UNIVERSITY OF CALIFORNIA, SAN DIEGO

Serial Post-Transcriptional Splicing Establishes Negative Feedback Regulation in HIV-1 Expression

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Bioinformatics and Systems Biology

by

Yi Wen

Committee in charge:

Professor Leor Weinberger, Chair
Professor Douglas Richman, Co-Chair
Professor John Guatelli
Professor Alexander Hoffmann
Professor Gurol Suel

2013

The Dissertation of Yi Wen is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Chair

University of California, San Diego

2013

DEDICATION

This dissertation is dedicated to my beloved family, my husband, and my future son. To my grandmother A-pen Yeh-Lee, who brought me up till I was 7 years old, for giving me an optimistic personality. To my father Chan-hwa Wen and mother His-ju Chen, for educating me and supporting me with their love and understanding. To my husband Panos, for always there for me when I felt discouraged with my research, and for supporting me with his love, patience, and cheerful "humor". I also want to delicate this dissertation to my future son, who is still in my belly at the moment, for making me feel strong and invincible to all the challenges in the last couple of months of my Ph.D. life.

I would like to delicate this dissertation to two other very special people in my life. First, to my junior high school teacher Chin-ling Lin, for her mentorship that continues to deeply influence my character for the rest of my life, even though I am not under her guidance any more. Last, I want to dedicate my dissertation to my deceased friend, Joanne Chen, who was once my best and closest friend. Without her, I would have had a very different life, probably a less happy one, and would not have come to UCSD for my Ph.D.

EPIGRAPH

Stop wasting your time being sorry,
just do it.

-----Daniele Micciancio

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

BlaM: Beta-Lactamase

CD: Cluster of Differentiation

CTD: Carboxy-Terminal Domain

CV: Coefficient of Variation

d2GFP or d2G: Destabilized Green Fluorescent Protein

DAPI: 4',6-Diamidino-2-Phenylindole

DC: Dendritic Cell

FBS: Fetal Bovine Serum

GM-CSF: Granulocyte Macrophage Colony Stimulating Factor

HIV-1: Human Immunodeficiency Virus Type 1

IRES: Internal Ribosome Entry Site

Lepto B: Leptomycin B

LTR: Long Terminal Repeat

MDM: Monocyte Derived Macrophage

MOI: Multiplicity of Infection

MS: Multiply Spliced

MSD: Major Splice Donor

ODE: Ordinary Differential Equation

ORF: Open Reading Frame

PBS: Phosphate Buffered Saline

PenStrep: Penicillin and Streptomycin

PFA: Paraformaldehyde

RNA: Ribonucleic Acid

RNAP II: RNA Polymerase II

RPMI: Roswell Park Memorial Institute

RRE: Rev-Responsive Element

SATURN: Splicing After Transfection of Unspliced pre-mRNA into Nucleus

smFISH: Single Molecule Fluorescent *in situ* Hybridization

snRNP: Small Nuclear Ribonucleic Particle

SS: Singly Spliced

TAMRA: Carboxytetramethylrhodamine

Tat-SF1: Tat Specific Factor 1

TC: Transcription Center

TNF-α: Tumor Necrotic Factor Alpha

TSA: Trichostatin A

US: Unspliced

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank all the people who have helped me in my past graduate years. First and most importantly, I want to thank my advisor Leor Weinberger for his support and mentorship for the past 5 years. It was a unique experience to work with a young and ambitious professor who had just started a new lab from scratch. I deeply appreciate his giving me lots of freedom in research and the chance to develop the ability of independent research.

I would like to acknowledge all the former and current Weinberger lab mates, Brandon Razooky, Melissa Teng, Roy Dar, Cynthia Bolovan-Fritts, Lisa Bishop, Igor Rouzine, Tim Notton, Jac Luna, Josep Sardanyes, Renee Ram, Grayson Kochi, Anand Pai, Rachel Tsai, Zhaleh Amini, Jeff Sasaki, and Marleen Kawahara, for their support and useful discussions, both scientific and non-scientific. I am also grateful to the people in the Gladstone Institute, especially Mauricio Montano, Marielle Cavrois, Simon Chu, Jason Neidleman, and Isa Munoz-Arias, who have helped me in learning several new techniques, or generously provided experimental reagents.

Additionally, I deeply appreciate the help from the collaborators I have worked with: John Young, Sumit Chanda, Kevin Olivieri, Lars Pache, Scott Rifkin, Alex Hoffmann, Vincent Shih, Ana Fernandez-Sesma, Dabeiba Bernal, for their advice and discussions, and especially the HINT (HIV Immune Networks Team) P01 project for the funding.

I want to say many thanks to my friends in my graduate life: to my Taiwanese friends, Wan-yen Lo, Alice Huang, Kuei-Chun Hsu, Chin-Yao Kuo, Yen-Lin Lee, Angel Lee, To-ju Huang, and Chi-chieh Chien for their help and friendship in my San Diego life; to my Bioinf pals, Josue Perez, Vipul Bhargava, Jian Wang for their friendship and support for the first few years in the Ph.D. program; to my "Taken" friends, Patrick Verkaik and Ruomei Gao, Marisol Chang and Nikolaos Trogkanis, Mary Pacold and Levon Budagyan, Gjergji Zyba and Mirjan Zyba, and Didem Unat, for their company and lively randomness; last but not least, to my Greek friends Lenia Dritsoula and Petros Venatis, for friendship, help, and participation in several very special events in my life.

Most importantly, I would like to present my deepest appreciation to my parents for supporting me and tolerating my occasional bad temper in the past few years. I would like to thank my aunt Carol Sui in Texas for being my amazing friend and "Mom in the US." I want to thank my husband, Panos, for being my best friend and biggest supporter during these years. Last, I want to thank my future son, for his cooperation in being a nice baby, so I can still work as a busy graduate student.

Chapters 2–5 are based on unpublished work, in which I was the primary researcher and Brandon Razooky, Roy Dar, and Leor Weinberger were critical contributors. Specific contributions made by each person will be detailed in the acknowledgement sections of each chapter. Special thanks go to Brandon Razooky, Roy Dar, and Mary Pacold for reviewing this dissertation.

VITA

2002	Bachelor of Science, Department of Botany, National University of Taiwan, Taipei, Taiwan
2002–2004	Research Assistant, Institute of Biochemical Sciences, National University of Taiwan, Taipei, Taiwan
2004	Master of Science, Institute of Biochemical Sciences, National University of Taiwan, Taipei, Taiwan
2004–2005	Research Assistant, Institute of Agricultural Research, Academia Sinica, Taipei, Taiwan
2006–2008	Teaching Assistant, University of California, San Diego
2008–2013	Research Assistant, University of California, San Diego
2013	Doctor of Philosophy, Bioinformatics Graduate Program, University of California, San Diego

FIELDS OF STUDY

Major Field: Biophysics (Systems Biology)

Studies in Quantitative Virology
Professor Leor Weinberger

ABSTRACT OF THE DISSERTATION

Serial Post-Transcriptional Splicing
Establishes Negative Feedback Regulation
in HIV-1 Expression

by

Yi Wen

Doctor of Philosophy in Bioinformatics and Systems Biology

University of California, San Diego, 2013

Professor Leor Weinberger, Chair
Professor Douglas Richman, Co-Chair

Alternative splicing is critical to many cellular functions and appears to be primarily regulated through co-transcriptional mechanisms, where spliced-transcript identity is determined during transcriptional elongation. For HIV-1, alternative

splicing is essential since the viral genome encodes multiple overlapping reading frames. Using quantitative single-cell imaging, we find that HIV-1 splicing is not co-transcriptionally regulated. Strikingly, HIV-1 mRNA appears to be spliced post-transcriptionally in a serial cascade. We propose a sensitive new assay to detect *in vivo* post-transcriptional splicing. The assay result shows that HIV-1 mRNA does not require transcription for spliceosome assembly. This serial splicing cascade appears crucial for the HIV-1 Rev negative-feedback circuit and in agreement with mathematical models. We experimentally verify the counter-intuitive prediction that Rev over-expression leads to reduced HIV-1 p24 expression. Serial post-transcriptional splicing may provide a novel target for antivirals and, more generally, an alternate gene-regulatory mechanism.

CHAPTER 1: INTRODUCTION

In eukaryotic cells, splicing is a general mechanism for pre-mRNA processing. Accumulating evidence suggests that the vast majority of pre-mRNA is spliced co-transcriptionally (Ameur et al., 2011; Carrillo Oesterreich et al., 2010; Girard et al., 2012; Khodor et al., 2011) (Figure 1), despite post-transcriptional introns being more enriched in alternative splicing (Khodor et al., 2012). While our understanding of the co-transcriptional splicing mechanism remains incomplete, splicing factors are recruited onto the polyphosphorylated C-terminal domain (CTD) of RNA polymerase II (RNAPII) during transcriptional elongation and subsequently facilitate loading of the spliceosome onto the elongating polymerase (David and Manley, 2011; Hsin and Manley, 2012; Martins et al., 2011). Despite this co-transcriptional recruitment of splicing factors, the splicing reaction is not necessarily an obligate co-transcriptional process (Listerman et al., 2006; Tardiff et al., 2006). A kinetic competition between the rates of transcriptional elongation and splicing may play a major role in determining if pre-mRNA splicing proceeds co-transcriptionally or post-transcriptionally (Braunschweig et al., 2013; Carrillo Oesterreich et al., 2011; Kornblihtt et al., 2013); transcriptional pausing is directly associated with co-transcriptional splicing (Alexander et al., 2010; Carrillo Oesterreich et al., 2010; Vargas et al., 2011). Determining the molecular mechanisms of splicing could lead to the identification of new classes of therapeutic targets.

Viruses, in particular, appear to exploit alternative splicing as a common strategy. Diverse viruses, including adenovirus (Davison et al., 2003), simian vacuolating virus 40 (SV40) (Butel and Lednicky, 1999), influenza virus (Lamb and Horvath, 1991), human cytomegalovirus (hCMV) (Fields et al., 2007; Karn and Stoltzfus, 2012), and human immunodeficiency virus type 1 (HIV-1) (Karn and Stoltzfus, 2012), utilize splicing to express their genetic programs. One obvious benefit of alternative splicing in viruses is that it increases the capacity of a small genome to encode multiple open reading frames (ORFs). For example, in HIV-1, nine genes are encoded on different reading frames in a ~9-kb-long genome, with the expression of each gene requiring splicing at a specific pattern of alternate splice donor-acceptor pairs. Nascent pre-mRNA transcripts are alternatively spliced into three classes of variants based on length: an ~9-kb long unspliced (US) class, ~4-kb long singly spliced (SS) class, and a ~2-kb long multiply spliced (MS) class (Figure 2). The splicing mechanism of HIV-1 pre-mRNA is thought to be co-transcriptional (Bohne and Krausslich, 2004; Fong and Zhou, 2001; Jablonski et al., 2010; Zhang et al., 1996). The pre-mRNA splicing of HIV-1 is coupled with transcription through human transcription elongation factor Tat-SF1. The complex of Tat-SF1 and U small nuclear ribonucleoproteins (U snRNPs) is reported to stimulate splicing *in vitro*, suggesting HIV-1 pre-mRNA splicing to be co-transcriptional (Fong and Zhou, 2001).

Despite the obvious benefit of alternative splicing, it also carries a cost, especially for viruses with RNA genomes: unspliced RNAs (i.e., full-length genomic transcripts) tend to be retained in the nucleus and nuclear export of these unspliced

genomic RNAs is far less efficient that spliced RNAs (i.e., sub-genomic transcripts) (Luo and Reed, 1999; Valencia et al., 2008). For example, HIV-1 requires that its unspliced RNA be exported to the cytoplasm for production of viral capsid and other structural proteins, and also so that genomic RNA can be packaged into progeny viruses. To overcome the RNA export problem, HIV-1 exploits the cellular RNA nuclear machinery using the viral Rev protein. Rev is a regulatory gene encoded on MS RNA, which homomultimerizes and binds to an RNA element, the Rev-responsive element (RRE), residing in the intron region of US and SS RNA to export US and SS RNA to the cytoplasm (Brice et al., 1999; Cook et al., 1991; Daelemans et al., 2004; Daugherty et al., 2010; Jain and Belasco, 2001; Malim and Cullen, 1991). Rev further interacts with the cellular export receptor I (CRM1) and facilitates the nuclear export of intron-containing RNA (Fornerod et al., 1997; Fukuda et al., 1997; Ossareh-Nazari et al., 1997), enabling the virus to express structure genes on the intron-containing mRNA, such as Env, Gag, and Pol, and also transport the unspliced genomic RNA into cytoplasm to be packed into progeny virions. As a consequence, Rev is required for HIV-1 replication (Figure 3).

Interestingly, this export of unspliced RNAs by Rev has been implicated in the formation of a negative-feedback loop. (Felber et al., 1990; Felber et al., 1989; Kim and Yin, 2005; Malim et al., 1988). The negative feedback is the result of a precursor-product relationship where US and SS RNA serve as templates for MS RNA. When Rev begins to deplete nuclear US and SS RNA by export, the level of MS RNA begins to decrease, thereby reducing the expression of Rev and the export capacity. Critically,

the negative-feedback model (Figure 5) and the conventional co-transcriptional splicing model (Figure 4) where the identity of an RNA (US, SS, or MS) is fixed during transcriptional elongation do not appear compatible.

To address this question, here we examine if the combinatorial effects of post-transcriptional splicing cascade and Rev RNA nuclear export generate negative feedback regulation or if HIV-1 pre-mRNA is spliced co-transcriptionally. An array of single-cell imaging approaches and a new assay, Splicing After Transfection of Unspliced pre-mRNA into Nucleus (SATURN), demonstrate that HIV-1 splicing occurs post-transcriptionally *in vivo* and is critical for maintaining Rev negative feedback. This finding provides novel insights into the development of new antiviral strategies, and more fundamentally, the eukaryotic splicing mechanisms.

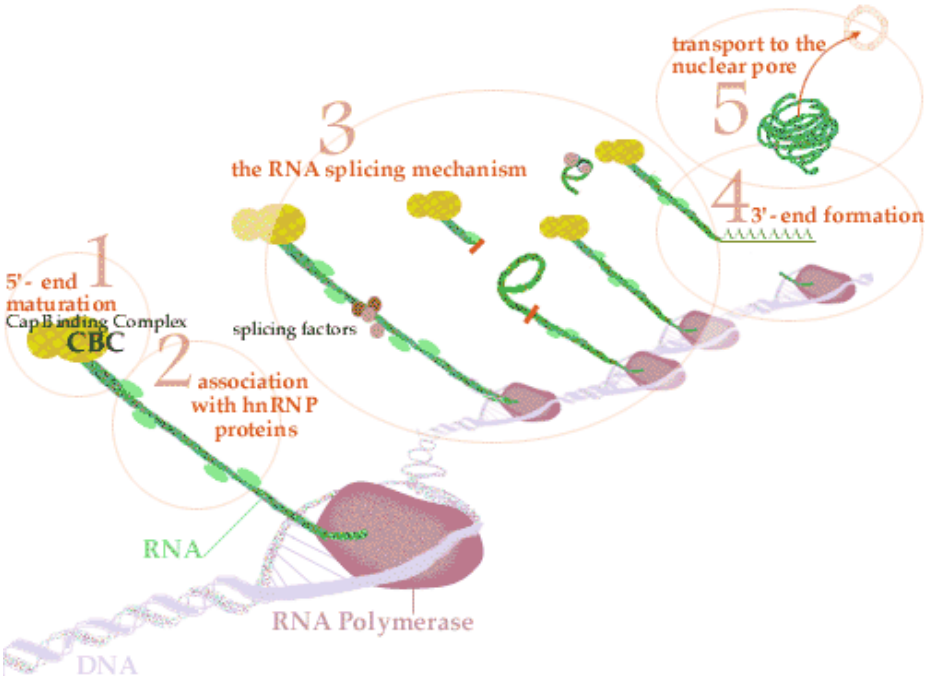

Figure 1. Splicing is co-transcriptional and functionally coupled to transcription.

The canonical biogenesis of eukaryotic mRNA has five steps. The 5' end of nascent RNA transcript is chemically modified (capped) during the initiation of transcription (step 1). Later hnRNA proteins are recruited to the C-terminal domain (CTD) of the elongating polymerase II, and then the intron is spliced upon transcription (steps 2 and 3). Last, 3' end is transcribed and polyadenylated (polyA-tailed) (step 4), and then the matured, processed mRNA is exported into the cytoplasm (step 5). (Figure courtesy of http://www.nobelprize.org/)

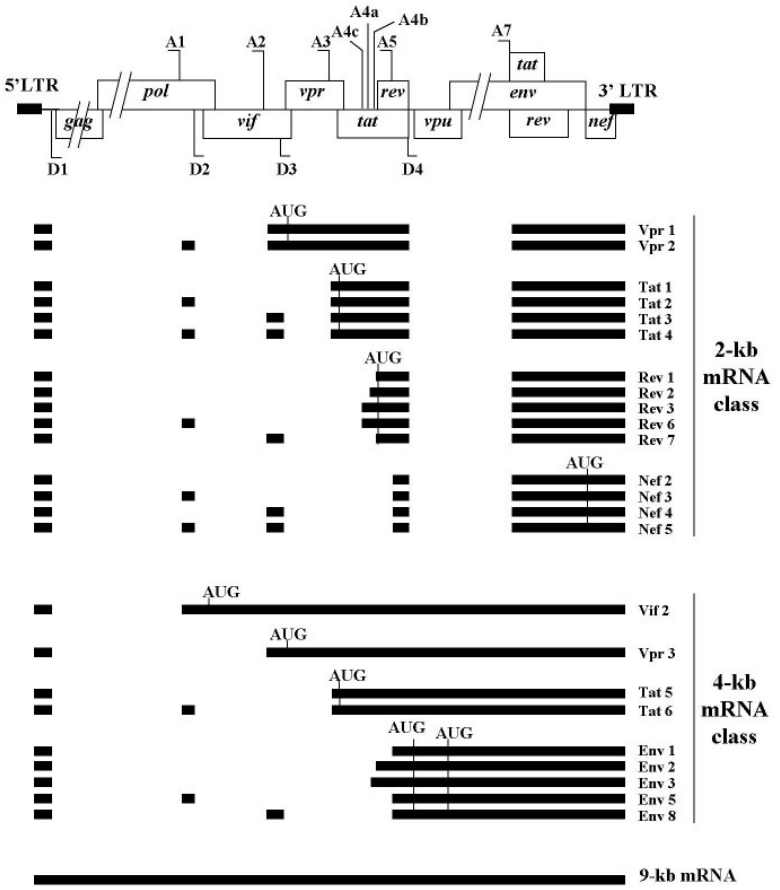

Figure 2. Genome structure of HIV-1 and its splicing variants.

The HIV-1 LTR (long terminal repeat) promoter drives the expression of nine genes encoded in different reading frames. Three classes of splicing variants are made from the provirus: 2-kb multiply spliced (MS) mRNA, 4-kb singly spliced (SS) mRNA, and 9-kb unspliced (US) mRNA. Figure courtesy of Jacquenet et al. (Jacquenet et al., 2005)

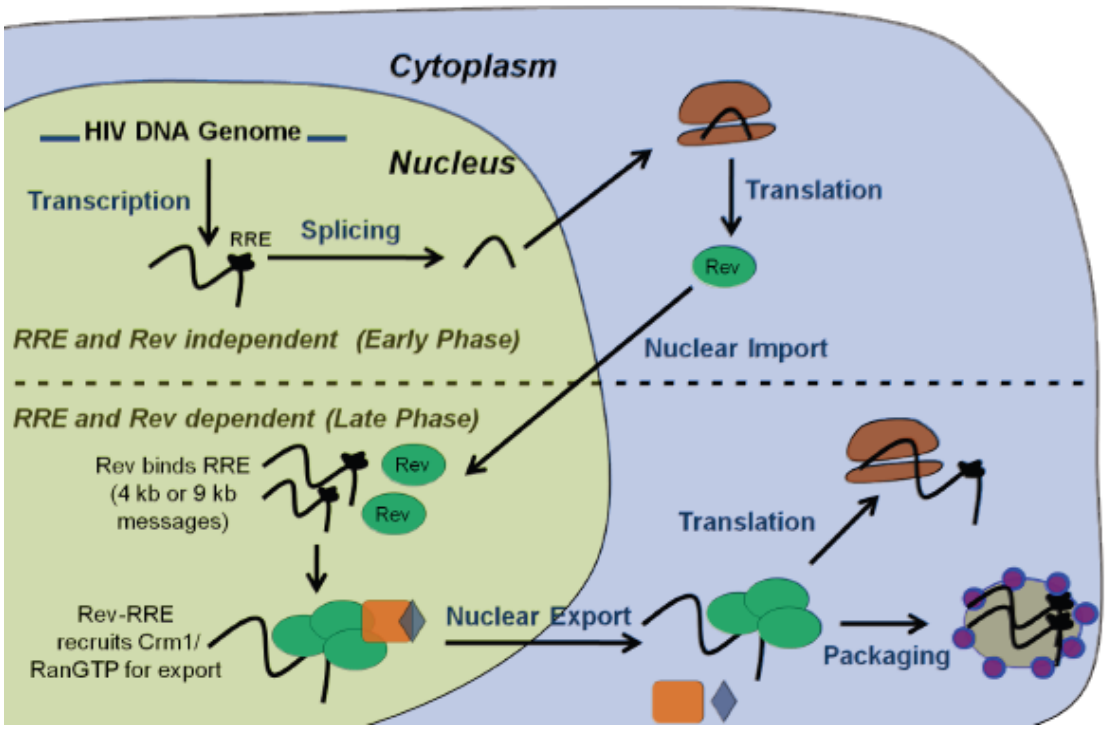

Figure 3. Rev/RRE-dependent RNA nuclear export is required for viral replication.

In the early phase of HIV-1 replication, Rev is expressed from the MS mRNA and shuttled into the nucleus to interact with the RRE, residing on 4-kb SS and 9-kb US mRNA. The Rev/RRE complex recruits CRM1 nuclear export machinery to translocate incompletely spliced RNA into the cytoplasm and enable downstream events of viral replication. Figure courtesy of Wikipedia (http://en.m.wikipedia.org/wiki/File:HIV_REV_RRE_FUNCTION.png)

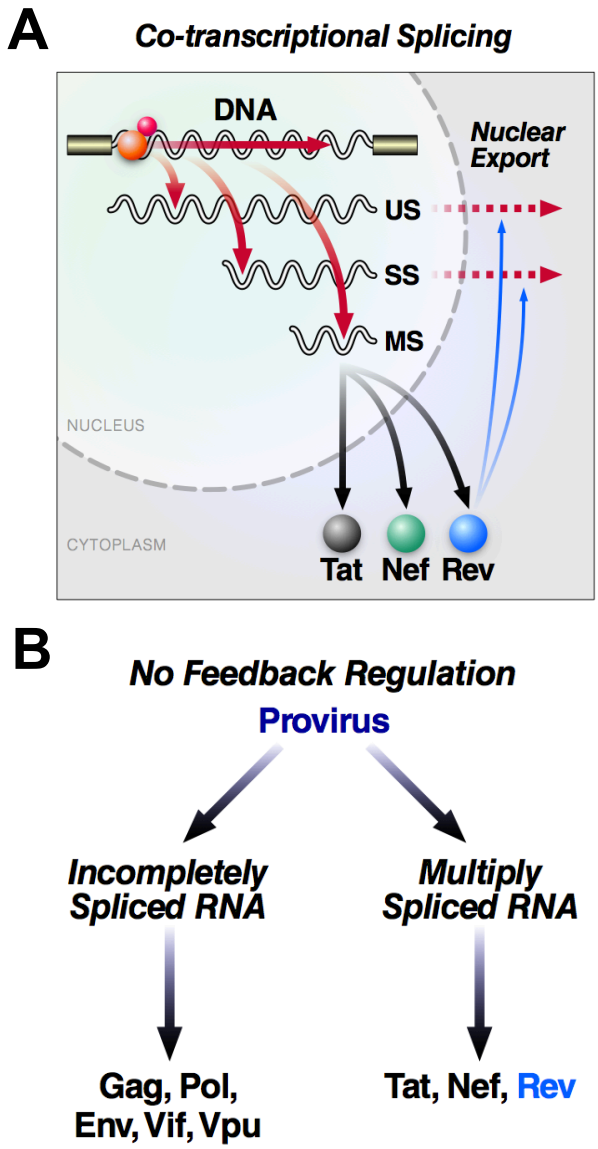

Figure 4. Alternate hypothesis 1: Co-transcriptional splicing can not generate feedback regulation on HIV-1 gene expression.

(A) In the 'co-transcriptional splicing' model, unspliced RNA (US), singly spliced (SS) RNA, and multiply spliced (MS) RNA are transcribed and spliced co-transcriptionally. As a result, the production of MS RNA is not dependent on any incompletely spliced forms of RNA, since pre-mRNA fate is determined during transcription. Tat, Nef, and Rev are translated from MS RNA. Rev facilitates the nuclear export of US and SS intron-containing RNA that is released after transcription. (B) Because the level of MS RNA is not dependent on the incompletely spliced RNA, when Rev starts to remove intron-containing US and SS RNA from nucleus, it does not affect the production of MS RNA, and thus no feedback regulation occurs.

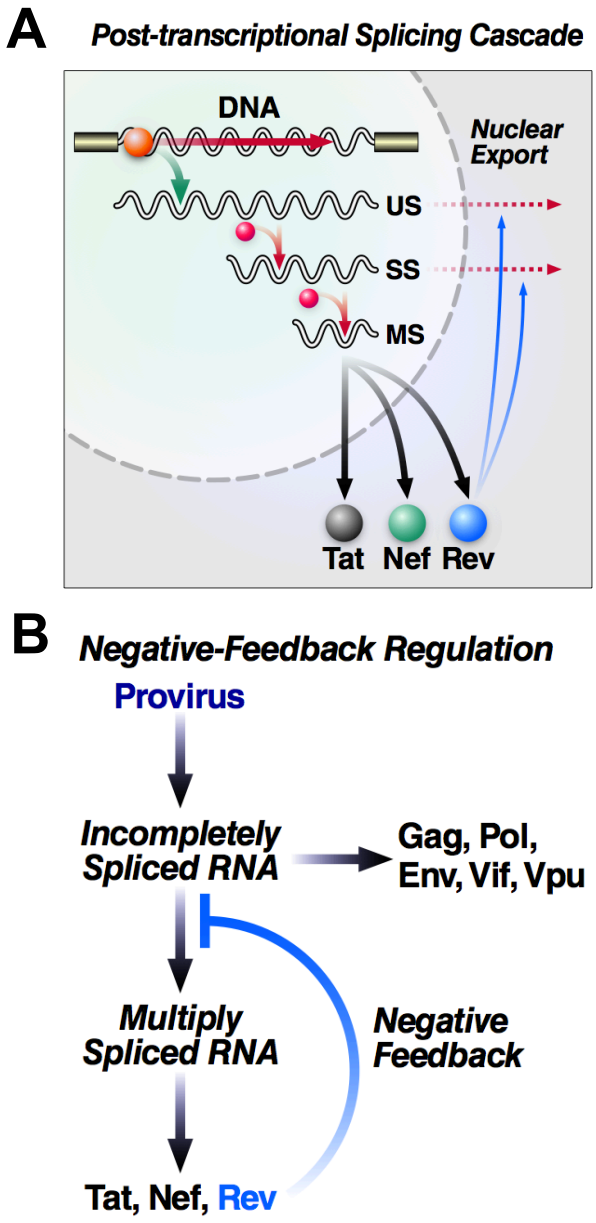

Figure 5. Alternate hypothesis 2: Post-transcriptional splicing cascade establishes negative-feedback regulation on HIV-1 gene expression.

(A) In the 'post-transcriptional splicing cascade' model, US RNA is transcribed from the proviral DNA template. SS RNA is spliced from US RNA post-transcriptionally, and further spliced into MS RNA. As a result, the production of spliced RNA is dependent on the previous incompletely spliced forms of RNA. (B) The production of MS RNA is dependent on incompletely spliced RNA. When Rev-mediated RNA nuclear export starts, the level of nuclear incompletely spliced RNA drops, and thus reduces the production of MS RNA. Ultimately the production of Rev decreases, generating negative feedback.

CHAPTER 2: POST-TRANSCRIPTIONAL SPLICING CASCADE IS REQUIRED TO GENERATE THE REV NEGATIVE FEEDBACK

Abstract

In this chapter, mathematical models representing the proposed co-transcriptional versus post-transcriptional splicing are described in detail. We tested the models against experimentally derived single-cell fluorescent time-lapse microscopy trajectories of viral gene-expression. Single-cell trajectories of HIV-1 gene expression in diverse cell types display the canonical "overshoot" feature of negative feedback. We further examine the properties of this overshoot and negative feedback via trajectory slope analysis and noise analysis. The results demonstrate that Rev-dependent RNA nuclear export establishes negative feedback regulation in HIV-1 gene expression.

Materials and Methods

Cloning

Lentiviral vector Ld2G, Ld2GT, dHIV-d2G and HIV-d2G were created as described (Razooky et al., 2012a). To create the mCherry-RRE lentiviral vector, BamHI-IRES2 fragment was PCR amplified and fused to mCherry (Clonetech)-XbaI PCR fragment. The BamHI/XbaI IRES2-mCherry PCR fusion fragment was inserted into the BamHI/XbaI restriction digested backbone of pNL-GFP-RRE (Wu et al., 2007b). For primer sequences, see Table 1 in the Appendix 2-1.

Cell Culture and Cell Lines

VSV-G pseudo-typed mCherry-RRE, Ld2G, Ld2GIT, HIVΔ-d2G and HIV-d2G lentiviruses were made and packaged in 293FT cells as described (Jordan et al., 2003; Weinberger et al., 2005). Isoclonal and polyclonal Jurkat populations were developed by infecting cells with mCherry-RRE, Ld2G, Ld2GIT, dHIV-d2G or HIV-d2G lentivirus at a low MOI and performing single-cell and bulk sorting as described (Jordan et al., 2003; Weinberger et al., 2005). Jurkat cells were cultured in RPMI-1640 medium + L-glutamine, 10% FBS, and 1% PenStrep at 37°C, 5% CO_2, in humidified conditions at between 2×10^5 to 2×10^6 cells/mL. To activate HIV-1 transduced Jurkat cells, 10 ng/mL TNF-α (Sigma) or 400 nM TSA (Sigma) was added into the culture. To inhibit the CRM1 nuclear export, 0.3 ng/mL Leptomycin B (Sigma) was added to TNF-α-treated cultures. Primary CD4+ T lymphocytes were isolated from peripheral blood, activated and infected as described (Razooky et al., 2012b). Primary

CD4+ T lymphocytes were cultured in RPMI 1640 + L-glutamine, 10% FBS, and 1% PenStrep at 37°C, 5% CO_2, in humidified conditions at between 1×10^6 to 5×10^6 cells/mL. Positively selected CD14+CD4+ primary monocytes were isolated from buffy coats of healthy donors (Roan et al., 2009) as generous gifts from Dr. Warner Greene. Primary monocytes were cultured at 5×10^5 cells/mL in DC medium (RPMI-1640 medium with 10% FBS, 1% PenStrep, 2 mM L-glutamine, and 1 mM sodium pyruvate) (Ramos et al., 2011). To induce the differentiation, 1000 U/mL human GM-CSF (Peprotech) was added to CD14+CD4+ primary monocytes culture every 2 or 3 days post isolation (Ramos et al., 2011). Fully differentiated primary monocyte-derived macrophages (primary MDMs) were used for imaging purpose at day 10 post isolation.

Infection and Imaging

Live-cell imaging of Jurkat cells was performed in humidified conditions at 37 °C and 5% CO_2 for 12–24 h with a 40X (1.3 N.A.) oil-immersion objective on an Zeiss Observer Z1 microscope equipped with 488-nm laser as described (Dar et al., 2012). Activation, infection, and imaging of primary CD4+ T lymphocytes were performed as described in a microdevice (Razooky et al., 2012b). Infection and imaging of primary MDMs was performed by differentiating ~200,000 primary monocytes (see above) in 8-well chambered cover-glass dish (Nunc Lab-Tek #155411). On day 10 post isolation, GM-CSF was washed off with cold DC medium. Cells were pre-chilled to 4 °C, and a high titer (MOI > 10) of HIV-d2G lentivirus was added on ice for 30 minutes to synchronize infection. Infected primary MDMs were

imaged in humidified conditions at 37 °C and 5% CO_2 for 72~100 h with a 20X/0.8 APO air objective on a Zeiss Observer Z1 microscope. Jurkat and primary CD4+ T lymphocyte images were segmented as described using a custom MATLAB code (Dar et al., 2012; Razooky et al., 2012b). Single-cell trajectories of infected primary MDMs were manually tracked and calculated using in-house MATLAB code (available upon request). The general trend lines shown in Figure 8–11, and 13–14 were obtained by synchronizing the initial GFP rise of each single-cell trajectory *in silico*.

Mathematical Modeling and Computational Analysis

Ordinary differential equation (ODE) models were constructed and simulated using Berkeley Madonna (http://www.berkeleymadonna.com/) and MATLAB. For details, see Appendix 2-2. General trend lines obtained from microscopic experiments were fit using Berkeley Madonna. Slope analysis and noise analysis of trajectories were performed with MATLAB.

Results and Discussion

Two scenarios of HIV-1 gene expression were proposed. In the first, if viral nascent mRNA is co-transcriptionally transcribed, the production of SS and MS RNA is not dependent on the nuclear levels of their precursor forms (i.e., US and SS RNA) (Figure 4A in Chapter 2). Consequently, Rev-mediated RNA nuclear export will not affect the processing of HIV-1 pre-mRNA (Figure 4B in Chapter 2). We first examined a simplified ODE model adopted from a previously proposed model (Kim and Yin, 2005) (Appendix 2-2) and predicted that the dynamics of gene products from MS RNA rises after activation and eventually reaches steady state (Figure 6A).

In the second scenario, HIV-1 pre-mRNA undergoes post-transcriptional splicing, and the production of MS RNA is dependent on the precursor US and SS RNA levels (Figure 5A in Chapter 2). Therefore, Rev-mediated RNA nuclear export causes decreased production of MS RNA, which leads to a reduction in Rev levels, resulting in negative-feedback control (Figure 5B in Chapter 2). Mathematical models (Appendix 2-2) predict that the dynamics of gene products of short transcripts (MS RNA) will exhibit the canonical "overshoot" feature of negative feedback (Figure 6B) (Alon, 2007).

HIV-1 Gene Expression Exhibits Rev Negative-Feedback Regulation in Diverse Cell Types

To investigate which mechanism HIV-1 utilizes, we used single-cell time-lapse fluorescent microscopy to examine the expression dynamics of HIV-1 infected cells.

Latently infected Jurkat cells transduced with HIV-d2G construct, a full-length HIV-1 carrying 2-hour half-life destabilized GFP (d2GFP or d2G) replacing Nef reading frame (Figure 7A), were stimulated with TNF-α. The single-cell expression trajectories display the overshoot-turnover trajectory (Figure 8). The averaged single-cell trajectory (n=67) indicates the expression dynamics follow the post-transcriptional splicing model (Figure 5).

We were next interested in what was minimally sufficient to generate the overshoot behavior. To test this, we examined the expression dynamics of Jurkat cells transduced with various HIV-1 constructs. The first simply encodes the HIV promoter driving the expression of a destabilized GFP reporter, LTR-d2G (Ld2G) (Figure 7D). The second construct, LTR-d2G-Tat (Ld2GT), only encodes for a well-characterized Tat positive-feedback loop (Weinberger et al., 2005; Weinberger and Shenk, 2007) (Figure 7C). The third construct, dHIV-d2G (Figure 7B), is a truncated version of HIV-1, which consists of LTR-Tat/Rev/Env/d2G (Pearson et al., 2008; Razooky et al., 2012a). The same imaging experiments were performed, and the results are shown in Figures 9, 10, and 11.

Comparing the general trajectories of Ld2G, Ld2GT, dHIV-d2G, and HIV-d2G, we find that the Tat/Rev/Env cassette is sufficient to produce the overshoot trajectories (Figure 11). The LTR promoter itself (Figure 9) or Tat positive feedback (Figure 10) can not produce the overshoot trajectories displayed in dHIV-d2G (Figure 11) or HIV-d2G (Figure 8) trajectories.

We next tested if the overshoot dynamics is due to the silencing of promoter activity (Pearson et al., 2008; Tyagi et al., 2010). The promoter was chronically exposed to TNF-α, and then boosted with either TNF-α or TSA 12 hours after the first TNF-α addition (Figure 12). If silencing was the cause of the overshoot, the second TNF-α or TSA addition would re-activate expression and rescue the GFP decay. The results showed that the GFP turnover feature is not affected by the second reactivation. Collectively, this indicates that the overshoot trajectory is an intrinsic feature of HIV gene regulation.

Finally, we determined if viral expression dynamics exhibit the same profile observed in Jurkat cell lines in natural host cells. CD34 pre-activated primary CD4+ T cells and primary monocyte-derived-macrophages (MDMs) were infected with VSV-G pseudo-typed HIV-d2G virus. Corroborating our previous findings, the expression dynamics in primary cells display the typical overshoot (Figure 13 and 14). Oscillating trajectories, the hallmark feature of negative feedback, were also observed in HIV-d2G and dHIV-d2G cells (Figure 15), but not Ld2G or Ld2GT cells. This highlights the existence of negative-feedback regulation only in HIV-1 constructs encoding for Rev.

Negative Feedback Is Generated by Rev-Dependent Nuclear Export of Unspliced RNA

To further examine the relationship of the turnover in expression dynamics and Rev-mediated RNA nuclear export, we utilized a two-colored reporter system. We transduced dHIV-d2G Jurkats with an additional lentivirus encoding LTR-mCherry-

RRE that acts as a reporter of Rev activity (Figure 16) (Wu et al., 2007a). GFP represents the gene expression from MS RNA, and mCherry represents the expression from US/SS RNA. We analyzed the slopes (i.e., rate of increase) of GFP and mCherry trajectories respectively (Figure 17A). The slope of the gene-expression trajectories is a direct measurement of the net production rate of the fluorescent protein and is a useful metric to probe the relationship between the two species of fluorescent protein production. In the early phase (0–4 hours) after TNF-α activation, only the GFP trajectory starts to increase (Figure 17B). The GFP slope remains positive and becomes larger. However, at the onset of mCherry expression, there is a drastic change in both GFP and mCherry slopes (4–12 hours). The mCherry slope starts to increase, and the GFP slope starts to decrease. Though the GFP trajectory is still rising and the slope still remains positive, the analysis shows that the GFP overshoot-turnover has started and that GFP production starts to slow down at the onset of mCherry expression. The concurrent transition of GFP and mCherry slopes indicates that the Rev-mediated RNA nuclear export is responsible for the overshoot-turnover feature of GFP trajectories. In the late phase (12 hours after activation), the GFP trajectory starts to fall and the slope becomes negative, while the mCherry trajectory is steadily rising. The slope phase plot demonstrates that the overshoot trajectory is a result of Rev/RRE function.

To confirm the relationship of Rev-mediated RNA nuclear export and negative feedback regulation, we used leptomycin B (Lepto B), an inhibitor of CRM1 nuclear export machinery that Rev uses to export viral RNA (Kudo et al., 1998; Otero et al.,

1998) (Figure 18). CRM1 inhibited cell turnover at a later time point post activation, when compared to control cells, consistent with the notion of Rev-mediated turnover. We also compared the ratio of CVs (coefficient of variation) between hour 12 and 5 after TNF activation for the various HIV-1 constructs (e.g. Ld2G, Ld2GT, and HIV-d2G) (Figure 19). Depending on the feedback regulation, there should be characteristic shifts in the noise of the system. For instance, positive feedback amplifies noise, while negative feedback reduces it (Alon, 2007; Raser and O'Shea, 2005; Savageau, 2011). The analysis shows that dHIV-d2G, which encodes both Tat and Rev, has the best control of noise propagation over time, compared to the positive-feedback-only Tat circuitry (Ld2GT), and no feedback Ld2G. The results are consistent with the hypothesis that Rev negatively feeds back on HIV-1 gene expression.

Acknowledgements

In this chapter, Brandon Razooky created the pNL-mCherryRRE vector and performed the imaging experiment of primary CD4+ T cells infected with HIV-d2G. Roy Dar performed the noise analysis. Simon Chou, Jason Neidleman, and Isa Munoz-Arias provided the primary monocytes. Ana Fernandez-Sesma and Dabeiba Bernal technically supported the experimental procedures of culturing primary MDMs. Leor Weinberger built the initial version of the preliminary ODE models, and Igor Rouzine helped with the refinement/fitting of ODE models. I performed the experiments, analyzed the results, modified and refined the models, and performed the computational analysis and simulations. Leor Weinberger directed the research.

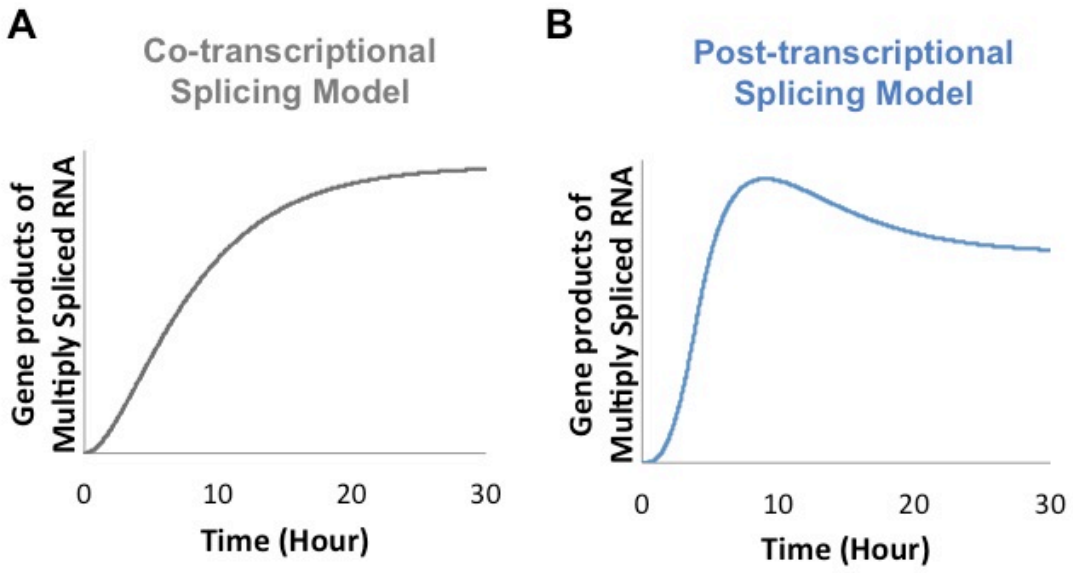

Figure 6. Simulated dynamics of MS RNA gene products from two alternate models

(A) Without feedback regulation, the co-transcriptional model shows that the expression level of MS RNA starts to grow after activation and eventually plateaus at steady state. (B) With the negative feedback predicted by the post-transcriptional model, the gene product of MS RNA accumulates, shortly followed by a turnover in expression level. The expression dynamics shows a characteristic "overshoot" trajectory of negative feedback.

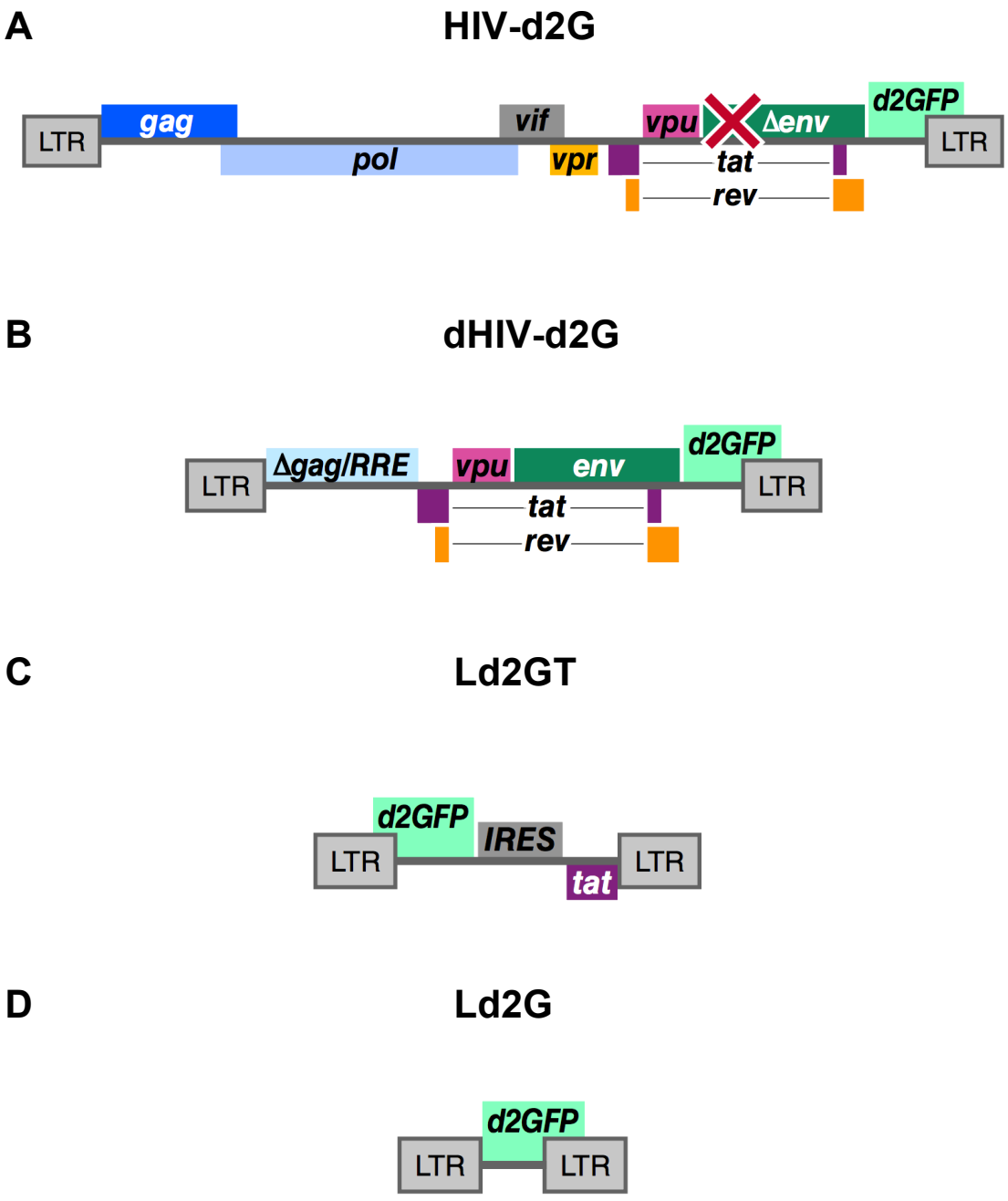

Figure 7. HIV-1 constructs used in Chapter 2

(A) Full-length construct. Env expression is abolished because of the nonsense mutation at the starting ATG. Nef is replaced with destabilized GFP (d2GFP). (B) A truncated version of HIV-1. Downstream the LTR promoter, it only carries the Tat/Rev/Env cassette. Nef is replaced with d2GFP. (C) The minimal LTR-Tat feedback loop construct. Destabilized GFP-IRES is inserted as the reporter gene. (D) The simple promoter LTR driving d2GFP construct.

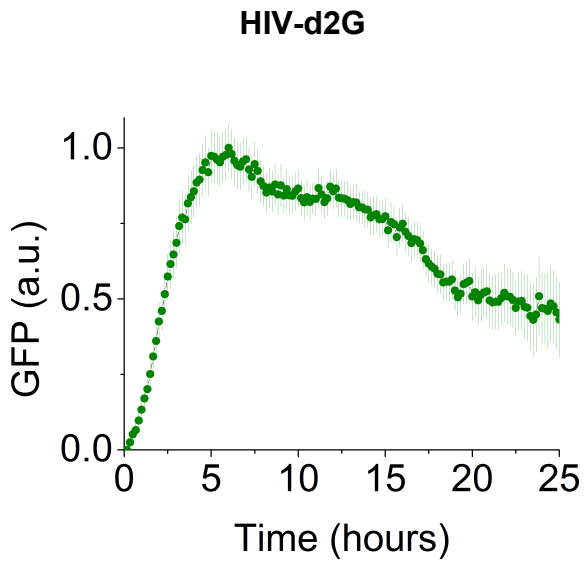

Figure 8. Dynamics of HIV-1 gene expression in Jurkat cells display overshoot trajectories

Jurkat cells were latently infected with HIV-d2G, and reactivated with 10 ng/mL TNF-α. Fluorescent images of cells were taken at 10-minute intervals, and the GFP levels of tracked single cells were quantified over time. The single-cell trajectories were synchronized *in silico* to create the general averaged trend line (n = 67). (Error bar represents the standard error.)

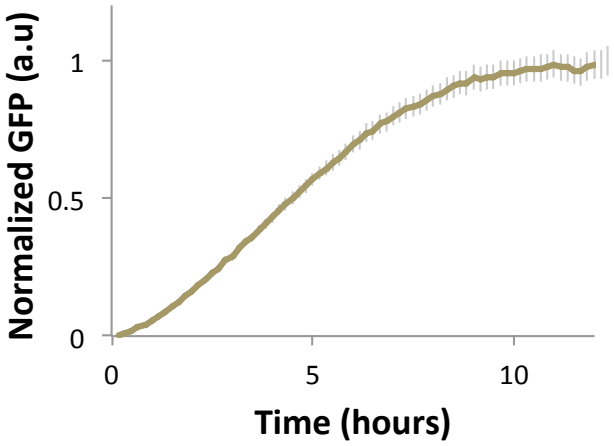

Figure 9. Expression dynamics of HIV-1 Ld2G in Jurkat cells

Transduced Ld2G Jurkat cells were reactivated with TNF-α. The general trend line (n=255) of this simple LTR promoter-driving d2GFP clone does not exhibit the same characteristic overshoot trajectories in cells transduced with full-length HIV. (Error bars represent the standard error.)

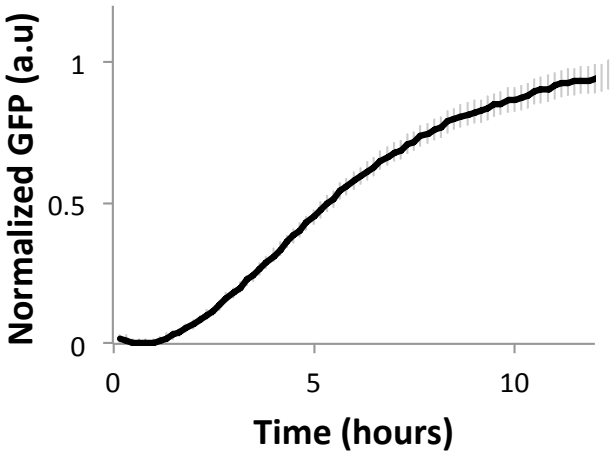

Figure 10. Expression dynamics of HIV-1 Ld2GT in Jurkat cells

Transduced Ld2GT Jurkat cells consist of the minimal Tat positive-feedback circuitry. The general trend line (n=135) does not show the typical characteristic overshoot either. (Error bars represent the standard error.)

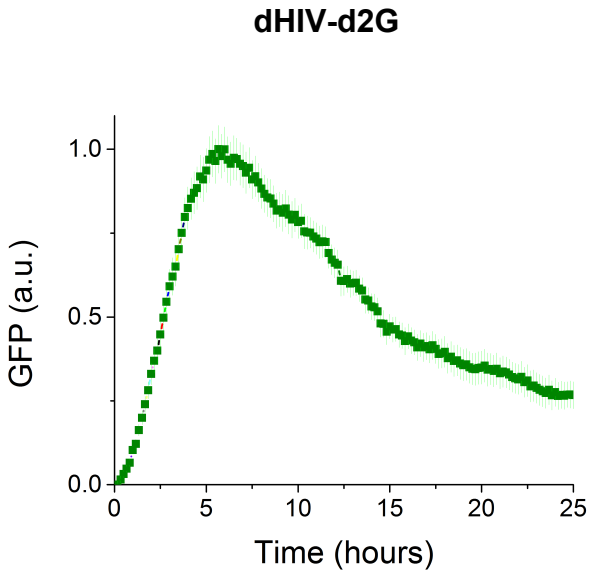

Figure 11. Expression dynamics of dHIV-d2G: Tat/Rev cassette is sufficient to generate the same overshoot dynamics as the full-length construct

Average trajectory (n=97) of Jurkat cells transduced with the dHIV-d2G. The results show that Tat/Rev/Env cassette is sufficient to generate the typical overshoot. When compared to the Ld2G and Ld2GT trajectories, Rev is the critical component to create the overshoot feature of negative feedback. (Error bars represent the standard error.)

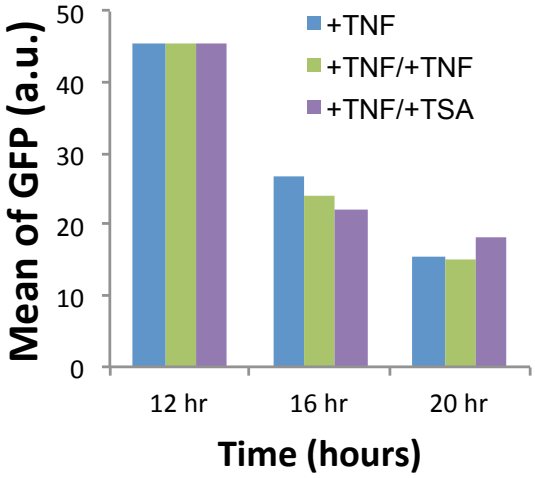

Figure 12. The decay in HIV-1 expression dynamics is not due to promoter silencing

Jurkat cells transduced with dHIV-d2G were activated with TNF-α. After 12 hours, another TNF-α or TSA was added to the culture, and GFP levels were quantified by flow cytometry. The results show that the second reactivation with TNF or TSA does not rescue the decaying GFP level, indicating the turnover of GFP is not due to the silencing of LTR promoter.

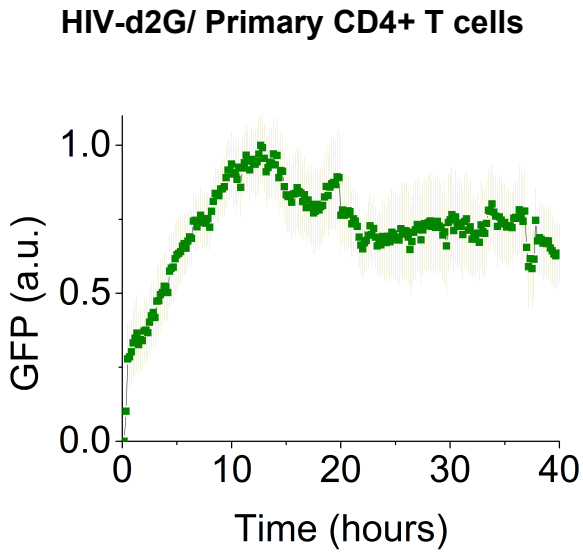

Figure 13. Gene expression dynamics of primary CD4+ T cells infected with full-length HIV

Pre-activated primary CD4+ cells were infected with VSV-G pseudo-typed HIV-d2G virus. After infection, cells were loaded onto a microfluidic device and imaged over time at 10-minutes intervals for more than 48 hours. The general trend line (n = 37) shows similar overshoot in the Jurkat cells. (Error bar represents the standard error.)

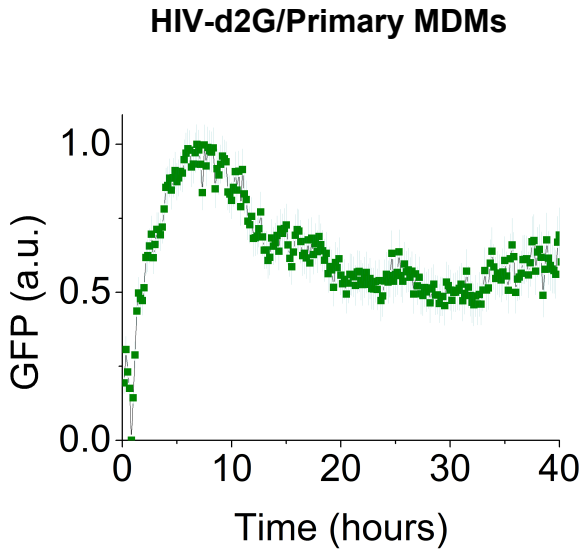

Figure 14. Gene expression dynamics of primary MDMs infected with full-length HIV

Primary MDMs were obtained by incubating primary CD14+/CD4+ monocytes with GM-CSF for 10 days. Differentiated primary MDMs were infected with VSV-G pseudo-typed HIV-d2G and imaged at 10-minute intervals for more than 96 hours. Averaged trend line (n = 489) exhibits the typical overshoot associated with negative feedback, indicating the existence of negative-feedback regulation in HIV-1 gene expression. (Error bars represent the standard error.)

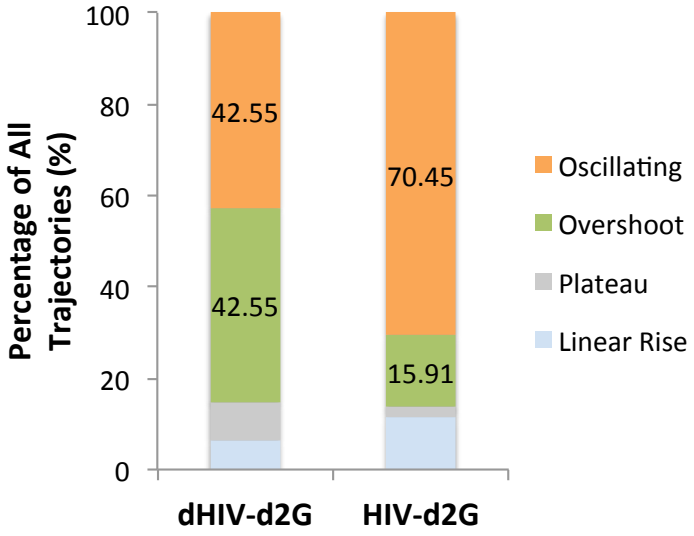

Figure 15. Significant percentages of dHIV-d2G or HIV-d2G single-cell trajectories are oscillating

Composition of single-cell trajectories of Jurkat dHIV-d2G (n=47) and HIV-d2G (n=44) were analyzed. More than 80 % of cells show the overshoot property (overshoot+oscillating) in 18 hours after the onset of GFP trajectory, and more than 40 % of cells display waving/oscillating features in the later hours after the overshoot peak. Oscillation is the signature property generated by negative feedback with delay. Notably, HIV-d2G (full-length HIV) cells have higher percentage of oscillating trajectories than dHIV-d2G (truncated HIV). HIV-d2G has higher tendency to oscillate may be a result of its extra genomic complexity.

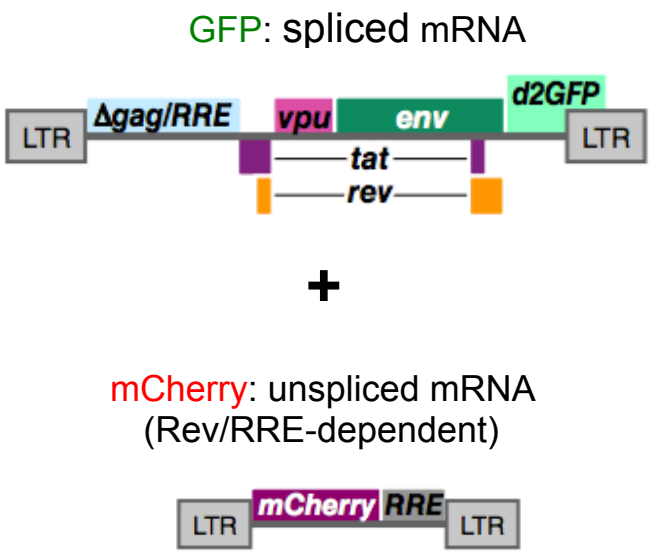

Figure 16. A two-color system for examining the relation between Rev-dependent expression and negative-feedback regulation

We used a two-color system to track the individual gene expression of MS and incompletely spliced RNA. A truncated version of HIV-1, dHIV-d2G, encodes the Tat/Rev/Env cassette, and the Nef is replaced with destabilized GFP as the reporter for MS RNA gene products. The second reporter gene is mCherry, encoded on another construct downstream the LTR promoter, and right upstream the RRE domain. The production of mCherry is dependent on the presence of Rev (Wu et al., 2007b).

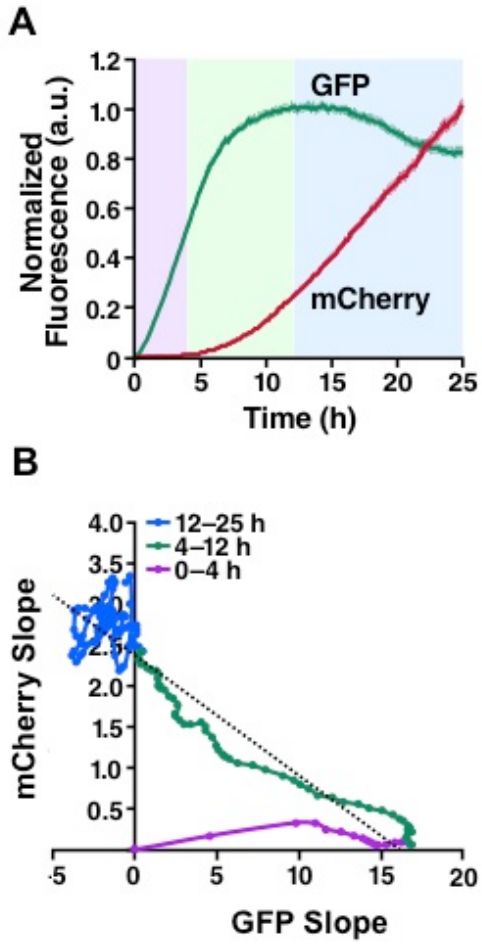

Figure 17. Negative feedback in HIV-1 gene expression is a consequence of Rev/RRE function

(A) Jurkat cells were co-transduced with both dHIV-d2G and pNL-mCherry RRE constructs (Figure 16), and the latently infected ones were selected. We reactivated the cells with TNF-α and monitored the fluorescent levels by time-lapse microscopy. Single cells were tracked, and the GFP and mCherry fluorescent levels were measured over time to generate the averaged trend line for each color. The GFP trend line shows the overshoot feature. There is an approximately 4-hour delay between the onset of GFP and mCherry expression. (B) The slope of each fluorescent protein level was further analyzed and plotted. In the early phase after reactivation (0–4 hours, purple), only GFP is expressed and the production is on an increasing rate. Later in the transition phase (4–12 hours, green), mCherry starts to be expressed at an increasing rate. Notably, the production of GFP starts to slow down when mCherry expression begins. In the late phase (12–25 hours, blue), GFP level starts to decrease, and thus the GFP rate of increase (net production) is negative. On the other hand, mCherry keeps its production at a positive rate. Eventually the production rates of both fluorescent proteins are nearly steady. The rate analysis indicates that the overshoot characteristic of negative feedback is a result of unspliced RNA being exported and translated, leading to the reduced expression of spliced RNA gene products. In addition, the transition phase of GFP/mCherry rate of increase suggests the "switch" of HIV early-to-late gene expression, is due to the Rev negative feedback.

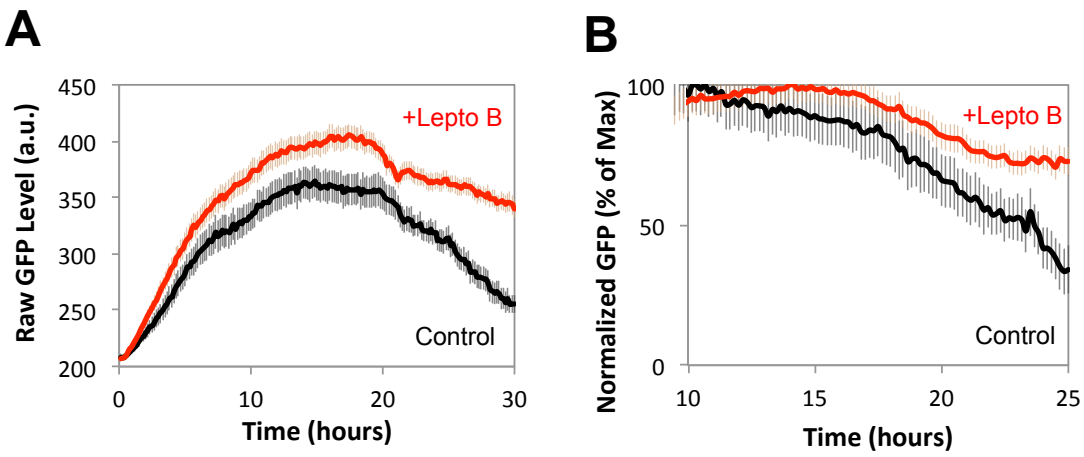

Figure 18. Leptomycin B-treated dHIV-d2G cells turnover at a slower rate after the overshoot

(A) Raw averaged single-cell trajectories of Lepto B treated and untreated dHIV-d2G cells (n=110 for +Lepto B; n=48 for control). With a mild concentration of the Rev nuclear export inhibitor (0.3 ng/mL Lepto B), the GFP level is upregulated, suggesting that Rev negative feedback is repressed. (B) The Lepto B–treated cells have a flatter turnover than the controls, indicating that the Rev-dependent RNA nuclear export contributes to the turnover of GFP trajectory.

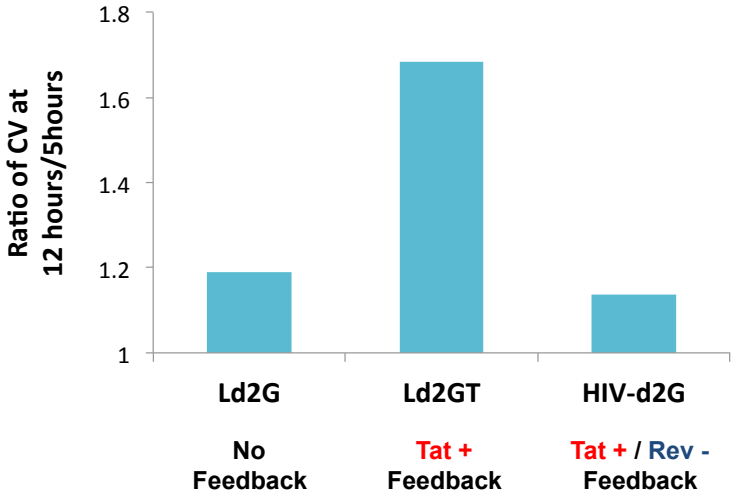

Figure 19. Coupling Rev negative feedback with Tat positive feedback rescues the HIV gene expression from propagation of noise over time

The coefficient of variation (CV) of single-cell trajectories (n=965 for Ld2G; n=194 for Ld2GT; n=166 for HIV-d2G) at hours 5 and 12 were measured. The ratio of CV_{12hr}/CV_{5hr} represents the change of noise magnitude from the early phase to the late phase of expression. Over time, expression noise accumulates in HIV-1 basal expression (Ld2G). Tat positive feedback (Ld2GT) amplifies noise in HIV. However, when Rev negative feedback is coupled with Tat positive feedback in the system (HIV-d2G), noise magnitude is reduced. The result implies that coupling Rev negative feedback with Tat positive feedback provides a method for HIV to limit expression noise.

Appendix 2-1

Table 1. Primer sequences for cloning in Chapter 2

Primer Name	Sequence	Purpose
BamHI-IRES F	attagtagtagcacccgggcggatccgcccct	For cloning pNL-mCherry-RRE
IRES-mCherry R	ttatcctcctcgcccttgctcaccatggttgtggccat attatcatcgtg	For cloning pNL-mCherry-RRE
IRES-mCherry F	cacgatgataatatggccacaaccatggtgagcaa gggcgaggaggataa	For cloning pNL-mCherry-RRE
mCherry-XbaI R	gcatggacgagctgtacaagtaaagcggccgcga ctctagatcataat	For cloning pNL-mCherry-RRE

Appendix 2-2

Construction of the HIV-1 Full-Length Gene Regulatory Models

We developed mathematical models of the HIV-1 full-length gene regulatory circuit to predict the overall architecture of HIV gene-expression. We do not attempt to accurately describe every biochemical step in HIV gene expression, but intend to build minimal model that captures the main traits of HIV gene expression. We first propose two schemes of HIV-1 regulatory circuits based on different splicing mechanisms (Figures 4 and 5 in Chapter 1).

The co-transcriptional model

We use a set of five ODE equations to describe the minimum full-length regulatory circuit of HIV-1 gene expression. To simplify the model, we only consider the two main classes of viral mRNA: unspliced (US) and multiply spliced (MS) RNA. Assuming the reaction of US RNA being spliced into singly spliced (SS) RNA is relatively fast and ignorable, SS RNA is lumped with US RNA. To facilitate the future expansion and prediction of the models, we further classify US RNA into two classes based on the cellular localization. USn stands for the nuclear US RNA, and USc stands for the cytoplasmic US RNA. Because we always use destabilized GFP (d2GFP) to replace the Nef in our experiment constructs as the reporter, we use d2GFP instead Nef in all of our models, and thus, the GFP level is representative of Nef levels. The differential equations that described this co-transcriptional splicing system are:

$$\frac{d}{dt}USn = b_{US} - \frac{kr \cdot USn \cdot Rev^h}{K_{Rev}{}^h + Rev^h} - drna \cdot USn \quad (1)$$

$$\frac{d}{dt}USc = \frac{kr \cdot USn \cdot Rev^h}{K_{Rev}{}^h + Rev^h} - drna \cdot USc \quad (2)$$

$$\frac{d}{dt}MS = b_{MS} - drna \cdot MS \quad (3)$$

$$\frac{d}{dt}Rev = p \cdot fr \cdot MS - dr \cdot Rev \quad (4)$$

$$\frac{d}{dt}GFP = p \cdot fg \cdot MS - dg \cdot GFP \quad (5)$$

where b_{US} is a lumped basal production rate consisting of US RNA transcription from the LTR promoter and Tat positive-feedback transactivation. kr is the maximum Rev-dependent RNA nuclear export rate, h is the Hill coefficient of the Rev-RRE interaction, K_{Rev} is the Michaelis-Menton saturation term for Rev nuclear export, drna is the RNA degradation rate, b_{MS} is the basal production rate of MS RNA directly from transcription and co-transcriptional splicing, p is the protein production rate per mRNA, fr is the fraction Rev-encoding RNA out of total MS RNA, dr is the degradation rate of Rev protein, fg is the fraction GFP-encoding RNA out of total MS RNA, and dg is the degradation rate of GFP protein. Among the rate parameters, we assume that the RNA degradation is uniform for all different RNA species and subcellular locations. Eq. (1) describes three biochemical reactions: the basal production of USn, the Rev-dependent RNA nuclear export of USn, and the degradation of USn. In the rest of the system, there are only two reactions in each equation: production and degradation. Usc is produced solely from the Rev-dependent

RNA nuclear export. Tat positive feedback is assumed to saturate early in the system, and thus the Tat positive-feedback term is simplified into a single rate parameter and lumped with the basal transcription rate. The reactions of MS RNA nuclear export and Rev nuclear-cytoplasmic shuttling are much faster than other kinetic rates of the system and are ignored.

The post-transcriptional model

We next constructed the ODE systems for the post-transcriptional model. This model is similar to the co-transcriptional one, but with two modified equations. First, Eq. (1), which describes USn dynamics in the co-transcriptional model, is replaced with the following:

$$\frac{d}{dt}\mathrm{USn} = b - sp \cdot \mathrm{USn} - \frac{kr \cdot \mathrm{USn} \cdot \mathrm{Rev}^h}{K_{\mathrm{Rev}}{}^h + \mathrm{Rev}^h} - drna \cdot \mathrm{USn} \quad (6)$$

In the new equation (6), b is the lumped production rate of basal LTR transcription and Tat positive-feedback transactivation, and sp is the rate of USn RNA spliced into MS RNA. This equation represents four biochemical reactions: transcription, post-transcriptional splicing, Rev-dependent nuclear export, and degradation.

Second, the Eq. (3) is modified as:

$$\frac{d}{dt}\mathrm{MS} = sp \cdot \mathrm{USn} - drna \cdot \mathrm{MS} \quad (7)$$

where MS RNA production is solely from the post-transcriptional splicing term in Eq (6). The other three house keeping equations in this post-transcriptional model are the same as the equations (2), (4), and (5).

Preliminary predictions of the ODE models

We first use parameter values obtained from literatures (Kim and Yin, 2005; Reddy and Yin, 1999), theoretical values, or direct measurements in our lab to model each system. The values of all parameters used in the simulations are summarized in the Table 2. All initial values are set zero. The results of preliminary predictions are shown in Figure 6.

Table 2. Parameters used in simulation results of preliminary ODE models

Parameter	Description	Values [Units]	Justification or Reference
b_{US}	Lumped transcription rate of basal LTR and Tat positive feedback	4.29 [μM h^{-1}] = 1188 [RNA copies $cell^{-1}$ h^{-1}]	80-fold basal transcription rate. (Bohan et al., 1992; Graeble et al., 1993; Laspia et al., 1993)
kr	Max Rev-dependent RNA nuclear export rate	20 minutes half-life = 2.08 [h^{-1}]	(Love et al., 1998)
K_{Rev}	Rev saturation threshold concentration of Rev-dependent RNA nuclear export	40000 [protein copies $cell^{-1}$]	(Reddy and Yin, 1999; Rempala et al., 2006)
h	Hill coefficient. cooperativity of Rev-dependent RNA nuclear export	12 [no unit]	Maximum number of Rev per RRE. (Mann et al., 1994)
drna	Degradation rate of HIV-1 mRNAs	4 hr half-life = 0.173 [h^{-1}]	(Felber et al., 1989; Malim and Cullen, 1993; Schwartz et al., 1992a)
b_{MS}	Basal production rate of MS RNA from transcription and co-transcriptional splicing	1.072 [μM h^{-1}] = 297 [RNA copies $cell^{-1}$ h^{-1}]	20-fold basal transcription rate. (Bohan et al., 1992; Graeble et al., 1993; Laspia et al., 1993)
p	General protein production rate	270 [protein copies h^{-1}]	(Kim and Yin, 2005)
fr	Fraction of Rev-encoding mRNA among all MS	0.19 [no unit]	(Purcell and Martin, 1993; Robert-Guroff et al., 1990)
dr	Degradation rate of Rev	4 hr half-life = 0.173 [h^{-1}]	(Kubota et al., 1996)
fg	Fraction of GFP-encoding mRNA among all MS	0.8	(Purcell and Martin, 1993; Robert-Guroff et al., 1990; Schwartz et al., 1992b)
dg	Degradation rate of d2GFP	0.346 [h^{-1}]	2-hr theoretical half-life
b	Lumped transcription rate of basal LTR and Tat positive feedback	5.36 [μM h^{-1}] = 1485 [RNA copies $cell^{-1}$ h^{-1}]	100-fold basal transcription rate. (Bohan et al., 1992; Graeble et al., 1993; Laspia et al., 1993)
sp	Splicing rate of US to MS	10-min lifespan = 0.116 [h^{-1}]	(Singh and Padgett, 2009)

CHAPTER 3: HIV-1 RNA IS POST-TRANSCRIPTIONALLY SPLICED

Abstract

In the previous chapter, Rev was shown to generate negative feedback in HIV-1 gene expression. Although it is suggested that post-transcriptional splicing cascade is essential for Rev to establish such feedback regulation, there is still no direct evidence for post-transcriptional splicing. Here we used single molecule fluorescence *in situ* hybridization (smFISH) to probe the *in vivo* splicing mechanism of HIV-1. Our results show the differential spatial distributions of HIV splicing variants, thereby indicating that HIV-1 pre-mRNA is spliced post-transcriptionally *in vivo*.

Materials and Methods

Single Molecule mRNA Fluorescence *in situ* Hybridization

Probes were developed using the designer tool from http://www.singlemoleculefish.com/. Three sets of probes were designed to detect regions the Pol/Vif splice junction (4548–5414 bp of pNL4-3), Env (7251–8251 bp of pNL4-3), and GFP (corresponding to Nef, 8799–8887 bp of pNL4-3) in HIV-d2G. Each set contain 34–45 probes, each probe is 20 nt long, and 2 bp apart between single probes (see binding region and probe sequence in Appendix 3). Probes are conjugated with Quasar 670 or TAMRA. A latently infected isoclonal Jurkat population was activated with 10 ng/ml TNF-α for 4 hr. Cells were washed with 10 mL of PBS solution and immobilized on Cel-Tak coated 8-well chambered image dish (Weinberger et al., 2008). Cells were fixed with PBS in 3.4% paraformaldehyde for 10 minutes. The process from washing off TNF-α to fixation in paraformaldehyde took less than 10 minutes. Fixed cells were stored in 70% EtOH at 4°C overnight to permeabilize the cell membranes. Probes were diluted 100-fold and allowed to hybridize at 37°C for 6 hours. Wash steps and DAPI staining were performed as described (https://www.biosearchtech.com/support/applications/stellaris-rna-fish). To minimize photo-bleaching, cells were imaged as described in buffer (50% glycerol, 75 μg/mL glucose oxidase, 520 μg/mL catalase, and 0.5 mg/mL Trolox) (Waks et al., 2011). Images were taken on a Nikon 6D System with Plan Apo VC 100x/1.4 oil objective in the UCSF Nikon Imaging Center. 10-30 xy locations were randomly

selected for each condition. For each xy location, Niquist sampling was performed by taking 44–51, 0.26-μm steps along the z-plane. The exposure times for Quasar 670, TAMRA, and DAPI channels were 3, 3, and 2 s for each image in a 3-D stack. Spot identification and counting was performed as described (Rifkin, 2011). Cells and transcription centers were segmented manually. DAPI image stacks were deconvoluted using Huygens software (http://www.svi.nl/HomePage) and used for nuclear segmentation based on a published edge-detection algorithm (http://www.mathworks.com/products/demos/image/ipexcell/ipexcell.html) in combination with in-house MATLAB programs (available upon request). For more detailed materials and methods, see Appendix 3.

Results and Discussion

Based on the hypothesis that establishment of Rev negative feedback requires HIV-1 pre-mRNA to be post-transcriptionally spliced, we examined the mechanism of HIV-1 pre-mRNA splicing by single molecule fluorescent *in situ* hybridization (smFISH) (Raj et al., 2008). This method has been used to detect the mechanism of alternative splicing by differentiating the spatial distribution of splice variants (Waks et al., 2011). We designed three sets of smFISH probes (US, SS, and MS probes) conjugated with fluorophores (Quasar 670 or TAMRA) to detect the *in vivo* splicing mechanism of HIV-1 (Figure 20). If HIV-1 pre-mRNA splicing occurs co-transcriptionally, the spatial distribution of all splicing variants will peal at the transcription center (TC) (Figure 21A). On the contrary, if pre-mRNA is post-transcriptionally spliced, only the unspliced form of mRNA will co-localized around the TC, and the other spliced forms of mRNA will diffuse or shuttle away and the distribution peak of spliced variants will shift away from the TC (Figure 21B). We fixed the HIV-d2G transduced Jurkat cells 4 hours post TNF-α activation and hybridized the cells with US/MS probe sets (Figure 22A) or SS/MS probe sets (Figure 22B). The data from other time points, such as 2 and 6 hours, were inconclusive due to difficulties with low expression levels (2-hour time point), or cells containing US RNA in the cytoplasm due to Rev activity (6-hour time point). We identified different splicing variants based on whether fluorescently labeled RNA colocalized to a single diffraction limited spot or were greater than 0.18 μm apart. We measured the distance between the RNA variants and the TC. We detected the spatial distribution of US

RNA versus the pool of SS and MS RNA by using US/MS probe combinations (n=146, Figure 23). The splicing center of SS/MS RNA is still at the TC. There is a slight shoulder in the distribution next to the center peak, implying there is a mixture of two different types of distribution. These results were difficult to interpret. However, the data for the MS RNA versus the pool of US and SS RNA (SS/MS probe sets, n=156) show a clear shift in the distribution. The unspliced/incompletely spliced forms of RNA is centered at the TC, but the splicing center of MS RNA is not co-localized at the TC (Figure 24). The result shows that in the early hours after HIV-1 active replication, the viral pre-mRNA splicing follows the route of post-transcriptional splicing cascade.

Acknowledgements

In this chapter, Brandon Razooky and I designed the RNA probe sets and the optimized the smFISH experimental protocols. Scott Rifkin and Brandon Razooky assisted me in using MATLAB program packages to perform the smFISH spot finding. I performed the experiment, developed the in-house MATLAB programs to process raw images and segment the nucleus region, and performed all the analyses. All the smFISH images in this chapter were taken at the Nikon Imaging Center (NIC) in UCSF with the help of Kurt Thorn. Leor Weinberger directed the research.

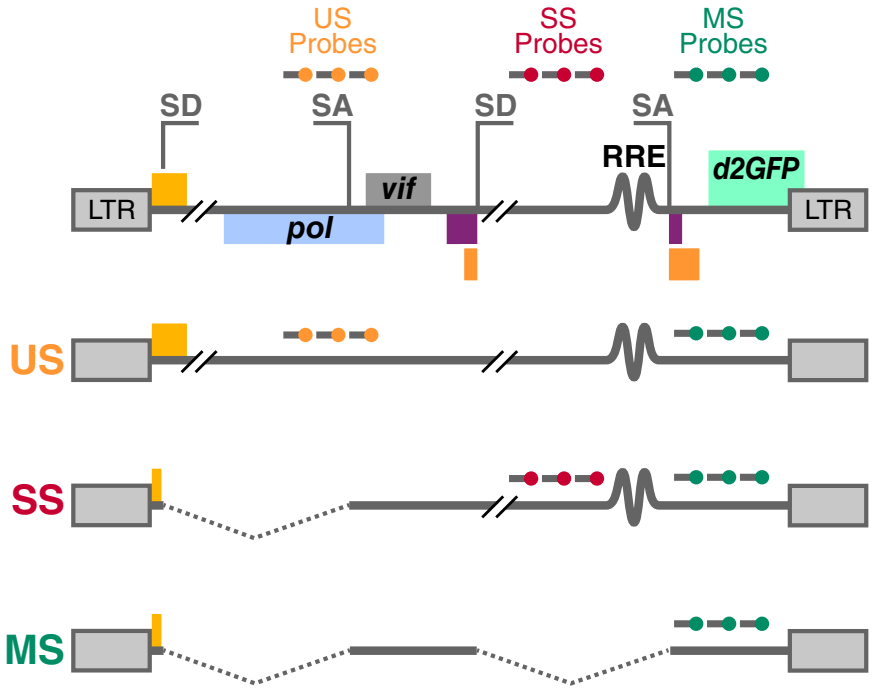

Figure 20. Illustration of probe designs

Three sets of probes, US, SS, and MS probes, were designed to bind to the Pol/Vif, Env, and d2GFP reading frames on HIV-d2G, respectively. Each probe set contains more than 34 single probes that conjugated with fluorophores. Using appropriate combinations of probe sets, we can detect different alternative splicing variants.

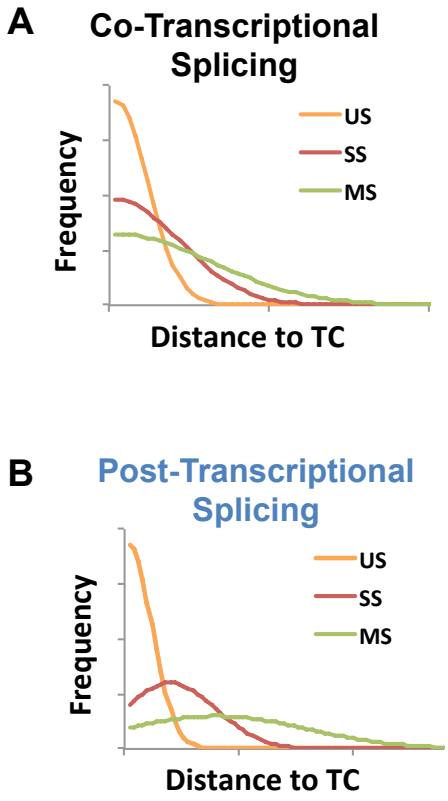

Figure 21. Co- or post-transcriptional splicing can be differentiated by the spatial distributions of the alternate splicing variants

(A) When the alternatively spliced RNA variants are spliced co-transcriptionally, the splicing center and transcription center (TC) co-localized, causing all the splicing variants enriched at the proximity of transcription center. (B) When the RNA is spliced post-transcriptionally, the splicing center is not co-localized with the transcription center. As a result, only US RNA is enriched at the proximity of transcription center. Other spliced RNA (SS and MS) is enriched farther away from the TC.

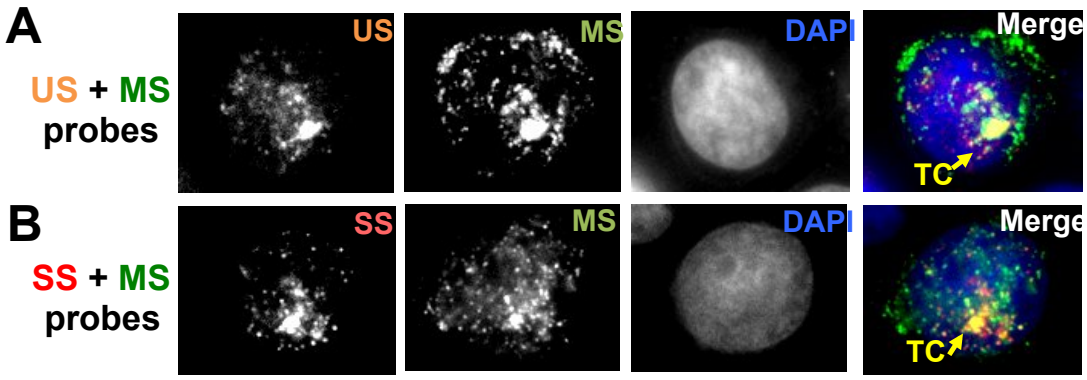

Figure 22. smFISH images of Jurkat HIV-d2G hybridized with US/MS or SS/MS probe sets

smFISH images of HIV-d2G latently infected Jurkat cells were taken 4 hours after TNF-α reactivation. Cells were hybridized with US/MS probe sets (A), or with SS/MS probe sets (B), together with DAPI stain to indicate the nucleus area. Yellow arrows in the merged images show the location of transcription center (TC).

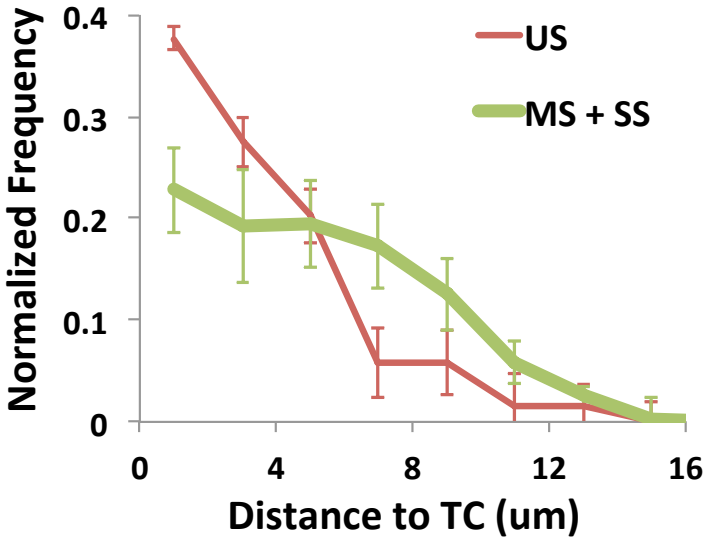

Figure 23. Spatial distribution of nuclear MS/SS RNA (green) and US RNA (red)

The peak of US RNA distribution is at the TC. Although the peak of MS/SS RNA distribution is still at the TC, but there is a shift and a shoulder in the distribution, suggesting it is a mixture of two distinct distributions (n=146).

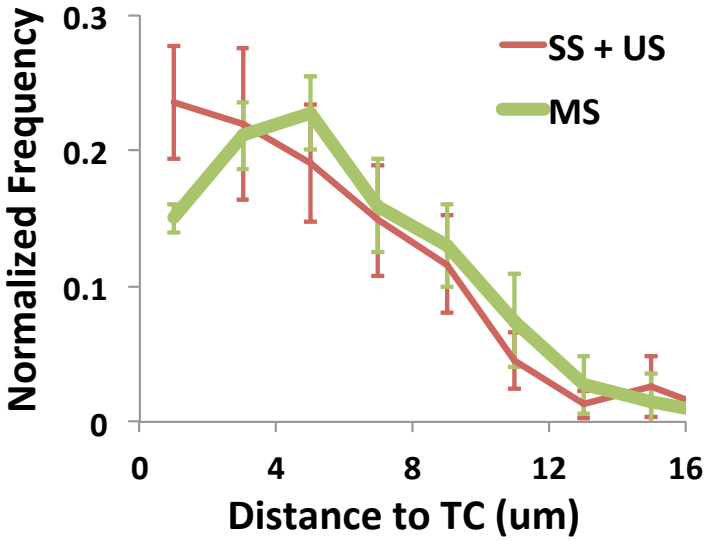

Figure 24. Spatial distribution of nuclear MS RNA (green) and SS/US RNA (red)

The distribution peak of MS RNA is clearly shifted away from the TC, indicating that MS RNA is post-transcriptionally spliced (n=156).

Appendix 3

Probe Designs of Single Molecule mRNA Fluorescence *in situ* Hybridization (smFISH)

US probe sets: Detecting Pol-Vif region

Target sequence:

agatggccagtaaaaacagtacatacagacaatggcagcaatttcaccagtactacagttaaggccgcctgttggtgggcg
gggatcaagcaggaatttggcattccctacaatccccaaagtcaaggagtaatagaatctatgaataaagaattaaagaaaat
tataggacaggtaagagatcaggctgaacatcttaagacagcagtacaaatggcagtattcatccacaattttaaaagaaaa
ggggggattgggggtacagtgcaggggaaagaatagtagacataatagcaacagacatacaaactaaagaattacaaa
aacaaattacaaaaattcaaaattttcgggtttattacagggacagcagagatccagtttggaaaggaccagcaaagctcctc
tggaaaggtgaaggggcagtagtaatacaagataatagtgacataaaagtagtgccaagaagaaaagcaaagatcatcag
ggattatggaaaacagatggcaggtgatgattgtgtggcaagtagacaggatgaggattaacacatggaaaagattagtaa
aacaccatatgtatatttcaaggaaagctaaggactggttttatagacatcactatgaaagtactaatccaaaaataagttcaga
agtacacatcccactaggggatgctaaattagtaataacaacatattggggtctgcatacaggagaaagagactggcatttg
ggtcagggagtctccatagaatggaggaaaaagagatatctgctataagaaataccatat

Probe sequence: One version of this US probe set is conjugated to TAMRA, and the other is conjugated to Quasar 670. The US-TAMRA probe set is used with SS-Quasar 670, and the US-Quasar 670 probe set is used with MS-TAMRA. For sequences, see Table 3.

Table 3. US probe sequences

US Probe #	Sequence	US Probe #	Sequence
1	actgtttttactggccatct	18	gctggtcctttccaaactgg
2	attgctgccattgtctgtat	19	ttcacctttccagaggagct
3	ccttaactgtagtactggtg	20	tatcttgtattactactgcc
4	tgatccccgcccaccaacag	21	ggcactacttttatgtcact
5	tagggaatgccaaattcctg	22	gatgatctttgcttttcttc
6	tactccttgactttggggat	23	ccatctgttttccataatcc
7	attctttattcatagattct	24	cttgccacacaatcatcacc
8	acctgtcctataattttctt	25	gtgttaatcctcatcctgtc
9	aagatgttcagcctgatctc	26	gtgttttactaatcttttcc
10	ctgccatttgtactgctgtc	27	gctttccttgaaatatacat
11	cttttaaaattgtggatgaa	28	atgtctataaaaccagtcct
12	gtacccccaatcccccctt	29	ttggattagtactttcatag
13	ctactattctttcccctgca	30	gatgtgtacttctgaactta
14	tgtatgtctgttgctattat	31	ctaatttagcatcccctagt
15	ttgtttttgtaattctttag	32	agaccccaatatgttgttat
16	ccgaaaattttgaatttttg	33	ccagtctctttctcctgtat
17	ctctgctgtccctgtaataa	34	tggagactccctgacccaaa

SS probe set: Detecting Env region

Target sequence:

ctagcaaattaagagaacaatttggaaataataaaacaataatctttaagcaatcctcaggaggggacccagaaattgtaacg
cacagttttaattgtggaggggaatttttctactgtaattcaacacaactgtttaatagtacttggtttaatagtacttggagtactg
aagggtcaaataacactgaaggaagtgacacaatcacactcccatgcagaataaaacaatttataaacatgtggcaggaag
taggaaaagcaatgtatgcccctcccatcagtggacaaattagatgttcatcaaatattactgggctgctattaacaagagatg
gtggtaataacaacaatgggtccgagatcttcagacctggaggaggcgatatgagggacaattggagaagtgaattatata
aatataaagtagtaaaaattgaaccattaggagtagcacccaccaaggcaaagagaagagtggtgcagagagaaaaaag
agcagtgggaataggagctttgttccttgggttcttgggagcagcaggaagcactatgggcgcacggtcaatgacgctgac
ggtacaggccagacaattattgtctgatatagtgcagcagcagaacaatttgctgagggctattgaggcgcaacagcatctg
ttgcaactcacagtctggggcatcaaacagctccaggcaagaatcctggctgtggaaagatacctaaaggatcaacagctc
ctggggatttggggttgctctggaaaactcatttgcaccactgctgtgccttggaatgctagttggagtaataaatctctggaac
agatttggaataacatgacctggatggagtgggacagagaaattaacaattacacaagcttaatacactccttaattgaagaat
cgcaaaaccagcaagaaaagaatgaacaagaattattggaattagataaatgggcaagtttgtggaattggtttaacataaca
aattggct

Probe sequences: Probes are all conjugated to Quasar 670. For sequences, see Table 4

Table 4. SS probe sequences

SS Probe #	Sequence	SS Probe #	Sequence
1	ttgttctcttaatttgctag	24	cccaagaacccaaggaacaa
2	ttattgttttattatttcca	25	gcccatagtgcttcctgctg
3	cctcctgaggattgcttaaa	26	ccgtcagcgtcattgaccgt
4	gtgcgttacaatttctgggt	27	gacaataattgtctggcctg
5	gaaaaattcccctccacaat	28	gttctgctgctgcactatat
6	acagttgtgttgaattacag	29	cctcaatagccctcagcaaa
7	gtactattaaaccaagtact	30	agttgcaacagatgctgttg
8	atttgacccttcagtactcc	31	ctgtttgatgccccagactg
9	ttgtgtcacttccttcagtg	32	cagccaggattcttgcctgg
10	tttattctgcatgggagtgt	33	tgatcctttaggtatctttc
11	cctgccacatgtttataaat	34	accccaaatccccaggagct
12	gcatacattgcttttcctac	35	tgcaaatgagttttccagag
13	aatttgtccactgatgggag	36	gcattccaaggcacagcagt
14	cagtaatatttgatgaacat	37	cagagatttattactccaac
15	ccatctcttgttaatagcag	38	tcatgttattccaaatctgt
16	ggacccattgttgttattac	39	tctctgtcccactccatcca
17	ctcctccaggtctgaagatc	40	taagcttgtgtaattgttaa
18	ctccaattgtccctcatatc	41	attcttcaattaaggagtgt
19	tttatatttatataattcac	42	ttcttttcttgctggttttg
20	ctaatggttcaatttttact	43	taattccaataattcttgtt
21	tttgccttggtgggtgctac	44	tccacaaacttgcccattta
22	ttctctctgcaccactcttc	45	caatttgttatgttaaacca
23	ctcctattcccactgctctt		

MS probe set: Detecting d2GFP (replacing Nef reading frame)

Target sequence:

atggtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccac
aagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggca
agctgcccgtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatg
aagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaa
ctacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaag
gaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcaga
agaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagca
gaacaccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagac
cccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgta
caagaagcttagccatggcttcccgccggaggtggaggagcaggatgatggcacgctgcccatgtcttgtgcccaggaga
gcgggatggaccgtcaccctgcagcctgtgcttctgctaggatcaatgtgtag

Probe sequences: Probes are all conjugated to TAMRA. For sequences, see Table 5.

Table 5. MS probe sequences

MS Probe #	Sequence	MS Probe #	Sequence
1	tcctcgcccttgctcaccat	20	tgtagttgtactccagcttg
2	gggcaccaccccggtgaaca	21	catgatatagacgttgtggc
3	cgccgtccagctcgaccagg	22	tgccgttcttctgcttgtcg
4	ctgaacttgtggccgtttac	23	cggatcttgaagttcacctt
5	gccctcgccctcgccggaca	24	gctgccgtcctcgatgttgt
6	tcagcttgccgtaggtggca	25	tggtagtggtcggcgagctg
7	gtggtgcagatgaacttcag	26	gtcgccgatgggggtgttct
8	ccagggcacgggcagcttgc	27	ttgtcgggcagcagcacggg
9	tcagggtggtcacgagggtg	28	ggactgggtgctcaggtagt
10	ctgaagcactgcacgccgta	29	cgttggggtctttgctcagg
11	cttcatgtggtcggggtagc	30	aggaccatgtgatcgcgctt
12	cggacttgaagaagtcgtgc	31	ggcggcggtcacgaactcca
13	tggacgtagccttcgggcat	32	tcgtccatgccgagagtgat
14	cttgaagaagatggtgcgct	33	catggctaagcttcttgtac
15	gggtcttgtagttgccgtcg	34	tcctccacctccggcgggaa
16	ccctcgaacttcacctcggc	35	atgggcagcgtgccatcatc
17	gatgcggttcaccagggtgt	36	tcccgctctcctgggcacaa
18	ttgaagtcgatgcccttcag	37	acaggctgcagggtgacggt
19	ccccaggatgttgccgtcct	38	acacattgatcctagcagaa

Image Processing and Data Analysis of smFISH

Raw image processing

Raw DAPI images stacks were deconvoluted along the x-y and z planes (in stack format) with Huygens with the following parameters: 20–30 for signal-to-noise ratio, and OFF for breaching. Deconvoluted DAPI images were resized (scaling factor 0.5) using MATLAB. Examples of DAPI images before and after deconvolution are shown in Figure 25.

Cellular and nuclear segmentation

Individual cell boundaries, the numbers of TCs for each single cells, and the x-y location of TCs were manually selected and segmented using in-house MATLAB programs (available upon request). The center positions of TCs were determined at the pixel position where the product of the intensity of the TAMRA and Quasar 670 channels was greatest in a 10 × 10 pixel that was created over all z-plans.

In-house MATLAB programs were used to determine the boundary between the nucleus and cytoplasm of individual cells with the deconvoluted DAPI images (available upon request). Note that due to the increasing error rate exhibited by the edge-detection algorithms near the top and bottom z-planes, only results from the middle sections of stacks (5–15 images along one stack) were considered as nuclear area. For example image of nuclear autosegmentation, see Figure 25.

Spot finding

The software package available at: http://labs.biology.ucsd.edu/rifkin/software.html (Rifkin, 2011) was used to determine spots. To ensure proper segmentation and control for day-to-day variation in imaging, images from the same channel, sample and experiment were used to train the program. All training processes generated the output spots with less than 5% false-positive and false-negative error rates. For examples of spot finding results, see Figure 26.

Spot classification and quality controls

Any signal within the same 2 × 2 × 2 pixel cube, was considered a single mRNA molecule, and one spot from the pair was randomly selected and excluded from further processing. If spots from two different channels were located in the same 3 × 3 × 3 pixel cube, the spots were considered co-localized. The locations of co-localized spots were computed by averaging the positions of the pairs of spots from two different channels. The distance between each spot and TC was described by the Euclidian distance from a two-dimensional projection of the image stack.

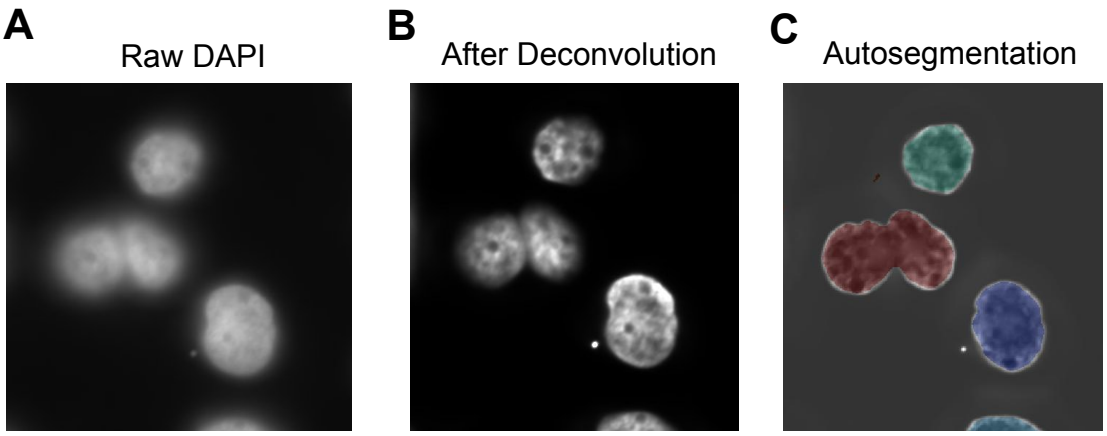

Figure 25. Raw DAPI imaging processing and nuclear segmentation in smFISH experiments

(A) Raw DAPI image. The edge of nucleus is blurry. (B) After deconvolution with Huygens, the edge of nucleus is more defined and easier to segment the nucleus with edge-detection algorithm.

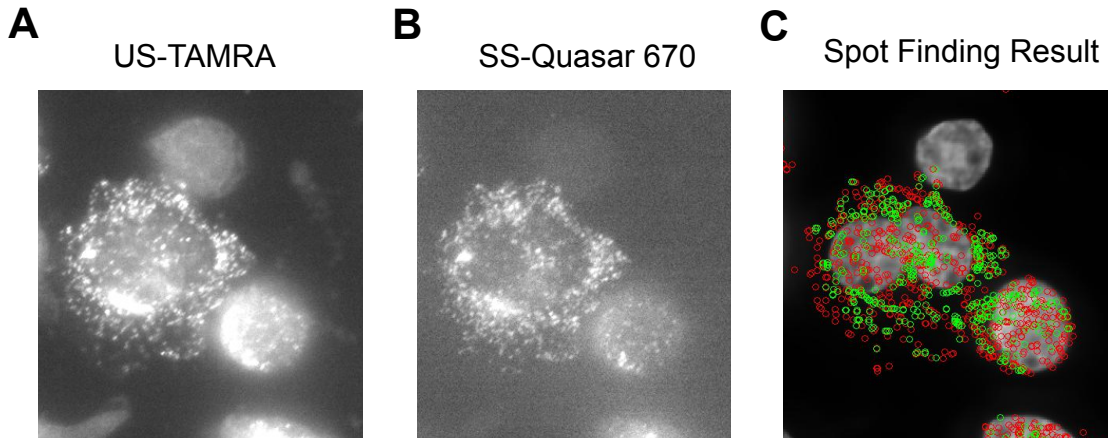

Figure 26. Example of automated spot finding results

Example images of cells stained with US/SS probe sets. (A) Raw image from the TAMRA channel (for detecting US probes) (B) Raw image from the Quasar 670 channel (for detecting SS probes). (C) Final spot finding results (green: spots found in US-TAMRA channel; red: spots found in SS Quasar 670 channel). Positions of identified spots form both channels will be projected onto one in-focus DAPI image. This is the quality control step to make sure the automated spot finding program identifies the majority of spots.

CHAPTER 4: DEVELOPEMENT OF SATURN (SPLICING AFTER TRANSFECTION OF UNSPLICED PRE-MRNA INTO NUCLEUS) ASSAY

Abstract

In this chapter, we further investigate if HIV-1 pre-mRNA can recruit spliceosome machinery post-transcriptionally. In our post-transcriptional model, HIV-1 pre-mRNA goes through a serial cascade of splicing. One key property of this model is that spliceosome loading is decoupled from transcription. To test our hypothesis, we *in vitro* transcribed HIV-1 pre-mRNA from a truncated HIV-1 template. To maximize the assay sensitivity, we used β-lactamase (BlaM) and its fluorescent substrate CCF2-AM as the reporter system. *In vitro* transcribed unspliced pre-mRNA was delivered into the nucleus using nucleofection. Surprisingly, we found that cells transfected with unspliced pre-mRNA produced gene products that could only be translated from a spliced RNA. The existence of this *in vivo* post-transcriptional splicing is confirmed by using a version of HIV that obliterates splicing by mutating the major splice domain (MSD). Cells transfected with the MSD mutated unspliced pre-mRNA are not able to efficiently produce the gene products from spliced RNA. This assay reveals that HIV-1 pre-mRNA does not require co-transcriptional recruitment of spliceosome. Additionally, these findings suggest that post-transcriptional recruitment of splicing factors may be a common mechanism to rescue the unspliced pre-mRNA in eukaryotes.

Materials and Methods

Cloning

Lentiviral vector dHIV-BlaM was created by inserting the PCR amplified MluI/XhoI fragment of BlaM (Cavrois et al., 2002; Zlokarnik et al., 1998) into the MluI/XhoI restriction–digested backbone of dHIV-d2G. The dHIV-BlaM MSD mutant (dHIV-dMSD-BlaM) was created using PCR site-directed mutagenesis.

To obtain the DNA templates for *in vitro* transcription, XmaI site and T7 promoter sequence were inserted into the upstream of TSS in the 5' LTR, and XmaI site was inserted into the downstream of PolyA signal in the 3' LTR of dHIV-BlaM and dHIV-dMSD-BlaM to create the XmaI-dHIV-BlaM and XmaI-dHIV-dMSD-BlaM plasmids. For insertion of XmaI in the 5' LTR, the AgeI/SalI-XmaI-T7 fusion PCR fragment was amplified using primer sets of dHIV-AgeI F, dHIV-XmaIT7TSS R, dHIV-XmalIT7TSS F, and dHIV-SalI R, and then inserted into the AgeI/SalI restriction digested backbone of dHIV-BlaM or dHIV-dMSD-BlaM. To insert the XmaI restriction site into the 3' LTR, the XhoI/PacI-PolyAXmaI fragment was first amplified by fusion PCR with primer sets of dHIV-XhoI F, dHIV-PolyAXmaI R, dHIV-PolyAXmaI F, and dHIV-PacI R. The XhoI/PacI-PolyAXmaI fusion PCR fragment was later inserted into the XhoI/PacI restriction digested dHIV-BlaM or dHIV-dMSD-BlaM (already XmaI/T7 modified). For primer sequences used in this Chapter, see Table 6 in Appendix 4.

In vitro Transcription

The DNA template for *in vitro* transcription was obtained by gel purification of XmaI restriction digested XmaI-dHIV-BlaM and XmaI-dHIV-dMSD-BlaM. We used mMESSAGE mMACHINE® T7 ULTRA Kit (Life technologies) to in vitro transcribe and poly-A tail the RNA. Approximate 0.3 μg of DNA template was used in each reaction, and cleaned up by turbo DNase, following the manufacturer's instruction. The RNA product was further purified with RNeasy Mini Kit (QIAGEN).

Transfection of RNA into the Nucleus

We used Cell Line Nucleofector® Kit R (Lonza) to transfect artificial mRNA into the nucleus of naive Jurkat cells. For each nucleofection, 2 μg of RNA and 1.6×10^6 Jurkat cells were used. For the electroporation step in the nucleofection, we used pre-set program O-028 in the Nucleofector device. Control transfection efficiency was estimated using 2 μg of pmaxGFP (Lonza) plasmid per nucleofection.

BlaM Assay and Flow Cytometry

We used LiveBLAzer™ FRET-B/G Loading Kit with CCF2-AM (#K1032, Life technologies) to detect β-lactamase positive (BlaM+) cells. Jurkat cells were harvested 2 hours post RNA nucleofection at 800 g for 5 minutes, and washed once with CO_2-independent medium (#18045-088, Life Technologies). Because the fluorescent substrate CCF2-AM is light sensitive, all the following procedures were performed with limited exposure to light. The pellet was resuspended in CCF2-AM loading medium (1 μM CCF2-AM, 1 mg/mL pluornic-F127, 0.001% acetic acid in CO_2-independent medium), followed by a 1-hour incubation on the bench top. Cells

were spun down and the pellet was resuspended in the development medium (2.5 mM probenecid, 10% FBS in CO_2-independent medium). The β-lactamase enzymatic reaction was performed by incubating the cells for 16 hours at room temperature. Cells were later washed once with PBS, and resuspended in fixation buffer (2% paraformaldehyde in PBS) at 4°C over night. We used LSR II cytometer (BD Biosciences) with 404-nm laser to detect the green-to-blue shift in emission fluorescence. The cytometry data were analyzed using FlowJo (http://www.flowjo.com/).

Results and Discussion

To further verify if HIV-1 pre-mRNA bears the capacity to be post-transcriptionally spliced, a new approach termed the SATURN assay, was developed (Figure 27). We utilized β-lactamase (BlaM) as the reporter gene to replace the d2GFP in the dHIV-d2G construct, and tested if unspliced pre-mRNA of this dHIV-BlaM can be post-transcriptionally spliced. The detection limit of BlaM enzymatic function is less than 100 BlaM molecules (Zlokarnik et al., 1998). As a result of its high sensitivity, BlaM has been used in other assays that require high sensitivity, such as the HIV-1 virion fusion assay (Cavrois et al., 2002). The 5' capped, 3' polyadenylated artificial dHIV-BlaM unspliced pre-mRNA was *in vitro* transcribed and transfected into the nucleus of naive Jurkat cells by nucleofection. Nucleofection incorporates electroporation and chemical reagents to deliver nucleic acids preferentially into the cell's nucleus. Without proper processing, the pre-mRNA remains unspliced, and there would be no observable BlaM expression. However, if the unspliced pre-mRNA is properly processed, the spliced form can express BlaM. After nucleofection, 22% of cells were BlaM positive (Figure 28), indicating that unspliced pre-mRNA can be post-transcriptionally spliced. The control transfection percentage was around 50%. Next, we transfected cells with the dHIV-BlaM artificial pre-mRNA only carrying one single-nucleotide mutation at the major splicing donor (MSD) motif. The MSD is required to obtain correct SS or MS RNA encoding Tat, Rev, or Env (Tazi et al., 2010). Remarkably, the percentage of BlaM-positive cells dropped to 7%. The result of dHIV-BlaM MSD mutant shows that the post-

transcriptional splicing of HIV-1 pre-mRNA does occur *in vivo*, and is responsible for the BlaM expression in the SATURN assay. More importantly, because there is no viral protein in the cells upon transfection, the assay shows that naked HIV-1 pre-mRNA can recruit spliceosomes post-transcriptionally. With the previous smFISH results and the single-cell imaging analysis, we conclude that post-transcriptional splicing of HIV-1 leads to Rev-mediated negative-feedback control on HIV-1 gene expression.

Acknowledgements

In this chapter, Marielle Cavrois provided the original BlaM plasmid and technically supported the BlaM assay. MSD mutated dHIV-BlaM was created by Jac Luna. Tim Notton and Leor Weinberger conceived the name of the SATURN assay. I designed and performed the experiment and analysis. Leor Weinberger directed the research.

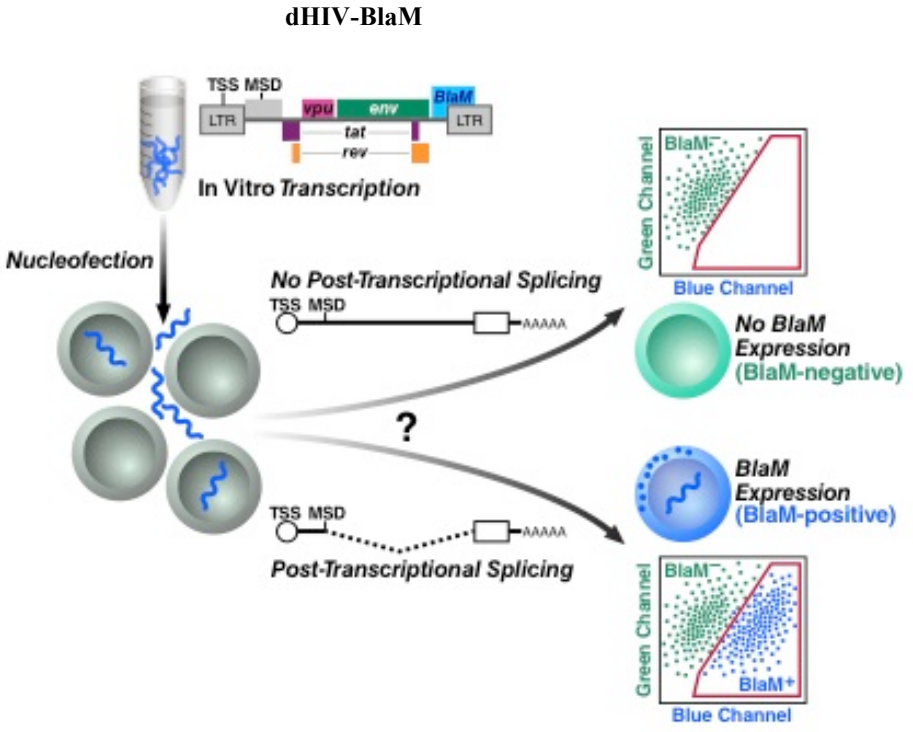

Figure 27. Experimental design of the SATURN assay

5'-Capped, and 3'-polyadenylated dHIV-BlaM (β-lactamase) pre-mRNA is *in vitro* transcribed. Purified pre-mRNA is nucleofected into naive Jurkat cells, and CCF2-AM, the fluorescent substrate of β-lactamase, is loaded after a 2-hour incubation of recovery from nucleofection. If cells uptake the unspliced pre-mRNA into the nucleus and splice the pre-mRNA post-transcriptionally, the spliced RNA can express BlaM to convert the green fluorescent substrate into the blue fluorescent product. Otherwise, the cells remain green fluorescent due to lack of BlaM expression. Flow cytometry is later utilized to determine the percentage of cells capable of post-transcriptional splicing.

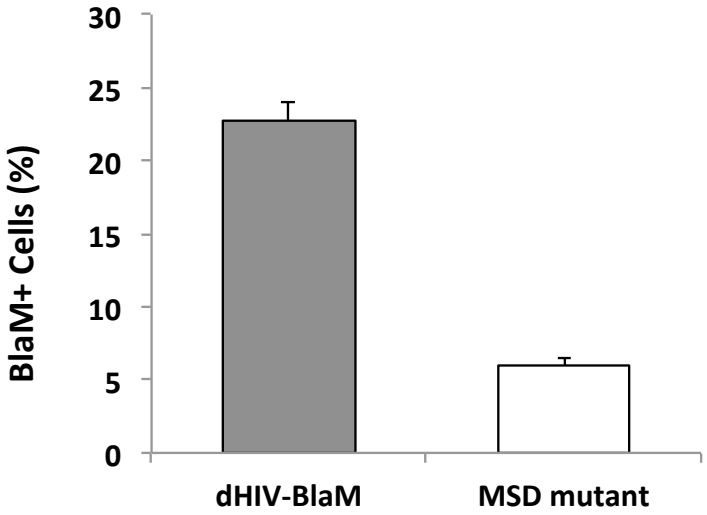

Figure 28. Results of SATURN assay show that unspliced HIV-1 pre-mRNA recruits spliceosome post-transcriptionally *in vivo*

Over 22% of cells were BlaM positive. However, a single-nucleotide mutation in the major splicing donor (MSD) motif drastically suppress the percentage of BlaM positive cells to 7%, suggesting unspliced HIV-1 pre-mRNA can recruit cellular splicing machinery post-transcriptionally in a viral protein-free context. (Control transfection efficiency is estimated ~50%.)

Appendix 4

Table 6. Primer sequences for cloning in Chapter 4

Primer Name	Sequence	Purpose
MluI-BlaM F	tataacacgcgtgccaccatggatcctgagaccctg gtgaaagtg	For cloning dHIV-BlaM
XhoI-BlaM R	ctaggtctcgagtcatcaccagtgcttgatcaggct	For cloning dHIV-BlaM
dHIV-AgeI F	gtggataccggtggtactgcgg	For inserting XmaI-T7 promoter sequence into dHIV-BlaM
dHIV-XmaIT7TSS R	agcagctgctttttgcctgtactggcccgggtaatac gactcactatagg	For inserting XmaI-T7 promoter sequence into dHIV-BlaM
dHIV-XmaIT7TSS F	tggcccgggtaatacgactcactatagggtctctctg gttagaccagatctgagcc	For inserting XmaI-T7 promoter sequence into dHIV-BlaM
dHIV-SalI R	gctctcctctgtcgagtaacgcctattctgctatgtcg ac	For inserting XmaI-T7 promoter sequence into dHIV-BlaM
dHIV-XhoI F	gtgtagctcgagacctagaaaaacatggagc	For inserting XmaI restriction site into the downstream of dHIV-BlaM polyA signal
dHIV-PolyAXmaI R	tagggaacccactgcttaagcctcaataaacccggg gcttgccttgag	For inserting XmaI restriction site into the downstream of dHIV-BlaM polyA signal
dHIV-PolyAXmaI F	aataaacccggggcttgccttgagtgcttcaagtagt gt	For inserting XmaI restriction site into the downstream of dHIV-BlaM polyA signal
dHIV-PacI R	ttgaggattaattaatatctgtcagaatagtaaattaga taagtttatac	For inserting XmaI restriction site into the downstream of dHIV-BlaM polyA signal
dHIV-dMSD- F	ggggcggcgactgatgagtacgc	For creating MSD mutated dHIV-BlaM
dHIV-dMSD-R	gcgtactcatcagtcgccgcccc	For creating MSD mutated dHIV-BlaM

CHAPTER 5: PREDICTION AND CHARACTERIZATION OF REV NEGATIVE FEEDBACK

Abstract

Here, we summarize the findings in previous chapters to characterize Rev negative feedback in detail. First, we fit the models to our experimental data and use the fits to generate informative predictions. Our model fitting results show that the post-transcriptional splicing cascade is requisite for generating the overshoot feature. Next, we examine the cooperativity of Rev/RRE-dependent RNA nuclear export. Our ODE simulations indicate that a high Hill coefficient is required, and the minimum Hill coefficient should be larger than 3 to fit our data. Stochastic simulations performed in parallel support this conclusion. Additionally, we used our model to predict HIV-1 expression dynamics under various perturbations. Our models suggested that Rev overexpression would lead to a decrease in the levels of both the MS and US gene products. We experimentally verified this prediction by transfecting Jurkat HIV-d2G cells with exogenous Rev expressing vector. More importantly, this counterintuitive finding implies that contradictory experimental results with perturbing Rev functions may be merely the consequence of a hidden coupled Tat/Rev gene regulatory circuitry.

Materials and Methods

Ordinary-Differential Equation Models and Stochastic Simulations

ODE models were developed and simulated using Berkeley Madonna and MATLAB. General trend lines obtained from microscopic experiments were fit in Berkeley Madonna. Sensitivity analysis and parameter plots for the ODE model were performed in MATLAB or Berkeley Madonna.

For stochastic simulations, chemical reaction schemes were coded in programming language c using the Gillespie algorithm (Gillespie, 2007; Gillespie et al., 2013). The simulation results were analyzed using MATLAB.

Rev Overexpression

We used Cell Line Nucleofector® Kit R (Lonza) to transfect pcRev plasmid DNA (Malim et al., 1988) into HIV-d2G Jurkat cells. For each transfection, 1.6×10^6 Jurkat cells and pre-set program O-028 were used. 10 ng/mL TNF-α was added into the culture 24 hours post transfection and incubated for another 24 hours. Intracellular p24 were stained using anti-p24 antibody KC57-RD1 (Coulter Clone). Intracellular GFP and p24 levels were measured using the flow cytometer High Throughput iQue Screener (IntelliCyt). Flow data were analyzed with FlowJo software.

Results and Discussion

Finally, we used our data of HIV-1 single-cell expression trajectories to fit the mathematical model. (For details, see Appendix 5). We compared the fitting results of co-transcriptional splicing model and post-transcriptional splicing model and found that only post-transcriptional splicing model provides a good fit (Figure 29). On top of that, we constructed a "mix" model including the elements of both co- and post-transcriptional splicing. The purpose of building this mix model is to investigate if leakage from co-transcriptional splicing would disrupt the negative feedback resulted from the post-transcriptional splicing cascade. The model fitting shows that post-transcriptional splicing is sufficient to generate the overshoot curve. Furthermore, we characterized the negative feedback circuitry by fitting the model with different Hill coefficients, the results show that the Hill coefficient of Rev RNA nuclear export should be larger than 3 (Figure 30).

We next used our model to predict the outcomes of exogenous Rev over-expression. If viral gene expression is under Rev negative-feedback control, MS RNA production will be repressed, leading to lower Tat levels. Tat activates HIV-1 transcription by 20–60 fold (Bohan et al., 1992). Therefore, the effect of reduced viral expression due to Rev-negative feedback would be amplified because of Tat transactivation. Accordingly, increased Rev-negative feedback would lead to decreased levels of p24 and Nef (Figure 31). To verify our predictions, we transfected latently infected HIV-d2G Jurkat cells with pcRev DNA (Malim et al., 1988), and

activated the replication with TNF-α 24 hours post-transfection. Surprisingly, with increasing Rev overexpression, levels of both GFP and p24 decreases at 24 hours post-activation (Figure 32). Our models successfully predict the counter-intuitive results, and these findings further strengthen the conclusion that serial post-transcriptional splicing and Rev nuclear export allow for negative-feedback regulation in HIV-1 gene expression.

Acknowledgements

In this chapter, Mauricio Montano technically supported the DNA transfection in the Rev over-expression experiment. Igor Rouzine and Leor Weinberger helped with the refinement/fitting of ODE models, Lisa Bishop helped with the stochastic simulations, and Panagiotis Voulgaris helped with the coding. I performed the experiments, analyzed the results, and performed the modeling, simulations, and computational analysis. Leor Weinberger directed the research.

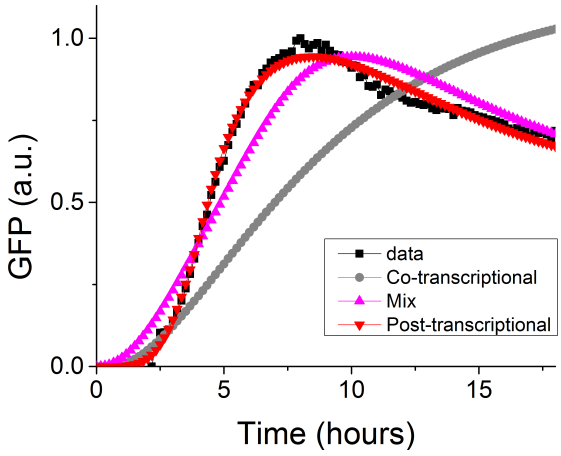

Figure 29. Post-transcriptional splicing ODE model provides the best fitting results

ODE models of HIV-1 gene expression circuitry were fit to the latently infected HIV-d2G data. Only the post-transcriptional splicing/Rev negative feedback model provides a good fit. Although the "mix" model can not fit the experimental data as well as the post-transcriptional model, it can still generate the overshoot curve. This suggests that post-transcriptional splicing is essential to generate negative feedback and is robust to the leakage from co-transcriptional splicing.

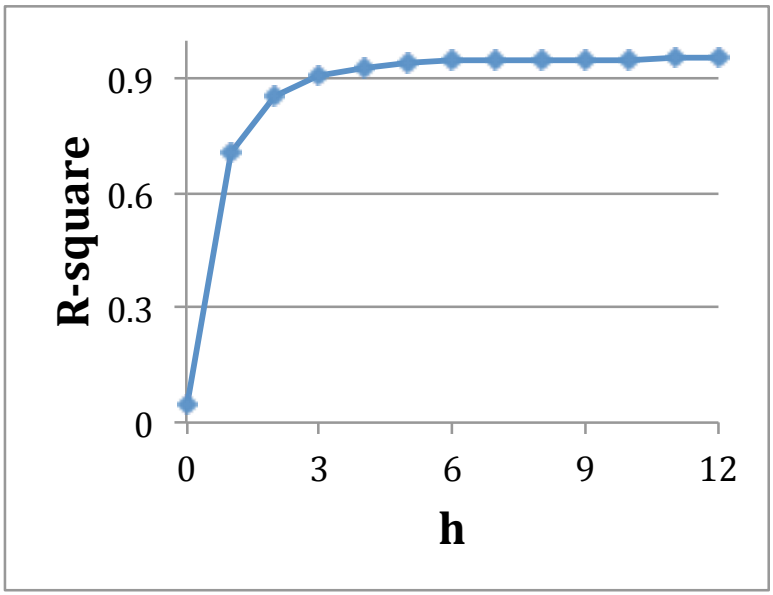

Figure 30. Hill coefficient h of Rev/RRE nuclear export should be larger than 3

When the h is lower than 3, the R-square of post-transcriptional model fitting is less than 0.9, and the trajectories lose the overshot feature. When the h is large (e.g., 12), other parameters can be made to fit the reference/physiologically reasonable range.

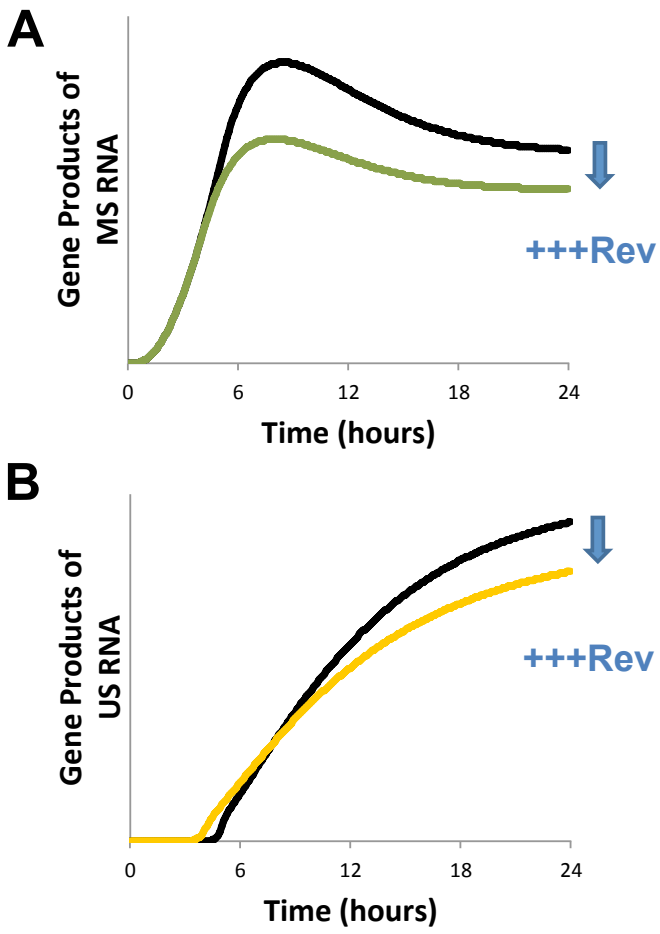

Figure 31. Prediction of the HIV-1 expression dynamics under Rev overexpression

Based on the post-transcriptional (Rev negative feedback) model, we predict how the dynamics of HIV gene expression respond to exogenous Rev overexpression. Surprisingly, expression of US (B) and MS (A) gene products will decrease when Rev is overexpressed. This prediction contradicts the common conclusion that Rev upregulates the expression of US RNA.

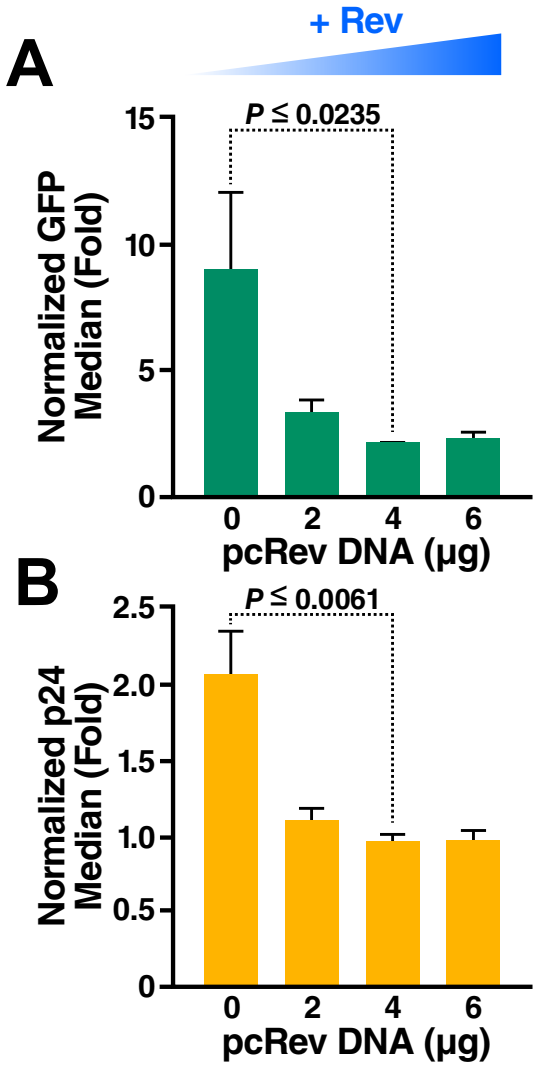

Figure 32. Overexpression of Rev decreases the level of p24 expression in full-length HIV-1

(A) Jurkat cells latently infected with HIV-d2G were transfected with pcRev plasmid DNA. At 24 hours post-transfection, TNF-α was added to activate viral replication. GFP and p24 levels were measured 24 hours after TNF-α treatment by immunostaining and flow cytometry, and the levels were normalized to the background fluorescence of naive Jurkat cells. As Rev overexpression was increased, GFP levels decreased. (B) The intracellular p24 level also decreased as Rev levels were increased.

Appendix 5

Fitting and Analysis of the ODE models

The "mix" model

To accommodate the possibility of co-transcriptional and post-transcriptional splicing coexisting in HIV-1 mRNA processing, we built a "mix" model. This system has five equations in total, and two are modified compared to the other models. The new equations are the following:

$$\frac{\mathrm{d}}{\mathrm{dt}}\mathrm{USn} = \mathrm{b_{US}} - \mathrm{sp}\cdot\mathrm{USn} - \frac{\mathrm{kr}\cdot\mathrm{USn}\cdot\mathrm{Rev^h}}{\mathrm{K_{Rev}{}^h}+\mathrm{Rev^h}} - \mathrm{drna}\cdot\mathrm{USn} \qquad (8)$$

$$\frac{\mathrm{d}}{\mathrm{dt}}\mathrm{MS} = \mathrm{b_{MS}} + \mathrm{sp}\cdot\mathrm{USn} - \mathrm{drna}\cdot\mathrm{MS} \qquad (9)$$

Eq. (8) is a mixture of Eq. (1) and (6) in Appendix 2-2, which consists of four biochemical events: the lumped co-transcriptional production of USn RNA, decay of USn due to post-transcriptional splicing, Rev-dependent nuclear export, and the RNA degradation. The Eq. (9) is a chimeric version of Eq. (3) and (7) in Appendix 2-2, where MS RNA can be produced either from the co-transcriptional production or splicing from USn RNA.

After obtaining gene expression trajectories (Chapter 2) from single-cell time-lapse imaging, we used Berkeley Madonna to fit the experimental data with the ODE models. We use the general trend line from a subset of HIV-d2G trajectories in figure

8, and relax the parameter values for b, kr, K_{Rev}, h, fr, and fg to fit, and with the goal of obtaining the smallest possible Hill Coefficient, h. The parameter values used for the post-transcriptional model after fitting are listed in the Table 7, and fitted plots are shown in the Figure 29.

From the fitting results, it is clear that only systems containing post-transcriptional splicing (the post-transcriptional model and the mix model) are able to produce the overshoot features shown in the data trajectory, and that the post-transcriptional splicing model provides the best fit.

Characterization of HIV-1 Gene Regulatory Circuitry

HIV-1 Rev homomultimerizes and interacts with the RRE region in the intron-containing RNA. However, the exact number of bound Rev oligomers per RRE is unclear. Previous reports suggest that Rev binds to the RRE in 2:1 (Daly et al., 1993), 3:1 (Cook et al., 1991), 4:1 (Daly et al., 1993; Mann et al., 1994; Pond et al., 2009), 6:1 (Daugherty et al., 2010), 8:1 (Cook et al., 1991; Daly et al., 1993), and up to 12:1 (Mann et al., 1994) ratios. For the Hill coefficient of Rev/RRE dependent RNA nuclear export, we observe that the Hill coefficient h needs to be high enough ($h \geq 3$) in order to get good fit ($R^2 \geq 0.9$) of the model (Figure 30).

Using Stochastic Models to Estimate the Cooperativity of Rev-Dependent Nuclear Export

To characterize the Rev-dependent RNA export reaction, we construct a simple stochastic model to describe the Rev-RRE interaction and the proceeding nuclear

export. The kinetics of Rev and RRE interaction at single-molecule level has been studied *in vitro*. Rev molecules binds to the RRE stem-loop one at a time, and the Rev-RRE interaction is reversible (Figure 33) (Pond et al., 2009). The maximum number of Rev oligomer bound to one single RRE stem-loop motif is 4. We assume that when four Rev molecules are bound to the RRE, it is sufficient to start the nuclear export process. Therefore, the system can be described as the following set of chemical reactions (Figure 33):

$$\mathrm{Rev} + \mathrm{RRE} \xrightarrow{k1} \mathrm{R1RRE} \qquad (10)$$

$$\mathrm{Rev} + \mathrm{R1RRE} \xrightarrow{k2} \mathrm{R2RRE} \qquad (11)$$

$$\mathrm{Rev} + \mathrm{R2RRE} \xrightarrow{k3} \mathrm{R3RRE} \qquad (12)$$

$$\mathrm{Rev} + \mathrm{R3RRE} \xrightarrow{k4} \mathrm{R4RRE} \qquad (13)$$

$$\mathrm{R1RRE} \xrightarrow{k5} \mathrm{Rev} + \mathrm{RRE} \qquad (14)$$

$$\mathrm{R2RRE} \xrightarrow{k6} \mathrm{Rev} + \mathrm{R1RRE} \qquad (15)$$

$$\mathrm{R3RRE} \xrightarrow{k7} \mathrm{Rev} + \mathrm{R2RRE} \qquad (16)$$

$$\mathrm{R4RRE} \xrightarrow{k8} \mathrm{Rev} + \mathrm{R3RRE} \qquad (17)$$

$$\mathrm{R4RRE} \xrightarrow{k9} \mathrm{RevRREc} \qquad (18)$$

In this minimal system, R1RRE stands for the Rev-RRE complex that one Rev molecule binds to the RRE, R2RRE stands for the complex of Rev-RRE complex that two Rev molecule binds to the RRE, and etc. Reaction (18) is the process that fully bound-nuclear Rev-RRE complex is exported to the cytoplasm, becoming the cytoplasmic form of Rev-RRE complex (RevRREc). This is the only irreversible reaction in the system, because Rev will eventually dissociate from the RRE, release the RRE-containing RNA in the cytoplasm, and recycle to shuttle/export viral RNA. Collectively, the overall reaction is

$$4\mathrm{Rev} + \mathrm{RRE} \rightarrow \mathrm{RevRREc} \qquad (19)$$

To simulate the Rev-RRE interaction and export, we used the Gillespie algorithm to code the chemical reaction in programming language c. The rates of reactions were obtained from previously published data (Love et al., 1998; Pond et al., 2009), and the values are converted into correct units for the simulation based on the reaction order. We also assume the reaction is taken place in an average size of Jurkat cells, which is the cell line most used in our study. The values of constants and parameter used in the simulation are summarized in Table 8. We iterate the simulation with the initial numbers of Rev and RRE molecules ranging from 10^1 to 10^6 with $10^{0.1}$ increments per iteration, and the initial values of all other species set to zero. For each initial value set, 1000 simulations were run, and the simulation time was approximately 120 minutes. In the case of long simulations, the set of with the same initial values was aborted.

We compute the overall reaction (19) half-time by fitting the average data of single simulation trajectories with the same initial conditions. The heat map of half-time is drawn in double log scale (Figure 34). The area at the bottom-right corner (low initial Rev-to-RRE ratio) is blank due to their very long reaction half-time (> 120 minutes simulation duration). A very drastic transition of the reaction half-time occurs on the heat map diagonally from the bottom-left to the top-right. Interestingly, the minimum Rev-to-RRE ratio in the "blue" area (half-time < 30 minutes) is approximately 3.1, which is very close to the minimum integer Hill coefficient ($h \geq 3$) required to fit in the ODE models. Although the stochastic simulation can not directly measure the cooperativity between Rev and RRE in previous studies, the simulation results support the observation that a 3:1 Rev-to-RRE ratio is required to initiate efficient Rev/RRE-dependent RNA nuclear export.

Prediction of HIV-1 Gene Regulatory Circuits with Rev Over-Expression

The final goal of constructing models for the HIV-1 full-length gene regulatory circuit is to assist the development of new antiviral therapies. A reliable model that gives correct prediction can provide valuable insight into this process, and help with interpretations of experimental results generated from antiviral drug screening. To test if our fitted model accurately predicts viral expression, we decided to over-express Rev in the system, and perform experiments to verify if our models generate accurate predictions.

We simulated Rev overexpression by modifying Eq. (4) to the following:

$$\frac{d}{dt}\mathrm{Rev} = b_{\mathrm{Rev}} + p \cdot fr \cdot MS - dr \cdot \mathrm{Rev} \qquad (20)$$

The extra parameter, b_{Rev}, is the production rate of Rev from an exogenous source other than HIV LTR. Experimentally, this can be performed by transfecting cells with Rev-encoding expression vectors. We overexpress Rev 24 hours before the onset of HIV-1 reactivation. Thus, the minimal Rev-only model will not be sufficient to describe the scenario, because the exogenous Rev will exist in the system before Tat positive feedback initiates HIV gene expression, in contrast to the previous minimal Rev negative-feedback model. As a result, the Tat positive feedback term in the ODE models can not be simplified and lumped with the basal transcription in this case. Therefore, we replace Eq. (6) with the following:

$$\frac{d}{dt}\mathrm{USn} = b_{\mathrm{LTR}} + \frac{kt \cdot \mathrm{Tat}}{K_{\mathrm{Tat}} + \mathrm{Tat}} - sp \cdot \mathrm{USn} - \frac{kr \cdot \mathrm{USn} \cdot \mathrm{Rev}^h}{K_{\mathrm{Rev}}{}^h + \mathrm{Rev}^h} - drna \cdot \mathrm{USn} \quad (21)$$

where b_{LTR} is basal expression rate from LTR promoter, kt is the maximum Tat transactivation rate, and K_{Tat} is the Michaelis-Menton saturation concentration of Tat transactivation. This new equation consists of the following elements: LTR basal transcription, Tat transactivation, post-transcriptional splicing, Rev-dependent nuclear export, and RNA degradation. We also add another equation to describe the Tat production and degradation:

$$\frac{d}{dt}\mathrm{Tat} = p \cdot ft \cdot MS - dtat \cdot \mathrm{Tat} \qquad (22)$$

where ft is the fraction of Tat encoding MS RNA among all MS RNA, and dtat is the degradation rate of Tat protein. We firstly fit the new Tat/Rev coupled feedback model

to the experimental data in Figure 8. The final parameters are listed in Table 9. Next, we simulate how expression dynamics change when Rev is overexpressed. To start with, we assume that the production of exogenous Rev reaches a steady state. Therefore, we can calculate the initial Rev concentration as the following formula:

$$\mathrm{Rev}_{\mathrm{Initial}} = \frac{b_{Rev}}{dr} \qquad (23)$$

We probe the system by tuning the b_{Rev}. Notably, when the b_{Rev} is not zero (i.e., Rev is overexpressing in the background), both GFP and USc levels decrease (Figure 31).

Table 7. Parameter values of the post-transcriptional model after fitting

Parameter	Description	Values [Units]	Justification or Reference
b	Lumped transcription rate of basal LTR and Tat positive feedback	120.98 [RNA copies $cell^{-1}$ h^{-1}]	Fit to data.
sp	Splicing rate of US to MS	1.5 [h^{-1}]	(Singh and Padgett, 2009)
kr	Max Rev-dependent RNA nuclear export rate	5.53 [h^{-1}]	Fit to data.
h	Hill coefficient. cooperativity of Rev-dependent RNA nuclear export	12 [no unit]	Maximum number of Rev per RRE. (Mann et al., 1994)
K_{Rev}	Rev saturation threshold concentration of Rev-dependent RNA nuclear export	45585 [protein copies $cell^{-1}$]	Fitting in literature range (Reddy and Yin, 1999; Rempala et al., 2006)
drna	Degradation rate of HIV-1 mRNAs	0.173 [h^{-1}]	(Felber et al., 1989; Malim and Cullen, 1993; Schwartz et al., 1992a)
p	General protein production rate	270 [protein copies $cell^{-1}$ RNA^{-1} h^{-1}]	(Kim and Yin, 2005)
fr	Fraction of Rev-encoding mRNA among all MS	0.046 [no unit]	Fit to data
dr	Degradation rate of Rev	0.173 [h^{-1}]	(Kubota et al., 1996)
fg	Fraction of GFP-encoding mRNA among all MS	5.48×10^{-6} [no unit]	Fit to data
dg	Degradation rate of d2GFP	0.346 [h^{-1}]	2 hr theoretical half-life

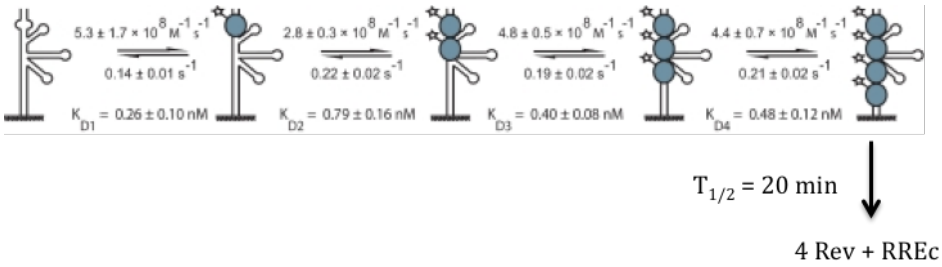

Figure 33. Reaction schemes of Rev-RRE mediated RNA nuclear export for stochastic simulations

In this reaction scheme, Rev binds to RRE stem-loop structure in a one-by-one manner (Pond et al., 2009). When the RRE stem-loop is fully loaded with Rev (4 Revs to 1 RRE), the Rev/RRE complex is removed from the nuclear reaction pool (Love et al., 1998). Figure courtesy of Pond et al. (Pond et al., 2009)

Table 8. Parameters and constants used in stochastic simulation of Rev-RRE nuclear export

Eq. No.	Parameter	Reactions	Values in literature	Reference
10	k1	Rev + RRE → R1RRE	1.91 molecule^{-1} sec^{-1}	(Pond et al., 2009)
11	k2	Rev + R1RRE → R2RRE	1.01 molecule^{-1} sec^{-1}	(Pond et al., 2009)
12	k3	Rev + R2RRE → R3RRE	1.73 molecule^{-1} sec^{-1}	(Pond et al., 2009)
13	k4	Rev + R3RRE → R4RRE	1.59 molecule^{-1} sec^{-1}	(Pond et al., 2009)
14	k5	R1RRE → Rev + RRE	R1RRE dissociation half time = 4.95 sec	(Pond et al., 2009)
15	k6	R2RRE → Rev + R1RRE	R1RRE dissociation half time = 3.15 sec	(Pond et al., 2009)
16	k7	R3RRE → Rev + R2RRE	R1RRE dissociation half time = 3.64 sec	(Pond et al., 2009)
17	k8	R4RRE → Rev + R3RRE	R1RRE dissociation half time = 3.30 sec	(Pond et al., 2009)
18	k9	R4RRE → RevRREc	Nuclear export and dissociation half-time = 20 minutes	(Love et al., 1998)

Avagadro's number (nA) = 6.023×10^{23}

$\pi = 3.14159265359$

Cell diameter (r_{cell}) = 5.8×10^{-6} m (measured in-house)

Cell volume (vol) = $0.75 \times \pi \times r_{cell}^3$ m^3 (assuming cell shape is a perfect sphere.)

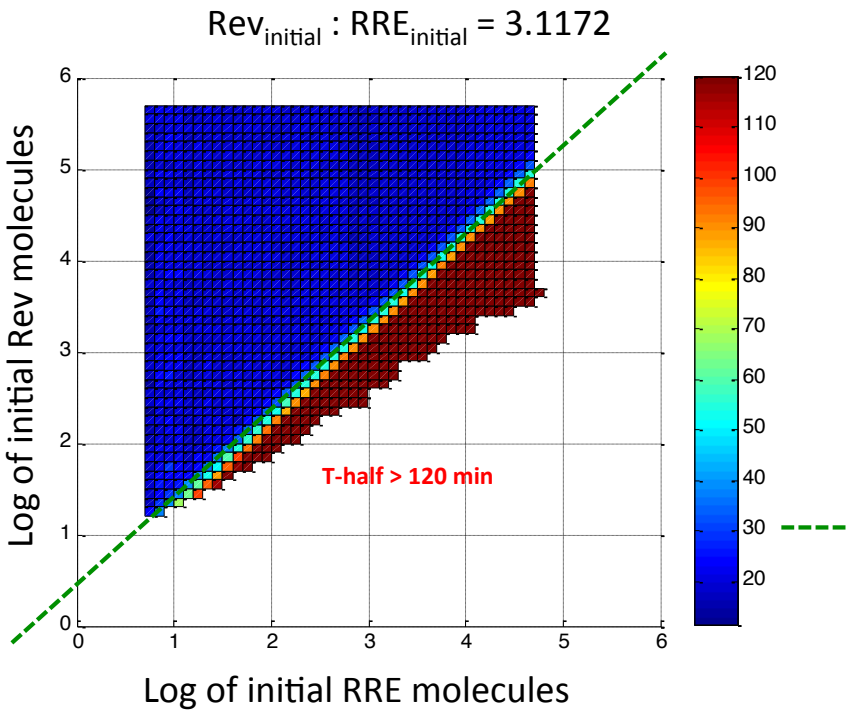

Figure 34. Heat map of reaction half-life of Rev-RRE nuclear export

The heat map shows that at least a 3.1 : 1 ($Rev_{initial}/RRE_{initial}$) ratio is required to ensure efficient Rev/RRE-mediated nuclear export.

Table 9. Parameter values of the Tat/Rev coupled feedback model after fitting

Parameter	Description	Values [Units]	Justification or Reference
b_{LTR}	Basal LTR transcription rate	0.87 [RNA copies $cell^{-1}$ h^{-1}]	Fitting in literature range. (Kim and Yin, 2005)
sp	Splicing rate of US to MS	1 [h^{-1}]	Fit to data; close to literature range. (Singh and Padgett, 2009)
kr	Max Rev-dependent RNA nuclear export rate	14.4 [h^{-1}]	Fit to data.
h	Hill coefficient. cooperativity of Rev-dependent RNA nuclear export	12 [no unit]	(Mann et al., 1994)
K_{Rev}	Rev saturation threshold concentration of Rev-dependent RNA nuclear export	247 [protein copies $cell^{-1}$]	Fit to data.
drna	Degradation rate of HIV-1 mRNAs	0.173 [h^{-1}]	(Felber et al., 1989; Malim and Cullen, 1993; Schwartz et al., 1992a)
p	General protein production rate	270 [protein copies $cell^{-1}$ RNA^{-1} h^{-1}]	(Kim and Yin, 2005)
fr	Fraction of Rev-encoding mRNA among all MS	0.19 [no unit]	(Purcell and Martin, 1993; Robert-Guroff et al., 1990)
dr	Degradation rate of Rev	0.173 [h^{-1}]	(Kubota et al., 1996)
fg	Fraction of GFP-encoding mRNA among all MS	0.79 [no unit]	(Purcell and Martin, 1993; Robert-Guroff et al., 1990)
dg	Degradation rate of d2GFP	0.346 [h^{-1}]	2 hr theoretical half-life
kt	Tat-induced transcription rate	42.4 [RNA copies $cell^{-1}$ h^{-1}]	48.7-fold of basal. Fitting in literature range. (Bohan et al., 1992; Graeble et al., 1993; Laspia et al., 1993)
K_{Tat}	Saturation concentration of Tat transactivation saturation	1152 [protein copies $cell^{-1}$ RNA^{-1} h^{-1}]	Fit to data
ft	Fraction of Tat-encoding mRNA among all MS	0.02 [no unit]	(Purcell and Martin, 1993; Robert-Guroff et al., 1990)
dtat	Degradation rate of Tat	0.087 [h^{-1}]	8 hr half-life (measured in-house)

CHAPTER 6: DISCUSSION

The data in previous chapters show that HIV-1 Rev negative-feedback circuitry requires a cascade of post-transcriptional splicing. While coupling transcription and splicing may allow splicing to be being governed by upstream transcriptional regulators and enhance the efficiency of producing fully matured mRNAs, post-transcriptional splicing enables an additional layer of feedback control if coupled with the other downstream events, such as RNA nuclear export or degradation.

Significance and Implications of HIV-1 Rev Negative Feedback

Because Tat and Rev are both encoded in the class of MS RNA, Rev-negative feedback is functionally coupled with Tat-positive feedback (Weinberger et al., 2005; Weinberger and Shenk, 2007). This coupled feedback-regulatory architecture makes it difficult to predict the outcomes of perturbing the system. Tat positive feedback is a strong amplifier and even a small decrease in Tat positive-feedback strength can generate a significant reduction in overall viral transcription. In other words, diminishing the expression of MS RNA through Rev-negative feedback will also repress the expression of Tat and thus weaken the overall transcription. As a result, mild/moderate inhibition of Rev function with drugs (Figure 18) actually can elevate the level of MS RNA gene products. Conversely, overexpression of Rev, which further reduced the MS RNA production, leads to down-regulation of US and MS gene products (Figure 32). The interpretation of the effects of antiviral compounds that target Rev may be complicated due to this counterintuitive result.

Negative feedback, in contrast to positive feedback, generally leads to acceleration (Alon, 2007; Savageau, 2011). In terms of noise, positive feedback increases the noise level in the system, and negative feedback reduces the noise (Cox et al., 2008). HIV-1 Rev-negative feedback, unlike conventional negative feedback circuits that control expression by regulating the promoter, functions at the downstream RNA level. A potential advantage of negative-feedback regulation post-transcriptionally is the ability to reduce stochastic fluctuations. Theoretical studies have compared the noise level of transcriptional negative feedback with post-transcriptional negative feedback (Singh, 2011) and have shown that negative feedback on the post-transcriptional steps confers lower noise. The noisy HIV-1 Tat positive feedback controls viral decision-making between active replication and a proviral latent state (Weinberger et al., 2005; Weinberger et al., 2008). While stochastic fluctuations are utilized in decision-making processes and provide beneficial effects, rigorous control of expression can facilitate the system to reach the final state with less variation. The Rev-dependent RNA nuclear export, which contributes to the expression of viral structural proteins and translocation of viral genomic RNA, is the key switch of the early-to-late gene expression. Extra reduction of expression noise in this step may contribute to better temporal or quantitative control of the viral expression, and thus assists viral replication or transmission in the competition with host immune response.

Implications of HIV-1 Splicing Mechanism and SATURN Assay

HIV-1 pre-mRNA splicing was generally believed to be co-transcriptional after the discovery of the functional role of Tat-SF1, that the Tat-SF1-U snRNP stimulates the HIV-1 RNA splicing *in vitro* (Fong and Zhou, 2001). In contrast, recent evidence shows that Tat-SF1 functions independently of transcription and splicing *in vivo* (Miller et al., 2011). How then is HIV pre-mRNA spliced? Our smFISH study shows that in the early stages of replication, HIV-1 pre-mRNA splicing is post-transcriptional. It remains unclear which mechanism or factor(s) cause HIV-1 to be spliced post-transcriptionally instead of the conventional co-transcriptional pathway. However, one possibility is that rate of HIV-1 transcription is faster than rate of splicing. The transcription elongation rate of HIV-1 is very fast in comparison with other genes (Boireau et al., 2007). The fast rate of transcription elongation may enable the completion of transcription to occur before splicing machinery can function or assemble. Moreover, HIV-1 pre-mRNA has many alternative splice sites and many of the alternative splice sites are weak compared the constitutive splice sites. Generally, it takes longer for splicing to complete at weak alternative splice sites (de la Mata et al., 2010; Pandya-Jones and Black, 2009). Therefore, HIV-1 pre-mRNA splicing may be too slow to be completed before transcription. Another possibility is that the virus hijacks cellular factors to deviate the viral splicing pathway from the co-transcriptional default. Last of all, nascent HIV-1 unspliced pre-mRNA might be shuttled to the nuclear speckles, which are nuclear dynamic structures enriched of splicing factors, act as the centers of post-transcriptional pre-mRNA splicing, and associate with RNA

nuclear export (Girard et al., 2012; Spector and Lamond, 2011). These speculations, nevertheless, need to be examined experimentally.

To dissociate the splicing step from transcription *in vivo*, we developed the SATURN assay to test if HIV-1 pre-mRNA can be post-transcriptionally spliced. The result of HIV-1 SATURN assay further supports our smFISH data. Particularly, the main concept of SATURN assay is very similar to the early cell-free *in vitro* splicing assays (Manley et al., 1980; Padgett et al., 1983). Both of them decouple splicing from transcription, but the SATURN assay detects the splicing events *in vivo* with high sensitivity. Our findings show that HIV-1 unspliced mRNA is capable of being spliced post-transcriptionally *in vivo* without the need of viral proteins. This suggests that the recruitment of spliceosomes is not necessarily co-transcriptional. Along with the early cell-free *in vitro* splicing studies, it is shown that there is an endogenous pathway for pre-mRNA splicing (i.e., that spliceosome recruitment and splicing can occur post-transcriptionally). This pathway is likely a viral exploitation of an endogenous cellular mechanism. It is important to further investigate if unspliced pre-mRNA from other endogenous genes is capable of being spliced post-transcriptionally *in vivo.* In addition, several questions should also be addressed regarding the post-transcriptional pathway. First, what factor(s) prevents the co-transcriptional recruitment of spliceosome? Second, what factor(s) is responsible for the post-transcriptional recruitment of spliceosome? Third, what factor(s) prevents incompletely spliced RNA variants from entering this pathway?

Our findings demonstrate that post-transcriptional splicing of HIV-1 pre-mRNA and Rev-dependent RNA nuclear export leads to Rev-negative feedback. The results suggest that HIV-1 exploits a minority host splicing mechanism to serve for viral replication. This study also provides novel insight into the development of new targets for antiviral agents. More fundamentally, our study indicates the existence of an underlying post-transcriptional splicing mechanism.

REFERENCES

.

Alexander, R.D., Innocente, S.A., Barrass, J.D., and Beggs, J.D. (2010). Splicing-dependent RNA polymerase pausing in yeast. Molecular cell *40*, 582-593.

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

Ameur, A., Zaghlool, A., Halvardson, J., Wetterbom, A., Gyllensten, U., Cavelier, L., and Feuk, L. (2011). Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. Nature structural & molecular biology *18*, 1435-1440.

Bohan, C.A., Kashanchi, F., Ensoli, B., Buonaguro, L., Boris-Lawrie, K.A., and Brady, J.N. (1992). Analysis of Tat transactivation of human immunodeficiency virus transcription in vitro. Gene expression *2*, 391-407.

Bohne, J., and Krausslich, H.G. (2004). Mutation of the major 5' splice site renders a CMV-driven HIV-1 proviral clone Tat-dependent: connections between transcription and splicing. FEBS letters *563*, 113-118.

Boireau, S., Maiuri, P., Basyuk, E., de la Mata, M., Knezevich, A., Pradet-Balade, B., Backer, V., Kornblihtt, A., Marcello, A., and Bertrand, E. (2007). The transcriptional cycle of HIV-1 in real-time and live cells. The Journal of cell biology *179*, 291-304.

Braunschweig, U., Gueroussov, S., Plocik, A.M., Graveley, B.R., and Blencowe, B.J. (2013). Dynamic integration of splicing within gene regulatory pathways. Cell *152*, 1252-1269.

Brice, P.C., Kelley, A.C., and Butler, P.J. (1999). Sensitive in vitro analysis of HIV-1 Rev multimerization. Nucleic acids research *27*, 2080-2085.

Butel, J.S., and Lednicky, J.A. (1999). Cell and molecular biology of simian virus 40: implications for human infections and disease. Journal of the National Cancer Institute *91*, 119-134.

Carrillo Oesterreich, F., Bieberstein, N., and Neugebauer, K.M. (2011). Pause locally, splice globally. Trends in cell biology *21*, 328-335.

Carrillo Oesterreich, F., Preibisch, S., and Neugebauer, K.M. (2010). Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. Molecular cell *40*, 571-581.

Cavrois, M., De Noronha, C., and Greene, W.C. (2002). A sensitive and specific enzyme-based assay detecting HIV-1 virion fusion in primary T lymphocytes. Nature biotechnology *20*, 1151-1154.

Cook, K.S., Fisk, G.J., Hauber, J., Usman, N., Daly, T.J., and Rusche, J.R. (1991). Characterization of HIV-1 REV protein: binding stoichiometry and minimal RNA substrate. Nucleic acids research *19*, 1577-1583.

Cox, C.D., McCollum, J.M., Allen, M.S., Dar, R.D., and Simpson, M.L. (2008). Using noise to probe and characterize gene circuits. Proceedings of the National Academy of Sciences of the United States of America *105*, 10809-10814.

Daelemans, D., Costes, S.V., Cho, E.H., Erwin-Cohen, R.A., Lockett, S., and Pavlakis, G.N. (2004). In vivo HIV-1 Rev multimerization in the nucleolus and cytoplasm identified by fluorescence resonance energy transfer. The Journal of biological chemistry *279*, 50167-50175.

Daly, T.J., Doten, R.C., Rennert, P., Auer, M., Jaksche, H., Donner, A., Fisk, G., and Rusche, J.R. (1993). Biochemical characterization of binding of multiple HIV-1 Rev monomeric proteins to the Rev responsive element. Biochemistry *32*, 10497-10505.

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.

Daugherty, M.D., Booth, D.S., Jayaraman, B., Cheng, Y., and Frankel, A.D. (2010). HIV Rev response element (RRE) directs assembly of the Rev homooligomer into discrete asymmetric complexes. Proceedings of the National Academy of Sciences of the United States of America *107*, 12481-12486.

David, C.J., and Manley, J.L. (2011). The RNA polymerase C-terminal domain: a new role in spliceosome assembly. Transcription *2*, 221-225.

Davison, A.J., Benko, M., and Harrach, B. (2003). Genetic content and evolution of adenoviruses. The Journal of general virology *84*, 2895-2908.

de la Mata, M., Lafaille, C., and Kornblihtt, A.R. (2010). First come, first served revisited: factors affecting the same alternative splicing event have different effects on the relative rates of intron removal. RNA *16*, 904-912.

Felber, B.K., Drysdale, C.M., and Pavlakis, G.N. (1990). Feedback regulation of human immunodeficiency virus type 1 expression by the Rev protein. Journal of virology *64*, 3734-3741.

Felber, B.K., Hadzopoulou-Cladaras, M., Cladaras, C., Copeland, T., and Pavlakis, G.N. (1989). rev protein of human immunodeficiency virus type 1 affects the stability and transport of the viral mRNA. Proceedings of the National Academy of Sciences of the United States of America *86*, 1495-1499.

Fields, B.N., Knipe, D.M., and Howley, P.M. (2007). Fields virology, 5th edn (Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins).

Fong, Y.W., and Zhou, Q. (2001). Stimulatory effect of splicing factors on transcriptional elongation. Nature *414*, 929-933.

Fornerod, M., Ohno, M., Yoshida, M., and Mattaj, I.W. (1997). CRM1 is an export receptor for leucine-rich nuclear export signals. Cell *90*, 1051-1060.

Fukuda, M., Asano, S., Nakamura, T., Adachi, M., Yoshida, M., Yanagida, M., and Nishida, E. (1997). CRM1 is responsible for intracellular transport mediated by the nuclear export signal. Nature *390*, 308-311.

Gillespie, D.T. (2007). Stochastic simulation of chemical kinetics. Annual review of physical chemistry *58*, 35-55.

Gillespie, D.T., Hellander, A., and Petzold, L.R. (2013). Perspective: Stochastic algorithms for chemical kinetics. The Journal of chemical physics *138*, 170901.

Girard, C., Will, C.L., Peng, J., Makarov, E.M., Kastner, B., Lemm, I., Urlaub, H., Hartmuth, K., and Luhrmann, R. (2012). Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. Nature communications *3*, 994.

Graeble, M.A., Churcher, M.J., Lowe, A.D., Gait, M.J., and Karn, J. (1993). Human immunodeficiency virus type 1 transactivator protein, tat, stimulates transcriptional read-through of distal terminator sequences in vitro. Proceedings of the National Academy of Sciences of the United States of America *90*, 6184-6188.

Hsin, J.P., and Manley, J.L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. Genes & development *26*, 2119-2137.

Jablonski, J.A., Amelio, A.L., Giacca, M., and Caputi, M. (2010). The transcriptional transactivator Tat selectively regulates viral splicing. Nucleic acids research *38*, 1249-1260.

Jacquenet, S., Decimo, D., Muriaux, D., and Darlix, J.L. (2005). Dual effect of the SR proteins ASF/SF2, SC35 and 9G8 on HIV-1 RNA splicing and virion production. Retrovirology *2*, 33.

Jain, C., and Belasco, J.G. (2001). Structural model for the cooperative assembly of HIV-1 Rev multimers on the RRE as deduced from analysis of assembly-defective mutants. Molecular cell *7*, 603-614.

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.

Karn, J., and Stoltzfus, C.M. (2012). Transcriptional and posttranscriptional regulation of HIV-1 gene expression. Cold Spring Harbor perspectives in medicine *2*, a006916.

Khodor, Y.L., Menet, J.S., Tolan, M., and Rosbash, M. (2012). Cotranscriptional splicing efficiency differs dramatically between Drosophila and mouse. RNA *18*, 2174-2186.

Khodor, Y.L., Rodriguez, J., Abruzzi, K.C., Tang, C.H., Marr, M.T., 2nd, and Rosbash, M. (2011). Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. Genes & development *25*, 2502-2512.

Kim, H., and Yin, J. (2005). Robust growth of human immunodeficiency virus type 1 (HIV-1). Biophysical journal *89*, 2210-2221.

Kornblihtt, A.R., Schor, I.E., Allo, M., Dujardin, G., Petrillo, E., and Munoz, M.J. (2013). Alternative splicing: a pivotal step between eukaryotic transcription and translation. Nature reviews Molecular cell biology *14*, 153-165.

Kubota, S., Duan, L., Furuta, R.A., Hatanaka, M., and Pomerantz, R.J. (1996). Nuclear preservation and cytoplasmic degradation of human immunodeficiency virus type 1 Rev protein. Journal of virology *70*, 1282-1287.

Kudo, N., Wolff, B., Sekimoto, T., Schreiner, E.P., Yoneda, Y., Yanagida, M., Horinouchi, S., and Yoshida, M. (1998). Leptomycin B inhibition of signal-mediated nuclear export by direct binding to CRM1. Experimental cell research *242*, 540-547.

Lamb, R.A., and Horvath, C.M. (1991). Diversity of coding strategies in influenza viruses. Trends in genetics : TIG *7*, 261-266.

Laspia, M.F., Wendel, P., and Mathews, M.B. (1993). HIV-1 Tat overcomes inefficient transcriptional elongation in vitro. Journal of molecular biology *232*, 732-746.

Listerman, I., Sapra, A.K., and Neugebauer, K.M. (2006). Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. Nature structural & molecular biology *13*, 815-822.

Love, D.C., Sweitzer, T.D., and Hanover, J.A. (1998). Reconstitution of HIV-1 rev nuclear export: independent requirements for nuclear import and export. Proceedings

of the National Academy of Sciences of the United States of America *95*, 10608-10613.

Luo, M.J., and Reed, R. (1999). Splicing is required for rapid and efficient mRNA export in metazoans. Proceedings of the National Academy of Sciences of the United States of America *96*, 14937-14942.

Malim, M.H., and Cullen, B.R. (1991). HIV-1 structural gene expression requires the binding of multiple Rev monomers to the viral RRE: implications for HIV-1 latency. Cell *65*, 241-248.

Malim, M.H., and Cullen, B.R. (1993). Rev and the fate of pre-mRNA in the nucleus: implications for the regulation of RNA processing in eukaryotes. Molecular and cellular biology *13*, 6180-6189.

Malim, M.H., Hauber, J., Fenrick, R., and Cullen, B.R. (1988). Immunodeficiency virus rev trans-activator modulates the expression of the viral regulatory genes. Nature *335*, 181-183.

Manley, J.L., Fire, A., Cano, A., Sharp, P.A., and Gefter, M.L. (1980). DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract. Proceedings of the National Academy of Sciences of the United States of America *77*, 3855-3859.

Mann, D.A., Mikaelian, I., Zemmel, R.W., Green, S.M., Lowe, A.D., Kimura, T., Singh, M., Butler, P.J., Gait, M.J., and Karn, J. (1994). A molecular rheostat. Co-operative rev binding to stem I of the rev-response element modulates human immunodeficiency virus type-1 late gene expression. Journal of molecular biology *241*, 193-207.

Martins, S.B., Rino, J., Carvalho, T., Carvalho, C., Yoshida, M., Klose, J.M., de Almeida, S.F., and Carmo-Fonseca, M. (2011). Spliceosome assembly is coupled to RNA polymerase II dynamics at the 3' end of human genes. Nature structural & molecular biology *18*, 1115-1123.

Miller, H.B., Robinson, T.J., Gordan, R., Hartemink, A.J., and Garcia-Blanco, M.A. (2011). Identification of Tat-SF1 cellular targets by exon array analysis reveals dual roles in transcription and splicing. RNA *17*, 665-674.

Ossareh-Nazari, B., Bachelerie, F., and Dargemont, C. (1997). Evidence for a role of CRM1 in signal-mediated nuclear protein export. Science *278*, 141-144.

Otero, G.C., Harris, M.E., Donello, J.E., and Hope, T.J. (1998). Leptomycin B inhibits equine infectious anemia virus Rev and feline immunodeficiency virus rev function but not the function of the hepatitis B virus posttranscriptional regulatory element. Journal of virology *72*, 7593-7597.

Padgett, R.A., Hardy, S.F., and Sharp, P.A. (1983). Splicing of adenovirus RNA in a cell-free transcription system. Proceedings of the National Academy of Sciences of the United States of America *80*, 5230-5234.

Pandya-Jones, A., and Black, D.L. (2009). Co-transcriptional splicing of constitutive and alternative exons. RNA *15*, 1896-1908.

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. Journal of virology *82*, 12291-12303.

Pond, S.J., Ridgeway, W.K., Robertson, R., Wang, J., and Millar, D.P. (2009). HIV-1 Rev protein assembles on viral RNA one molecule at a time. Proceedings of the National Academy of Sciences of the United States of America *106*, 1404-1408.

Purcell, D.F., and Martin, M.A. (1993). Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. Journal of virology *67*, 6365-6378.

Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. Nature methods *5*, 877-879.

Ramos, I., Bernal-Rubio, D., Durham, N., Belicha-Villanueva, A., Lowen, A.C., Steel, J., and Fernandez-Sesma, A. (2011). Effects of receptor binding specificity of avian influenza virus on the human innate immune response. Journal of virology *85*, 4421-4431.

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. (2012a). Microwell Devices with Finger-like Channels for Long-Term Imaging of HIV-1 Expression Kinetics in Primary Human Lymphocytes. Under Revision.

Razooky, B.S., Gutierrez, E., Terry, V.H., Spina, C.A., Groisman, A., and Weinberger, L.S. (2012b). Microwell devices with finger-like channels for long-term imaging of HIV-1 expression kinetics in primary human lymphocytes. Lab Chip *12*, 4305-4312.

Reddy, B., and Yin, J. (1999). Quantitative intracellular kinetics of HIV type 1. AIDS research and human retroviruses *15*, 273-283.

Rempala, G.A., Ramos, K.S., and Kalbfleisch, T. (2006). A stochastic model of gene transcription: an application to L1 retrotransposition events. Journal of theoretical biology *242*, 101-116.

Rifkin, S.A. (2011). Identifying fluorescently labeled single molecules in image stacks using machine learning. Methods Mol Biol *772*, 329-348.

Roan, N.R., Munch, J., Arhel, N., Mothes, W., Neidleman, J., Kobayashi, A., Smith-McCune, K., Kirchhoff, F., and Greene, W.C. (2009). The cationic properties of SEVI underlie its ability to enhance human immunodeficiency virus infection. Journal of virology *83*, 73-80.

Robert-Guroff, M., Popovic, M., Gartner, S., Markham, P., Gallo, R.C., and Reitz, M.S. (1990). Structure and expression of tat-, rev-, and nef-specific transcripts of human immunodeficiency virus type 1 in infected lymphocytes and macrophages. Journal of virology *64*, 3391-3398.

Savageau, M.A. (2011). Biochemical systems analysis: a study of function and design in molecular biology (CreateSpace Independent Publishing Platform).

Schwartz, S., Felber, B.K., and Pavlakis, G.N. (1992a). Distinct RNA sequences in the gag region of human immunodeficiency virus type 1 decrease RNA stability and inhibit expression in the absence of Rev protein. Journal of virology *66*, 150-159.

Schwartz, S., Felber, B.K., and Pavlakis, G.N. (1992b). Mechanism of translation of monocistronic and multicistronic human immunodeficiency virus type 1 mRNAs. Molecular and cellular biology *12*, 207-219.

Singh, A. (2011). Negative feedback through mRNA provides the best control of gene-expression noise. IEEE transactions on nanobioscience *10*, 194-200.

Singh, J., and Padgett, R.A. (2009). Rates of in situ transcription and splicing in large human genes. Nature structural & molecular biology *16*, 1128-1133.

Spector, D.L., and Lamond, A.I. (2011). Nuclear speckles. Cold Spring Harbor perspectives in biology *3*.

Tardiff, D.F., Lacadie, S.A., and Rosbash, M. (2006). A genome-wide analysis indicates that yeast pre-mRNA splicing is predominantly posttranscriptional. Molecular cell *24*, 917-929.

Tazi, J., Bakkour, N., Marchand, V., Ayadi, L., Aboufirassi, A., and Branlant, C. (2010). Alternative splicing: regulation of HIV-1 multiplication as a target for therapeutic action. The FEBS journal *277*, 867-876.

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. Journal of virology *84*, 6425-6437.

Valencia, P., Dias, A.P., and Reed, R. (2008). Splicing promotes rapid and efficient mRNA export in mammalian cells. Proceedings of the National Academy of Sciences of the United States of America *105*, 3386-3391.

Vargas, D.Y., Shah, K., Batish, M., Levandoski, M., Sinha, S., Marras, S.A., Schedl, P., and Tyagi, S. (2011). Single-molecule imaging of transcriptionally coupled and uncoupled splicing. Cell *147*, 1054-1065.

Waks, Z., Klein, A.M., and Silver, P.A. (2011). Cell-to-cell variability of alternative RNA splicing. Molecular systems biology *7*, 506.

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. (2008). Transient-mediated fate determination in a transcriptional circuit of HIV. Nat Genet *40*, 466-470.

Weinberger, L.S., and Shenk, T. (2007). An HIV feedback resistor: auto-regulatory circuit deactivator and noise buffer. PLoS Biol *5*, e9.

Wu, Y., Beddall, M.H., and Marsh, J.W. (2007a). Rev-dependent indicator T cell line. Current HIV research *5*, 394-402.

Wu, Y., Beddall, M.H., and Marsh, J.W. (2007b). Rev-dependent lentiviral expression vector. Retrovirology *4*, 12.

Zhang, G., Zapp, M.L., Yan, G., and Green, M.R. (1996). Localization of HIV-1 RNA in mammalian nuclei. The Journal of cell biology *135*, 9-18.

Zlokarnik, G., Negulescu, P.A., Knapp, T.E., Mere, L., Burres, N., Feng, L., Whitney, M., Roemer, K., and Tsien, R.Y. (1998). Quantitation of transcription and clonal selection of single living cells with beta-lactamase as reporter. Science *279*, 84-88.