline and Shield-1, expression from the HIV LTR can be controlled through tunable binding of the rTta transactivator and Tat-Dendra-FKBP expression (right). (B) Example probability density plots of an isoclonal cell line expressing the construct from (A) in the absence or presence of TNF-α. Each histogram in the Dendra and Cherry probability density plots represents a different combination of doxycycline and Shield-1 (48 conditions total). Expression from the HIV LTR is broken down into a 'Basal LTR State', BLS, and a high expressing state termed 'Tat Induced State', TIS, where the gray line represents the boundary and is different for each clone and shifts upon TNF-α addition. (C) Plots of the Cherry Fluorescence of TIS versus Dendra fluorescence for four isoclonal populations in the presence (green squares) or absence (blue triangles) of TNF-α. (D) Plots of the percentage of cells in the TIS

versus Dendra fluorescence for the same four isoclonal populations in the presence (green squares) or absence (blue triangles) of TNF-α.

3. Discussion

While Tat is obligate for HIV replication (Huang et al., 1994), there is only minimal evidence for how Tat influences decision-making (Donahue et al., 2013; Donahue et al., 2012), expression kinetics, and steady-state behavior (Razooky and Weinberger, 2011). Here, we have provided a comprehensive study of Tat feedback circuitry by manipulating Tat inputs and half-life. Surprisingly, HIV positive feedback displays three unique properties: (*i*) robustness in expression kinetics, (*ii*) robustness in steady-state output, and (*iii*) conversion of graded inputs into a binary response. The three behaviors occur as a result of Tat positive feedback saturation. Intriguingly, many simple feedback circuits could acquire these hallmark characteristics if they operated in the saturation regime. These properties could be ideal in synthetic systems, where a certain portion of the circuit would be required to have absolute concentration robustness (Shinar and Feinberg, 2010).

These results call attention to the potential pitfalls of the HIV shock and kill approach. Stimulation of latently infected cells can lead to activation of Tat positive feedback and robust transitioning of the virus from the latent to actively replicating fate, however, the stochastic nature of latent reactivation suggests that stimulating 100% of latently infected cells would be extremely difficult. Even under conditions where Tat was supplied exogenously in the presence of potent transcriptional activators of the LTR, sub-100% activation was observed (Fig. 52). Since purging 100% of the latent reservoir is necessary for successful treatment, more efficacious intervention mechanisms are

necessary. Additionally, these data implicate that possible targeting of Tat positive feedback to inhibit HIV replication would alter viral dynamics, not by slowing the kinetic rates of viral progression, but by modulating the fraction of infections that enter the active or latent state.

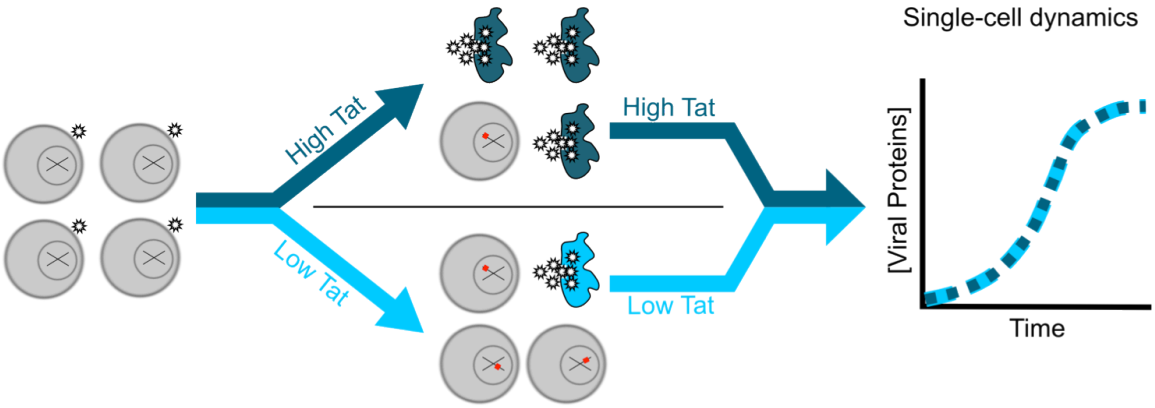

Figure 54. HIV decision-making as an early stochastic and late robust process.

A population of cells infected with HIV will have a higher proportion of active infections dependent on the availability of Tat. The ratio of latent-to-active infections will scale inversely with Tat availability. However, gene-expression dynamics and the steady-state output from the HIV LTR will be the same irrespective of varying Tat concentrations.

Chapter 10: Discussion on HIV latency

1. The HIV latency problem and potential solutions

Due to the prevailing notion that cellular state controls HIV latency, current therapeutic strategies aimed towards eradication of latently infected populations focus on activating cellular state to purge the virus from the latent reservoir (Deeks, 2012). The 'Shock and Kill' approach has shown moderate success with HDACi's (Archin et al., 2012), but previous attempts with cytokines such as IL-2, IL-7, and TNF-α, and other cell-state modifiers proved less fruitful (Ruelas and Greene, 2013). These purging strategies may prove more efficacious if mechanisms to reactivate the latent provirus were also directed towards viral mechanisms that would stabilize expression from the LTR. Perhaps a combinatorial approach attacking both cell state and viral circuitry would improve the effectiveness of these purging strategies.

2. Rethinking HIV latency as a two-stage mechanism

Contrary to current dogma, viruses encoding for two different fluorophores—expression of one fluorophore tracks latent infections, while expression of both fluorophores indicates active replication of the virus—have shown that HIV can enter the latent state immediately after infection (Calvanese et al., 2013; Dahabieh et al., 2013). Previous hypotheses suggested that latency occurs only as cells transition from the activated to resting state (Siliciano and Greene, 2011). Does the establishment of latency and maintenance of the latent state occur through fundamentally distinct mechanisms? While interruption of Tat positive feedback is sufficient for entrance into the latent state

(Weinberger et al., 2005), the cellular based hypothesis of restrictive chromatin, suppressive methylation, promoter occlusion, and limiting cellular resources most likely stabilize and maintain the latent state (Siliciano and Greene, 2011). This two-pronged mechanism could describe the disparate results in the attributed mechanisms of latency establishment because observation of a latently infected population immediately after infection or later on may yield different results. It would be of interest to follow a latent population from the first moment of entrance into the latent state through to longer time points to test if the most optimal reactivation strategy may in fact change over time. For example, activators of cellular state could destabilize the host-cell factors that stabilize latency to reawaken latent provirus, while activators of viral circuitry could inhibit the immediate establishment of a latent population.

References

Ackers, G.K., Johnson, A.D., and Shea, M.A. (1982). Quantitative model for gene regulation by lambda phage repressor. Proc Natl Acad Sci U S A *79*, 1129-1133.

Adachi, A., Gendelman, H.E., Koenig, S., Folks, T., Willey, R., Rabson, A., and Martin, M.A. (1986). Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. J Virol *59*, 284-291.

Adalsteinsson, D., McMillen, D., and Elston, T.C. (2004). Biochemical Network Stochastic Simulator (BioNetS): software for stochastic modeling of biochemical networks. BMC Bioinformatics *5*, 24.

Alm, E., and Baker, D. (1999). Matching theory and experiment in protein folding. Curr Opin Struct Biol *9*, 189-196.

Almogy, G., and Nolan, G.P. (2009). Conditional protein stabilization via the small molecules Shld-1 and rapamycin increases the signal-to-noise ratio with tet-inducible gene expression. Biotechniques *46*, 44-50.

Alon, U. (2007). An introduction to systems biology : design principles of biological circuits (Boca Raton, FL: Chapman & Hall/CRC).

Alon, U., Surette, M.G., Barkai, N., and Leibler, S. (1999). Robustness in bacterial chemotaxis. Nature *397*, 168-171.

Althaus, C.L., De Vos, A.S., and De Boer, R.J. (2009). Reassessing the human immunodeficiency virus type 1 life cycle through age-structured modeling: life span of infected cells, viral generation time, and basic reproductive number, R0. J Virol *83*, 7659-7667.

Anderson, R.M., Gupta, S., and May, R.M. (1991a). Potential of community-wide chemotherapy or immunotherapy to control the spread of HIV-1. Nature *350*, 356-359.

Anderson, R.M., and May, R.M. (1988). Epidemiological parameters of HIV transmission. Nature *333*, 514-519.

Anderson, R.M., and May, R.M. (1996). The population biology of the interaction between HIV-1 and HIV-2: coexistence or competitive exclusion? Aids *10*, 1663-1673.

Anderson, R.M., May, R.M., Boily, M.C., Garnett, G.P., and Rowley, J.T. (1991b). The spread of HIV-1 in Africa: sexual contact patterns and the predicted demographic impact of AIDS. Nature *352*, 581-589.

Anderson, R.M., May, R.M., and McLean, A.R. (1988). Possible demographic consequences of AIDS in developing countries. Nature *332*, 228-234.

Archin, N.M., Liberty, A.L., Kashuba, A.D., Choudhary, S.K., Kuruc, J.D., Crooks, A.M., Parker, D.C., Anderson, E.M., Kearney, M.F., Strain, M.C.*, et al.* (2012). Administration of vorinostat disrupts HIV-1 latency in patients on antiretroviral therapy. Nature *487*, 482-485.

Arkin, A., Ross, J., and McAdams, H.H. (1998). Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected Escherichia coli cells. Genetics *149*, 1633-1648.

Austin, D.W., Allen, M.S., McCollum, J.M., Dar, R.D., Wilgus, J.R., Sayler, G.S., Samatova, N.F., Cox, C.D., and Simpson, M.L. (2006). Gene network shaping of inherent noise spectra. Nature *439*, 608-611.

Bahar, R., Hartmann, C.H., Rodriguez, K.A., Denny, A.D., Busuttil, R.A., Dolle, M.E., Calder, R.B., Chisholm, G.B., Pollock, B.H., Klein, C.A., *et al.* (2006). Increased cell-to-cell variation in gene expression in ageing mouse heart. Nature *441*, 1011-1014.

Balaban, N. (2011). Persistence: mechanisms for triggering and enhancing phenotypic variability. Curr Opin Genet Dev.

Balaban, N.Q., Merrin, J., Chait, R., Kowalik, L., and Leibler, S. (2004). Bacterial persistence as a phenotypic switch. Science *305*, 1622-1625.

Balazsi, G., van Oudenaarden, A., and Collins, J.J. (2011). Cellular decision making and biological noise: from microbes to mammals. Cell *144*, 910-925.

Banaszynski, L.A., Chen, L.C., Maynard-Smith, L.A., Ooi, A.G., and Wandless, T.J. (2006). A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. Cell *126*, 995-1004.

Banerjee, C., Archin, N., Michaels, D., Belkina, A.C., Denis, G.V., Bradner, J., Sebastiani, P., Margolis, D.M., and Montano, M. (2012). BET bromodomain inhibition as a novel strategy for reactivation of HIV-1. J Leukoc Biol *92*, 1147-1154.

Bar-Even, A., Paulsson, J., Maheshri, N., Carmi, M., O'Shea, E., Pilpel, Y., and Barkai, N. (2006a). Noise in protein expression scales with natural protein abundance. Nat Genet *38*, 636-643.

Bar-Even, A., Paulsson, J., Maheshri, N., Carmi, M., O'Shea, E., Pilpel, Y., and Barkai, N. (2006b). Noise in protein expression scales with natural protein abundance. Nature Genet *38*, 636-643.

Barber, G.N. (2001). Host defense, viruses and apoptosis. Cell Death Differ *8*, 113-126.

Barboric, M., Nissen, R.M., Kanazawa, S., Jabrane-Ferrat, N., and Peterlin, B.M. (2001). NF-kappaB binds P-TEFb to stimulate transcriptional elongation by RNA polymerase II. Molecular Cell *8*, 327-337.

Barkai, N., and Leibler, S. (1997). Robustness in simple biochemical networks. Nature *387*, 913-917.

Bartholomeeusen, K., Xiang, Y., Fujinaga, K., and Peterlin, B.M. (2012). Bromodomain and extra-terminal (BET) bromodomain inhibition activate transcription via transient release of positive transcription elongation factor b (P-TEFb) from 7SK small nuclear ribonucleoprotein. J Biol Chem *287*, 36609-36616.

Becskei, A., Seraphin, B., and Serrano, L. (2001). Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. EMBO J *20*, 2528-2535.

Bednarik, D.P., Cook, J.A., and Pitha, P.M. (1990). Inactivation of the HIV LTR by DNA CpG methylation: evidence for a role in latency. EMBO J *9*, 1157-1164.

Blake, W.J., Balazsi, G., Kohanski, M.A., Isaacs, F.J., Murphy, K.F., Kuang, Y., Cantor, C.R., Walt, D.R., and Collins, J.J. (2006). Phenotypic consequences of promoter-mediated transcriptional noise. Mol Cell *24*, 853-865.

Blake, W.J., M, K.A., Cantor, C.R., and Collins, J.J. (2003). Noise in eukaryotic gene expression. Nature *422*, 633-637.

Blazkova, J., Murray, D., Justement, J.S., Funk, E.K., Nelson, A.K., Moir, S., Chun, T.W., and Fauci, A.S. (2012). Paucity of HIV DNA methylation in latently infected, resting CD4+ T cells from infected individuals receiving antiretroviral therapy. J Virol *86*, 5390-5392.

Blazkova, J., Trejbalova, K., Gondois-Rey, F., Halfon, P., Philibert, P., Guiguen, A., Verdin, E., Olive, D., Van Lint, C., Hejnar, J.*, et al.* (2009). CpG methylation controls reactivation of HIV from latency. PLoS Pathog *5*, e1000554.

Boehm, D., Calvanese, V., Dar, R.D., Xing, S., Schroeder, S., Martins, L., Aull, K., Li, P.C., Planelles, V., Bradner, J.E.*, et al.* (2013). BET bromodomain-targeting compounds reactivate HIV from latency via a Tat-independent mechanism. Cell Cycle *12*, 452-462.

Bull, J.J., and Vogt, R.C. (1979). Temperature-dependent sex determination in turtles. Science *206*, 1186-1188.

Burnett, J.C., Miller-Jensen, K., Shah, P.S., Arkin, A.P., and Schaffer, D.V. (2009). Control of stochastic gene expression by host factors at the HIV promoter. PLoS Pathog *5*, e1000260.

Cagatay, T., Turcotte, M., Elowitz, M.B., Garcia-Ojalvo, J., and Suel, G.M. (2009). Architecture-dependent noise discriminates functionally analogous differentiation circuits. Cell *139*, 512-522.

Cai, L., Dalal, C.K., and Elowitz, M.B. (2008). Frequency-modulated nuclear localization bursts coordinate gene regulation. Nature *455*, 485-490.

Cai, L., Friedman, N., and Xie, X.S. (2006). Stochastic protein expression in individual cells at the single molecule level. Nature *440*, 358-362.

Calvanese, V., Chavez, L., Laurent, T., Ding, S., and Verdin, E. (2013). Dual-color HIV reporters trace a population of latently infected cells and enable their purification. Virology *446*, 283-292.

Carvalho, A., Olson, S.K., Gutierrez, E., Zhang, K., Noble, L.B., Zanin, E., Desai, A., Groisman, A., and Oegema, K. (2011). Acute drug treatment in the early C. elegans embryo. PLoS One *6*, e24656.

Chang, H.H., Hemberg, M., Barahona, M., Ingber, D.E., and Huang, S. (2008). Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature *453*, 544-547.

Chao, S.H., and Price, D.H. (2001). Flavopiridol inactivates P-TEFb and blocks most RNA polymerase II transcription in vivo. J Biol Chem *276*, 31793-31799.

Chen, J.J., Huang, J.C., Shirtliff, M., Briscoe, E., Ali, S., Cesani, F., Paar, D., and Cloyd, M.W. (2002). CD4 lymphocytes in the blood of HIV(+) individuals migrate rapidly to lymph nodes and bone marrow: support for homing theory of CD4 cell depletion. J Leukoc Biol *72*, 271-278.

Chu, B.W., Banaszynski, L.A., Chen, L.C., and Wandless, T.J. (2008). Recent progress with FKBP-derived destabilizing domains. Bioorg Med Chem Lett *18*, 5941-5944.

Chun, T.W., Finzi, D., Margolick, J., Chadwick, K., Schwartz, D., and Siliciano, R.F. (1995). In vivo fate of HIV-1-infected T cells: quantitative analysis of the transition to stable latency. Nat Med *1*, 1284-1290.

Coffin, J., and Swanstrom, R. (2013). HIV pathogenesis: dynamics and genetics of viral populations and infected cells. Cold Spring Harb Perspect Med *3*, a012526.

Cohen, A.A., Geva-Zatorsky, N., Eden, E., Frenkel-Morgenstern, M., Issaeva, I., Sigal, A., Milo, R., Cohen-Saidon, C., Liron, Y., Kam, Z.*, et al.* (2008). Dynamic proteomics of individual cancer cells in response to a drug. Science *322*, 1511-1516.
Cook, D.L., Gerber, A.N., and Tapscott, S.J. (1998). Modeling stochastic gene expression: implications for haploinsufficiency. Proc Natl Acad Sci U S A *95*, 15641-15646.
Coull, J.J., Romerio, F., Sun, J.M., Volker, J.L., Galvin, K.M., Davie, J.R., Shi, Y., Hansen, U., and Margolis, D.M. (2000). The human factors YY1 and LSF repress the human immunodeficiency virus type 1 long terminal repeat via recruitment of histone deacetylase 1. J Virol *74*, 6790-6799.
Cox, C.D., McCollum, J.M., Allen, M.S., Dar, R.D., and Simpson, M.L. (2008a). Using noise to probe and characterize gene circuits. Proc Natl Acad Sci U S A *105*, 10809-10814.
Cox, C.D., McCollum, J.M., Allen, M.S., Dar, R.D., and Simpson, M.L. (2008b). Using noise to probe and characterize gene circuits. Proc Natl Acad Sci U S A *105*, 10809-10814.
D'Orso, I., and Frankel, A.D. (2009). Tat acetylation modulates assembly of a viral-host RNA-protein transcription complex. Proc Natl Acad Sci U S A *106*, 3101-3106.
Dahabieh, M.S., Ooms, M., Simon, V., and Sadowski, I. (2013). A doubly fluorescent HIV-1 reporter shows that the majority of integrated HIV-1 is latent shortly after infection. J Virol *87*, 4716-4727.
Dar, R.D., Razooky, B.S., Singh, A., Trimeloni, T.V., McCollum, J.M., Cox, C.D., Simpson, M.L., and Weinberger, L.S. (2012). Transcriptional burst frequency and burst size are equally modulated across the human genome. Proc Natl Acad Sci U S A *109*, 17454-17459.
Deeks, S.G. (2012). HIV: Shock and kill. Nature *487*, 439-440.
Deutsch, M., Deutsch, A., Shirihai, O., Hurevich, I., Afrimzon, E., Shafran, Y., and Zurgil, N. (2006). A novel miniature cell retainer for correlative high-content analysis of individual untethered non-adherent cells. Lab Chip *6*, 995-1000.
Di Carlo, D., Aghdam, N., and Lee, L.P. (2006a). Single-cell enzyme concentrations, kinetics, and inhibition analysis using high-density hydrodynamic cell isolation arrays. Anal Chem *78*, 4925-4930.
Di Carlo, D., Wu, L.Y., and Lee, L.P. (2006b). Dynamic single cell culture array. Lab Chip *6*, 1445-1449.
Dodd, I.B., Perkins, A.J., Tsemitsidis, D., and Egan, J.B. (2001). Octamerization of lambda CI repressor is needed for effective repression of P(RM) and efficient switching from lysogeny. Genes Dev *15*, 3013-3022.
Donahue, D.A., Bastarache, S.M., Sloan, R.D., and Wainberg, M.A. (2013). Latent HIV-1 Can Be Reactivated by Cellular Superinfection in a Tat-Dependent Manner, Which Can Lead to the Emergence of Multidrug-Resistant Recombinant Viruses. J Virol *87*, 9620-9632.
Donahue, D.A., Kuhl, B.D., Sloan, R.D., and Wainberg, M.A. (2012). The viral protein Tat can inhibit the establishment of HIV-1 latency. J Virol *86*, 3253-3263.
Duggal, N.K., and Emerman, M. (2012). Evolutionary conflicts between viruses and restriction factors shape immunity. Nat Rev Immunol *12*, 687-695.

Dull, T., Zufferey, R., Kelly, M., Mandel, R.J., Nguyen, M., Trono, D., and Naldini, L. (1998). A third-generation lentivirus vector with a conditional packaging system. J Virol *72*, 8463-8471.
Dunlop, M.J., Cox, R.S., 3rd, Levine, J.H., Murray, R.M., and Elowitz, M.B. (2008). Regulatory activity revealed by dynamic correlations in gene expression noise. Nat Genet *40*, 1493-1498.
Edelstein-Keshet, L. (1988). Mathematical models in biology, 1st edn (New York: Random House).
Eisele, E., and Siliciano, R.F. (2012). Redefining the viral reservoirs that prevent HIV-1 eradication. Immunity *37*, 377-388.
Eldar, A., Chary, V.K., Xenopoulos, P., Fontes, M.E., Loson, O.C., Dworkin, J., Piggot, P.J., and Elowitz, M.B. (2009). Partial penetrance facilitates developmental evolution in bacteria. Nature *460*, 510-514.
Elowitz, M.B., and Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. Nature *403*, 335-338.
Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. Science *297*, 1183-1186.
Emiliani, S., Fischle, W., Ott, M., Van Lint, C., Amella, C.A., and Verdin, E. (1998). Mutations in the tat gene are responsible for human immunodeficiency virus type 1 postintegration latency in the U1 cell line. J Virol *72*, 1666-1670.
Faley, S., Seale, K., Hughey, J., Schaffer, D.K., VanCompernolle, S., McKinney, B., Baudenbacher, F., Unutmaz, D., and Wikswo, J.P. (2008). Microfluidic platform for real-time signaling analysis of multiple single T cells in parallel. Lab Chip *8*, 1700-1712.
Farrell, M.J., Dobson, A.T., and Feldman, L.T. (1991). Herpes simplex virus latency-associated transcript is a stable intron. Proc Natl Acad Sci U S A *88*, 790-794.
Feinberg, M.B., Baltimore, D., and Frankel, A.D. (1991). The role of Tat in the human immunodeficiency virus life cycle indicates a primary effect on transcriptional elongation. Proc Natl Acad Sci U S A *88*, 4045-4049.
Ferrell, J.E., Jr. (2002). Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. Curr Opin Cell Biol *14*, 140-148.
Ferrell, J.E., Jr. (2008). Feedback regulation of opposing enzymes generates robust, all-or-none bistable responses. Curr Biol *18*, R244-245.
Finzi, D., Blankson, J., Siliciano, J.D., Margolick, J.B., Chadwick, K., Pierson, T., Smith, K., Lisziewicz, J., Lori, F., Flexner, C.*, et al.* (1999). Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. Nat Med *5*, 512-517.
Frankel, A.D. (1992). Activation of HIV transcription by Tat. Curr Opin Genet Dev *2*, 293-298.
Franz, K., Singh, A., and Weinberger, L.S. (2010). Lentiviral vectors to study stochastic noise in gene expression. Methods in Enzymology *accepted*.
Franz, K., Singh, A., and Weinberger, L.S. (2011). Lentiviral vectors to study stochastic noise in gene expression. Methods Enzymol *497*, 603-622.
Fraser, H.B., Hirsh, A.E., Giaever, G., Kumm, J., and Eisen, M.B. (2004). Noise minimization in eukaryotic gene expression. PLoS Biol *2*, e137.

Fujinaga, K., Cujec, T.P., Peng, J., Garriga, J., Price, D.H., Grana, X., and Peterlin, B.M. (1998). The ability of positive transcription elongation factor B to transactivate human immunodeficiency virus transcription depends on a functional kinase domain, cyclin T1, and Tat. J Virol *72*, 7154-7159.
Gardner, T.S., Cantor, C.R., and Collins, J.J. (2000). Construction of a genetic toggle switch in Escherichia coli. Nature *403*, 339-342.
Gatignol, A., Buckler-White, A., Berkhout, B., and Jeang, K.T. (1991). Characterization of a human TAR RNA-binding protein that activates the HIV-1 LTR. Science *251*, 1597-1600.
Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods *6*, 343-345.
Gillespie, D.T. (1976). General Method for Numerically Simulating Stochastic Time Evolution of Coupled Chemical-Reactions. Journal of Computational Physics *22*, 403-434.
Gillespie, D.T. (1977). Exact Stochastic Simulation of Coupled Chemical-Reactions. Journal of Physical Chemistry *81*, 2340-2361.
Gillespie, D.T. (2009). A diffusional bimolecular propensity function. Journal of Chemical Physics *131*, 164109.
Golding, I., Paulsson, J., Zawilski, S.M., and Cox, E.C. (2005). Real-time kinetics of gene activity in individual bacteria. Cell *123*, 1025-1036.
Gong, Y., Ogunniyi, A.O., and Love, J.C. (2010). Massively parallel detection of gene expression in single cells using subnanolitre wells. Lab Chip *10*, 2334-2337.
Gossen, M., and Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proc Natl Acad Sci U S A *89*, 5547-5551.
Greger, I.H., Demarchi, F., Giacca, M., and Proudfoot, N.J. (1998). Transcriptional interference perturbs the binding of Sp1 to the HIV-1 promoter. Nucleic Acids Res *26*, 1294-1301.
Groisman, A., Lobo, C., Cho, H., Campbell, J.K., Dufour, Y.S., Stevens, A.M., and Levchenko, A. (2005). A microfluidic chemostat for experiments with bacterial and yeast cells. Nat Methods *2*, 685-689.
Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77-88.
Guldevall, K., Vanherberghen, B., Frisk, T., Hurtig, J., Christakou, A.E., Manneberg, O., Lindstrom, S., Andersson-Svahn, H., Wiklund, M., and Onfelt, B. (2010). Imaging immune surveillance of individual natural killer cells confined in microwell arrays. PLoS One *5*, e15453.
Gurskaya, N.G., Verkhusha, V.V., Shcheglov, A.S., Staroverov, D.B., Chepurnykh, T.V., Fradkov, A.F., Lukyanov, S., and Lukyanov, K.A. (2006). Engineering of a monomeric green-to-red photoactivatable fluorescent protein induced by blue light. Nat Biotechnol *24*, 461-465.
Han, Q., Bradshaw, E.M., Nilsson, B., Hafler, D.A., and Love, J.C. (2010). Multidimensional analysis of the frequencies and rates of cytokine secretion from single cells by quantitative microengraving. Lab Chip *10*, 1391-1400.

Han, Y., Wind-Rotolo, M., Yang, H.C., Siliciano, J.D., and Siliciano, R.F. (2007). Experimental approaches to the study of HIV-1 latency. Nat Rev Microbiol *5*, 95-106.
Hao, S., and Baltimore, D. (2009). The stability of mRNA influences the temporal order of the induction of genes encoding inflammatory molecules. Nat Immunol *10*, 281-288.
Harper, C.V., Finkenstadt, B., Woodcock, D.J., Friedrichsen, S., Semprini, S., Ashall, L., Spiller, D.G., Mullins, J.J., Rand, D.A., Davis, J.R*., et al.* (2011). Dynamic analysis of stochastic transcription cycles. Plos Biology *9*, e1000607.
Hayden, M.S., and Ghosh, S. (2004a). Signaling to NF-kappaB. Genes Dev *18*, 2195-2224.
Hayden, M.S., and Ghosh, S. (2004b). Signaling to NF-kappaB. Genes & development *18*, 2195-2224.
Herz, A.V., Bonhoeffer, S., Anderson, R.M., May, R.M., and Nowak, M.A. (1996). Viral dynamics in vivo: limitations on estimates of intracellular delay and virus decay. Proc Natl Acad Sci U S A *93*, 7247-7251.
Ho, D.D., Neumann, A.U., Perelson, A.S., Chen, W., Leonard, J.M., and Markowitz, M. (1995). Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature *373*, 123-126.
Ho, Y.C., Shan, L., Hosmane, N.N., Wang, J., Laskey, S.B., Rosenbloom, D.I., Lai, J., Blankson, J.N., Siliciano, J.D., and Siliciano, R.F. (2013). Replication-Competent Noninduced Proviruses in the Latent Reservoir Increase Barrier to HIV-1 Cure. Cell *155*, 540-551.
Hochschild, A., Douhan, J., 3rd, and Ptashne, M. (1986). How lambda repressor and lambda Cro distinguish between OR1 and OR3. Cell *47*, 807-816.
Hochschild, A., and Ptashne, M. (1988). Interaction at a distance between lambda repressors disrupts gene activation. Nature *336*, 353-357.
Hoffmann, A., Levchenko, A., Scott, M.L., and Baltimore, D. (2002). The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. Science *298*, 1241-1245.
Huang, L.M., Joshi, A., Willey, R., Orenstein, J., and Jeang, K.T. (1994). Human immunodeficiency viruses regulated by alternative trans-activators: genetic evidence for a novel non-transcriptional function of Tat in virion infectivity. EMBO J *13*, 2886-2896.
Huh, D., and Paulsson, J. (2011). Non-genetic heterogeneity from stochastic partitioning at cell division. Nat Genet *43*, 95-100.
Ingram, P.J., Stumpf, M.P., and Stark, J. (2008). Nonidentifiability of the source of intrinsic noise in gene expression from single-burst data. PLoS Comput Biol *4*, e1000192.
Jeeninga, R.E., Westerhout, E.M., van Gerven, M.L., and Berkhout, B. (2008). HIV-1 latency in actively dividing human T cell lines. Retrovirology *5*, 37.
Johnson, A.D., Poteete, A.R., Lauer, G., Sauer, R.T., Ackers, G.K., and Ptashne, M. (1981). lambda Repressor and cro--components of an efficient molecular switch. Nature *294*, 217-223.
Jones, T.R., Kang, I.H., Wheeler, D.B., Lindquist, R.A., Papallo, A., Sabatini, D.M., Golland, P., and Carpenter, A.E. (2008). CellProfiler Analyst: data exploration and analysis software for complex image-based screens. BMC Bioinformatics *9*, 482.

Jordan, A., Bisgrove, D., and Verdin, E. (2003). HIV reproducibly establishes a latent infection after acute infection of T cells in vitro. EMBO J *22*, 1868-1877.
Jordan, A., Defechereux, P., and Verdin, E. (2001). The site of HIV-1 integration in the human genome determines basal transcriptional activity and response to Tat transactivation. EMBO J *20*, 1726-1738.
Kaern, M., Elston, T.C., Blake, W.J., and Collins, J.J. (2005). Stochasticity in gene expression: from theories to phenotypes. Nat Rev Genet *6*, 451-464.
Kagan, J.C., Su, T., Horng, T., Chow, A., Akira, S., and Medzhitov, R. (2008). TRAM couples endocytosis of Toll-like receptor 4 to the induction of interferon-beta. Nat Immunol *9*, 361-368.
Kao, S.Y., Calman, A.F., Luciw, P.A., and Peterlin, B.M. (1987). Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. Nature *330*, 489-493.
Kartalov, E.P., Walker, C., Taylor, C.R., Anderson, W.F., and Scherer, A. (2006). Microfluidic vias enable nested bioarrays and autoregulatory devices in Newtonian fluids. Proc Natl Acad Sci U S A *103*, 12280-12284.
Kauder, S.E., Bosque, A., Lindqvist, A., Planelles, V., and Verdin, E. (2009). Epigenetic regulation of HIV-1 latency by cytosine methylation. PLoS Pathog *5*, e1000495.
Kemkemer, R., Schrank, S., Vogel, W., Gruler, H., and Kaufmann, D. (2002). Increased noise as an effect of haploinsufficiency of the tumor-suppressor gene neurofibromatosis type 1 in vitro. Proc Natl Acad Sci U S A *99*, 13783-13788.
Kepler, T.B., and Elston, T.C. (2001a). Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations. Biophys J *81*, 3116-3136.
Kepler, T.B., and Elston, T.C. (2001b). Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. Biophys J *81*, 3116-3136.
Khademhosseini, A., Yeh, J., Jon, S., Eng, G., Suh, K.Y., Burdick, J.A., and Langer, R. (2004). Molded polyethylene glycol microstructures for capturing cells within microfluidic channels. Lab Chip *4*, 425-430.
Kim, D.W., Uetsuki, T., Kaziro, Y., Yamaguchi, N., and Sugano, S. (1990). Use of the human elongation factor 1 alpha promoter as a versatile and efficient expression system. Gene *91*, 217-223.
Kim, S., Ikeuchi, K., Byrn, R., Groopman, J., and Baltimore, D. (1989). Lack of a negative influence on viral growth by the nef gene of human immunodeficiency virus type 1. Proc Natl Acad Sci U S A *86*, 9544-9548.
Kobel, S., Valero, A., Latt, J., Renaud, P., and Lutolf, M. (2010). Optimization of microfluidic single cell trapping for long-term on-chip culture. Lab Chip *10*, 857-863.
Kuerbitz, S.J., Plunkett, B.S., Walsh, W.V., and Kastan, M.B. (1992). Wild-type p53 is a cell cycle checkpoint determinant following irradiation. Proc Natl Acad Sci U S A *89*, 7491-7495.
Kwon, H.S., Brent, M.M., Getachew, R., Jayakumar, P., Chen, L.F., Schnolzer, M., McBurney, M.W., Marmorstein, R., Greene, W.C., and Ott, M. (2008). Human immunodeficiency virus type 1 Tat protein inhibits the SIRT1 deacetylase and induces T cell hyperactivation. Cell Host Microbe *3*, 158-167.
Lahav, G., Rosenfeld, N., Sigal, A., Geva-Zatorsky, N., Levine, A.J., Elowitz, M.B., and Alon, U. (2004). Dynamics of the p53-Mdm2 feedback loop in individual cells. Nat Genet *36*, 147-150.

Laspia, M.F., Rice, A.P., and Mathews, M.B. (1989). HIV-1 Tat protein increases transcriptional initiation and stabilizes elongation. Cell *59*, 283-292.
Lassen, K., Han, Y., Zhou, Y., Siliciano, J., and Siliciano, R.F. (2004). The multifactorial nature of HIV-1 latency. Trends Mol Med *10*, 525-531.
Lassen, K.G., Ramyar, K.X., Bailey, J.R., Zhou, Y., and Siliciano, R.F. (2006). Nuclear retention of multiply spliced HIV-1 RNA in resting CD4+ T cells. PLoS Pathog *2*, e68.
Lauffenburger, D.A., and Linderman, J.J. (1993). Receptors : models for binding, trafficking, and signaling (New York: Oxford University Press).
Lenasi, T., Contreras, X., and Peterlin, B.M. (2008). Transcriptional interference antagonizes proviral gene expression to promote HIV latency. Cell Host Microbe *4*, 123-133.
Li, P.C., de Camprieu, L., Cai, J., and Sangar, M. (2004). Transport, retention and fluorescent measurement of single biological cells studied in microfluidic chips. Lab Chip *4*, 174-180.
Li, X., Zhao, X., Fang, Y., Jiang, X., Duong, T., Fan, C., Huang, C.C., and Kain, S.R. (1998). Generation of destabilized green fluorescent protein as a transcription reporter. J Biol Chem *273*, 34970-34975.
Lindstrom, S., Mori, K., Ohashi, T., and Andersson-Svahn, H. (2009). A microwell array device with integrated microfluidic components for enhanced single-cell analysis. Electrophoresis *30*, 4166-4171.
Lo, M.Y., Rival-Gervier, S., Pasceri, P., and Ellis, J. (2012). Rapid transcriptional pulsing dynamics of high expressing retroviral transgenes in embryonic stem cells. PLoS ONE *7*, e37130.
Locke, J.C., and Elowitz, M.B. (2009). Using movies to analyse gene circuit dynamics in single cells. Nat Rev Microbiol *7*, 383-392.
Losick, R., and Desplan, C. (2008). Stochasticity and cell fate. Science *320*, 65-68.
Love, J.C., Ronan, J.L., Grotenbreg, G.M., van der Veen, A.G., and Ploegh, H.L. (2006). A microengraving method for rapid selection of single cells producing antigen-specific antibodies. Nat Biotechnol *24*, 703-707.
Maamar, H., Raj, A., and Dubnau, D. (2007). Noise in gene expression determines cell fate in Bacillus subtilis. Science *317*, 526-529.
Maeda, Y.T., and Sano, M. (2006). Regulatory dynamics of synthetic gene networks with positive feedback. J Mol Biol *359*, 1107-1124.
May, R.M., and Anderson, R.M. (1987). Transmission dynamics of HIV infection. Nature *326*, 137-142.
Mayo, A.E., Setty, Y., Shavit, S., Zaslaver, A., and Alon, U. (2006). Plasticity of the cis-regulatory input function of a gene. PLoS Biol *4*, e45.
McAdams, H.H., and Arkin, A. (1997). Stochastic mechanisms in gene expression. Proc Natl Acad Sci U S A *94*, 814-819.
Mcquarri, D.A. (1967). Stochastic Approach to Chemical Kinetics. Journal of Applied Probability *4*, 413-&.
Miller-Jensen, K., Skupsky, R., Shah, P.S., Arkin, A.P., and Schaffer, D.V. (2013). Genetic selection for context-dependent stochastic phenotypes: Sp1 and TATA mutations increase phenotypic noise in HIV-1 gene expression. PLoS Comput Biol *9*, e1003135.

Mitchell, R.S., Beitzel, B.F., Schroder, A.R., Shinn, P., Chen, H., Berry, C.C., Ecker, J.R., and Bushman, F.D. (2004). Retroviral DNA Integration: ASLV, HIV, and MLV Show Distinct Target Site Preferences. PLoS Biol *2*, E234.
Mizuguchi, H., Xu, Z., Ishii-Watabe, A., Uchida, E., and Hayakawa, T. (2000). IRES-dependent second gene expression is significantly lower than cap-dependent first gene expression in a bicistronic vector. Mol Ther *1*, 376-382.
Mizutani, T., Ishizaka, A., Tomizawa, M., Okazaki, T., Yamamichi, N., Kawana-Tachikawa, A., Iwamoto, A., and Iba, H. (2009). Loss of the Brm-type SWI/SNF chromatin remodeling complex is a strong barrier to the Tat-independent transcriptional elongation of human immunodeficiency virus type 1 transcripts. J Virol *83*, 11569-11580.
Munsky, B., Neuert, G., and van Oudenaarden, A. (2012). Using gene expression noise to understand gene regulation. Science *336*, 183-187.
Munsky, B., Trinh, B., and Khammash, M. (2009). Listening to the noise: random fluctuations reveal gene network parameters. Mol Syst Biol *5*, 318.
Muramoto, T., Cannon, D., Gierlinski, M., Corrigan, A., Barton, G.J., and Chubb, J.R. (2012). Live imaging of nascent RNA dynamics reveals distinct types of transcriptional pulse regulation. Proc Natl Acad Sci U S A.
Murray, J.D. (2002). Mathematical biology, 3rd edn (New York: Springer).
Nelson, W.G., and Kastan, M.B. (1994). DNA strand breaks: the DNA template alterations that trigger p53-dependent DNA damage response pathways. Mol Cell Biol *14*, 1815-1823.
Newman, J.R., Ghaemmaghami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L., and Weissman, J.S. (2006). Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature *441*, 840-846.
Nowak, M.A., Anderson, R.M., Boerlijst, M.C., Bonhoeffer, S., May, R.M., and McMichael, A.J. (1996). HIV-1 evolution and disease progression. Science *274*, 1008-1011.
Nowak, M.A., and May, R.M. (2000). Virus dynamics : mathematical principles of immunology and virology (Oxford ; New York: Oxford University Press).
Ogunniyi, A.O., Story, C.M., Papa, E., Guillen, E., and Love, J.C. (2009). Screening individual hybridomas by microengraving to discover monoclonal antibodies. Nat Protoc *4*, 767-782.
Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.
Osborn, L., Kunkel, S., and Nabel, G.J. (1989). Tumor necrosis factor alpha and interleukin 1 stimulate the human immunodeficiency virus enhancer by activation of the nuclear factor kappa B. Proc Natl Acad Sci U S A *86*, 2336-2340.
Ott, M., Geyer, M., and Zhou, Q. (2011). The control of HIV transcription: keeping RNA polymerase II on track. Cell Host Microbe *10*, 426-435.
Ozawa, T., Kinoshita, K., Kadowaki, S., Tajiri, K., Kondo, S., Honda, R., Ikemoto, M., Piao, L., Morisato, A., Fukurotani, K*., et al.* (2009). MAC-CCD system: a novel lymphocyte microwell-array chip system equipped with CCD scanner to generate human monoclonal antibodies against influenza virus. Lab Chip *9*, 158-163.

Ozbudak, E.M., Thattai, M., Kurtser, I., Grossman, A.D., and van Oudenaarden, A. (2002). Regulation of noise in the expression of a single gene. Nat Genet *31*, 69-73.
Ozbudak, E.M., Thattai, M., Lim, H.N., Shraiman, B.I., and Van Oudenaarden, A. (2004). Multistability in the lactose utilization network of Escherichia coli. Nature *427*, 737-740.
Pagans, S., Pedal, A., North, B.J., Kaehlcke, K., Marshall, B.L., Dorr, A., Hetzer-Egger, C., Henklein, P., Frye, R., McBurney, M.W.*, et al.* (2005). SIRT1 regulates HIV transcription via Tat deacetylation. PLoS Biol *3*, e41.
Palsson, B.O., Keasling, J.D., and Emerson, S.G. (1990). The regulatory mechanisms of human immunodeficiency virus replication predict multiple expression rates. Proc Natl Acad Sci U S A *87*, 772-776.
Paulsson, J. (2004). Summing up the noise in gene networks. Nature *427*, 415-418.
Paulsson, J. (2005). Models of stochastic gene expression. Physics of Life Reviews *2*, 157-175.
Pazin, M.J., Sheridan, P.L., Cannon, K., Cao, Z., Keck, J.G., Kadonaga, J.T., and Jones, K.A. (1996). NF-kappa B-mediated chromatin reconfiguration and transcriptional activation of the HIV-1 enhancer in vitro. Genes Dev *10*, 37-49.
Pearson, R., Kim, Y.K., Hokello, J., Lassen, K., Friedman, J., Tyagi, M., and Karn, J. (2008). Epigenetic silencing of human immunodeficiency virus (HIV) transcription by formation of restrictive chromatin structures at the viral long terminal repeat drives the progressive entry of HIV into latency. J Virol *82*, 12291-12303.
Pedraza, J.M., and Paulsson, J. (2008). Effects of molecular memory and bursting on fluctuations in gene expression. Science *319*, 339-343.
Pelletier, J., and Sonenberg, N. (1988). Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. Nature *334*, 320-325.
Perelson, A.S., Essunger, P., Cao, Y., Vesanen, M., Hurley, A., Saksela, K., Markowitz, M., and Ho, D.D. (1997a). Decay characteristics of HIV-1-infected compartments during combination therapy. Nature *387*, 188-191.
Perelson, A.S., Essunger, P., and Ho, D.D. (1997b). Dynamics of HIV-1 and CD4+ lymphocytes in vivo. AIDS *11 Suppl A*, S17-24.
Perelson, A.S., Neumann, A.U., Markowitz, M., Leonard, J.M., and Ho, D.D. (1996). HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. Science *271*, 1582-1586.
Pessler, F., and Cron, R.Q. (2004). Reciprocal regulation of the nuclear factor of activated T cells and HIV-1. Genes Immun *5*, 158-167.
Polinkovsky, M., Gutierrez, E., Levchenko, A., and Groisman, A. (2009). Fine temporal control of the medium gas content and acidity and on-chip generation of series of oxygen concentrations for cell cultures. Lab Chip *9*, 1073-1084.
Ptashne, M. (2004). A genetic switch : phage lambda revisited, 3rd edn (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press).
Ptashne, M., and Gann, A. (2002). Genes & signals (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).
Ptashne, M., Johnson, A.D., and Pabo, C.O. (1982). A genetic switch in a bacterial virus. Sci Am *247*, 128-130, 132, 134-140.
Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. PLoS Biol *4*, e309.

Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. Cell *135*, 216-226.
Ramezani, A., Hawley, T.S., and Hawley, R.G. (2000). Lentiviral vectors for enhanced gene expression in human hematopoietic cells. Mol Ther *2*, 458-469.
Raser, J.M., and O'Shea, E.K. (2004). Control of stochasticity in eukaryotic gene expression. Science *304*, 1811-1814.
Raser, J.M., and O'Shea, E.K. (2005). Noise in gene expression: origins, consequences, and control. Science *309*, 2010-2013.
Razooky, B.S., Gutierrez, E., Terry, V.H., Spina, C.A., Groisman, A., and Weinberger, L.S. (2012). Microwell devices with finger-like channels for long-term imaging of HIV-1 expression kinetics in primary human lymphocytes. Lab Chip *12*, 4305-4312.
Razooky, B.S., and Weinberger, L.S. (2011). Mapping the architecture of the HIV-1 Tat circuit: A decision-making circuit that lacks bistability and exploits stochastic noise. Methods *53*, 68-77.
Reddy, B., and Yin, J. (1999). Quantitative intracellular kinetics of HIV type 1. AIDS Res Hum Retroviruses *15*, 273-283.
Reuse, S., Calao, M., Kabeya, K., Guiguen, A., Gatot, J.S., Quivy, V., Vanhulle, C., Lamine, A., Vaira, D., Demonte, D.*, et al.* (2009). Synergistic activation of HIV-1 expression by deacetylase inhibitors and prostratin: implications for treatment of latent infection. PLoS One *4*, e6093.
Richman, D.D., Margolis, D.M., Delaney, M., Greene, W.C., Hazuda, D., and Pomerantz, R.J. (2009). The challenge of finding a cure for HIV infection. Science *323*, 1304-1307.
Ridley, A.J., Schwartz, M.A., Burridge, K., Firtel, R.A., Ginsberg, M.H., Borisy, G., Parsons, J.T., and Horwitz, A.R. (2003). Cell migration: integrating signals from front to back. Science *302*, 1704-1709.
Ruelas, D.S., and Greene, W.C. (2013). An integrated overview of HIV-1 latency. Cell *155*, 519-529.
Saez-Cirion, A., Nicola, M.A., Pancino, G., and Shorte, S.L. (2006). Quantitative real-time analysis of HIV-1 gene expression dynamics in single living primary cells. Biotechnol J *1*, 682-689.
Sakane, N., Kwon, H.S., Pagans, S., Kaehlcke, K., Mizusawa, Y., Kamada, M., Lassen, K.G., Chan, J., Greene, W.C., Schnoelzer, M.*, et al.* (2011). Activation of HIV transcription by the viral Tat protein requires a demethylation step mediated by lysine-specific demethylase 1 (LSD1/KDM1). PLoS pathogens *7*, e1002184.
Sanchez, A., Garcia, H.G., Jones, D., Phillips, R., and Kondev, J. (2011). Effect of promoter architecture on the cell-to-cell variability in gene expression. PLoS Comput Biol *7*, e1001100.
Savageau, M.A. (1976). Biochemical systems analysis : a study of function and design in molecular biology (Reading, Mass.: Addison-Wesley Pub. Co., Advanced Book Program).
Schiffenbauer, Y.S., Kalma, Y., Trubniykov, E., Gal-Garber, O., Weisz, L., Halamish, A., Sister, M., and Berke, G. (2009). A cell chip for sequential imaging of individual non-adherent live cells reveals transients and oscillations. Lab Chip *9*, 2965-2972.
Schroder, A.R., Shinn, P., Chen, H., Berry, C., Ecker, J.R., and Bushman, F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. Cell *110*, 521-529.

Schrump, D.S., Matthews, W., Chen, G.A., Mixon, A., and Altorki, N.K. (1998). Flavopiridol mediates cell cycle arrest and apoptosis in esophageal cancer cells. Clin Cancer Res *4*, 2885-2890.
Seth, N., Kaufmann, D., Lahey, T., Rosenberg, E.S., and Wucherpfennig, K.W. (2005). Expansion and contraction of HIV-specific CD4 T cells with short bursts of viremia, but physical loss of the majority of these cells with sustained viral replication. J Immunol *175*, 6948-6958.
Shan, L., Yang, H.C., Rabi, S.A., Bravo, H.C., Shroff, N.S., Irizarry, R.A., Zhang, H., Margolick, J.B., Siliciano, J.D., and Siliciano, R.F. (2011). Influence of host gene transcription level and orientation on HIV-1 latency in a primary-cell model. J Virol *85*, 5384-5393.
Shapiro, G.I., Koestner, D.A., Matranga, C.B., and Rollins, B.J. (1999). Flavopiridol induces cell cycle arrest and p53-independent apoptosis in non-small cell lung cancer cell lines. Clin Cancer Res *5*, 2925-2938.
Shinar, G., and Feinberg, M. (2010). Structural sources of robustness in biochemical reaction networks. Science *327*, 1389-1391.
Sible, J.C., and Tyson, J.J. (2007). Mathematical modeling as a tool for investigating cell cycle control networks. Methods *41*, 238-247.
Siciliano, J.D., and Siciliano, R.F. (2004). A long-term latent reservoir for HIV-1: discovery and clinical implications. J Antimicrob Chemother *54*, 6-9.
Sigal, A., Milo, R., Cohen, A., Geva-Zatorsky, N., Klein, Y., Liron, Y., Rosenfeld, N., Danon, T., Perzov, N., and Alon, U. (2006). Variability and memory of protein levels in human cells. Nature *444*, 643-646.
Siliciano, R.F., and Greene, W.C. (2011). HIV Latency. Cold Spring Harb Perspect Med *1*, a007096.
Simpson, M.L., Cox, C.D., and Sayler, G.S. (2004a). Frequency domain chemical Langevin analysis of stochasticity in gene transcriptional regulation. J Theor Biol *229*, 383-394.
Simpson, M.L., Cox, C.D., and Sayler, G.S. (2004b). Frequency domain chemical Langevin analysis of stochasticity in gene transcriptional regulation. J Theor Biol *229*, 383-394.
Singh, A., Razooky, B., Cox, C.D., Simpson, M.L., and Weinberger, L.S. (2010a). Transcriptional bursting from the HIV-1 promoter is a significant source of stochastic noise in HIV-1 gene expression. Biophys J *98*, L32-34.
Singh, A., Razooky, B., Cox, C.D., Simpson, M.L., and Weinberger, L.S. (2010b). Transcriptional Bursting from the HIV-1 Promoter Is a Significant Source of Stochastic Noise in HIV-1 Gene Expression. Biophys J *98*, L32-L34.
Singh, A., Razooky, B., Dar, R.D., and Weinberger, L.S. (In Review, 2012). Dynamics of protein noise can distinguish between alternative sources of gene-expression variability Mol Syst Biol.
Singh, A., Razooky, B.S., Dar, R.D., and Weinberger, L.S. (2012). Dynamics of protein noise can distinguish between alternate sources of gene-expression variability. Mol Syst Biol *8*, 607.
Singh, A., and Weinberger, L.S. (2009). Stochastic gene expression as a molecular switch for viral latency. Curr Opin Microbiol *12*, 460-466.

Skupsky, R., Burnett, J.C., Foley, J.E., Schaffer, D.V., and Arkin, A.P. (2010). HIV promoter integration site primarily modulates transcriptional burst size rather than frequency. PLoS computational biology *6*.
So, L.H., Ghosh, A., Zong, C., Sepulveda, L.A., Segev, R., and Golding, I. (2011). General properties of transcriptional time series in Escherichia coli. Nature Genetics *43*, 554-560.
Sobell, H.M. (1985). Actinomycin and DNA transcription. Proc Natl Acad Sci U S A *82*, 5328-5331.
Spina, C.A., Anderson, J., Archin, N.M., Bosque, A., Chan, J., Famiglietti, M., Greene, W.C., Kashuba, A., Lewin, S.R., Margolis, D.M.*, et al.* (2013). An in-depth comparison of latent HIV-1 reactivation in multiple cell model systems and resting CD4+ T cells from aviremic patients. PLoS Pathog *9*, e1003834.
Srivastava, R., You, L., Summers, J., and Yin, J. (2002). Stochastic vs. deterministic modeling of intracellular viral kinetics. J Theor Biol *218*, 309-321.
St-Pierre, F., and Endy, D. (2008). Determination of cell fate selection during phage lambda infection. Proc Natl Acad Sci U S A *105*, 20705-20710.
Suel, G.M., Garcia-Ojalvo, J., Liberman, L.M., and Elowitz, M.B. (2006). An excitable gene regulatory circuit induces transient cellular differentiation. Nature *440*, 545-550.
Sureka, K., Ghosh, B., Dasgupta, A., Basu, J., Kundu, M., and Bose, I. (2008). Positive feedback and noise activate the stringent response regulator rel in mycobacteria. PLoS One *3*, e1771.
Suter, D.M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., and Naef, F. (2011a). Mammalian genes are transcribed with widely different bursting kinetics. Science *332*, 472-474.
Suter, D.M., Molina, N., Naef, F., and Schibler, U. (2011b). Origins and consequences of transcriptional discontinuity. Current opinion in cell biology *23*, 657-662.
Swain, P.S., Elowitz, M.B., and Siggia, E.D. (2002). Intrinsic and extrinsic contributions to stochasticity in gene expression. Proc Natl Acad Sci U S A *99*, 12795-12800.
Tan, R.Z., and van Oudenaarden, A. (2010). Transcript counting in single cells reveals dynamics of rDNA transcription. Mol Syst Biol *6*, 358.
Taniguchi, Y., Choi, P.J., Li, G.W., Chen, H.Y., Babu, M., Hearn, J., Emili, A., and Xie, X.S. (2010). Quantifying E-coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. Science *329*, 533-538.
Teng, M.W., Bolovan-Fritts, C., Dar, R.D., Womack, A., Simpson, M.L., Shenk, T., and Weinberger, L.S. (2012). An endogenous accelerator for viral gene expression confers a fitness advantage. Cell *151*, 1569-1580.
Terry, V.H., Johnston, I.C., and Spina, C.A. (2009). CD44 microbeads accelerate HIV-1 infection in T cells. Virology *388*, 294-304.
Thattai, M., and van Oudenaarden, A. (2001). Intrinsic noise in gene regulatory networks. Proc Natl Acad Sci U S A *98*, 8614-8619.
Thorslund, S., Larsson, R., Bergquist, J., Nikolajeff, F., and Sanchez, J. (2008). A PDMS-based disposable microfluidic sensor for CD4+ lymphocyte counting. Biomed Microdevices *10*, 851-857.

Ting, A.Y., Kain, K.H., Klemke, R.L., and Tsien, R.Y. (2001). Genetically encoded fluorescent reporters of protein tyrosine kinase activities in living cells. Proc Natl Acad Sci U S A *98*, 15003-15008.
To, T.L., and Maheshri, N. (2010). Noise can induce bimodality in positive transcriptional feedback loops without bistability. Science *327*, 1142-1145.
Tokimitsu, Y., Kishi, H., Kondo, S., Honda, R., Tajiri, K., Motoki, K., Ozawa, T., Kadowaki, S., Obata, T., Fujiki, S.*, et al.* (2007). Single lymphocyte analysis with a microwell array chip. Cytometry A *71*, 1003-1010.
Tyagi, M., Pearson, R.J., and Karn, J. (2010). Establishment of HIV latency in primary CD4+ cells is due to epigenetic transcriptional silencing and P-TEFb restriction. J Virol *84*, 6425-6437.
Upton, J.P., Valentijn, A.J., Zhang, L., and Gilmore, A.P. (2007). The N-terminal conformation of Bax regulates cell commitment to apoptosis. Cell Death Differ *14*, 932-942.
Vallabhapurapu, S., and Karin, M. (2009). Regulation and function of NF-kappaB transcription factors in the immune system. Annu Rev Immunol *27*, 693-733.
Van Lint, C., Emiliani, S., Ott, M., and Verdin, E. (1996). Transcriptional activation and chromatin remodeling of the HIV-1 promoter in response to histone acetylation. EMBO J *15*, 1112-1120.
VanLint, C., Emiliani, S., Ott, M., and Verdin, E. (1996). Transcriptional activation and chromatin remodeling of the HIV-1 promoter in response to histone acetylation. Embo Journal *15*, 1112-1120.
Volkov, Y., Long, A., and Kelleher, D. (1998). Inside the crawling T cell: leukocyte function-associated antigen-1 cross-linking is associated with microtubule-directed translocation of protein kinase C isoenzymes beta(I) and delta. J Immunol *161*, 6487-6495.
von Andrian, U.H., and Mempel, T.R. (2003). Homing and cellular traffic in lymph nodes. Nat Rev Immunol *3*, 867-878.
Warmflash, A., and Dinner, A.R. (2008). Signatures of combinatorial regulation in intrinsic biological noise. Proc Natl Acad Sci U S A *105*, 17262-17267.
Wei, P., Garber, M.E., Fang, S.M., Fischer, W.H., and Jones, K.A. (1998). A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. Cell *92*, 451-462.
Wei, X., Ghosh, S.K., Taylor, M.E., Johnson, V.A., Emini, E.A., Deutsch, P., Lifson, J.D., Bonhoeffer, S., Nowak, M.A., Hahn, B.H.*, et al.* (1995). Viral dynamics in human immunodeficiency virus type 1 infection. Nature *373*, 117-122.
Weinberger, A.D., Perelson, A.S., Ribeiro, R.M., and Weinberger, L.S. (2009). Accelerated immunodeficiency by anti-CCR5 treatment in HIV infection. PLoS Comput Biol *5*, e1000467.
Weinberger, L.S., Burnett, J.C., Toettcher, J.E., Arkin, A.P., and Schaffer, D.V. (2005). Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. Cell *122*, 169-182.
Weinberger, L.S., Dar, R.D., and Simpson, M.L. (2008a). Transient-mediated fate determination in a transcriptional circuit of HIV. Nat Genet *40*, 466-470.
Weinberger, L.S., Dar, R.D., and Simpson, M.L. (2008b). Transient-mediated fate determination in a transcriptional circuit of HIV. Nature Genet *40*, 466-470.

Weinberger, L.S., Schaffer, D.V., and Arkin, A.P. (2003). Theoretical design of a gene therapy to prevent AIDS but not human immunodeficiency virus type 1 infection. J Virol *77*, 10028-10036.
Weinberger, L.S., and Shenk, T. (2007). An HIV feedback resistor: auto-regulatory circuit deactivator and noise buffer. PLoS Biol *5*, e9.
West, M.J., Lowe, A.D., and Karn, J. (2001). Activation of human immunodeficiency virus transcription in T cells revisited: NF-kappaB p65 stimulates transcriptional elongation. Journal of virology *75*, 8524-8537.
Williams, S.A., Chen, L.F., Kwon, H., Ruiz-Jarabo, C.M., Verdin, E., and Greene, W.C. (2006a). NF-kappaB p50 promotes HIV latency through HDAC recruitment and repression of transcriptional initiation. EMBO J *25*, 139-149.
Williams, S.A., Chen, L.F., Kwon, H., Ruiz-Jarabo, C.M., Verdin, E., and Greene, W.C. (2006b). NF-kappaB p50 promotes HIV latency through HDAC recruitment and repression of transcriptional initiation. The EMBO journal *25*, 139-149.
Williams, S.A., Kwon, H., Chen, L.F., and Greene, W.C. (2007a). Sustained induction of NF-kappa B is required for efficient expression of latent human immunodeficiency virus type 1. Journal of virology *81*, 6043-6056.
Williams, S.A., Kwon, H., Chen, L.F., and Greene, W.C. (2007b). Sustained induction of NF-kappa B is required for efficient expression of latent human immunodeficiency virus type 1. J Virol *81*, 6043-6056.
Wilson, D.P., Coplan, P.M., Wainberg, M.A., and Blower, S.M. (2008). The paradoxical effects of using antiretroviral-based microbicides to control HIV epidemics. Proc Natl Acad Sci U S A *105*, 9835-9840.
Wolthers, K.C., Noest, A.J., Otto, S.A., Miedema, F., and De Boer, R.J. (1999). Normal telomere lengths in naive and memory CD4+ T cells in HIV type 1 infection: a mathematical interpretation. AIDS Res Hum Retroviruses *15*, 1053-1062.
Wong, J.V., Yao, G., Nevins, J.R., and You, L. (2011). Viral-mediated noisy gene expression reveals biphasic E2f1 response to MYC. Mol Cell *41*, 275-285.
Yamamura, S., Kishi, H., Tokimitsu, Y., Kondo, S., Honda, R., Rao, S.R., Omori, M., Tamiya, E., and Muraguchi, A. (2005). Single-cell microarray for analyzing cellular response. Anal Chem *77*, 8050-8056.
Yang, X., Forouzan, O., Burns, J.M., and Shevkoplyas, S.S. (2011). Traffic of leukocytes in microfluidic channels with rectangular and rounded cross-sections. Lab Chip.
Yen, H.C., Xu, Q., Chou, D.M., Zhao, Z., and Elledge, S.J. (2008). Global protein stability profiling in mammalian cells. Science *322*, 918-923.
Ylisastigui, L., Archin, N.M., Lehrman, G., Bosch, R.J., and Margolis, D.M. (2004). Coaxing HIV-1 from resting CD4 T cells: histone deacetylase inhibition allows latent viral expression. Aids *18*, 1101-1108.
Yoshida, M., Kijima, M., Akita, M., and Beppu, T. (1990). Potent and specific inhibition of mammalian histone deacetylase both in vivo and in vitro by trichostatin A. J Biol Chem *265*, 17174-17179.
Yukl, S., Pillai, S., Li, P., Chang, K., Pasutti, W., Ahlgren, C., Havlir, D., Strain, M., Gunthard, H., Richman, D*., et al.* (2009). Latently-infected CD4+ T cells are enriched for HIV-1 Tat variants with impaired transactivation activity. Virology *387*, 98-108.
Yunger, S., Rosenfeld, L., Garini, Y., and Shav-Tal, Y. (2010). Single-allele analysis of transcription kinetics in living mammalian cells. Nature Methods *7*, 631-U641.

Zeng, L., Skinner, S.O., Zong, C., Sippy, J., Feiss, M., and Golding, I. (2010). Decision making at a subcellular level determines the outcome of bacteriophage infection. Cell *141*, 682-691.
Zenklusen, D., Larson, D.R., and Singer, R.H. (2008). Single-RNA counting reveals alternative modes of gene expression in yeast. Nat Struct Mol Biol *15*, 1263-1271.
Zhang, L., Gurskaya, N.G., Merzlyak, E.M., Staroverov, D.B., Mudrik, N.N., Samarkina, O.N., Vinokurov, L.M., Lukyanov, S., and Lukyanov, K.A. (2007). Method for real-time monitoring of protein degradation at the single cell level. Biotechniques *42*, 446, 448, 450.
Zhou, P. (2004). Determining protein half-lives. Methods Mol Biol *284*, 67-77.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

BL RG

Author Signature

4/28/14

Date