High Throughput Analysis of Integrant-specific HIV-1 Expression Patterns, Persistence, and Latency Reactivation

by

Edmond Akugbire Atindaana

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Microbiology and Immunology)
in the University of Michigan
2022

Doctoral Committee:

Professor Alice Telesnitsky, Chair
Professor Kathleen Collins
Associate Professor Jeffrey Kidd
Professor Akira Ono
Professor Mailini Raghavan

Edmond Akugbire Atindaana

atindaan@umich.edu

ORCID iD: 0000-0003-2355-5730

© Edmond Atindaana 2022

Dedication

In loving memory of Atindaana Akugre, Ayinpoka Akunumzuure, Joyce Asampana Akugre ("Mma Pee"), Tindaana Alanbanya Asaamaligo Asibi Tindaana, Vivian Nsomah Akugre, Elliot Kaplan, Deborah Bull, and Rodney Aloyine Anasigre.

Acknowledgements

I am thankful to my mentor Dr. Alice Telesnitsky for the excellent scientific training and guidance on a very exciting project during my time as a graduate student. I am indebted to all past and present members of the Telesnitsky lab especially, Dr. Feng Yang who contributed immensely to the foundation of this work, David Read for co-authoring one of the articles presented here, John Collins and Dr. Siarhei Kharytonchyk for the intellectual discussions and for serving as sounding boards for my ideas. I am thankful to our collaborators, Dr. Jeffrey M. Kidd, and Dr. Cheong-Hee Chang for the invaluable contributions to my thesis.

To the members of my cohort, Stephanie Theide, Austin Campbell, Anna-Lisa Lawrence, Yolanda Rivera-Cuevas, Matt Schnizlein, and adjunct cohort member, Zachary Mendel, I could not have asked for a better group to go through graduate school. Thank you for being such incredible friends and a wonderful support group.

I am grateful to the members of my thesis committee Dr. Akira Ono, Dr. Kathleen Collins, Dr. Malini Raghavan and Dr. Jeffrey M. Kidd for the scientific guidance on all the projects I worked on for this thesis.

This work will not be possible without the love from the Atindaana family and the support from my fathers; Pius Nsobila Atindaana, Robert Kaplan, and Gerald Andrews, Mothers; Veronica Nkebihi Atindaana, Veronica Akurugu Atindaana, Kim Kaplan, and Rose Andrews.

Finally, to my fiancée and best friend, Jasmine Smith, thank you for the unwavering love and enduring life's trials with me.

Table of Contents

List of Tables

List of Figures

Abstract

The residual HIV-1 reservoir is the major source of viral rebound in patients on antiretroviral therapy (ART). However, the dynamics of reservoir formation, its expression properties, and persistence are not completely characterized. *In vitro* models for studying the properties of HIV-1 persistence are necessary since direct characterization from virally suppressed patients is limited due to the rarity of infected cells. *In vitro* models utilize monoclonal or polyclonal proviral populations and results from such studies are usually integration site specific or report aggregate effects of many integration sites respectively. Moreover, some of these models do not include the viral protein R (Vpr) gene despite its conservation among lentiviruses. Therefore, the contribution of *vpr* to reservoir dynamics is less defined.

My work establishes a system that incorporates unique genetic barcodes into the U3 region of HIV-1 genomes that enable clonal tracking. This system allowed for high throughput analysis of the expression properties of hundreds to thousands of integrant clones, and the effect of Vpr on *in vitro* proviral landscapes to be determined. Using *vpr- env-* barcoded viruses, my work demonstrates that upon stable integration into the host genome and subsequent cell division, each provirus establishes its own pattern of gene expression consisting of waves of LTR-activity and inactivity that is stably transmitted to daughter cells. Integrant clones differed in the percentages of LTR-active daughter cells, with some clones containing few to no LTR-active cells while almost all daughter cells were LTR-active for others. Clones with high percentage LTR-active daughter cells were integrated closer to H3K9me3, H3K27ac, and H3K4me3 marks than those with low

LTR-active daughter cells. High LTR-active clones were rapidly lost in the presence of Vpr but there was no discrimination among clones based on the amount of virus released per LTR-active cell, despite burst size variation spanning four orders of magnitude among clones. Differences in clonal reactivation patterns were observed with or without Vpr. These studies provide insight into the expression properties of individual proviruses, the impact of Vpr on selective loss of infected subpopulations, and the spectra of clonal reactivation patterns.

Chapter 1 Introduction

Over thirty-seven million people are still afflicted by HIV/AIDS globally (WHO, 2020). Antiretroviral therapy (ART) suppresses viral levels below detectable limits, however, ARTs are not curative. Upon treatment cessation or interruption of antiretroviral therapy, virus rebounds and rekindles the spread of the virus (Sneller, Huiting et al. 2020). The source of viral rebound is thought to be from a small fraction of infected cells harboring transcriptionally silent replication-competent proviruses known as the latent reservoir, which persists for the entire lifetime of the patient (Bailey, Blankson et al. 2004). Despite the existence and use of antiretroviral drugs, the access, side effects, and adherence to these drugs in some parts of the world still pose major barriers to eradication and cure efforts. Therefore, understanding the mechanism of reservoir formation, proviral expression properties and persistence will help inform cure strategies. My dissertation sets up a high throughput approach to examine the expression properties of hundreds to thousands of individual proviruses. I apply this approach to address how the integration-site landscape changes in the presence of the viral protein R (Vpr) over time and make observations that highlight differences in the spectra of clonal reactivation between latent pools with or without Vpr. This chapter lays out the background and basis for the research presented in this thesis.

1.1 ***HIV-1 Replication cycle***

HIV-1 primarily infects CD4+ T lymphocytes and other CD4+ cell types such as macrophages and microglia (Epstein, Pantaleo et al. 1993, Burger and Poles 2003). Infection is initiated by the binding of the envelope glycoprotein gp120 to its CD4 receptor. This initial interaction induces conformational changes in gp120 and enables tight interaction through subsequent binding of the co-receptor CXCR4/CCR5. The tight interaction leads to the insertion of envelope's gp41 subunit into the target cell membrane, which causes the fusion of the viral envelope with the target cell membrane and release of the capsid core into the cytoplasm.

In the cytoplasm, the virus-encapsidated reverse transcriptase (RT) initiates reverse transcription of the viral genome in a process now believed to be completed in the nucleus, forming a double stranded DNA copy of the genome with duplicated long terminal repeat (LTR) regions on both the 5' and 3' ends (Hu and Hughes 2012). The viral DNA (vDNA) is then integrated into the host genome with a preference for active gene regions through the activity of the viral integrase (Craigie and Bushman 2012). The integrated HIV DNA is called the provirus, and is expressed similarly to host genes using its own promoter that is located within the 5' LTR (De Crignis and Mahmoudi 2017). Virus release, maturation, and spread is preceded by transcription of the provirus, splicing of some of the transcribed RNAs, translation of virion proteins, and packaging of the viral genomic RNAs into the assembling viral particle. A schematic representation of the replication cycle is shown in Figure 1.1.

1.2 ***Reverse transcription and integration***

The double-stranded DNA copy of the viral genome has duplicated LTRs at both the 5' and 3' ends that is generated by the multistep reverse transcription process and inserted into the host cell's DNA. Following the partial uncoating of the capsid core in the cytoplasm, reverse transcription is initiated using host tRNA-Lys3 that is bound to the primer binding site (PBS) as a primer. While single-stranded complementary DNA (cDNA) synthesis continues through U5-R, the RNA strand of the DNA-RNA heteroduplex is degraded by the DNA dependent ribonuclease activity of reverse transcriptase (Figure 1.2). The complementary DNA formed is called minus strand strong stop cDNA and is a chimera of the newly synthesized cDNA and tRNA-Lys3. The first strand transfer occurs when the strong stop cDNA formed at the RNA's 5' end binds to the 3' repeat region (R) of the viral genomic RNA. RT then synthesizes the remainder of the genome and the viral genomic RNA portion of the heteroduplex is degraded, except for the RNaseH resistant polypurine track (PPT). RT, which also has DNA-dependent DNA polymerase activity, initiates the second strand cDNA synthesis. The PPT is used to prime second strand synthesis by RT, which continues through to the 3' of its template, including the tRNA region that was originally bound to PBS of the viral genomic RNA (gRNA). Following the displacement of the original tRNA region of the heteroduplex, a second strand transfer occurs. The resulting single-stranded positive strand is complimentary to the PBS of the first strand cDNA, and thus this segment of nascent DNA can bind to the PBS. RT then synthesizes the remainder of both strands resulting in generation of the U3-R-U5 (long terminal repeat; LTR) that is duplicated at both the 5' and 3' ends of the double stranded DNA intermediate.

The process of reverse transcription starts in the cytoplasm but is believed to complete in the nucleus of the host cell (Arhel 2020, Dharan, Bachmann et al. 2020). During this process, some capsid (CA) protein sheds off, while viral nucleocapsid (NC), matrix (MA), Vpr, and integrase (IN) proteins remain associated with the newly synthesized viral DNA. Host factors such as barrier to autointegration factor 1, BANF1, associate with the vDNA protein complex before host chromosome association to form the pre-integration complex (PIC) (Raghavendra, Shkriabai et al. 2010). The viral DNA is then inserted into the host cell's DNA (Arhel 2010, Craigie and Bushman 2012). The DNA cutting and the joining of the viral DNA from the PIC to the host cell's DNA is catalyzed by IN. The 3' end processing first occurs, involving the removal of two bases from each blunt end of the linear viral DNA to form a two-nucleotide overhang (usually CA-3'). Next, the vDNA transfer occurs, where the 3' ends attack a pair of phosphodiester bonds on opposite strands of the target DNA and becomes covalently linked. The single stranded two nucleotide gaps are then filled, and the nicks are repaired by cellular DNA repair machinery.

HIV-1 DNA integrates into the host DNA with preference to transcriptionally active regions. Early studies with gammaretroviruses suggested that viral DNA integration was associated with open chromatin (Bowerman, Brown et al. 1989, Coffin, Hughes et al. 1997). However, recent studies show that host factors such as the lens epithelium-derived growth factor F variant 75 (LEDGF/p75) and the cleavage and polyadenylation specificity factor 6 (CPSF6) mediate integration site target selection for HIV-1. LEDGF/p75 is a ubiquitously expressed 530 amino acid spliced variant of the *PSIP1* gene that has been shown to be a co-activator of transcription with restricted nuclear localization (Nishizawa, Usukura et al. 2001). HIV-1 IN associates with LEDGF/p75. This IN-LEDGF/p75 interaction tethers and targets PICs to active gene units (De Rijck, Bartholomeeusen

et al. 2010). HIV-1 based lentiviral vectors were integrated into heterochromatin regions in experiments using hybrid LEDGF/p75 where the DNA binding domain was replaced with a heterochromatin-binding domain. CPSF6 targets PICs to transcriptionally active speckle-associated domains (SPAD) (Li, Singh et al. 2020). In contrast to LEDGF/p75, CPSF6 associates with PICs through CA. Without this interaction, PICs generally accumulate at the nuclear periphery and are destined to heterochromatic lamina-associated domains (LADs) while integrations into SPADs are disfavored. Therefore, CPSF6 contributes to integration-site selection.

1.3 ***Bimodal expression outcome of HIV-1 DNA integration***

Cells containing a stably integrated intact proviral DNA can either express the provirus in an active infection or remain transcriptionally silent. To complete the replication cycle, an active provirus is expressed to produce progeny virions that bud off the surface of the host cell. These newly formed viral particles then mature and are capable of infecting new cells in the absence of antiretroviral therapy. Shortly after infection, the majority of active cells have an estimated half-life of about two days due to viral cytotoxic effects and immune detection (Perelson, Neumann et al. 1996, Stewart, Poon et al. 1997, McMichael and Rowland-Jones 2001, Andersen, Le Rouzic et al. 2008). Despite the relatively short half-life of most infected cells, some infected cells can persist for the entire lifetime of the patient even in the presence of ART. Part of the persistent reservoir is generally referred to as latent and is comprised of a small fraction of infected cells that harbor transcriptionally silent replication-competent proviruses (Finzi, Hermankova et al. 1997). The establishment and maintenance of the HIV-1 latent reservoir is multifaceted and may involve intracellular depletion of transcription factors, chromatin distribution, integration position effects, epigenetic variation, and viral regulatory protein deficiencies, among other factors (Emiliani, Van

Lint et al. 1996, Williams and Greene 2007, Lenasi, Contreras et al. 2008, Mbonye, Wang et al. 2018). Because the current drugs block viral replication by inhibiting key steps in the replication cycle, ART is potent only when the virus is actively replicating. For this reason, HIV-1 latency remains the main hurdle to eradication efforts. Therefore, a deep understanding of establishment and maintenance of HIV-1 latency will help in developing strategies for a cure.

1.4 ***Regulation of HIV-1 gene expression by host and viral factors***

HIV-1 proviral expression is regulated by both host and viral factors. The promoter of HIV-1 is located in the 5' LTR and requires the interaction of constitutive and inducible host factors to form stable transcription complexes for expression (Figure 1.3). The binding of RNA polymerase II, and transcription factors such as nuclear factor kappa B(NFkB), SP1, AP-1, and NFAT to the promoter and regulatory elements within the 5' LTR, are required for transcriptional activity. Transcription is initiated by RNA polymerase II but transcriptional elongation depends on the phosphorylation of the serine 2 residue in the heptapeptide repeat of Pol II's carboxy-terminal domain (CTD) by the positive transcription factor B (P-TEFb) (Price 2000, Garriga and Graña 2004, Peterlin and Price 2006). Ordinarily, the majority of P-TEFb is sequestered in an inactive form either tethered to acetylated histone motifs or in a HEXIM-7SK snRNP complex. The HIV-1 transactivator of transcription (Tat) binds P-TEFb and the *trans*-activation response (TAR) element of the growing RNA transcript, thus recruiting P-TEFb to stalled Pol II and releasing it from the promoter region. Without the viral protein Tat, proviral expression is repressed due to abortive transcription (Sedore, Byers et al. 2007).

Chromatin modulation adds an extra layer of regulation of HIV-1 gene expression. The promoter of HIV-1 is wrapped around three distinct nucleosomes, Nuc0, Nuc1 and Nuc2, that are positioned at -415/-255, +10/+155, and +256/+412 respectively. These nucleosomes are separated by two regions containing consensus transcription factor binding sites, DHS1 and DHS2, and are repositioned upon activation stimuli (Verdin 1991, Verdin, Paras et al. 1993, Steger and Workman 1997). Additionally, chromatin and nucleosomes serve as marks for selectively recruiting histone modifying enzymes that can read, write or erase epigenetic marks. The action of such enzymes modulates chromatin structure and can act to either activate or repress HIV-1 gene expression. For example (Figure 1.4), reversible acetylation of histones relies on histone acetyl transferases (HATs) and histone deacetyl transferases (HDACs) which can have distinct transcriptional consequences depending on which H3K (histone-3 Lysine) is modified (Tripathy, Abbas et al. 2011, Van Lint, Bouchat et al. 2013, Kumar, Darcis et al. 2015).

1.5 ***Role of transcription start site heterogeneity and splicing in HIV-1 gene expression***

Transcription start site heterogeneity contributes to the regulation of HIV-1 gene expression. Transcription can be initiated at any one of three contiguous guanosines at the junction of U3 and R in the 5' LTR. Depending on which guanosine is capped, the transcript can either be destined for translation or packaged in particles as viral gRNA (Kharytonchyk, Monti et al. 2016). Further studies provide evidence for the structural basis of HIV-1 RNA fate determination (Kharytonchyk, Monti et al. 2016, Brown, Kharytonchyk et al. 2020). Evidence provided by Esquiaqui et. al., shows that 3G HIV-1 RNAs are enriched among spliced RNAs and suggests that the choice of transcription start site helps define which transcripts are selected to be spliced (Esquiaqui, Kharytonchyk et al. 2020).

HIV-1 relies on a fine balance of spliced and unspliced RNA for expression. Most transcripts result in ~9.2kb full-length genomic RNA and either serve as the genome to be packaged into virions or serve as mRNA for translation of gag/pro/pol (Ocwieja, Sherrill-Mix et al. 2012). Some HIV-1 RNA can either be partially or completely spliced after the recognition of paired splice donor (D1-D5) and acceptor (A1-A7) sites by cellular small ribonuclear proteins snRNP U1 and snRNP U2 (Stoltzfus 2009). All spliced RNA uses the splice donor site D1 and splice to downstream acceptor A1 to A5 to form partially or incompletely spliced ~4kb class of mRNAs for *vpr, vpu, env* and *vif*. Once splicing at D1 has occurred, partially spliced RNA may be further spliced, for example, by removing the D4 to A7 fragment, to form ~1.8kb transcripts required to encode accessory proteins including Tat and Rev. These shorter transcripts are called completely spliced. The viral accessory proteins Tat and Rev are essential for transcriptional elongation and nuclear export of unspliced viral RNA respectively and therefore add a layer of regulation to HIV-1 gene expression. HIV-1 spliced RNA variants are known to associate with different cellular proteins. Limiting the cellular levels of some members of the interactome by knockdown experiments greatly reduces HIV-1 gene expression (Knoener, Evans III et al. 2021). Thus, unspliced, partially spliced, and completely spliced HIV-1 RNA variants are the three distinct classes of RNA required for productive viral replication and infectivity (Stoltzfus 2009, LeBlanc, Weil et al. 2013). Despite the importance of splicing in HIV-1 gene expression, and that HIV-1 integrates into different gene rich regions, it is yet to be determined whether the balance of splicing differs among integrants.

1.6 ***Stochasticity of HIV-1 gene expression***

The interplay of several mechanisms determines how, when, and to what extent the virus is expressed. Although the outcome of an infection is simply described as active or latent, many layers of regulation cooperate to determine this fate. Even among a monoclonal population of cells harboring a latent provirus at the same integration site, daughter cells exhibit nonuniform basal and stimuli-induced expression using a GFP reporter system (Weinberger, Burnett et al. 2005). This non-uniform basal expression results from stochastic noise in switching between off and on transcription states (Singh, Razooky et al. 2010, Dar, Razooky et al. 2012, Knoener, Evans III et al. 2021). The contribution of noise to the stochasticity of gene expression may be due to fluctuations in transcriptional bursting caused by the asymmetric distribution of transcription factors among daughter cells during cell division and differences in the availability and levels of transcription factors (Swain, Elowitz et al. 2002). Stochastic mechanisms can yield either random or robust outcomes. For example, eukaryotic developmental programs incorporate stochasticity to produce a wide variety of cell types, and stochastic decisions must be maintained to yield reproducible outcomes (Johnston Jr and Desplan 2010). Tight regulation may be required to accomplish this. Previous studies of a clonal HIV-1 population revealed inherent randomness in on or off expression fate decisions that were stabilized by a positive feedback loop (Weinberger, Burnett et al. 2005, Weinberger and Weinberger 2013). Whether such fate decisions can result in stable reproducible outcomes is not known. This would require studying thousands of clones, with HIV-1 stably integrated into different loci and comparing gene expression across clonal populations over time. Such a study would require a high throughput approach to analyze clone-specific HIV-1 gene expression.

1.7 ***The role of Vpr in HIV-1 gene expression***

The enigmatic accessory protein Vpr is 14 KDa with 98 amino acids and is highly conserved among primate lentiviruses (Wong-Staal, Chanda et al. 1987, Beaumont, van Nuenen et al. 2001). Vpr causes G2 to M phase cell-cycle arrest in dividing cells. During the G2 phase of the cell cycle, cellular proteins and factors are synthesized to about twice the level usually contained in a cell in preparation for mitotic cell division. This makes the G2 phase an essential stage of the cell cycle where gene expression is most efficient. In CD4+ T cells, Vpr may enhance viral replication by causing G2/M cell cycle arrest and exploiting the availability of cellular factors for HIV-1 gene expression (Goh, Rogel et al. 1998). Cell cycle arrest by Vpr is achieved by interaction with DDB1 and Cul4 Associated Factor 1 (DCAF1), a conserved substrate-binding subunit of the CLT4 (Cul4a-Ddb1-Roc1) ubiquitin ligase complex. In cells, DCAF1 regulates the G2/M transition by regulating protein poly-ubiquitination and mono-ubiquitination (Hrecka, Gierszewska et al. 2007, Wen, Duus et al. 2007). With DCAF1 as a cofactor, Vpr interacts with and targets CLT4 to Vpr-interacting partners such as PLK1, Srk1, and CCDC137 for ubiquitination and degradation. Degradation of these cellular proteins or the direct association of Vpr with the nuclear serine/threonine kinase, Wee1 prevents them from either directly or indirectly activating cdc2, a regulator of DNA damage checkpoint (Zimmerman, Sherman et al. 2006, Hrecka, Gierszewska et al. 2007, Le Rouzic, Belaïdouni et al. 2007, Gérard, Yang et al. 2014, Zhang and Bieniasz 2020). Despite extensive evidence showing beneficial effects of G2/M cell cycle arrest to the virus, the cellular response to this arrest is apoptosis (Stewart, Poon et al. 1997). How can Vpr enhance viral expression and at the same time kill cells when it is expressed?

Our understanding of the function of Vpr is still incomplete, but some studies show that its function can be cell type-specific or dependent on the expressed isoform of its interacting partner. Vpr can interact with the mitochondrial adenine nucleotide translocator (*ANT*) and can induce caspase-dependent apoptosis (Sabbah, Druillennec et al. 2006). Vpr induces apoptosis in TZM-bl cells but fails to do so in 293T cells. This is because *ANT-3* is predominantly expressed in TZM-bl cells while *ANT-1* and *ANT-2* are the predominant isoforms expressed in 293T cells (Du, Wu et al. 2021). Therefore, at least in cell lines, the interaction of Vpr with different *ANT* isoforms can result in pro or anti-apoptotic effects.

1.8 The role of Vpr in macrophage infection

Vpr enhances the infection of macrophages. Although the virus replicates and establishes latency in CD4+ T cells without Vpr, the highly conserved nature of the *vpr* gene suggests that it may be critical *in vivo*. HIV-1 infects tissue-resident macrophages, but unlike in T cells, Vpr is required for infection and replication (Eckstein, Sherman et al. 2001). To infect and replicate in macrophages, several host restriction factors must be surpassed. One such factor is the lysosomal-associated transmembrane protein 5 (LAPTM5), which transports HIV-1 Env to the lysosomes for degradation and therefore inhibiting the infectivity of virions produced. Vpr counteracts this restriction by degrading LAPTM5 via DCAF1. In addition, the macrophage-specific mannose receptor (MR) is one of the pattern recognition receptors that binds HIV-1 Env at its basic mannose patch. Lubow et al, recently showed that Vpr reduces the expression of the MR and enhances HIV-1 Env expression and virus release (Lubow, Virgilio et al. 2020). These findings demonstrate that Vpr plays a critical role in the infection of macrophages *in vivo* despite the fact that it is removed in most in vitro experiments.

1.9 HIV-1 Persistence

HIV-1 persists for decades following initial infection. Within the first nine weeks after viral infection, HIV-1 spreads and infiltrates lymphoid organs, the central nervous system (CNS) and other areas of the body. This stage is clinically known as the acute phase and is characterized by normal CD4+ T cell count and high viremia. The immune system kicks in and brings viremia under control by killing infected CD4+ T cells. Over the next decade (variable among patients), which is referred to as the clinical latent phase, virus spread continues and CD4+ T cell counts gradually decline due to CD8+ T cell-mediated killing and death by pyroptosis until CD4+ T cell counts fall below 200 cells/ul (Doitsh, Galloway et al. 2014). Once CD4+ T cell-counts diminish, the individual starts to develop symptoms such as prolonged diarrhea, rapid weight loss, recurring fever and night sweats, prolonged swelling of lymph nodes among others, and is prone to opportunistic infections. Viral spread rapidly increases and the individual eventually dies from AIDS. (Figure 1.5). HIV-1 can continue to replicate for as long as the untreated infected individual lives.

Decades of research have led to the discovery and development of antiretroviral therapy (ART) that effectively blocks viral replication, spread, and progression to AIDS. Therefore, infected individuals can live a near normal life but must remain on ART for the remainder of their lifespan because of the latent reservoir. The latent reservoir remains poorly defined. HIV-1 establishes stable reservoirs in patients treated with ART, which consist of cells containing replication competent proviruses that are not cleared by the immune system and that have the potential to rekindle spreading infection (Bailey, Blankson et al. 2004). It is generally assumed that this latent

reservoir consists of persistent proviruses that are transcriptionally silent. The phenomenon is called HIV-1 viral latency and differs from clinical latency already described.

The best described mechanism of latency is that which relies on the biology of CD4+ T cells. Once activated by receptor engagement, naïve CD4+ T cells proliferate and differentiate into effector T cells in response to invasion by tissue damage or pathogen. Once the invader is cleared, most effector cells die but a few of them can further differentiate into resting memory T cells, which are not permissive to viral replication (T cell biology is reviewed here (Kaech, Wherry et al. 2002). *In vitro,* the direct infection of memory CD4+ T cells leads to a labile state known as pre-integration latency where viral DNA is made by reverse transcription but subsequent replication steps fail (Kaech, Wherry et al. 2002).

The lack of productive infection of resting memory CD4+ T cells led to questions about how the latent reservoir is established. A generally accepted hypothesis is that initial infection occurs when the T cell is activated and that subsequent differentiation to resting memory causes silencing of the provirus. Reservoir establishment and maintenance is multifaceted and may involve intracellular depletion of transcription factors, integration position effects, epigenetic variation, and viral regulatory protein deficiencies, among other factors (Verdin, Paras et al. 1993, Emiliani, Van Lint et al. 1996, Emiliani, Fischle et al. 1998, Kulkosky, Sullivan et al. 2004, Lassen, Han et al. 2004, Archin and Margolis 2006, Williams and Greene 2007, Han, Lin et al. 2008, Lenasi, Contreras et al. 2008, Blazkova, Trejbalova et al. 2009, Mbonye, Wang et al. 2018). The oligoclonal nature of ART-suppressed patients' integrant populations suggests the reservoir's long-lived nature is at least partially due to infected cells' proliferation, which may be either

homeostatic or driven by T cell receptor engagement (Wang, Gurule et al. 2018, Mendoza, Jackson et al. 2020, Simonetti, Zhang et al. 2021).

HIV-1 latency may not be the only contributor to persistence. In immune privileged sites such as lymphoid organs or the CNS, drug concentrations are generally sub-optimal and lower than in the blood. Hence low level viral replication and spread may occur at these sites (Zhang, Ramratnam et al. 1999). Similarly, macrophages require higher concentrations of antiretroviral drugs to block viral replication. Macrophages can cross the blood brain barrier and enable macrophage-T cell spread (Burdo, Lackner et al. 2013, Duncan, Russell et al. 2013). Hematopoietic stem cells in the bone marrow are long-lived and can contribute to viral persistence by harboring proviruses (McNamara and Collins 2011). Since the reservoir is the main barrier to a cure for HIV-1, a complete characterization and understanding may be required in order to develop strategies to eliminate it.

1.10 ***Viable strategies for reservoir elimination***

Much of the current studies have focused on ways to rid individuals of the residual reservoir. Some of the proposed strategies include the shock and kill, block and lock, CAR T cell approaches, immunomodulation, gene editing, and bone marrow transplantation.

Shock and Kill

The shock and kill method involves inducing virus expression with the intention that it will lead either to cytopathic death of reactivated cells or to immune-mediated clearance while ART blocks new infections (Deeks 2012). Because HIV-1 latency is established and maintained by multiple

mechanisms (Lassen, Han et al. 2004), the candidate approaches that perturb cellular pathway or complement intracellular deficiencies in experimental models are also diverse. These include the use of prostratin, which stimulates T cells without inducing cellular proliferation and increases the level of NFkB (Korin, Brooks et al. 2002, Williams, Chen et al. 2004). Other latency-reversing agents (LRAs) include those that act to increase the level of P-TEFb, including the BET bromodomain inhibitor JQ1, (Boehm, Calvanese et al. 2013) as well as treatments that modify the chromatin environment, such as histone deacetylase inhibitors like suberoyl anilide hydroxamic acid (SAHA) and entinostat (Archin, Espeseth et al. 2009, Contreras, Schweneker et al. 2009, Edelstein, Micheva-Viteva et al. 2009, Zhu, Gaiha et al. 2012, Jiang, Mendes et al. 2014).

Block and lock

In contrast to latency reversal, the block and lock strategy is a functional cure that seeks to permanently silence residual proviruses instead of eliminating them. Several groups have demonstrated the use of inhibitors to block HIV-1 transcription. These include Tat-inhibitor *Didehydro cortistatin A* which potently suppresses HIV-1 transcription and reactivation by LRA in primary CD4+ T cells and cell lines, and reduces viral rebound in BLT (Bone Marrow-Liver-Thymus) humanized mouse models of latency (Mousseau, Clementz et al. 2012, Mousseau, Kessing et al. 2015, Kessing, Nixon et al. 2017). The use of the histone chaperone, FACT (facilitates chromatin transcription) inhibitor, curaxin blocks NFkB-mediated transcription or the heat shock protein 90 (HSP90) that is required for viral protein expression (Vozzolo, Loh et al. 2010, Gasparian, Burkhart et al. 2011, Joshi and Stoddart 2011). Others have explored the use of RNA-induced epigenetic silencing to suppress HIV-1 reactivation. For example, by designing a short hairpin (sh) RNA that targets transcription factor binding sites in the 5'LTR, Mendez et. al.,

showed that it can recruit Argonaut 1, histone deacetylase 1, and histone methyl transferases to the HIV-1 promoter to maintain a heterochromatin landscape, and prevented reactivation (Méndez, Ledger et al. 2018). This has led to the proposition that RNA interference could be used as a block and lock strategy.

CAR-T cell therapy

Chimeric antigen receptor (CAR) engineering of T cells to target HIV-1 specific proteins is one of the strategies that is being explored for a cure. In this approach, cytolytic CD8+ T cells are engineered to express chimeric proteins composed of an extracellular domain that recognizes HIV antigens such as Env with an intracellular signaling domain. These HIV-1 CAR T cells lysed envelope-expressing cells *in vitro (Masiero, Del Vecchio et al. 2005, Liu, Patel et al. 2015). Recent advances in CAR T cell technology have led to the development of a CD4+ HIV-specific CAR that not only reduces rebound viremia, but also helps mitigate disease progression in vivo by conferring CD4+ T cell help (Maldini, Gayout et al. 2020). Improvements to HIV-1 specific CAR T cell therapy are still ongoing.*

Bone Marrow Transplant

The only approach that has been reported to have successfully cured HIV patients is allogenic bone marrow transplant. An HIV-1 patient who was on suppressive ART had developed acute leukemia unrelated to HIV and required myeloablative allogenic hematopoietic stem cell transfer (HSCT). After stem cell transplantation from a human leukocyte antigen (HLA)-matching donor who was also homozygous for CCR5 delta 32, a 32bp deletion in the *CCR5* gene that renders cells resistant to HIV infection, no evidence of viral rebound was found even though ART was discontinued

(Hütter, Nowak et al. 2009). This first case of HIV remission came almost a decade after Lui and colleagues reported that the genetic basis for rare individuals who remained uninfected despite multiple exposures to HIV-1 was a deletion in the *CKR-5 gene* (now known as CCR5) (Liu, Paxton et al. 1996). Other HSCTs involving HLA-matched donors without the CCR5 delta 32 deletion only led to a delay in rebound of the virus and reduction of the viral reservoir (Henrich, Hu et al. 2013, Henrich, Hanhauser et al. 2014), which highlights the importance of the CCR5 mutation in remission from HIV-1 infections. Unfortunately, HLA-matched donors who have the CCR5 delta 32 homozygous mutation are extremely rare. The second case of remission from HIV-1 by HSCT was reported in 2019 (Gupta, Abdul-Jawad et al. 2019) and another report of a possible cure was announced at the conference on retroviruses and opportunistic infections (CROI) 2022 in Denver by Bryson and colleagues (JingMei Hsu 2022). Altogether, these provide evidence that remission by HSCT is possible, but pursuing it on a large scale may not be feasible due to difficulties in finding donors who have matching HLA and the corresponding CCR5 mutation.

1.11 ***Models for HIV-1 latency studies***

In vitro models are required due to the rarity of latent cells in patients on ART. Evidence for the existence of a latent reservoir came from the interrupted treatment studies in patients who had no detectable viremia. Cessation of antiretroviral treatment resulted in viral rebound (Davey, Bhat et al. 1999). Estimates suggest that the size of a patient's latent reservoir is between one in a million and three in ten thousand resting memory CD4+ T cells (Eriksson, Graf et al. 2013). The rarity of latently infected cells in virally suppressed individuals makes them difficult to identify and isolate for further studies. Therefore, *in vitro* models that provide a better understanding of molecular mechanisms regulating viral latency are necessary.

Several groups have developed different cell culture models for studying different aspects of the replication cycle of HIV-1. These culture systems differ in cell types, the source of the cells, the cell cycle status during infection, vector used, and the readout of virus expression (Spina, Anderson et al. 2013). The closest culture models to natural infections are those based on primary CD4+ T cells. However, even among these models, the subset of CD4+ T cells and vector used differ. For example, the Greene model infects resting CD4+ T cells with a near wild-type (WT) vector based on the NL4-3 strain of HIV-1 that is modified to contain a luciferase reporter and virus expression is measured by relative luciferase units (RLU) (Lassen, Hebbeler et al. 2012). Similarly, the Lewin and Spina models infect resting CD4+ T cells with WT NL4-3 but measure virus expression by quantifying soluble RT and Tat mRNA copies respectively (Spina, Guatelli et al. 1995, Saleh, Solomon et al. 2007, Mousseau, Clementz et al. 2012). The Planelles and Siliciano models utilize primary cells that are initially differentiated to form central memory T cells (T_{CM}) and effector memory T cells respectively. Primary CD4+ T cells are first activated to divide into T_{CM} by stimulating the T cell receptor (TCR) in the presence of TGF-β, and α-IL-4 and α-IL-12 monoclonal antibodies. These T_{CM} cells are infected with an *env- nef-GFP* NL4-3- based vector to establish a single round of infection in the Planelles model, where virus expression is measured by either percent GFP or intracellular Gag staining. In the Siliciano model, primary CD4+ T cells are first activated by TCR stimulation and transduced with a lentiviral vector carrying the human bcl-2 gene (EB-FLV), then are cultured in IL-2 and allowed to return to a resting state. The resting cells are infected with NL4-3-Δ6-drEGFP reporter virus following reactivation, and are cultured for 3-4 weeks. The effector memory cells are then isolated by flow cytometry for further studies. Virus expression in the Siliciano model is measured by percent GFP (Yang, Xing et al. 2009).

A limitation of CD4+ T cell models in culture is that they need stimulation before or after infection or both. Since T cell stimulation reactivates the virus and perturbs the basal expression state of infected cells, the requirement for stimulation makes primary CD4+ T cell models not ideal for studying persistence *in vitro*.

The widely used J-Lat clones assess LTR promoter activity by GFP reporter gene expression and are established in the Jurkat cell line. However, the vector used to establish these clones lacks HIV-1 genetic elements believed to be dispensable in HIV-1 latency (Jordan, Bisgrove et al. 2003, Pace, Agosto et al. 2011, Darcis, Kula et al. 2015). Many latency models use *vpr*-defective proviruses which, when cultured over time to allow proviral silencing, can be used for reactivation studies (Jordan, Defechereux et al. 2001, Chen, Martinez et al. 2017, Fennessey, Pinkevych et al. 2017, Jefferys, Burgos et al. 2021). In such systems, the extent to which proviral quiescence represents the silencing and outgrowth of a subset of integrant clones versus global proviral silencing is unknown. Moreover, the lack of viral cytopathic proteins prevents the selective pressure these proteins exert on proviral populations *in vitro* and enables clones that would otherwise not survive to be chosen for mechanistic studies.

Animal models of HIV-1 infection exist (Hatziioannou and Evans 2012). However, they do not recapitulate infection in humans since HIV only infects and causes AIDS in humans. The closest animal model to HIV-1 in a human is a hybrid of HIV-1 and the simian immunodeficiency virus (SIV). SIV causes a disease that is similar to AIDS in non-human primates. Other animal models for HIV-1 research include severe combined immunodeficient (SCID) humanized mice, *Rag2*$^{-/-}$*Il2rg*$^{-/-}$ mice engrafted with human CD34^{+} stem cells (Gao, Bailes et al. 1999) and others

(Hatziioannou and Evans 2012). The choice of animal model used ultimately depends on the goal of the research. A review of the animal models and the purposes for which they are used is reviewed by Fujinaga et. al. (Fujinaga and Cary 2020).

1.12 ***Reservoir Quantification***

One of the hurdles to an HIV cure is the lack of an assay to accurately quantify the latent reservoir. The quantitative viral outgrowth assay (QVOA) has for years been regarded as the gold standard for quantification of the replication-competent reservoir (Laird, Bullen et al. 2015). However, recent studies suggest that QVOAs underestimate the latent reservoir because not all replication-competent proviruses can be induced by a single round of T cell activation. In contrast, PCR-based assays that quantify proviral DNA overestimate reservoir sizes (Ho, Shan et al. 2013) because these assays do not typically discriminate between replication-competent and defective proviruses, which tend to predominate patient viruses. Assays that characterize proviruses as defective or intact include full-length sequencing assays or intact proviral DNA assays (IPDA), which are high-throughput assays in which multiple segments of the HIV-1 genome are multiplexed in a digital droplet PCR (ddPCR) assay (Hiener, Horsburgh et al. 2017, Bruner, Wang et al. 2019, Levy, Hughes et al. 2021, Cassidy, Fish et al. 2022). The effectiveness of cure strategies depends on the ability to reliably measure the reservoir and any changes that may occur after treatment has been administered. Therefore, method improvement for reservoir quantification is ongoing.

1.13 ***Summary of dissertation***

The work presented here addresses some of the gaps in our understanding of HIV-1 replication outcomes of individual integrants, their expression properties, and the proviral population structure of persistent viruses. In Chapter 2, the expression properties of hundreds of individual proviral clones were investigated. Using a zip coded proviruses approach that enabled tracking of individual integrants over time, the experiments demonstrated that each integrant clone exhibited a unique pattern of bimodal expression that was stably transmitted through cell generations. Chapter 3 extends this work by investigating how the bimodal expression patterns, proviral integration sites, population structure, and latency reversal potential is influenced by Vpr. The work here demonstrates that the cytotoxic effect of Vpr selects against integrant clones in which the majority of daughter cells' proviruses are LTR-active (actively expressing HIV genes). As a consequence, persisting viruses are integrated farther away from chromatin marks associated with active gene expression. For integrant clones in which only a minority of daughter cells' proviruses were LTR-active, clones' burst sizes were higher by about three orders of magnitude than high LTR-active clones'. Drug-induced reactivation of *vpr*+ and *vpr*- proviral pools revealed that despite similar population-wide trends, clonal responses differed between these pools.

Chapter 4 provides a discussion of these results, their relevance, and future experiments that will further elucidate the expression properties of the persistent HIV-1 reservoir *in vitro*.

Chapter 4 also presents and discusses preliminary evidence that suggests that that burse size variation among clones may be due to differences in the unspliced to spliced HIV-1 RNA ratios.

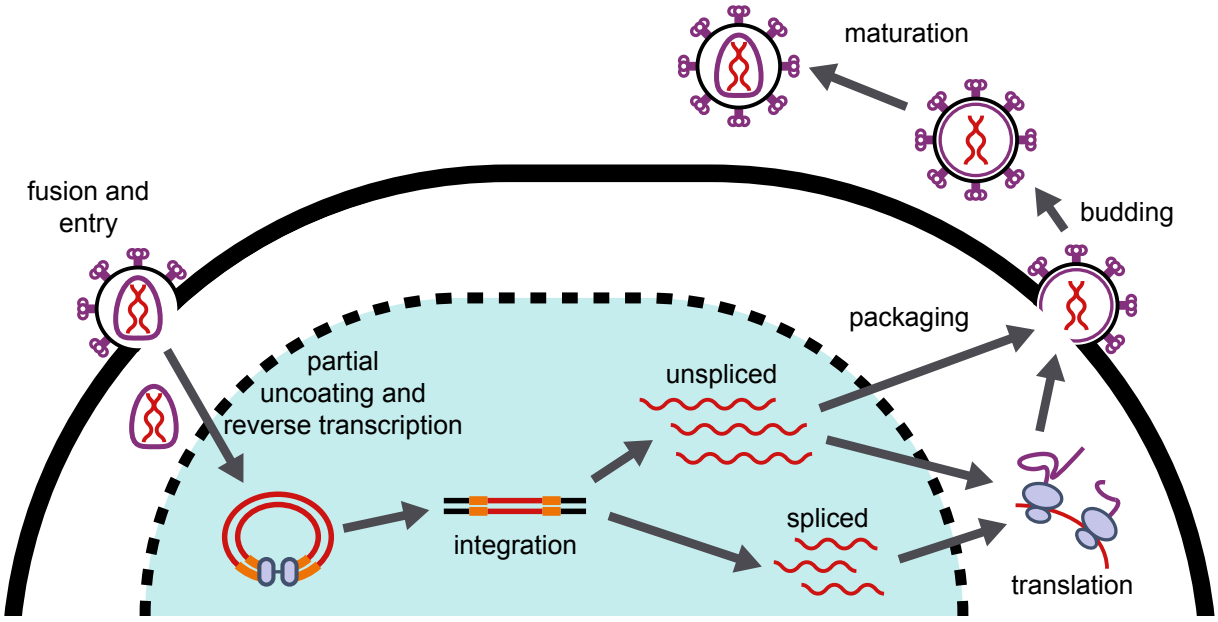


Figure 1. 1 A schematic diagram of the HIV-1 replication cycle. From left to right and the bottom to the top. The viral envelope fuses with the cellular membrane and releases the capsid core into the cytoplasm through viral envelope protein-CD4 receptor interaction. In the cytoplasm, partial uncoating of the capsid occurs and reverse transcription is initiated. The process of reverse transcription completes in the nucleus to generate a double-stranded viral DNA. The viral integrase (blue ovals) then catalyzes the cutting and joining of the linear viral DNA to the chromosome of the host cell (the circular viral DNA depicts how integrates brings the ends of the linear viral DNA together to mediate integration). The transcription machinery of the host cell transcribes and splices some of the transcripts for translation. Two copies of some of the unspliced viral RNAs are packaged into the assembling viral particle that buds off the surface of the cell. Maturation of the viral particle occurs to form an infectious virion

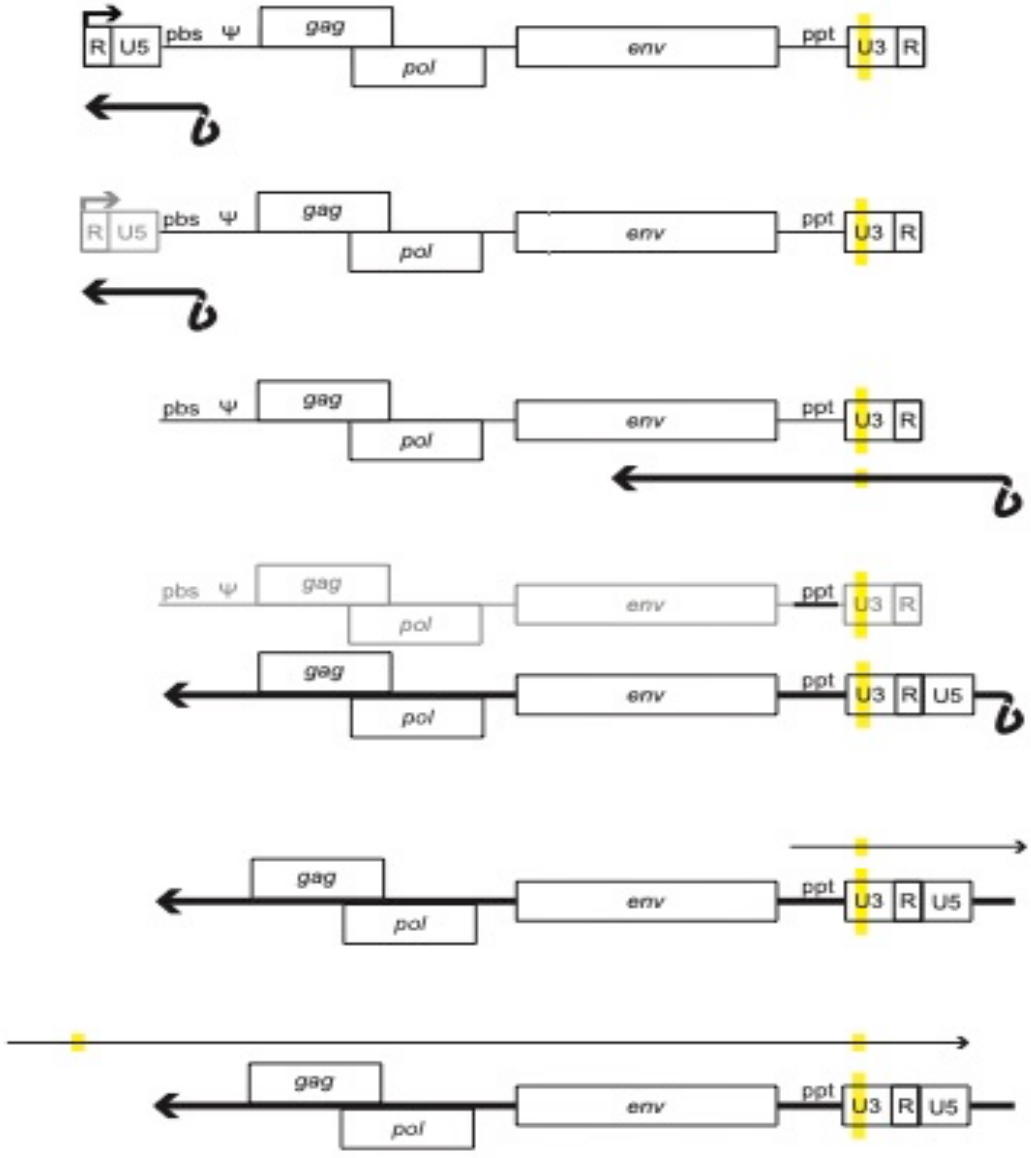


Figure 1. 2. Mechanism of HIV-1 reverse transcription. A schematic representation of the multistep process of reverse transcription detailing from the top to the bottom, the synthesis of the viral RNA (light black) to the double-stranded viral DNA (bold black) with duplicated long terminal repeats (LTR). Reverse transcriptase synthesizes the strong stop complementary DNA (cDNA) in the 5' to 3' direction using the tRNA bound to the primer binding site. The strong stop cDNA anneals to complementary R sequence in the 3' of the viral RNA, enabling synthesis of the first strand of the cDNA. The ribonuclease activity of RT degrades the viral RNA template of the viral RNA-cDNA heteroduplex leaving behind the polypurine track (PPT). The PPT serves as a primer for second strand synthesis that results in duplication of the U3-R-U5 at the 5' and 3' of the double stranded HIV-1 DNA.

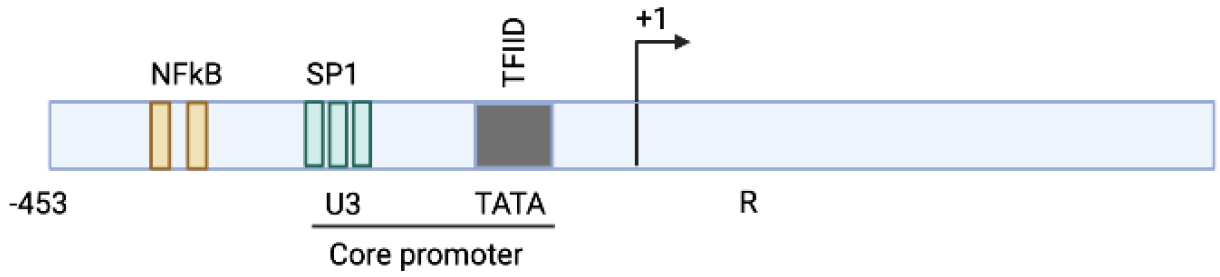


Figure 1. 3. A schematic diagram of the HIV-1 promoter. A graphical representation of the U3-R DNA of the LTR. The U3 contains two NFkB binding sites (yellow), three SP1 sites (green), and a TATA box (grey). The +1 site indicates the transcription start site.

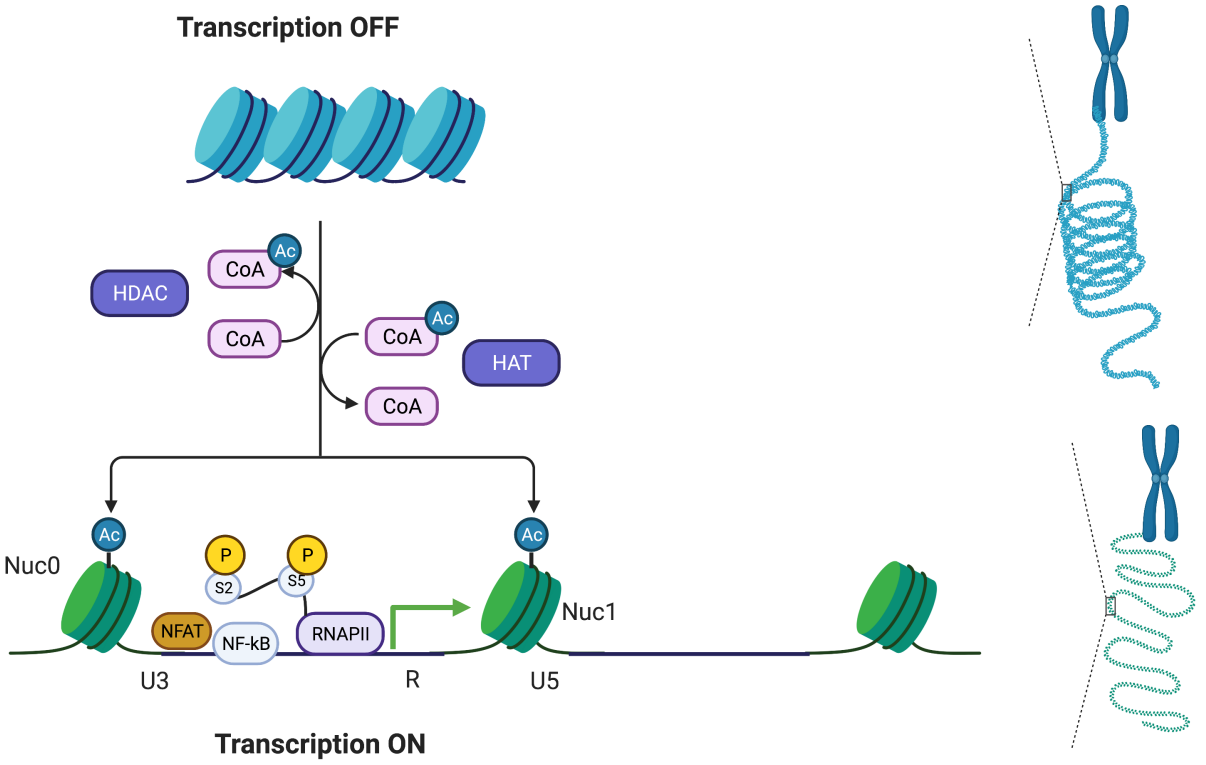


Figure 1. 4. Chromatin modulation regulates HIV-1 proviral expression. Graphical illustration of the control of the transcriptional inactive (Transcription OFF: top) and active (Transcription ON: bottom) states by chromatin modulation. The tightly packed nucleosomes (blue) prevent access to the underlying HIV-1 promoter by transcription factors; thus, transcription is turned off. Histone acetyl transferases (HATs) catalyzes the transfer of acetyl groups (Ac) from acetyl coenzyme A (CoA) to lysine residues (not shown) of histone tails. As a result, the chromatin relaxes, the nucleosomes (green) open, and the HIV-1 promoter is accessible by transcription factors, turning on transcription. Histone deacetylases (HDACs) catalyzes the reversible removal of Ac from the tail of histones to change the chromatin state from an open (euchromatin) to a closed or tightly packed (heterochromatin) state. This figure was created by me in bioRender.

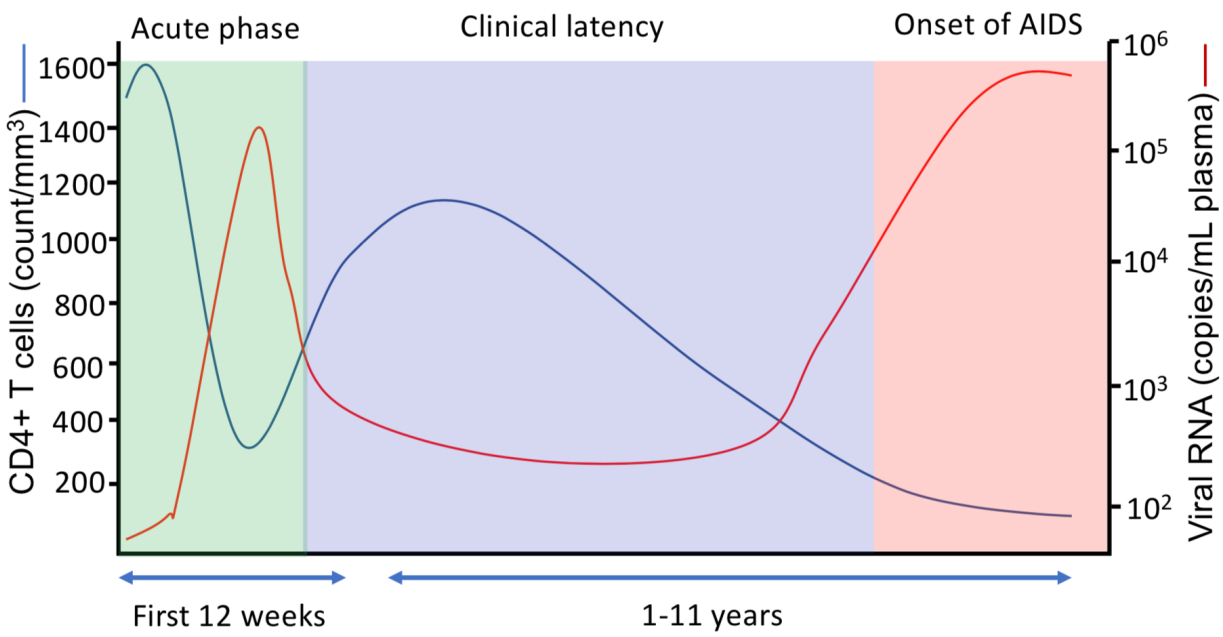


Figure 1. 5. Time course of HIV-1 infection and progression to AIDS. Within the first 12 weeks after initial infection (acute phase: green rectangle), CD4+ T cell count (left X-axis: blue line) drops sharply as viremia (right X-axis: red line) in the plasma rises as the virus continues to replicate. Gradually, the CD4+ levels rise due to immune control of the virus replication and the virus levels decline. Over the next decade (clinical latent phase: blue rectangle), CD4+ T cell count slowly diminishes until the individual is immunocompromised and can no longer control the virus replication (Doitsh, Galloway et al. 2014). Virus levels rise sharply, and the symptoms of AIDS begins (pink rectangle). This figure was created by me based on available information on: https://www.cdc.gov/hiv/basics/whatishiv.html.

References

Andersen, J. L., E. Le Rouzic and V. Planelles (2008). "HIV-1 Vpr: mechanisms of G2 arrest and apoptosis." Experimental and molecular pathology **85**(1): 2-10.

Archin, N. M., A. Espeseth, D. Parker, M. Cheema, D. Hazuda and D. M. Margolis (2009). "Expression of latent HIV induced by the potent HDAC inhibitor suberoylanilide hydroxamic acid." AIDS research and human retroviruses **25**(2): 207-212.

Archin, N. M. and D. M. Margolis (2006). "Attacking latent HIV provirus: from mechanism to therapeutic strategies." Current Opinion in HIV and AIDS **1**(2): 134-140.

Arhel, N. (2010). "Revisiting HIV-1 uncoating." Retrovirology **7**(1): 1-10.

Arhel, N. J. (2020). "Early HIV replication revisited." Nature Microbiology **5**(9): 1065-1066.

Bailey, J., J. N. Blankson, M. Wind-Rotolo and R. F. Siliciano (2004). "Mechanisms of HIV-1 escape from immune responses and antiretroviral drugs." Current opinion in immunology **16**(4): 470-476.

Beaumont, T., A. van Nuenen, S. Broersen, W. A. Blattner, V. V. Lukashov and H. Schuitemaker (2001). "Reversal of human immunodeficiency virus type 1 IIIB to a neutralization-resistant phenotype in an accidentally infected laboratory worker with a progressive clinical course." Journal of Virology **75**(5): 2246-2252.

Blazkova, J., K. Trejbalova, F. Gondois-Rey, P. Halfon, P. Philibert, A. Guiguen, E. Verdin, D. Olive, C. Van Lint and J. Hejnar (2009). "CpG methylation controls reactivation of HIV from latency." PLoS pathogens **5**(8): e1000554.

Boehm, D., V. Calvanese, R. D. Dar, S. Xing, S. Schroeder, L. Martins, K. Aull, P.-C. Li, V. Planelles and J. E. Bradner (2013). "BET bromodomain-targeting compounds reactivate HIV from latency via a Tat-independent mechanism." Cell cycle **12**(3): 452-462.

Bowerman, B., P. O. Brown, J. M. Bishop and H. E. Varmus (1989). "A nucleoprotein complex mediates the integration of retroviral DNA." Genes & development **3**(4): 469-478.

Brown, J. D., S. Kharytonchyk, I. Chaudry, A. S. Iyer, H. Carter, G. Becker, Y. Desai, L. Glang, S. H. Choi and K. Singh (2020). "Structural basis for transcriptional start site control of HIV-1 RNA fate." Science **368**(6489): 413-417.

Bruner, K. M., Z. Wang, F. R. Simonetti, A. M. Bender, K. J. Kwon, S. Sengupta, E. J. Fray, S. A. Beg, A. A. Antar and K. M. Jenike (2019). "A quantitative approach for measuring the reservoir of latent HIV-1 proviruses." Nature **566**(7742): 120-125.

Burdo, T. H., A. Lackner and K. C. Williams (2013). "Monocyte/macrophages and their role in HIV neuropathogenesis." Immunological reviews **254**(1): 102-113.

Burger, S. and M. A. Poles (2003). Natural history and pathogenesis of human immunodeficiency virus infection. Seminars in liver disease, Copyright© 2002 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New ….

Cassidy, N. A., C. S. Fish, C. N. Levy, P. Roychoudhury, D. B. Reeves, S. M. Hughes, J. T. Schiffer, S. Benki-Nugent, G. John-Stewart and D. Wamalwa (2022). "HIV reservoir quantification using cross-subtype multiplex ddPCR." iScience **25**(1): 103615.

Chen, H.-C., J. P. Martinez, E. Zorita, A. Meyerhans and G. J. Filion (2017). "Position effects influence HIV latency reversal." Nature Structural and Molecular Biology **24**(1): 47.

Coffin, J. M., S. H. Hughes and H. E. Varmus (1997). "Retroviruses."

Contreras, X., M. Schweneker, C.-S. Chen, J. M. McCune, S. G. Deeks, J. Martin and B. M. Peterlin (2009). "Suberoylanilide hydroxamic acid reactivates HIV from latently infected cells." Journal of Biological Chemistry **284**(11): 6782-6789.

Craigie, R. and F. D. Bushman (2012). "Hiv dna integration." Cold Spring Harbor perspectives in medicine **2**(7): a006890.

Dar, R. D., B. S. Razooky, A. Singh, T. V. Trimeloni, J. M. McCollum, C. D. Cox, M. L. Simpson and L. S. Weinberger (2012). "Transcriptional burst frequency and burst size are equally modulated across the human genome." Proceedings of the National Academy of Sciences **109**(43): 17454-17459.

Darcis, G., A. Kula, S. Bouchat, K. Fujinaga, F. Corazza, A. Ait-Ammar, N. Delacourt, A. Melard, K. Kabeya and C. Vanhulle (2015). "An in-depth comparison of latency-reversing agent combinations in various in vitro and ex vivo HIV-1 latency models identified bryostatin-1+ JQ1 and ingenol-B+ JQ1 to potently reactivate viral gene expression." PLoS pathogens **11**(7): e1005063.

Davey, R. T., N. Bhat, C. Yoder, T.-W. Chun, J. A. Metcalf, R. Dewar, V. Natarajan, R. A. Lempicki, J. W. Adelsberger and K. D. Miller (1999). "HIV-1 and T cell dynamics after interruption of highly active antiretroviral therapy (HAART) in patients with a history of sustained viral suppression." Proceedings of the National Academy of Sciences **96**(26): 15109-15114.

De Crignis, E. and T. Mahmoudi (2017). "The multifaceted contributions of chromatin to HIV-1 integration, transcription, and latency." International Review of Cell and Molecular Biology **328**: 197-252.

De Rijck, J., K. Bartholomeeusen, H. Ceulemans, Z. Debyser and R. Gijsbers (2010). "High-resolution profiling of the LEDGF/p75 chromatin interaction in the ENCODE region." Nucleic acids research **38**(18): 6135-6147.

Deeks, S. G. (2012). "HIV: Shock and kill." Nature **487**(7408): 439.

Dharan, A., N. Bachmann, S. Talley, V. Zwikelmaier and E. M. Campbell (2020). "Nuclear pore blockade reveals that HIV-1 completes reverse transcription and uncoating in the nucleus." Nature Microbiology **5**(9): 1088-1095.

Doitsh, G., N. L. Galloway, X. Geng, Z. Yang, K. M. Monroe, O. Zepeda, P. W. Hunt, H. Hatano, S. Sowinski and I. Muñoz-Arias (2014). "Cell death by pyroptosis drives CD4 T-cell depletion in HIV-1 infection." Nature **505**(7484): 509-514.

Du, L., C.-S. Wu, J. Sun, T. Yu, P.-P. Lyu, S.-F. Han, C. Qiu and Z.-F. Meng (2021). "Multifactorial role of HIV-Vpr in cell apoptosis revealed by a naturally truncated 54aa variant." Chinese Medical Journal **134**(7): 845.

Duncan, C. J., R. A. Russell and Q. J. Sattentau (2013). "High multiplicity HIV-1 cell-to-cell transmission from macrophages to CD4+ T cells limits antiretroviral efficacy." AIDS (London, England) **27**(14): 2201.

Edelstein, L. C., S. Micheva-Viteva, B. D. Phelan and J. P. Dougherty (2009). "Activation of latent HIV type 1 gene expression by suberoylanilide hydroxamic acid (SAHA), an HDAC inhibitor approved for use to treat cutaneous T cell lymphoma." AIDS research and human retroviruses **25**(9): 883-887.

Emiliani, S., W. Fischle, M. Ott, C. Van Lint, C. A. Amella and E. Verdin (1998). "Mutations in the tat gene are responsible for human immunodeficiency virus type 1 postintegration latency in the U1 cell line." Journal of virology **72**(2): 1666-1670.

Emiliani, S., C. Van Lint, W. Fischle, P. Paras, M. Ott, J. Brady and E. Verdin (1996). "A point mutation in the HIV-1 Tat responsive element is associated with postintegration latency." Proceedings of the National Academy of Sciences **93**(13): 6377-6381.

Epstein, F., G. Pantaleo, C. Graziosi and A. Fauci (1993). "The immunopathogenesis of human immunodeficiency virus infection." N. Engl. J. Med **328**(5): 327-335.

Eriksson, S., E. H. Graf, V. Dahl, M. C. Strain, S. A. Yukl, E. S. Lysenko, R. J. Bosch, J. Lai, S. Chioma and F. Emad (2013). "Comparative analysis of measures of viral reservoirs in HIV-1 eradication studies." PLoS pathogens **9**(2): e1003174.

Esquiaqui, J. M., S. Kharytonchyk, D. Drucker and A. Telesnitsky (2020). "HIV-1 spliced RNAs display transcription start site bias." RNA **26**(6): 708-714.

Fennessey, C. M., M. Pinkevych, T. T. Immonen, A. Reynaldi, V. Venturi, P. Nadella, C. Reid, L. Newman, L. Lipkey and K. Oswald (2017). "Genetically-barcoded SIV facilitates enumeration of rebound variants and estimation of reactivation rates in nonhuman primates following interruption of suppressive antiretroviral therapy." PLoS pathogens **13**(5): e1006359.

Finzi, D., M. Hermankova, T. Pierson, L. M. Carruth, C. Buck, R. E. Chaisson, T. C. Quinn, K. Chadwick, J. Margolick and R. Brookmeyer (1997). "Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy." Science **278**(5341): 1295-1300.

Fujinaga, K. and D. C. Cary (2020). "Experimental Systems for Measuring HIV Latency and Reactivation." Viruses **12**(11): 1279.

Gao, F., E. Bailes, D. L. Robertson, Y. Chen, C. M. Rodenburg, S. F. Michael, L. B. Cummins, L. O. Arthur, M. Peeters and G. M. Shaw (1999). "Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes." Nature **397**(6718): 436-441.

Garriga, J. and X. Graña (2004). "Cellular control of gene expression by T-type cyclin/CDK9 complexes." Gene **337**: 15-23.

Gasparian, A. V., C. A. Burkhart, A. A. Purmal, L. Brodsky, M. Pal, M. Saranadasa, D. A. Bosykh, M. Commane, O. A. Guryanova and S. Pal (2011). "Curaxins: anticancer compounds that simultaneously suppress NF-κB and activate p53 by targeting FACT." Science translational medicine **3**(95): 95ra74-95ra74.

Gérard, F. C., R. Yang, B. Romani, A. Poisson, J.-P. Belzile, N. Rougeau and É. A. Cohen (2014). "Defining the interactions and role of DCAF1/VPRBP in the DDB1-cullin4A E3 ubiquitin ligase complex engaged by HIV-1 Vpr to induce a G2 cell cycle arrest." PLoS One **9**(2): e89195.

Goh, W. C., M. E. Rogel, C. M. Kinsey, S. F. Michael, P. N. Fultz, M. A. Nowak, B. H. Hahn and M. Emerman (1998). "HIV-1 Vpr increases viral expression by manipulation of the cell cycle: a mechanism for selection of Vpr in vivo." Nature medicine **4**(1): 65-71.

Gupta, R. K., S. Abdul-Jawad, L. E. McCoy, H. P. Mok, D. Peppa, M. Salgado, J. Martinez-Picado, M. Nijhuis, A. M. Wensing and H. Lee (2019). "HIV-1 remission following CCR5Δ32/Δ32 haematopoietic stem-cell transplantation." Nature **568**(7751): 244-248.

Han, Y., Y. B. Lin, W. An, J. Xu, H.-C. Yang, K. O'Connell, D. Dordai, J. D. Boeke, J. D. Siliciano and R. F. Siliciano (2008). "Orientation-dependent regulation of integrated HIV-1 expression by host gene transcriptional readthrough." Cell host & microbe **4**(2): 134-146.

Hatziioannou, T. and D. T. Evans (2012). "Animal models for HIV/AIDS research." Nature Reviews Microbiology **10**(12): 852-867.

Henrich, T. J., E. Hanhauser, F. M. Marty, M. N. Sirignano, S. Keating, T.-H. Lee, Y. P. Robles, B. T. Davis, J. Z. Li and A. Heisey (2014). "Antiretroviral-free HIV-1 remission and viral rebound after allogeneic stem cell transplantation: report of 2 cases." Annals of internal medicine **161**(5): 319-327.

Henrich, T. J., Z. Hu, J. Z. Li, G. Sciaranghella, M. P. Busch, S. M. Keating, S. Gallien, N. H. Lin, F. F. Giguel and L. Lavoie (2013). "Long-term reduction in peripheral blood HIV type 1 reservoirs following reduced-intensity conditioning allogeneic stem cell transplantation." The Journal of infectious diseases **207**(11): 1694-1702.

Hiener, B., B. A. Horsburgh, J.-S. Eden, K. Barton, T. E. Schlub, E. Lee, S. von Stockenstrom, L. Odevall, J. M. Milush and T. Liegler (2017). "Identification of genetically intact HIV-1

proviruses in specific CD4+ T cells from effectively treated participants." Cell reports **21**(3): 813-822.

Ho, Y.-C., L. Shan, N. N. Hosmane, J. Wang, S. B. Laskey, D. I. Rosenbloom, J. Lai, J. N. Blankson, J. D. Siliciano and R. F. Siliciano (2013). "Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure." Cell **155**(3): 540-551.

Hrecka, K., M. Gierszewska, S. Srivastava, L. Kozaczkiewicz, S. K. Swanson, L. Florens, M. P. Washburn and J. Skowronski (2007). "Lentiviral Vpr usurps Cul4–DDB1 [VprBP] E3 ubiquitin ligase to modulate cell cycle." Proceedings of the National Academy of Sciences **104**(28): 11778-11783.

Hu, W.-S. and S. H. Hughes (2012). "HIV-1 reverse transcription." Cold Spring Harbor perspectives in medicine **2**(10): a006882.

Hütter, G., D. Nowak, M. Mossner, S. Ganepola, A. Müßig, K. Allers, T. Schneider, J. Hofmann, C. Kücherer and O. Blau (2009). "Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation." New England Journal of Medicine **360**(7): 692-698.

Jefferys, S. R., S. D. Burgos, J. J. Peterson, S. R. Selitsky, A.-M. W. Turner, L. I. James, Y.-H. Tsai, A. R. Coffey, D. M. Margolis and J. Parker (2021). "Epigenomic characterization of latent HIV infection identifies latency regulating transcription factors." PLoS pathogens **17**(2): e1009346.

Jiang, G., E. A. Mendes, P. Kaiser, S. Sankaran-Walters, Y. Tang, M. G. Weber, G. P. Melcher, G. R. Thompson III, A. Tanuri and L. F. Pianowski (2014). "Reactivation of HIV latency by a newly modified Ingenol derivative via protein kinase Cδ–NF-κB signaling." AIDS (London, England) **28**(11): 1555.

JingMei Hsu, K. V. B., Marshall J. Glesby, Anne Coletti, Savita G. Pahwa, Meredith Warshaw, Amanda Golner, Frederic Bone, Nicole Tobin, Marcie Riches, John W. Mellors, Renee Browning, Deborah Persaud, Yvonne Bryson. (2022). "HIV-1 REMISSION WITH CCR5?32?32 HAPLO-CORD TRANSPLANT IN A US WOMAN: IMPAACT P1107."

Johnston Jr, R. J. and C. Desplan (2010). "Stochastic mechanisms of cell fate specification that yield random or robust outcomes." Annual review of cell and developmental biology **26**: 689-719.

Jordan, A., D. Bisgrove and E. Verdin (2003). "HIV reproducibly establishes a latent infection after acute infection of T cells in vitro." The EMBO journal **22**(8): 1868-1877.

Jordan, A., P. Defechereux and E. Verdin (2001). "The site of HIV-1 integration in the human genome determines basal transcriptional activity and response to Tat transactivation." The EMBO journal **20**(7): 1726-1738.

Joshi, P. and C. A. Stoddart (2011). "Impaired infectivity of ritonavir-resistant HIV is rescued by heat shock protein 90AB1." Journal of Biological Chemistry **286**(28): 24581-24592.

Kaech, S. M., E. J. Wherry and R. Ahmed (2002). "Effector and memory T-cell differentiation: implications for vaccine development." Nature Reviews Immunology **2**(4): 251-262.

Kessing, C. F., C. C. Nixon, C. Li, P. Tsai, H. Takata, G. Mousseau, P. T. Ho, J. B. Honeycutt, M. Fallahi and L. Trautmann (2017). "In vivo suppression of HIV rebound by didehydro-cortistatin A, a "block-and-lock" strategy for HIV-1 treatment." Cell reports **21**(3): 600-611.

Kharytonchyk, S., S. Monti, P. J. Smaldino, V. Van, N. C. Bolden, J. D. Brown, E. Russo, C. Swanson, A. Shuey and A. Telesnitsky (2016). "Transcriptional start site heterogeneity modulates the structure and function of the HIV-1 genome." Proceedings of the National Academy of Sciences **113**(47): 13378-13383.

Knoener, R., E. Evans III, J. T. Becker, M. Scalf, B. Benner, N. M. Sherer and L. M. Smith (2021). "Identification of host proteins differentially associated with HIV-1 RNA splice variants." Elife **10**: e62470.

Korin, Y. D., D. G. Brooks, S. Brown, A. Korotzer and J. A. Zack (2002). "Effects of prostratin on T-cell activation and human immunodeficiency virus latency." Journal of virology **76**(16): 8118-8123.

Kulkosky, J., J. Sullivan, Y. Xu, E. Souder, D. H. Hamer and R. J. Pomerantz (2004). "Expression of latent HAART-persistent HIV type 1 induced by novel cellular activating agents." AIDS research and human retroviruses **20**(5): 497-505.

Kumar, A., G. Darcis, C. Van Lint and G. Herbein (2015). "Epigenetic control of HIV-1 post integration latency: implications for therapy." Clinical Epigenetics **7**(1): 1-12.

Laird, G. M., C. K. Bullen, D. I. Rosenbloom, A. R. Martin, A. L. Hill, C. M. Durand, J. D. Siliciano and R. F. Siliciano (2015). "Ex vivo analysis identifies effective HIV-1 latency–reversing drug combinations." The Journal of clinical investigation **125**(5): 1901-1912.

Lassen, K., Y. Han, Y. Zhou, J. Siliciano and R. F. Siliciano (2004). "The multifactorial nature of HIV-1 latency." Trends in molecular medicine **10**(11): 525-531.

Lassen, K. G., A. M. Hebbeler, D. Bhattacharyya, M. A. Lobritz and W. C. Greene (2012). "A flexible model of HIV-1 latency permitting evaluation of many primary CD4 T-cell reservoirs." PloS one **7**(1): e30176.

Le Rouzic, E., N. Belaïdouni, E. Estrabaud, M. Morel, J.-C. Rain, C. Transy and F. Margottin-Goguet (2007). "HIV1 Vpr arrests the cell cycle by recruiting DCAF1/VprBP, a receptor of the Cul4-DDB1 ubiquitin ligase." Cell cycle **6**(2): 182-188.

LeBlanc, J., J. Weil and K. Beemon (2013). "Posttranscriptional regulation of retroviral gene expression: primary RNA transcripts play three roles as pre-mRNA, mRNA, and genomic RNA." Wiley Interdisciplinary Reviews: RNA **4**(5): 567-580.

Lenasi, T., X. Contreras and B. M. Peterlin (2008). "Transcriptional interference antagonizes proviral gene expression to promote HIV latency." Cell host & microbe **4**(2): 123-133.

Levy, C. N., S. M. Hughes, P. Roychoudhury, D. B. Reeves, C. Amstuz, H. Zhu, M.-L. Huang, Y. Wei, M. E. Bull and N. A. Cassidy (2021). "A highly multiplexed droplet digital PCR assay to measure the intact HIV-1 proviral reservoir." Cell Reports Medicine **2**(4): 100243.

Li, W., P. K. Singh, G. A. Sowd, G. J. Bedwell, S. Jang, V. Achuthan, A. V. Oleru, D. Wong, H. J. Fadel and K. Lee (2020). "CPSF6-dependent targeting of speckle-associated domains distinguishes primate from nonprimate lentiviral integration." MBio **11**(5): e02254-02220.

Liu, L., B. Patel, M. H. Ghanem, V. Bundoc, Z. Zheng, R. A. Morgan, S. A. Rosenberg, B. Dey and E. A. Berger (2015). "Novel CD4-based bispecific chimeric antigen receptor designed for enhanced anti-HIV potency and absence of HIV entry receptor activity." Journal of virology **89**(13): 6685-6694.

Liu, R., W. A. Paxton, S. Choe, D. Ceradini, S. R. Martin, R. Horuk, M. E. MacDonald, H. Stuhlmann, R. A. Koup and N. R. Landau (1996). "Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection." Cell **86**(3): 367-377.

Maldini, C. R., K. Gayout, R. S. Leibman, D. L. Dopkin, J. P. Mills, X. Shan, J. A. Glover and J. L. Riley (2020). "HIV-resistant and HIV-specific CAR-modified CD4+ T cells mitigate HIV disease progression and confer CD4+ T cell help in vivo." Molecular Therapy **28**(7): 1585-1599.

Masiero, S., C. Del Vecchio, R. Gavioli, G. Mattiuzzo, M. Cusi, L. Micheli, F. Gennari, A. Siccardi, W. Marasco and G. Palu (2005). "T-cell engineering by a chimeric T-cell receptor with antibody-type specificity for the HIV-1 gp120." Gene therapy **12**(4): 299-310.

Mbonye, U., B. Wang, G. Gokulrangan, W. Shi, S. Yang and J. Karn (2018). "Cyclin-dependent kinase 7 (CDK7)-mediated phosphorylation of the CDK9 activation loop promotes P-TEFb assembly with Tat and proviral HIV reactivation." Journal of Biological Chemistry **293**(26): 10009-10025.

McMichael, A. J. and S. L. Rowland-Jones (2001). "Cellular immune responses to HIV." Nature **410**(6831): 980-987.

McNamara, L. A. and K. L. Collins (2011). "Hematopoietic stem/precursor cells as HIV reservoirs." Current Opinion in HIV and AIDS **6**(1): 43.

Méndez, C., S. Ledger, K. Petoumenos, C. Ahlenstiel and A. D. Kelleher (2018). "RNA-induced epigenetic silencing inhibits HIV-1 reactivation from latency." Retrovirology **15**(1): 1-18.

Mendoza, P., J. R. Jackson, T. Y. Oliveira, C. Gaebler, V. Ramos, M. Caskey, M. Jankovic, M. C. Nussenzweig and L. B. Cohn (2020). "Antigen-responsive CD4+ T cell clones contribute to the HIV-1 latent reservoir." Journal of Experimental Medicine **217**(7).

Mousseau, G., M. A. Clementz, W. N. Bakeman, N. Nagarsheth, M. Cameron, J. Shi, P. Baran, R. Fromentin, N. Chomont and S. T. Valente (2012). "An analog of the natural steroidal alkaloid cortistatin A potently suppresses Tat-dependent HIV transcription." Cell host & microbe **12**(1): 97-108.

Mousseau, G., C. F. Kessing, R. Fromentin, L. Trautmann, N. Chomont and S. T. Valente (2015). "The Tat inhibitor didehydro-cortistatin A prevents HIV-1 reactivation from latency." MBio **6**(4): e00465-00415.

Nishizawa, Y., J. Usukura, D. P. Singh, L. T. Chylack and T. Shinohara (2001). "Spatial and temporal dynamics of two alternatively spliced regulatory factors, lens epithelium-derived growth factor (ledgf/p75) and p52, in the nucleus." Cell and tissue research **305**(1): 107-114.

Ocwieja, K. E., S. Sherrill-Mix, R. Mukherjee, R. Custers-Allen, P. David, M. Brown, S. Wang, D. R. Link, J. Olson and K. Travers (2012). "Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing." Nucleic acids research **40**(20): 10345-10355.

Pace, M. J., L. Agosto, E. H. Graf and U. O'Doherty (2011). "HIV reservoirs and latency models." Virology **411**(2): 344-354.

Perelson, A. S., A. U. Neumann, M. Markowitz, J. M. Leonard and D. D. Ho (1996). "HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time." Science **271**(5255): 1582-1586.

Peterlin, B. M. and D. H. Price (2006). "Controlling the elongation phase of transcription with P-TEFb." Molecular cell **23**(3): 297-305.

Price, D. H. (2000). "P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II." Molecular and cellular biology **20**(8): 2629-2634.

Raghavendra, N. K., N. Shkriabai, R. L. Graham, S. Hess, M. Kvaratskhelia and L. Wu (2010). "Identification of host proteins associated with HIV-1 preintegration complexes isolated from infected CD4+ cells." Retrovirology **7**(1): 1-7.

Sabbah, E. N., S. Druillennec, N. Morellet, S. Bouaziz, G. Kroemer and B. Roques (2006). "Interaction between the HIV-1 protein Vpr and the adenine nucleotide translocator." Chemical biology & drug design **67**(2): 145-154.

Saleh, S., A. Solomon, F. Wightman, M. Xhilaga, P. U. Cameron and S. R. Lewin (2007). "CCR7 ligands CCL19 and CCL21 increase permissiveness of resting memory CD4+ T cells to HIV-1 infection: a novel model of HIV-1 latency." Blood, The Journal of the American Society of Hematology **110**(13): 4161-4164.

Sedore, S. C., S. A. Byers, S. Biglione, J. P. Price, W. J. Maury and D. H. Price (2007). "Manipulation of P-TEFb control machinery by HIV: recruitment of P-TEFb from the large form by Tat and binding of HEXIM1 to TAR." Nucleic acids research **35**(13): 4347-4358.

Simonetti, F. R., H. Zhang, G. P. Soroosh, J. Duan, K. Rhodehouse, A. L. Hill, S. A. Beg, K. McCormick, H. E. Raymond and C. L. Nobles (2021). "Antigen-driven clonal selection shapes the persistence of HIV-1–infected CD4+ T cells in vivo." The Journal of clinical investigation **131**(3).

Singh, A., B. Razooky, C. D. Cox, M. L. Simpson and L. S. Weinberger (2010). "Transcriptional bursting from the HIV-1 promoter is a significant source of stochastic noise in HIV-1 gene expression." Biophysical journal **98**(8): L32-L34.

Sneller, M. C., E. D. Huiting, K. E. Clarridge, C. Seamon, J. Blazkova, J. S. Justement, V. Shi, E. J. Whitehead, R. F. Schneck and M. Proschan (2020). "Kinetics of plasma HIV rebound in the era of modern antiretroviral therapy." The Journal of infectious diseases **222**(10): 1655-1659.

Spina, C. A., J. Anderson, N. M. Archin, A. Bosque, J. Chan, M. Famiglietti, W. C. Greene, A. Kashuba, S. R. Lewin and D. M. Margolis (2013). "An in-depth comparison of latent HIV-1 reactivation in multiple cell model systems and resting CD4+ T cells from aviremic patients." PLoS pathogens **9**(12): e1003834.

Spina, C. A., J. C. Guatelli and D. D. Richman (1995). "Establishment of a stable, inducible form of human immunodeficiency virus type 1 DNA in quiescent CD4 lymphocytes in vitro." Journal of virology **69**(5): 2977-2988.

Steger, D. J. and J. L. Workman (1997). "Stable co-occupancy of transcription factors and histones at the HIV-1 enhancer." The EMBO Journal **16**(9): 2463-2472.

Stewart, S. A., B. Poon, J. Jowett and I. Chen (1997). "Human immunodeficiency virus type 1 Vpr induces apoptosis following cell cycle arrest." Journal of Virology **71**(7): 5579-5592.

Stoltzfus, C. M. (2009). "Regulation of HIV-1 alternative RNA splicing and its role in virus replication." Advances in virus research **74**: 1-40.

Swain, P. S., M. B. Elowitz and E. D. Siggia (2002). "Intrinsic and extrinsic contributions to stochasticity in gene expression." Proceedings of the National Academy of Sciences **99**(20): 12795-12800.

Tripathy, M. K., W. Abbas and G. Herbein (2011). "Epigenetic regulation of HIV-1 transcription." Epigenomics **3**(4): 487-502.

Van Lint, C., S. Bouchat and A. Marcello (2013). "HIV-1 transcription and latency: an update." Retrovirology **10**(1): 1-38.

Verdin, E. (1991). "DNase I-hypersensitive sites are associated with both long terminal repeats and with the intragenic enhancer of integrated human immunodeficiency virus type 1." Journal of virology **65**(12): 6790-6799.

Verdin, E., P. Paras and C. Van Lint (1993). "Chromatin disruption in the promoter of human immunodeficiency virus type 1 during transcriptional activation." The EMBO journal **12**(8): 3249-3259.

Vozzolo, L., B. Loh, P. J. Gane, M. Tribak, L. Zhou, I. Anderson, E. Nyakatura, R. G. Jenner, D. Selwood and A. Fassati (2010). "Gyrase B inhibitor impairs HIV-1 replication by targeting Hsp90 and the capsid protein." Journal of Biological Chemistry **285**(50): 39314-39328.

Wang, Z., E. E. Gurule, T. P. Brennan, J. M. Gerold, K. J. Kwon, N. N. Hosmane, M. R. Kumar, S. A. Beg, A. A. Capoferri and S. C. Ray (2018). "Expanded cellular clones carrying replication-competent HIV-1 persist, wax, and wane." Proceedings of the National Academy of Sciences **115**(11): E2575-E2584.

Weinberger, A. D. and L. S. Weinberger (2013). "Stochastic fate selection in HIV-infected patients." Cell **155**(3): 497-499.

Weinberger, L. S., J. C. Burnett, J. E. Toettcher, A. P. Arkin and D. V. Schaffer (2005). "Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity." Cell **122**(2): 169-182.

Wen, X., K. M. Duus, T. D. Friedrich and C. M. de Noronha (2007). "The HIV1 protein Vpr acts to promote G2 cell cycle arrest by engaging a DDB1 and Cullin4A-containing ubiquitin ligase complex using VprBP/DCAF1 as an adaptor." Journal of Biological Chemistry **282**(37): 27046-27057.

Williams, S. A., L.-F. Chen, H. Kwon, D. Fenard, D. Bisgrove, E. Verdin and W. C. Greene (2004). "Prostratin antagonizes HIV latency by activating NF-κB." Journal of Biological Chemistry **279**(40): 42008-42017.

Williams, S. A. and W. C. Greene (2007). "Regulation of HIV-1 latency by T-cell activation." Cytokine **39**(1): 63-74.

Wong-Staal, F., P. K. Chanda and J. Ghrayeb (1987). "Human immunodeficiency virus: the eighth gene." AIDS research and human retroviruses **3**(1): 33-39.

Yang, H.-C., S. Xing, L. Shan, K. O'Connell, J. Dinoso, A. Shen, Y. Zhou, C. K. Shrum, Y. Han and J. O. Liu (2009). "Small-molecule screening using a human primary cell model of HIV latency identifies compounds that reverse latency without cellular activation." The Journal of clinical investigation **119**(11): 3473-3486.

Zhang, F. and P. D. Bieniasz (2020). "HIV-1 Vpr induces cell cycle arrest and enhances viral gene expression by depleting CCDC137." Elife **9**: e55806.

Zhang, L., B. Ramratnam, K. Tenner-Racz, Y. He, M. Vesanen, S. Lewin, A. Talal, P. Racz, A. S. Perelson and B. T. Korber (1999). "Quantifying residual HIV-1 replication in patients receiving combination antiretroviral therapy." New England Journal of Medicine **340**(21): 1605-1613.

Zhu, J., G. D. Gaiha, S. P. John, T. Pertel, C. R. Chin, G. Gao, H. Qu, B. D. Walker, S. J. Elledge and A. L. Brass (2012). "Reactivation of latent HIV-1 by inhibition of BRD4." Cell reports **2**(4): 807-816.

Zimmerman, E. S., M. P. Sherman, J. L. Blackett, J. A. Neidleman, C. Kreis, P. Mundt, S. A. Williams, M. Warmerdam, J. Kahn and F. M. Hecht (2006). "Human Immunodeficiency Virus Type 1 Vpr Induces DNA Replication Stress InVitro and In Vivo." Journal of virology **80**(21): 10407-10418.

Chapter 2 [1]Stable Integrant-specific Differences in Bimodal HIV-1 Expression Patterns Revealed by High-throughput Analysis

Abstract

HIV-1 gene expression is regulated by host and viral factors that interact with viral motifs and is influenced by proviral integration sites. Here, expression variation among integrants was followed for hundreds of individual proviral clones within polyclonal populations throughout successive rounds of virus and cultured cell replication, with limited findings using CD4+ cells from donor blood consistent with observations in immortalized cells. Tracking clonal behavior by proviral "zip codes" indicated that mutational inactivation during reverse transcription was rare, while clonal expansion and proviral expression states varied widely. By sorting for provirus expression using a GFP reporter in the *nef* open reading frame, distinct clone-specific variation in on/off proportions were observed that spanned three orders of magnitude. Tracking GFP phenotypes over time revealed that as cells divided, their progeny alternated between HIV transcriptional activity and non-activity. Despite these phenotypic oscillations, the overall GFP+ population within each clone was remarkably stable, with clones maintaining clone-specific equilibrium mixtures of GFP+ and GFP- cells. Integration sites were analyzed for correlations between genomic features and the epigenetic phenomena described here. Integrants inserted in the sense orientation of genes were more frequently found to be GFP negative than those in the antisense orientation, and clones

[1] This chapter has been published as Read, D.F., Atindaana, E., Pyaram, K., Yang, F., Emery, S., Cheong, A., Nakama, K.R., Burnett, C., Larragoite, E.T., Battivelli, E. and Verdin, E., 2019. Stable integrant-specific differences in bimodal HIV-1 expression patterns revealed by high-throughput analysis. *PLoS pathogens*, *15*(10), p.e1007903.

with high GFP+ proportions were more distal to repressive H3K9me3 peaks than low GFP+ clones. Clones with low frequencies of GFP positivity appeared to expand more rapidly than clones for which most cells were GFP+, even though the tested proviruses were Vpr-. Thus, much of the increase in the GFP- population in these polyclonal pools over time reflected differential clonal expansion. Together, these results underscore the temporal and quantitative variability in HIV-1 gene expression among proviral clones that are conferred in the absence of metabolic or cell-type dependent variability, and shed light on cell-intrinsic layers of regulation that affect HIV-1 population dynamics.

Summary

Very few HIV-1 infected cells persist in patients for more than a couple days, but those that do pose life-long health risks. Strategies designed to eliminate these cells have been based on assumptions about what viral properties allow infected cell survival. However, such approaches for HIV-1 eradication have not yet shown therapeutic promise, possibly because many assumptions about virus persistence are based on studies involving a limited number of infected cell types, the averaged behavior of cells in diverse populations, or snapshot views. Here, we developed a high-throughput approach to study hundreds of distinct HIV-1 infected cells and their progeny over time in an unbiased way. This revealed that each virus established its own pattern of gene expression that, upon infected cell division, was stably transmitted to all progeny cells. Expression patterns consisted of alternating waves of activity and inactivity, with the extent of activity differing among infected cell families over a 1000-fold range. The dynamics and variability among infected cells and within complex populations that the work here revealed has

not previously been evident, and may help establish more accurate correlates of persistent HIV-1 infection.

Introduction

Early in the HIV-1 replication cycle, a DNA intermediate integrates into the host cell's genome. HIV-1 replication ordinarily progresses into its late phases, with viral gene expression, virion production, and cell death. However, some proviruses can remain dormant upon integration. In patients, the resulting latently infected cells persist throughout antiretroviral treatment, and their sporadic reactivation can lead to virus rebound after antiretroviral cessation.

This source of persistent virus is called the latent reservoir and is believed to consist largely of transcriptionally silent proviruses integrated into resting memory T cells (Chun, Stuyver et al. 1997, Finzi, Hermankova et al. 1997, Wong, Hezareh et al. 1997). Experimentally, infectious virus can be produced by T lymphocytes from such patients when the cells are activated or treated with certain chromatin remodeling drugs *ex vivo*. These observations inspired "shock and kill" HIV cure strategies, which involve pharmacologically inducing provirus expression to promote the recognition and clearance of latently infected cells (Archin, Liberty et al. 2012, Deeks 2012). However, while intervention that reactivates silenced proviruses can activate HIV-1 gene expression in cell culture models of latency, such treatments have thus far failed to fulfill their promise in the clinic, suggesting much remains to be learned about the establishment and maintenance of the latent reservoir (Rasmussen and Lewin 2016, Mbonye and Karn 2017, Spivak and Planelles 2018).

HIV-1 gene expression requires sequence motifs within proviral sequences that specify nucleosome positioning and allow HIV-1 to respond to host factor differences among infected cell types (Verdin, Paras et al. 1993, Kaczmarek, Morales et al. 2013, Ne, Palstra et al. 2018). HIV-1 has a marked preference for integration in transcriptionally active genome regions (Jordan, Defechereux et al. 2001, Schroder, Shinn et al. 2002), and certain host chromatin binding factors as well as nuclear architecture further bias the distribution of integration sites (Ciuffi, Llano et al. 2005, Wong, Mamede et al. 2015). Integration sites influence HIV gene expression (Lewinski, Bisgrove et al. 2005, Sherrill-Mix, Lewinski et al. 2013, Sunshine, Kirchner et al. 2016, Chen, Martinez et al. 2017), and it has been postulated that integration sites may affect the odds of a provirus establishing long-lived latency (Dahabieh, Battivelli et al. 2015). Differences in HIV-1 expression due to integration site features likely influence the extent to which cells survive and proliferate after HIV-1 integration, and in turn contribute to the expression profile of persistent HIV-1 (Anderson and Maldarelli 2018).

Recent work with patient samples has demonstrated that for at least some suppressed patients, residual provirus-containing cells are polyclonal yet dominated by a limited number of clonal subsets (Mullins and Frenkel 2017), and similar observations of clonal expansion have been made during HIV-1 infection of humanized mice (Satou, Katsuya et al. 2017). Thus, the integration sites represented in persistent proviruses are probably a limited subset of the spectrum initially generated (Anderson and Maldarelli 2018).

Recent evidence indicates that latent proviruses differ in the extents to which they can be reactivated, and that a large majority of cells harboring latent proviruses may be refractory to our

current arsenal of reactivation agents (Ho, Shan et al. 2013, Battivelli, Dahabieh et al. 2018). Work using dual color reporter viruses in primary cells has shown that proviruses differ in their reactivation potential depending on their sites of integration, with chromatin context as maintained within the confines of the nucleus being a significant contributing factor (Battivelli, Dahabieh et al. 2018). Additional work monitoring HIV-1 expression in individual cells has questioned the earlier view that complete proviral silencing is necessary for infected cell persistence during antiretroviral therapy (Wiegand, Spindler et al. 2017, Pinzone, VanBelzen et al. 2019).

The majority of proviruses detectable in suppressed patients are replication defective (Bruner, Murray et al. 2016, Pinzone, VanBelzen et al. 2019). Although such proviruses are incapable of rekindling infection, emerging evidence suggests they can be expressed and may contribute to pathogenesis (Pollack, Jones et al. 2017, Pinzone, VanBelzen et al. 2019).

In this study, we developed a high throughput approach to monitor cellular and viral progeny of individual integration events within complex populations, and used it to address the frequency of defective provirus formation and the extent to which provirus integration sites affect provirus expression levels. Initial work was performed using transformed cell lines, where selective pressures and variation of intracellular factors should be lower than in primary cells, with additional experiments performed in CD4+ lymphocytes from donor blood. Examining the extent of expression variation within and among cellular progeny of large panels of individual HIV-1 integration events indicated that in all these cell types, epigenetic differences among proviral clones led to the establishment of distinct heritable patterns of HIV-1 gene expression.

Results

Nearly 90% of zip coded proviruses supported a second round of replication

We developed a system to identify individual HIV-1 proviral lineages within polyclonal populations, track proviral gene expression, and monitor replication properties of individual cell clones and their viral progeny. To achieve this, NL4-3 strain-derived vectors that encoded Gag, Pol, Tat, Rev and a puromycin reporter (pNL4-3 GPP (Lu, Heng et al. 2011); Figure 2.1A) were modified to each contain a unique 20-base randomized sequence tag. Once integrated, these were called "zip coded" proviruses because the tags reported provirus locations. Tags were inserted into the upstream edge of U3, downstream of integrase recognition sequences and upstream of the site of nuc0 nucleosome binding (Verdin, Paras et al. 1993). Vector RNAs were transcribed from uncloned DNA template libraries generated by *in vitro* assembly without amplification by plasmid replication. High throughput sequencing confirmed that the tag complexity of the starting library vastly exceeded the analyzed provirus population size (Figure S2.1.). Because the process of reverse transcription duplicates U3, each progeny provirus contained the same randomized tag in both LTRs, and each provirus's tags differed from those in every other integrant.

To validate this approach, adherent 293T cells were transduced at a very low multiplicity of infection (<0.00005) and the randomized regions amplified from ten individual puromycin-resistant colonies were sequenced. The results showed that no two colonies contained the same 20-mer (Table S2.1).

An initial pooled-clone pilot study was then performed, which addressed the frequency of defective provirus formation during a single replication round (Figure 2.1B). 71 well-separated

puromycin-resistant colonies were combined to generate an F_1 cell pool. After expansion, pseudotyped virions ("F_1 virus") generated from F_1 cells were used to infect fresh 293T cells. Because the number of colonies pooled to generate the F_2 cell pool--roughly 1000-- was significantly greater than the F_1 pool's zip code complexity, any infectious zip code present in the F_1 pool was predicted to generate multiple F_2 integrants.

The ability of each F_1 provirus to complete a second replication round was addressed by comparing F_1 and F_2 virion cDNA and F_1 and F_2 cell DNA zip codes using high throughput sequencing (Figure 2.1B). How zip codes were analyzed and quantified is described in Materials and Methods, and included ranking zip codes based on sequencing read frequencies, beginning with the most abundant. F_1 pool cells were found to contain 74 unique zip codes, which accounted for 99.87% of total sequencing reads (Figure S2.2). Although the possibility of dual infection cannot be ruled out, the low multiplicity of infection used here suggested the discrepancy between this value and the 71 colonies visualized was likely due to miscounting double colonies as single expanded clones. Because 65 out of the 74 zip codes found in F_1 cell DNA were also observed in the F_2 cell library, these 65 (88% of F_1 cell zip codes) unambiguously represented proviruses capable of completing a second round of replication (Figure 2.1C). The remaining 9 zip codes were candidate non-infectious proviruses. If a first-round provirus could assemble but not replicate, its zip code might be detectable in F_1 virus but not F_2 cells. Seven zip codes were candidates for this class of defective proviruses (green lines in Figure 2.1C).

The remaining two clones were initially enigmatic. The number of colonies pooled to make the F2 library suggested it contained roughly twenty re-transduced copies of each F1 zip code. Based on

how frequently replication competence was maintained after the first round of replication, any fully infectious F_1 provirus was expected to display a roughly 90% second-round success rate. Thus, the likelihood that all ~20 sibling F_2 progeny of any infectious F_1 provirus would be defective seemed exceptionally low. However, among the 65 replication-competent zip codes in the F_2 cell library, two were not observed in F_2 virus RNA.

Integrant clone expansion and provirus expression levels varied widely among zip coded 293T cell clones

To address whether the absence of two F_2 cell zip codes from the F_2 virus library might reflect a population bottleneck, the number of sequencing reads associated with each zip code was compared within and across libraries. Unexpectedly, reads per zip code were observed to vary over three orders of magnitude within the F_1 cell library (Figure 2.1 D). Although variation in the expansion rates of provirus-containing cells have been reported previously (Nolan-Stevaux, Tedesco et al. 2013), the wide range in cell clone sizes observed here had not been anticipated.

Clone-specific differences in the amount of virus released per cell were also observed (Figure 2.1D, y axis). When normalized to the number of F_1 cells harboring a given zip code, differences in virion release per cell spanned two orders of magnitude or more. Because of this, zip code abundance in the F_1 cell and F_1 virus libraries were only moderately correlated (Figure 2.1D) (Spearman $\rho = .639$, $p=8.7*10^{-10}$). In contrast, the correlation between cell count and virion production was strong in the F_2 generation (Spearman $\rho = .890$, $p=3.75*10^{-23}$) where each zip code was polyclonal (Figure 2.1 F), suggesting that virus-per-cell ratios were fairly consistent when averaged across many cell clones.

Looking specifically at sequencing read data for the two F_2 cell zip codes that were missing from F_2 virus revealed that these lineages were scarce in both the F_1 virus and in F_2 cells (red points, Figure 2.1 E). Similarly, read frequency trends for the seven F_1 zip codes not observed in F_2 cells (green points, Figure 2.1 D) suggested that population bottlenecks, and not loss of infectivity, may account for the absence of some of these in F_2 cells.

The pilot studies above validated assessing multiple proviral lineages within cultured cell populations by tracking read counts in high throughput sequencing libraries. However, because the cells were not physically cloned, it remains possible that experimental procedures may have introduced unintended variation and skewed read counts. Nonetheless, the fact that amplicons were the same length for all library members, paired with the practice of performing all experiments in duplicate, provides general support for the assumption applied below: namely, that zip code read frequencies in sequencing libraries reflected the abundance of that zip code within the cell population used to generate the library.

Clonal expansion in Jurkat cells

Larger zip coded integrant populations were then established using Jurkat cells. The vector in these experiments (HIV GPV$^-$) expressed all HIV-1 genes except *env*, *nef*, and *vpr*, contained GFP in the *nef* open reading frame, and expressed a selectable marker (for puromycin resistance) from a secondary, internal promoter, as has been done in previous strategies (Dahabieh, Ooms et al. 2013) (Figure 2.1A, lower construct). Selective concentrations of puromycin were applied briefly, and cells were subsequently maintained without drug. Cell pools infected with differing

amounts of virus were analyzed by high throughput sequencing, and one of these pools, which was determined to contain roughly 1,000 zip coded clones, was used in subsequent studies.

Sequencing duplicate aliquots of this pool revealed that many zip codes were shared in both replicates, but lower abundance zip codes were sampled unevenly. To better address the complexity of the pool and differential clone expansion, ten technical replicates were combined to provide evidence for 706 zip codes, which together accounted for 97.8% of the total reads. Based on zip code sequencing read frequencies (Figure S2.3), the pool displayed clonal abundances spanning over two orders of magnitude, with the most prevalent half-dozen zip codes each accounting for >1% of the total reads while the lower 10% of the 706 zip codes each containing <0.01%of the total reads.

Significant clone-by-clone differences in HIV-1 expression in both Jurkat and primary cells

Detecting GFP by flow cytometry allowed binary (on/off) monitoring of LTR expression in individual cells, and work here used GFP as a surrogate for HIV-1 gene expression. Portions of the total Jurkat pool characterized above, designated Pools 1 and 2, were independently sorted into GFP positive "GFP+" and negative "GFP-" sub-pools (Figure 2.2A; Figure S2.4 shows how sorting was gated). As a control, cells were also analyzed by FACS based on their p24 content using an anti-p24 antibody, and a strong correlation between GFP expression and p24 content was observed (Figure S2.5).

An expression value termed the "GFP+ proportion" was determined for each zip coded clone. GFP+ proportions were calculated by dividing the read frequency of each zip code in GFP+ sorted

cells by the sum of the abundance of that zip code in GFP+ and GFP- sorted cells after weighting values to reflect the proportions of total cells in the GFP+ and GFP- sub-pools. A sample calculation is provided in Materials and Methods. Consistent with clonal variation in virus release per cell observed in the pilot experiment above, GFP+ proportions differed substantially among Jurkat cell clones, with the GFP+ proportions of individual clones ranging from >99% to <1% (Figure 2.2).

To test if the broad range of clones' GFP+ proportions reflected clone-specific properties or were a result of sampling, we compared data for duplicate experimental samples, with the GFP+ proportions calculated for each zip code in Pool 1 compared to those independently determined for Pool 2. As shown in Figure 2.2B, when GFP+ proportion data were plotted against each other, most clones displayed similar values, suggesting that each clone possessed a distinct GFP+ proportion that was not defined by sampling (Spearman $\rho = 0.474$, $p=8.23*10^{-40}$ for the 688 zip codes detected in each pool). GFP+ proportions were particularly well correlated for the most abundant zip codes (Figure 2.2C, Spearman $\rho = 0.716$, $p=1.22*10^{-36}$ for the 225 zip codes with fractional abundance > 0.001 in the parental pool), suggesting that at least 200 clones were sufficiently abundant in the total population to be reproducibly well sampled in repeated sub-pools.

The experiments above were performed with cell lines, where within-experiment differences in environment and *trans*-acting factors should be minimized (Chomont, El-Far et al. 2009). In an initial test of whether primary cells also displayed integrant-specific differences in our system, CD4+ cells were isolated from donor blood, stimulated, and transduced with VSV-G pseudotyped zip coded GPV^- (Figure S2.6). Six days post infection, the cells were divided into 2 sub-pools that

were each sorted into GFP- and GFP+ cell fractions, and the GFP+ proportions of individual clones in each sub-pool were compared. Three independent experimental repetitions were performed, each using cells from a different blood donor. The approaches are described in the methods and Figure S2.6. The results showed that in these primary cell experiments, most zip codes were represented by very few sequencing reads, possibly due to variation in primary cell division rates and to the retention of significant amounts of unintegrated viral DNA during the relatively short duration of primary cell propagation. Additionally, the number of clones that were sampled sufficiently to meet inclusion criteria (that the clone was detectable with fractional abundance > 0.0001 in each sub-pool) was low. Nonetheless, significant correlations were observed when the GFP+ proportions for these primary cell zip codes were calculated for each independently analyzed sub-pool and values for the two replicate sorts within each experimental repetition were plotted against one another (Figure S2.6), indicating that the provirus-containing progeny of primary cells have clonal differences in HIV-1 gene expression levels.

GFP+ proportions of clones are a stable, heritable phenotype

Longitudinal studies were then performed with the zip coded Jurkat pool established in Figure 2.2A, to monitor GFP+ proportions throughout cell generations. After sampling for sequencing library preparation, aliquots of the GFP+ and GFP- sub-pools of both Pool 1 and Pool 2 were passaged separately for an additional 8 to 9 days, at which time point each of these four sub-pools were again sorted by FACS (Figure 2.3A and 3D). The results showed that the cellular descendants of Pool 1 and Pool 2 GFP+ sub-pools did not all remain GFP+ (Figure 2.3B and 3E), nor did the descendants of the GFP- sub-pools remain all GFP- (Figure 2.3C and 3F). Instead, some cells from each sub-pool had switched expression phenotypes during passaging. This suggested that the HIV-

1 expression pattern in any individual cell was not stably inherited by all of its progeny, but that instead expression "flickered" (alternated between LTR expression and silencing) during cell propagation.

Integrant specific, intrinsic rates of expression that are maintained across cell generations have previously been reported for basal expression from the HIV-1 promoter (Jordan, Defechereux et al. 2001). To test whether or not the expression patterns studied here also were stable over time, the GFP+ proportions determined for the GFP+ or GFP- pools in the second sort were combined after weighting to reconstitute the original unfractionated population (reconstructed second sort pools in Figure 2.3A and 3D; described in Materials and Methods). The GFP+ proportions in Pool 1 and Pool 2 at the time of the first sort were then compared to the GFP+ proportions of the reconstructed pools at the time of the second sort (Figure 2.3G and 3H). Consistent with the stable inheritance of clone-specific intrinsic expression patterns, here the data indicated that the weighted GFP+ proportions for each integrant following the second sort showed a strong correlation with its GFP+ proportion in the first sort (Spearman $\rho = 0.939$, $p=1.1*10^{-105}$ for Pool 1 and 0.806, $p=1.2*10^{-52}$ for Pool 2; Figure 2.3 first sort *vs.* reconstructed second sort GFP+ proportion values).

Correlates of integration site features and provirus activity

Integration site features were compared to address whether these features affected the viral gene expression patterns observed here. Integration sites were determined using a linker-mediated nested PCR strategy applied to genomic DNA from the original unsorted Jurkat pool. Primers were designed so that sequencing reads included integration site sequences and U3 resident zip codes. Initial analysis indicated variable rates of assignment of a single zip code to multiple genomic

locations, likely reflecting the formation of chimeric molecules during PCR (Quail, Kozarewa et al. 2008, Kircher, Sawyer et al. 2012). We therefore implemented an algorithm that removed minor assignments presumed to be PCR artifacts and selected abundant, redundantly implicated integration sites. This strategy assigned genomic location to each of the 225 high abundance zip codes (Table S2.2). As expected (Serrao and Engelman 2016), integrants were substantially enriched for annotated genes and genes expressed in Jurkat cells (Figure 2.4A and 4B), with 58% having the same orientation as the intersecting transcript (109 of 188 that intersect with single genes, p=0.034, binomial test).

To search for factors that may affect set point expression levels, we assigned each of the 225 zip codes to one of three classes based on their balance of bimodal expression: those with a GFP+ proportion of at least 0.6 in both pools ('mostly GFP+'; 157 clones), those with a GFP+ proportion less than 0.4 in both pools ('mostly GFP-'; 48) and those with mixed levels of GFP expression ('mixed'; 20). Ignoring integrants that intersect with no genes or with genes having overlapping expression in divergent directions, we found no orientation preference for the 'mostly GFP+' integrants (65 of 129 with single intersection have same orientation; p= 0.99 binomial test) (Figure 2.4C), whereas both the mostly GFP- and mixed populations were enriched for integration in the same orientation as gene transcription (30 out of 40; p=0.002 and 15 out of 19; p=0.019). The GFP+ proportion of each integrant had a strong negative correlation with original abundance in the pool (Spearman ρ=-0.289, p=$1.08*10^{-5}$).

We additionally compared the distance of integrants to enhancer associated (H3K27ac) and repressive (H3K9me3) chromatin marks previously determined in Jurkat cell lines (Reeder, Kwak

et al. 2015, Hnisz, Weintraub et al. 2016) (Figure 2.5), again based on their balance of bimodal expression. Distance to H3K27ac peaks had a negative but non-significant correlation to GFP+ proportion (Spearman ρ=-0.105, p=0.118). Distance to existing H3K9me3 repressive marks in Jurkat cells was also negatively correlated with GFP+ proportion (Spearman ρ=-0.195, p=0.0034). Thus, these results conflictingly showed that integrants with higher GFP expression states were on average closer to both existing repressive and enhancer chromatin marks. Comparing the range of values across classes revealed the modest nature of these enrichments (Figure 2.5), with mostly GFP+ and mostly GFP- clones having a significant difference in original clone abundance (p=0.044, Mann Whitney U 2-sided test) and a nearly significant difference in distance to H3K9me3 peaks (p=0.07, Mann Whitney U 2-sided test), while the distance to H3K9me3 peaks was significantly different between the GFP+ and 'mixed' classes (p=0.004, Mann Whitney U 2-sided test).

Discussion

Here, persistence and HIV-1 expression profiles of individual integrant clones were compared within polyclonal populations using "zip coded" proviruses, each tagged to identify the genomic neighborhood where the provirus had integrated. The results revealed a complex array of heritable differences among clones in population sizes and expression characteristics.

Marking libraries with randomized sequence tags has been used in many systems including SIV and HIV-1 (Mei, Nourbakhsh et al. 1997, Chen, Martinez et al. 2017, Fennessey, Pinkevych et al. 2017). One group reported infectious SIV derivatives barcoded to track population dynamics during treatment and rebound (Fennessey, Pinkevych et al. 2017). Unlike those SIV derivatives

(Fennessey, Pinkevych et al. 2017), our vectors lacked Env and (except when remobilized by pseudotyping) were limited to single replication cycles. Barcodes were inserted toward the center of the virus in the SIV work, while ours were inserted near provirus edges to facilitate integration site determination. Another group described barcoded HIV-derived vectors called B-HIVE, with barcodes inserted in HIV-1's multifunctional 5' untranslated region. (Chen, Martinez et al. 2017). We chose to leave the 5' leader region intact because it modulates HIV-1 expression by specifying nucleosome and transcription factor binding (Ne, Palstra et al. 2018), folds into a finely-balanced equilibrium of RNA elements that regulate RNA fates (Bieniasz and Telesnitsky 2018), and is highly sensitive to mutation (Kharytonchyk, Brown et al. 2018). B-HIVE vectors encode LTR-driven GFP but no virus structural proteins. In contrast, our vectors retained *gag* and *pol*, thus allowing progeny virus production and the tracking of both virions and cellular nucleic acids. B-HIVE experiments were performed at a multiplicity of infection of 0.5 and likely included dually infected clones, while we used a much lower MOI. Additionally, we assessed expression in both unsorted cell pools and in serial sub-pools sorted for LTR reporter expression, and observed both dynamic and heritable aspects of clone-specific expression not evident in the B-HIVE work (Chen, Martinez et al. 2017).

We benchmarked our system using a small (74 clones) pilot study that addressed replication fidelity. Zip code abundance varied widely in this pool, as did virus release per cell. Zip code survival rates suggested a single replication cycle lethal mutation rate in transformed cells of about 10%, but the true rate was likely lower than 10% because our vectors included non-viral sequences and the assay design introduced transmission bottlenecks. Most zip codes lost during the second cycle of replication were significantly less abundant in the initially infected cell pool than were

those that persisted for the second round of replication, suggesting that population bottlenecks contributed to zip code extinction. Thus, in contrast to observations that the majority of patients' persistent proviruses are defective even when sampled less than 60 days after initial infection (Bruner, Murray et al. 2016, Pinzone, VanBelzen et al. 2019), the rate of mutational inactivation observed here was in the range predicted by previous work that suggests roughly one in three HIV-1 genomes accumulates any sort of reverse transcriptase-generated mutation per cycle of replication (Menendez-Arias 2009). Thus, the difference between the relatively low rate of mutational inactivation here and the high prevalence of defective proviruses *in vivo* is consistent with the notion that the proviral landscape *in vivo* reflects selective pressures more than reverse transcriptase infidelity (Finzi, Plaeger et al. 2006, Bruner, Murray et al. 2016).

Subsequent experiments were performed in Jurkat cells, using larger zip code libraries and proviruses with GFP in the *nef* open reading frame. In these experiments, we relied on GFP expression as a reporter of LTR activity, and did not assess expression by other means such as quantifying intracellular HIV-1 RNA or measuring virion release. HIV-1 proteins including Vpr and Env, which kill or inhibit cultured cells, were absent by design (Re, Braaten et al. 1995, Costin 2007). Within the unsorted polyclonal Jurkat pool, GFP+ cells were more numerous than GFP- cells and virion release remained robust. As previously demonstrated with similar vectors, populations were readily separable into GFP+ and GFP- pools (Carter, Onafuwa-Nuga et al. 2010, Pace, Agosto et al. 2011, Hakre, Chavez et al. 2012). GFP+ pools displayed high levels of virion release while there was a near-absence of virus from GFP- cells. All abundant zip codes were reproducibly present in both GFP+ and GFP- cell sub-pools, but to widely varying extents. Using "GFP+ proportion" to represent the fraction of each clone's cells that sorted GFP+, most clones

were either “mostly GFP-” (with GFP+ proportions ≤0.4) or “mostly GFP+” (≥0.6). Although most cells in the unsorted pool were GFP+, the average number of cells per mostly GFP- Jurkat clone was significantly greater than for mostly GFP+ clones. This suggests that caution is appropriate when interpreting findings based on latency models that use GFP reporters and that passage cells until GFP activity largely disappears. Specifically, our results suggest that some of the apparent increases in latency over time may reflect outgrowth by clones with low GFP+ proportions rather than proviral silencing (Tyagi and Romerio 2011).

The stability of GFP+ proportions over time was addressed by re-sorting separately passaged GFP+ and GFP- sub-pools. Daughter cells did not always adopt a parental phenotype, but instead “flickered” between GFP+ and GFP-. When overall GFP+ values from the secondary sorts were compared to those from the sort 1 time point, the GFP+ proportions for each clone were remarkably similar over time. It is unclear whether the flickering observed here differs from the intraclonal expression variation described previously within individual retroviral vector cell clones, which was interpreted to indicate integration site-dependent differences in silencing rather than alternating waves of expression (Zentilin, Qin et al. 2000, Jordan, Defechereux et al. 2001).

Heritable high levels of variation among HIV-1 integrant clones have been reported previously. However, unlike the flickering we observed, within-clone HIV-1 expression level variation has appeared relatively narrow using previous approaches (Zentilin, Qin et al. 2000, Jordan, Defechereux et al. 2001). For example, wide inter-clone variation was reported in the B-HIVE study, but HIV-1 expression was quantified as intracellular RNA copies per cell barcode using an unsorted cell pool, and it was assumed that every cell within a given clone expressed LTR-driven

RNAs to the same extent (Chen, Martinez et al. 2017). In contrast, because we determined that GFP positivity and intracellular p24 co-occurred in the polyclonal population, but that clones differed widely in their GFP+ proportions, our results suggest that at least part of the expression differences among clones reflects that each clone consists of a phenotypic mixture of cells—some that release virus and others that do not—in heritable clone-specific proportions.

What is responsible for these clone-specific stable equilibrium mixtures of GFP+ and GFP- cells? Intrinsic fluctuations in transcription factor availability and other stochastic events contribute significantly to gene expression, and can cause genetically identical cells propagated under uniform conditions to display a spectrum of phenotypes (Kaern, Elston et al. 2005). The sources, regulatory mechanisms, and implications of this genetic noise are active areas of investigation (Coulon, Chow et al. 2013, Battich, Stoeger et al. 2015). Phenotypic bifurcation for HIV-1 infected cells, in which intrinsic noise in Tat expression leads to the co-existence within individual integrant clones of some cells that display high levels of expression and others that display essentially none, has previously been described (Weinberger, Burnett et al. 2005). Transcriptional bursting from the HIV-1 promoter is a significant source of stochastic noise (Singh, Razooky et al. 2010), the bursting behavior of the export factor Rev may further exacerbate noise due to Tat (Pocock, Zimdars et al. 2017), and the phase of the cell cycle may also exert influence (Kok, Schmutz et al. 2018). These and other parameters likely contributed to the broad range of GFP+ proportion set points that differentiated clones here, even though our system was carried out in transformed cells with the intention of minimizing extrinsic variability (Swain, Elowitz et al. 2002).

The simplest explanation for why each clone adopted a unique GFP+ proportion set point may be that multiple inputs—some stochastic and others deterministic-- combined in a clone-intrinsic manner to skew the probability that a given cell would reach the Tat threshold needed for GFP expression. The deterministic components could in concept be of either host or viral origin. However, our initial pilot experiment suggests that the principal differences were not within proviral sequences, but instead of host origin. Specifically, the amount of virus release per cell differed among zip codes when all cells with the same zip code were progeny of a single integration event, but virus release per cell was fairly uniform in a second generation when zip codes were polyclonal.

To explore host contributions due to integration site features, virus/host junctions were sequenced, integration sites determined, and the characteristics of mostly GFP+ and mostly GFP- clones compared. The results indicated that mostly GFP+ and mostly GFP- clones differed significantly in proviral orientation relative to host transcription. This may reflect transcriptional repression, which has been reported for HIV-1 (Lewinski, Bisgrove et al. 2005 , Gallastegui, Millan-Zambrano et al. 2011), although one study reported an opposite orientation bias (Han, Lin et al. 2008). We also assessed the correlations between repressive or activating chromatin marks previously determined in Jurkat cells (Reeder, Kwak et al. 2015, Hnisz, Weintraub et al. 2016) and observed modest differences in proximity to H3K9me3 marks. Some previous studies appeared to find more conclusive correlates between epigenetic features and HIV expression (Chen, Martinez et al. 2017) (Lewinski, Bisgrove et al. 2005). The less definitive trends reported here may be due to different approaches in measuring expression (RNA quantification vs. GFP+ proportions) or limited sample sets, and in some cases reflects how significance thresholds were defined. The magnitude of effects

evident here suggest that our understanding of the roles of integration site features to robustly discriminate latency or viral expression remains incomplete.

Speculatively, some component of the observations here may reflect epigenetic marks introduced at the time of integration: due either to stochastic events or to differences in the intracellular environment or architecture of specific integration sites. It is generally assumed that most of the latent reservoir results from the rare infection of activated cells that transition to a memory state. However, HIV-1 can enter cells at any phase of the cell cycle. Histone biogenesis is cell cycle dependent (Ma, Kanakousaki et al. 2015) and many histone post-translational modifications are faithfully introduced onto nascent strands at the time of DNA replication. Although all epigenetic marks appear regenerated within the course of a single cell generation, some marks are copied with the replication fork while others (including H3K9me3 and H3K27me3) are deposited throughout the cell cycle (Alabert, Barth et al. 2015, Reveron-Gomez, Gonzalez-Aguilera et al. 2018). Because HIV can infect dividing or resting T cells, and the cell's chromatin modification machinery displays cell cycle-dependent regulation, it is possible that integration at differing phases of the cell cycle results in distinct patterns of chromatin decoration (Pace, Graf et al. 2012, Chavez, Calvanese et al. 2015, Ma, Kanakousaki et al. 2015).

It seems plausible that the HIV-1 expression variation reported here may cause some of the differences among experimental models for latency (Sherrill-Mix, Lewinski et al. 2013) and that expression flickering and differential set points of expression may be a fairly common outcome during the establishment of polyclonal HIV-1 populations. As such, these properties may contribute to defining the nascent proviral populations within infected people that are subsequently

culled by immune and other selective pressures. Understanding how patterns of expression that persist compare to the palette of outcomes in the absence of selection may aid efforts to identify HIV-1's epigenetic havens, and to the design of fruitful strategies for proviral eradication.

Materials and Methods

Cell line propagation

293T cells were grown from a master cell bank (Yang, Delgado et al. 1999) and Jurkat (Clone E6-1) cells were obtained from ATCC. Both cell lines were maintained as lab frozen stocks and validated at the time of study by tandem repeat analysis using the Applied Biosystems AmpFLSTR™ Identifiler™ Plus PCR Amplification Kit (Thermo Fisher Scientific, Carlsbad, CA). Jurkat cells were cultured in RPMI supplemented with 10% FBS (Gemini), 100 U/mL penicillin, 100 µg/mL streptomycin, 2mM glutamine and 55µM β-mercaptoethanol at 1 x 10^6 cell/ml, while Human Embryonic Kidney (HEK) 293 T cells were grown in DMEM supplemented with 10% FBS (Gemini) and 125 µM gentamycin. Both cell lines were maintained in a 37°C incubator containing 5% CO_2.

Construction of zip coded vectors

All HIV-1 vectors were templated by derivatives of the NL4-3 strain plasmid NL4-3 GPP (Lu, Heng et al. 2011) or by HIV-GPV$^-$, which was derived from the GKO (Battivelli, Dahabieh et al. 2018) provided by Emilie Battivelli and Eric Verdin (University of California San Francisco). HIV-GPV$^-$ was constructed by replacing mKO2 in GKO with the puromycin resistance gene from NL4-3 GPP. After initial work with standard two-LTR vectors, including the pilot fidelity study described here, subsequent zip coded vector preparation used single LTR versions of these vectors.

For this, both vectors were modified into single “inside out” LTR forms containing the 5’ terminal 49 bases of U3 with an engineered Cla I site plus a second unique site (either Xho I or Mlu I) in U3, and inserted into pBR322 as previously described (Kharytonchyk, King et al. 2016). To generate zip coded HIV-1 vector templates, the single LTR plasmid versions of NL4-3 GPP and GPV- were digested with ClaI plus Xho I or Mlu I, respectively. The resulting 11.4kb HIV vector-containing fragments free of plasmid backbone were purified from agarose using QIAquick Gel Extraction Kit (Cat No./ID: 28706 Qiagen, Germantown, MD). A 304 bp zip code-containing insert fragment pool was generated by PCR using NL4-3 GPP or GPV^- as template, Phusion® High-Fidelity DNA Polymerase (New England Biolabs, Inc., Ipswich, MA), and primers 5’-GACAAGATATCCTTGATCTGNNNNNNNNNNNNNNNNNNNGCCATCGATGTGGATC TACCACACACAAGGC-3’ and 5’-CGGTGCCTGATTAATTAAACGCGTGCTCGAGACCTGGAAAAAC-3’ for GPV^- and 5’ GTGTGGTAGATCCACATCGATGGCNNNNNNNNNNNNNNNNNNNCAGATCAAGGA TATCTTGTCTTC-3’ and 5’- ATG CCA CGT AAG CGA AAC TCT CTG GAA GGG CTA ATT CAC TCC-3’ for NL4-3 GPP.

To generate the uncloned vector template library, the 11.4 kb fragments of GPV- or HIV-GPP were joined with their cognate 304 bp zip coded partial U3 inserts via Gibson Assembly in a molar ratio of 1:5 per reaction using HiFi DNA assembly mix (New England Biolabs) following the manufacturer’s protocol. The assembled DNA was then cleaned and concentrated using Zymo Clean and Concentrator-5 kit (SKU D4013 Zymo Research, Irvine, CA), quantified by Nanodrop (Thermo Fisher Scientific), and used directly in transfections.

Virion production

Fresh monolayers of HEK 293T cells, in 10 cm diameter plates and approximately 70% confluent, were co-transfected with 3 μg Gibson Assembly product DNA plus 330 ng pHEF-VSV-G using polyethylenimine (Polysciences, Inc., Warrington, PA) at a ratio of 1 μg total DNA to 4 μg polyethylenimine in 800 μl of 150 mM NaCl (Keene, King et al. 2010). 24 hours post-transfection, DMEM was replaced with 4 ml RPMI1640 medium with 10% FBS and 1% Pen/strep. Culture supernatant was harvested at 48 hours post-transfection and filtered through a 0.22 μm filter (Fisher Scientific. Cat. No. 09-720-511). Released virus was quantified using a real-time reverse-transcription PCR assay and normalized for p24 level based on p24 protein values determined in parallel for reference samples (Kharytonchyk, King et al. 2016). Zip coded virus stocks were titered by infecting 90% confluent HEK 293 T cells and selecting in puromycin. Colony forming units per milliliter of viral media as determined on 293T cells was the standard for defining infectious titer in this work.

Infection of HEK 293 T and Jurkat cells

The media on 10 cm plates of 90% confluent HEK 293 T cells was replaced with 2000 μl infection mix comprised of the indicated amount of virus-containing medium plus additional DMEM in 1 μg/ml polybrene, then incubated at 37 °C with 5% CO_2 for 5 hours. After incubation, the infection mix was replaced with 10 ml of fresh media. Twenty-four hours post-infection, cells were placed in media containing puromycin at a concentration of 1μg/ml, which was replaced every three days for 2 weeks. Following this, colonies were individually cloned, pooled together for subsequent experiments, or stained with crystal violet and counted.

For Jurkat cell infections, virus-containing media and polybrene at a final concentration of 0.5 μg/ml were brought to a total volume of 1000 μl. This infection mixture was added to 1.5 x 10^6 Jurkat cells and incubated in one well of a 12 well tissue culture plate (Fisher Scientific, Cat. 150628) at 37 °C with 5% CO_2 for 5 hours. Infected cells were then transferred to Eppendorf tubes and centrifuged for 5 minutes at 2500 rpm at 4°C. Following centrifugation, supernatants were replaced with fresh media and cell pellets were resuspended and cultured at 37 °C with 5% CO_2. At 24 hours post-infection, puromycin was added to a final concentration of 0.5 μg/ml. The infected cells were expanded into 6 cm culture plates without puromycin on day 5. Ten days post-infection, the culture supernatant was replaced with fresh media and the cultures were divided into aliquots, to be either frozen or further expanded for subsequent experiments.

Primary T cell isolation and infection

Peripheral blood mononuclear cells (PBMCs) were isolated from fresh human blood from healthy donors provided by the Department of Pathology at the University of Michigan using Ficoll Histopaque as described earlier (Kim, Zhu et al. 2018). All use of human samples was approved by the Institutional Review Board at the University of Michigan. Total CD4+ T cells were then purified from PBMCs. Because selective drugs were not applied in these primary cell experiments, the GFP- sub-pools included uninfected cells. Thus, the fraction of infected primary cells that were GFP+ was not experimentally determined and a 5% GFP+ value was used in calculations. 10% GFP+ is on the upper edge of previously reported values for primary cells infected with similar vectors, with values less than 3% more typical, possibly reflecting donor-dependent variation or survival of some non-transduced cells (Martins, Bonczkowski et al. 2016). Thus a 5% value was used as a conservative measure, to spread data points that would have appeared similar if a value

<3% were used. Importantly, note that although absolute values would change if true GFP- value were higher or lower than this assumed value, correlation values and their interpretation would not be affected. MACS beads (Miltenyi Biotec Bergisch Gladbach, Germany) as per the manufacturer's instructions. On day 0, a total of 5 x 10^6 cells were seeded in complete culture medium composed of RPMI supplemented with 10% FBS, 100 U/mL penicillin, 100 µg/mL streptomycin, 2mM glutamine and 55µM β-mercaptoethanol at 1 x 10^6 cell/ml. The cells were stimulated using plate-bound anti-CD3 (5 µg/mL; eBioscience, Thermo Fisher Scientific) and soluble anti-CD28 (1 µg/mL; eBioscience, Thermo Fisher Scientific) antibodies in the presence of 50 U/ml IL-2 (PeproTech, Inc., Rocky Hill, NJ). On day 2 of activation, the cells were infected by spinoculation at 2500 rpm for 90 minutes at 37^{o}C with 125 µL zip coded viral media and 0.4 µg/ml polybrene (Sigma Aldrich, St. Louis, MO) in 2.5 ml of supplemented RPMI. After spinoculation, media containing virus was replaced with fresh supplemented RPMI and cells were cultured further and expanded as needed. On day 6 or 7 post-activation, cells were harvested and sorted into GFP^+ and GFP^- sub-pools by flow cytometry using FACS Aria II (BD Biosciences, Franklin Lakes, NJ) or iCyt Synergy SY3200 (Sony Biotechnology, San Jose, CA) cell sorter.

Flow cytometry

For flow cytometry analysis and sorting, Jurkat or primary T cells were suspended in phosphate buffered saline (PBS) containing 1% FBS (FACS buffer). Dead cells were excluded in all analyses and sorting experiments using propidium iodide (PI). Intracellular Gag staining was carried out using a Gag monoclonal antibody conjugated to Phycoerythrin (KC-57 RD1 Beckman Coulter). 1x10^5 cells from a HIV GPV- zip coded library were washed once with FACS buffer and fixed with 100 µl of BD cytofix for 10 minutes at room temperature in the dark. Cells were then washed

twice with FACS buffer then once with BD perm/wash buffer. Staining was carried out at a 1:200 dilution of antibody in 1x BD perm buffer. The cells were incubated in the dark at room temperature for 15 minutes, washed twice, then resuspended in 200 µl FACS buffer. Acquisition was carried out on the FITC channel for GFP and PE channel for Gag. Cell fluorescence was assessed using FACSCanto II (BD Biosciences) and data were analyzed using FlowJo software, version 9.9 (FlowJo, LLC., Ashland, Oregon).

PCR amplification of zip codes from zip coded cells and virus

Genomic DNA was extracted from zip coded cell libraries using Qiagen DNeasy Blood & Tissue Kit (Qiagen, Germantown, MD). Zip codes were amplified from 100 ng of genomic DNA using primers flanking the zip code region (primers: 5‘-NNACGAAGACAAGATATCCTTGATC-3’ and 5’-NNTGTGTGGTAGATCCACATCG-3’) using Phusion® High-Fidelity DNA Polymerase (New England Biolabs) in HF Buffer. For zip code amplification, we designed multiple primers complementary to the template binding site that included two known, random nucleotides at the 5’ end for use in separate reactions. By comparing the primers used for amplification and the nucleotides at the end of each amplicon, we could confirm that PCR cross contamination had not occurred. Reactions were cycled 26-35 times with 30 second extension at 72^0 and a 59^0 annealing temperature. Zip coded amplicons were purified with DNA Clean & Concentrator-5 (Zymo Research, CA. Cat. No. D4013) and eluted in 20µl of H_2O. To amplify zip codes from virus, virus-containing media was filtered through a 0.22 µm filter, concentrated by ultracentrifugation at 25,000 rpm through a 20% sucrose cushion, and RNA extracted with Invitrogen TRIzol Reagent (Thermo Fisher Scientific). The dissolved RNA was treated with RQ1 DNase (Promega, Fitchburg, WI) to remove possible DNA traces, re-extracted with phenol-chloroform, and stored

at −80 °C. cDNA was synthesized using M-MLV RT (H–) (Promega) and U3 antisense primer 5'-TGTGTGGTAGATCCACATCG-3'. Zip codes were amplified from this cDNA using conditions outlined above.

For library construction, protocols and reagents from NEBNext® Ultra™ DNA Library Prep Kit for Illumina® (New England Biolabs) were used for end repair, dA-tailing, and to ligate Nextflex adapters (Perkin Elmer, Waltham, MA) onto amplicons. After ligation, reactions were diluted up to 100 µl with H_2O, purified with 0.85x SPRIselect beads, washed twice in 70% ethanol, and eluted into H_2O. PCR enrichment of adapter-ligated amplicons was done for 7 cycles using NEBNext® Ultra™ DNA Library Prep Kit, reactions were diluted up to 100 µl with H_2O, and purified with 0.85x SPRIselect beads (Beckman Coulter) as outlined above. Libraries were quantitated with KAPA Library Quantification Kits for Next-Generation Sequencing (Roche Sequencing Solutions, Inc., Pleasanton, CA) and Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific), pooled equally, and sequenced with MiSeq Reagent Kit v3, 150 cycle PE on MiSeq sequencer (Illumina, San Diego, CA).

Calculating GFP+ proportions

GFP+ proportions were calculated by dividing the read frequency of each zip code within GFP+ sorted cells by the summed abundance of the zip code in both GFP+ and GFP- sorted cells, after weighting values to reflect the fractions of total cells that sorted into GFP+ and GFP- sub-pools. For example, GFP+ read frequency of a clone would be the proportion of GFP+ total reads that contained that zip code. If the total pool was 75% GFP+ and 25% GFP- cells and a given zip code were 2% of the GFP+ cells and 3% of the GFP- cells, the GFP+ proportion of that clone would be (2% of 75%) / (2% of 75% + 3% of 25%) = 67%.

HIV integration-site sequencing

Template for hemi-specific ligation mediated PCR of insertion sites was obtained by linear PCR and biotin enrichment of sheared, genomic DNA with linkers ligated on each end. Linker was synthesized by mixing oligo 5'–GTAATACGACTCACTATAGGGCTCCGCTTAAGGGACT-3' and 5'–PO4- GTCCCTTAAGCGGAG-3'-C6 (Maldarelli, Wu et al. 2014) at a final concentration of 40 μM each in 100 μl volume. Oligo mixture was heated in a PCR block for 5 minutes at 95ºC, th PCR machine was immediately shut off, and the block was allowed to cool for 2 hours to room temperature. Genomic DNA was extracted from cells using Qiagen DNeasy Blood & Tissue kit (Qiagen) and 200 ng of DNA was sheared to 1 kb fragments using Covaris M220 and micro-TUBE according to manufacturer's recommended settings (Covaris, Woburn, MA). Sheared DNA was purified with 1x SPRIselect beads according to manufacturer's instructions (Beckman Coulter) and sheared ends were repaired with NEBNext® Ultra™ End Repair/dA-Tailing Module (New England Biolabs) according to manufacturer's protocol. Repaired, dA-tailed DNA was purified with 0.7x SPRIselect beads (Beckman Coulter) and the partially double stranded DNA linker with dT overhang was ligated in a 60 μl reaction containing 6ul of 10X T4 DNA Ligase Buffer, 1.33 μM linker DNA, and 3600U Ultrapure T4 DNA ligase (Qiagen) at 16ºC for 16 hours followed by 70ºC incubation for 10 minutes. Ligated DNA was purified with 0.7 x SPRIselect beads (Beckman Coulter) and used for template in linear PCR reaction containing 1x Expand Long Range Buffer, 500 μM dNTPs, 3% DMSO, 3.5U Long Range Enzyme Mix, and a 500 μM biotinylated primer that anneals to the HIV LTR in our construct, 5'- /52-Bio/CAAAGGTCAGTGGATATCTGACCCC-3'. Cycling parameters were 95^0 C for 5 minutes, 40 cycles of 95^0 C for 45 seconds, 60^0 C for 1 minute, and 68^0 C for 1.5 minutes, followed by a 10

minute incubation at 68^0 C. PCR product was purified with 1x SPRIselect beads (Beckman Coulter), resuspended in 20 µl H2O, and biotinylated fragments were captured using Dynabeads kilobase BINDER kit (Thermo Fisher Scientific) according to manufacturer's instructions. DNA captured by beads was used as template in a hemi-specific PCR reaction containing 1x Expand Long Range Buffer, 500 µM dNTPs, 3% DMSO, 3.5U Long Range Enzyme Mix, 500 µM of a nested primer that anneals to HIV LTR in our construct, 5'-GCCAATCAGGGAAGTAGCCTTGTGTGTGG-3', and 500 µM of a primer that anneals to the linker, 5'-AGGGCTCCGCTTAAGGGAC-3'. Cycling parameters were 95^0 C for 5 minutes, 30 cycles of 95^0 C for 45 seconds, 60^0 C for 1 minute, and 68^0 C for 1.5 minutes, followed by 10 minutes' incubation at 68^0 C. PCR product was purified with 0.7x SPRIselect beads (Beckman Coulter), then protocol and reagents from NEBNext® Ultra™ DNA Library Prep Kit for Illumina (New England Biolabs) were used to end repair, dA-tail, and ligate Nextflex sequencing adapters (Perkin Elmer) onto amplicons. Ligation reaction was purified with 0.65x SPRIselect beads (Beckman Coulter) and 7 cycles of PCR to enrich for ligated product was done with NEBNext® Ultra™ DNA Library Prep Kit for Illumina (New England Biolabs). Libraries were quantitated with KAPA Library Quantification Kits for Next-Generation Sequencing (Roche Sequencing Solutions, Inc., Pleasanton, CA) and Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific), pooled equally, and sequenced with MiSeq Reagent Kit v3, 600 cycle PE on MiSeq sequencer (Illumina, San Diego, CA). All generated sequence data has been deposited to the Sequence Read Archive (SRA) under project accession PRJNA531502

Zip code analysis and quantification

Zip codes were identified and quantified from Illumina sequencing reads using a custom suite of tools implemented in Python (https://github.com/KiddLab/hiv-zipcode-tools). First, 2x75 bp paired reads were merged together using *flash* v1.2.11 (Magoc and Salzberg 2011). Zip codes were identified by searching for known flanking sequence (with up to 1 mismatch). Only candidate zip codes with a length of 17-23 nucleotides were considered and the read count for each unique zip code was tabulated. To identify the set of zip codes for further analysis, zip code families which account for PCR and sequencing errors were determined by clustering together the observed unique zip codes. Comparisons among zip codes were calculated using a full Needleman-Wunch alignment tabulated with a score of +1 for sequence matches, -1 for mismatches, and a constant gap score of -1. Comparisons with two of fewer mismatches (counting a gap as a mismatch) were accepted as a match. Using this criteria clusters were then identified. First, unique zip codes were sorted by abundance. Then, beginning with the most abundant zip code, each sequence was compared with all of the previous zip codes. If no previous zip code had two or fewer mismatches that zip code was accepted as a cluster and then the next most abundant zip code was considered. This process was continued until the first unique zip code having a match to a more abundant zip code was identified. This defined the set of families for consideration. Abundance for the families was then determined by assigning unique zip codes to the most abundant family whose sequence was within 2 mismatches and summing their associated read counts.

In sorting experiments, the GFP+ proportion for each zip code was determined as $F_i = (G_i * P)/(G_i * P + W_i * Q)$ where F_i is the GFP+ fraction of zip code i, G_i is the fraction abundance of zip code i in the GFP+ sorted pool, W_i is the fraction abundance of zip code i in the GFP- sorted pool, P is the fraction of cells that sorted into the GFP+ pool and Q is the fraction of cells that sorted

into the GFP- pool. In the Jurkat pool 1, the initial GFP+ fraction was 0.524 and the initial GFP- fraction was 0.36. Of the GFP+ sort from pool 1 the GFP+ fraction was 0.887 and the GFP- fraction was 0.079 GFP- while in the GFP- sort from pool 1 the GFP+ fraction was 0.046 and the GFP- fraction was 0.928. In the Jurkat pool 2, the initial GFP+ fraction was 0.518 and the initial GFP- fraction was 0.364. Of the GFP+ sort from pool 2 the GFP+ fraction was 0.915 and the GFP- fraction was 0.082 GFP- while in the GFP- sort from pool 2 the GFP+ fraction was 0.063 and the GFP- fraction was 0.923. For primary cell data analysis, the abundance of each zip codes in the GFP+ and GFP- pools summed, and only those zip codes with summed abundance greater than 0.0001 in both replicates were considered, and a GFP+ fraction of 0.95 and a GFP- fraction of 0.05 were assumed.

Analysis of integration sites occurred in two stages. First, read-pairs were analyzed to identify which read derived from the LTR sequence and which from the genomic linker. Zip code sequences were extracted from the LTR-derived read based on matches to flanking sequence in the vector as described above. The linker sequence and LTR sequence flanking the zip code were removed and the extracted zip code sequence was then associated with the remaining portion of each read pair. Second, the trimmed read pairs were aligned to a version of the hg19 genome that included the sequence of the utilized HIV vector using bwa mem version 0.7.15. The resulting alignments were then parsed to identify the shear point (DNA adjacent to where the linker was ligated) and integration point (the DNA location adjacent to the LTR sequence). The zip codes were then assigned to previously identified zip code families, and the number of unique shear points and total reads supporting a integration site for each zip code were tabulated. Only reads with a mapping quality greater than 10 were considered, and sonication breakpoints that appear

within 3 nucleotides of one another were considered to represent the same shear point (Maldarelli, Wu et al. 2014). A greedy algorithm was then used to associate each zip code with a genomic location, to remove minor assignments presumed to be artifactual chimeras generated during PCR. "Greedy strategy" is a term from computer science that refers to an algorithm which solves a multi-part problem by dividing the problem into separate states or pieces and then selecting the outcome that maximizes an indicated criteria at each stage (Corman, Leiserson et al. 2009). We assigned zip codes to genome locations based on the number of supporting fragments. First, we assign the zip code with the largest number of fragments to the location supported by the most fragments. Next, other fragments associated with that zip code are removed from consideration. This process is then repeated for the remaining zip code with the largest number of supporting fragments.

Determination of chromatin marks and expressed genes

Gene annotations were determined based on Ensembl release 75. Jurkat gene expression data produced by Encode (Consortium 2012) was used (accession ENCSR000BXX), and genes with TPM counts greater than 5 in both replicated were considered to be expressed. H3K27ac peaks were identified using data from (Hnisz, Weintraub et al. 2016) (GSM1697880 and GSM1697882). Chip-seq and control data were aligned to hg19 using bwa mem and peaks were identified using macs2 v 2.1.0 (Zhang, Liu et al. 2008) with the --nomodel option. For H3K9me3 peaks, data from (Reeder, Kwak et al. 2015) (GSM1603227) were aligned to hg19 using bwa mem and processed using macs2 without a control sequence set. For both marks a p value cutoff of $1*10^{-9}$ was used.

Ethics statement

Peripheral blood mononuclear cells (PBMCs) were isolated from fresh human blood from healthy donors provided by the Department of Pathology at the University of Michigan. All samples were anonymized and all use of human samples was approved by the Institutional Review Board at the University of Michigan.

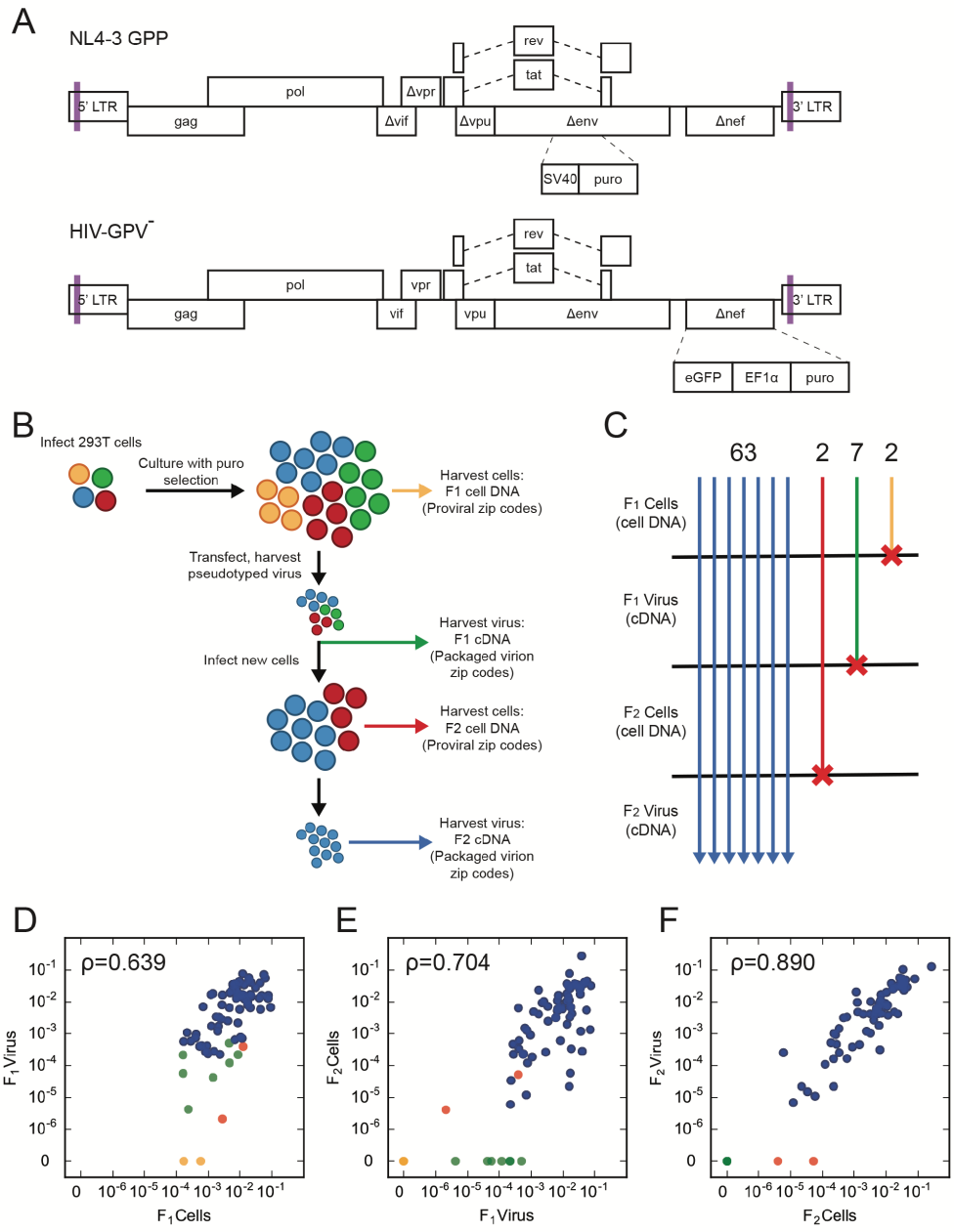


Figure 2. 1. Monitoring proviral replication competence across generations. (A) Schematic illustrations of the vectors used in this paper. Lavender bars represent the sites of randomized sequence insertions. Features and construction are described in Materials and Methods. (B) Schematic of the experimental flow of the replication competence experiment, depicting the analysis of genomic DNA and viral cDNA harvested from the F1 and F2 generations. Each color represents members of a clone: on initial integration, one cell (represented by a circle) per color. The loss of colors over the course of the experiment represents predicted outcomes of mutational inactivation. (C) Summary of the number of independent zip codes detectable at different stages of the experiment. A total of 63 zip codes were detected in all four pools. The number of clones present at the indicated stage but not later are indicated at the top. (D, E, F) Scatter plots of zip code read proportions across indicated stages of the experiment, as outlined in (B). Each clone is represented by a single point, colored to reflect that clone's persistence based upon the progression pattern depicted in (C). The Spearman correlation for each comparison is given.

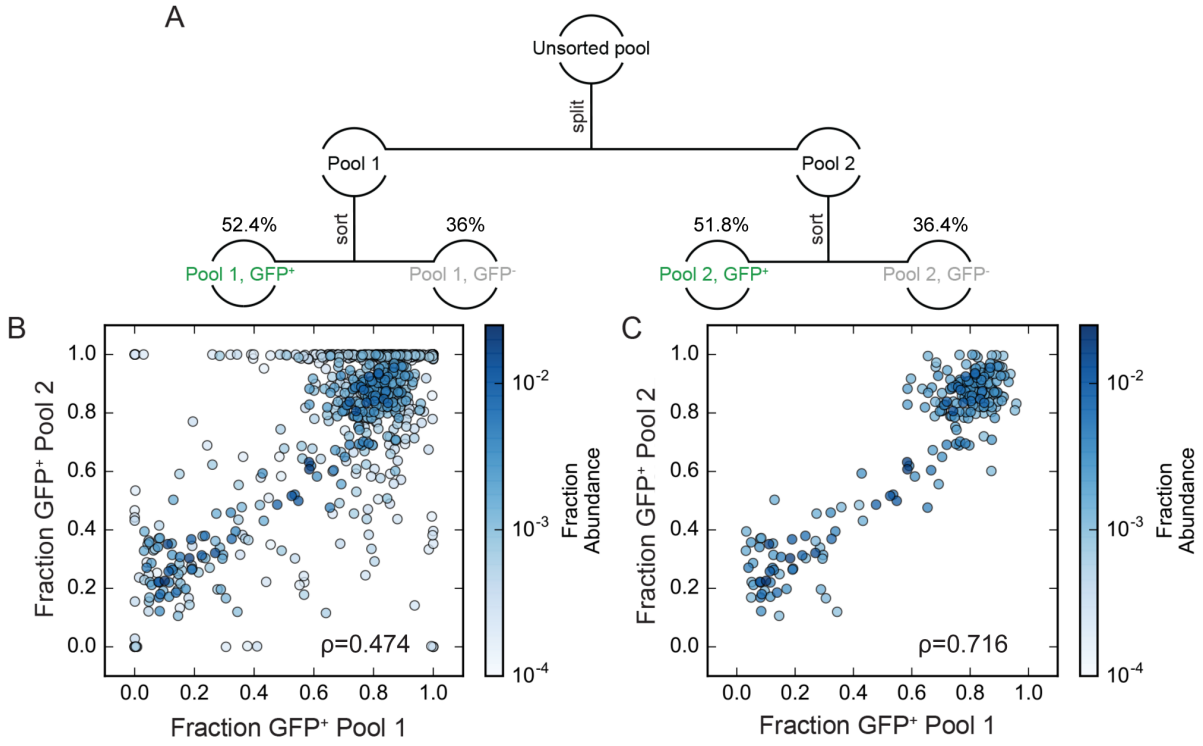


Figure 2. 2. GFP+ proportions for independent clonal lines within a complex population. (A) Schematic description of the cell pool splitting and sorting procedures performed. GFP+ proportions were determined as described in the text. (B) comparison of fraction GFP+ determined for each zip code in Pool 1 and Pool 2. Each point represents a single zip coded cell clone. Individual clones are colored based on their fractional abundance in the original unsorted pool as indicated by the color bar on the right side of the panel. (C) as in (B), but with data for the less abundant clones removed to show only the 225 zip codes with fractional abundance > 0.001.

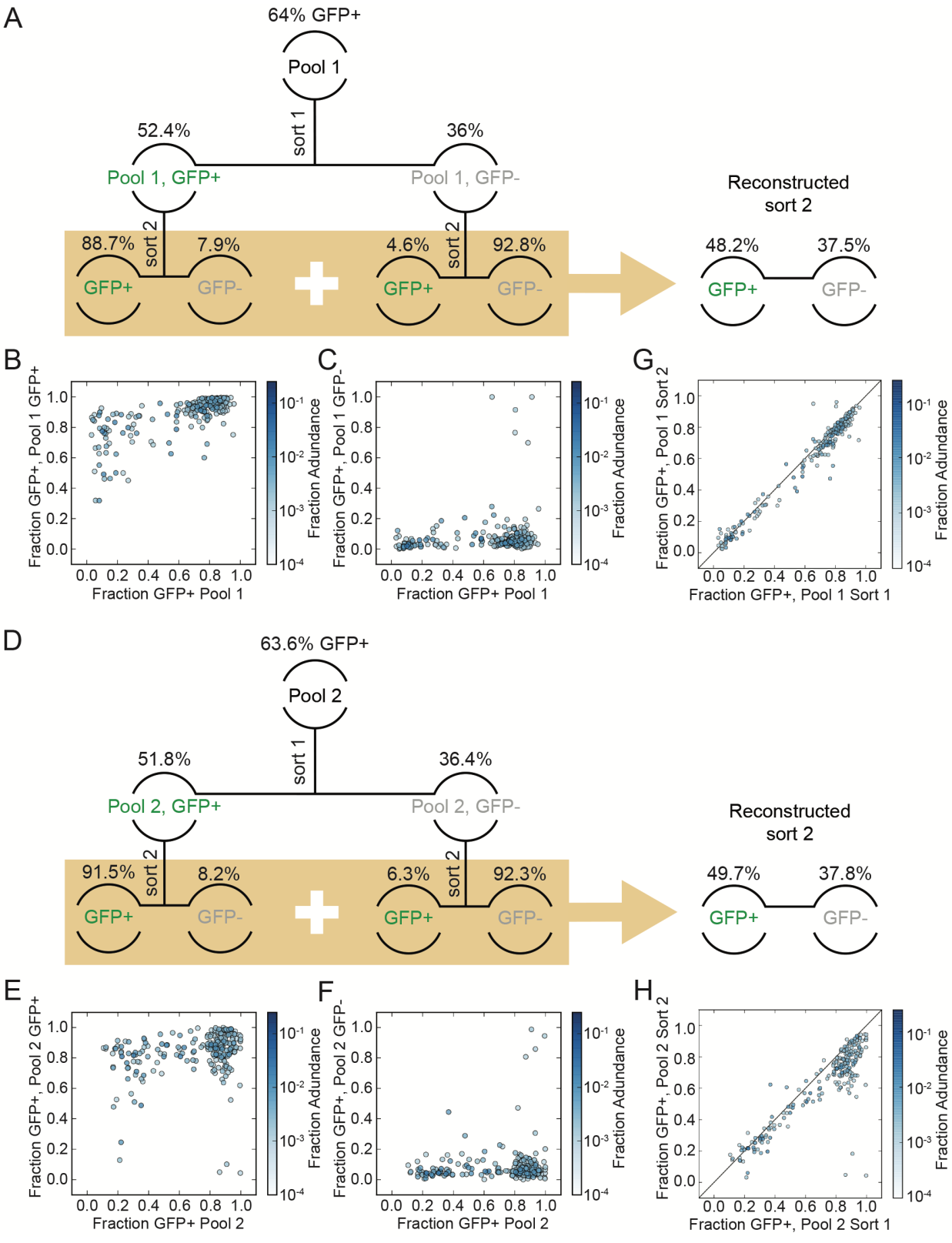

A
64% GFP+
Pool 1
sort 1
52.4%
36%
Pool 1, GFP+
Pool 1, GFP-
sort 2
sort 2
88.7%
7.9%
4.6%
92.8%
GFP+
GFP-
GFP+
GFP-
Reconstructed sort 2
48.2%
37.5%
GFP+
GFP-
B
C
G
Fraction GFP+, Pool 1 GFP+
Fraction GFP+, Pool 1 GFP-
Fraction GFP+, Pool 1 Sort 2
Fraction GFP+ Pool 1
Fraction GFP+ Pool 1
Fraction GFP+, Pool 1 Sort 1
Fraction Adundance
D
63.6% GFP+
Pool 2
sort 1
51.8%
36.4%
Pool 2, GFP+
Pool 2, GFP-
sort 2
sort 2
91.5%
8.2%
6.3%
92.3%
GFP+
GFP-
GFP+
GFP-
Reconstructed sort 2
49.7%
37.8%
GFP+
GFP-
E
F
H
Fraction GFP+, Pool 2 GFP+
Fraction GFP+, Pool 2 GFP-
Fraction GFP+, Pool 2 Sort 2
Fraction GFP+ Pool 2
Fraction GFP+ Pool 2
Fraction GFP+, Pool 2 Sort 1
Fraction Adundance

Figure 2. 3. GFP+ proportions of passaged and re-sorted GFP+ and GFP- cell pools. (A) Depiction of the cells' passaging and sorting scheme, with the initial sorted pools characterized in Figure 2.2 at the top, followed by the re-sorted sub-pools analyzed here. The % GFP+ above Pool 1 or Pool 2 in (A) or (B) respectively represents the % GFP+ cells in that Pool prior to sorting. The percentages listed for the sorted cells below that indicate the proportion of the unsorted pool that were sorted into the indicated, after gating as described in S4 Fig. (B) Analysis of zip codes that sorted GFP+ in Pool 1. Y axis indicates the GFP+ proportions determined in the second sort (within beige shaded box; this was a re-sorting of the sub-pool that had sorted GFP+ in the first sort and had been passaged separately for > I week) and X axis is GFP+ proportions from Pool I first sort (eg: Figure 2.2C X axis) (C) Analysis of zip codes that sorted GFP- in Pool 1. Y axis indicates the GFP+ proportions determined in the second sort (within beige shaded box this was a re-sorting of the sub-pool that had sorted GFP- in the first sort and had been passaged separately for > I week) and X axis is GFP+ proportions from Pool I first sort (eg: Figure 2.2C X axis) (D, E, F) Analysis performed as in Figure 2.3A, 3B, and 3C, for zip codes in Pool 2. (G) Stability of GFP+ proportions in Pool 1 over time. GFP+ proportions determined in the first sort (Figure 2.3 data) plotted against reconstructed, as assessed by comparing GFP+ proportions for each zip code derived from data in Pool 1 at the second sort (Y axis, data from panels B and C) vs the first sort (Figure 2.2C X axis). Second sort GFP+ proportions were reconstructed by weighting the GFP+ and GFP- sub-pool values determined in panels A and B as described in Materials and Methods. (H) Stability of GFP+ proportions in Pool 2 over time, performed as described in panel G for Pool 1.

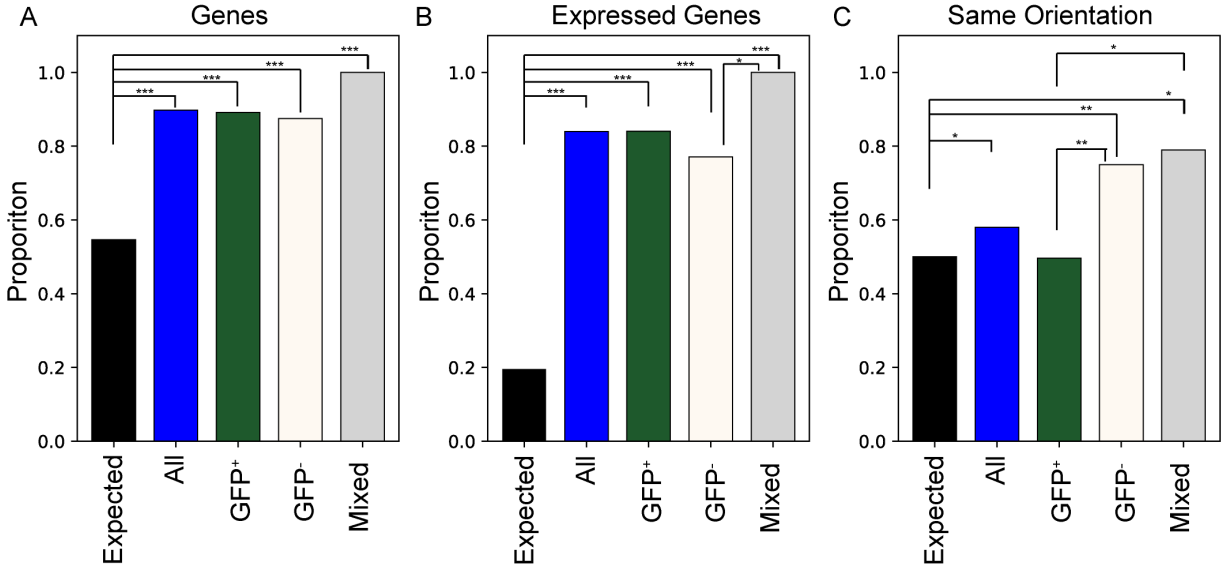


*Figure 2. 4. Integration site features. Integration site properties are shown for each zip code. In each panel, the 'expected' bar shows the proportions that would arise if integration occurred uniformly at random positions throughout the genome, 'all' represents proportions for all examined zip coded integrants, and 'GFP+', 'GFP-', and 'Mixed' represent zip codes assigned to the mostly GFP+, mostly GFP-, or mixed, as described in the text. (A) Comparison of proportions of each category of integrants that resides in annotated genes. (B) Proportions within genes annotated as expressed in Jurkat cells (Serrao and Engelman 2016) (C) Comparison of proportions of each category of integrants that resides in the same orientation as gene transcription. Statistically significant pair-wise differences are indicated by bracket lines and an asterisk symbol with * indicating $p<0.05$, ** $p< 0.01$, and *** $p< 0.001$. Results for (A) and (B) were determined by permutation while results for (C) are based on a binomial test. Nominal p-values are indicated without correction for the number of tests performed.*

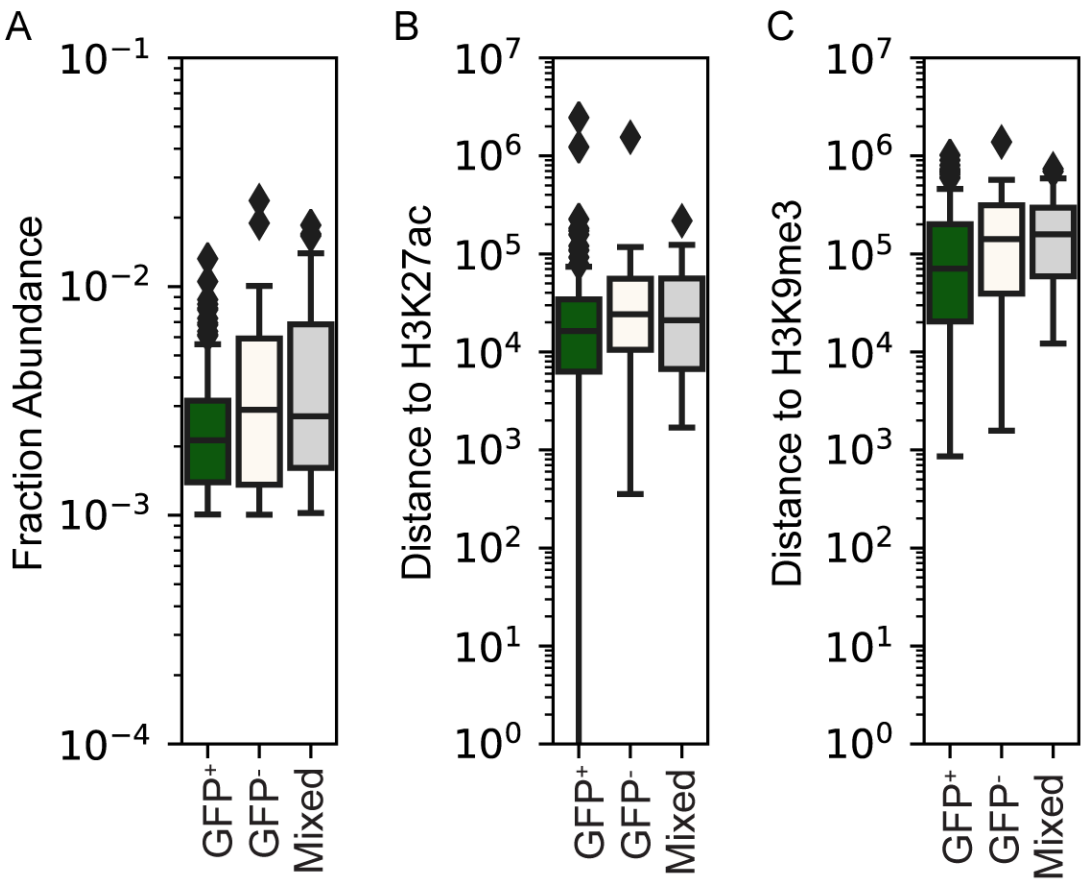


*Figure 2. 5. Correlations between GFP+ proportions and mapped epigenetic features. Each of the 225 zip codes were binned into one of three categories (mostly GFP+, mostly GFP-, or mixed, and as described in the text). (A) box plots show the fractional abundance of each zip code residing in that category of clones, as determined in the unsorted Jurkat pool (Figure 2.2 data). (B) distances to H3k27ac and (C) H3k9me3 peaks, respectively, for the mostly GFP+, mostly GFP-, and mixed expression pattern zip codes. For each box plot the median and interquartile range is depicted. Pairwise comparisons with significant differences based on a Mann-Whitney U two-sided test are indicated, *= p<0.5, ** = p<0.01. Nominal p-values are indicated without correction for the number of tests performed.*

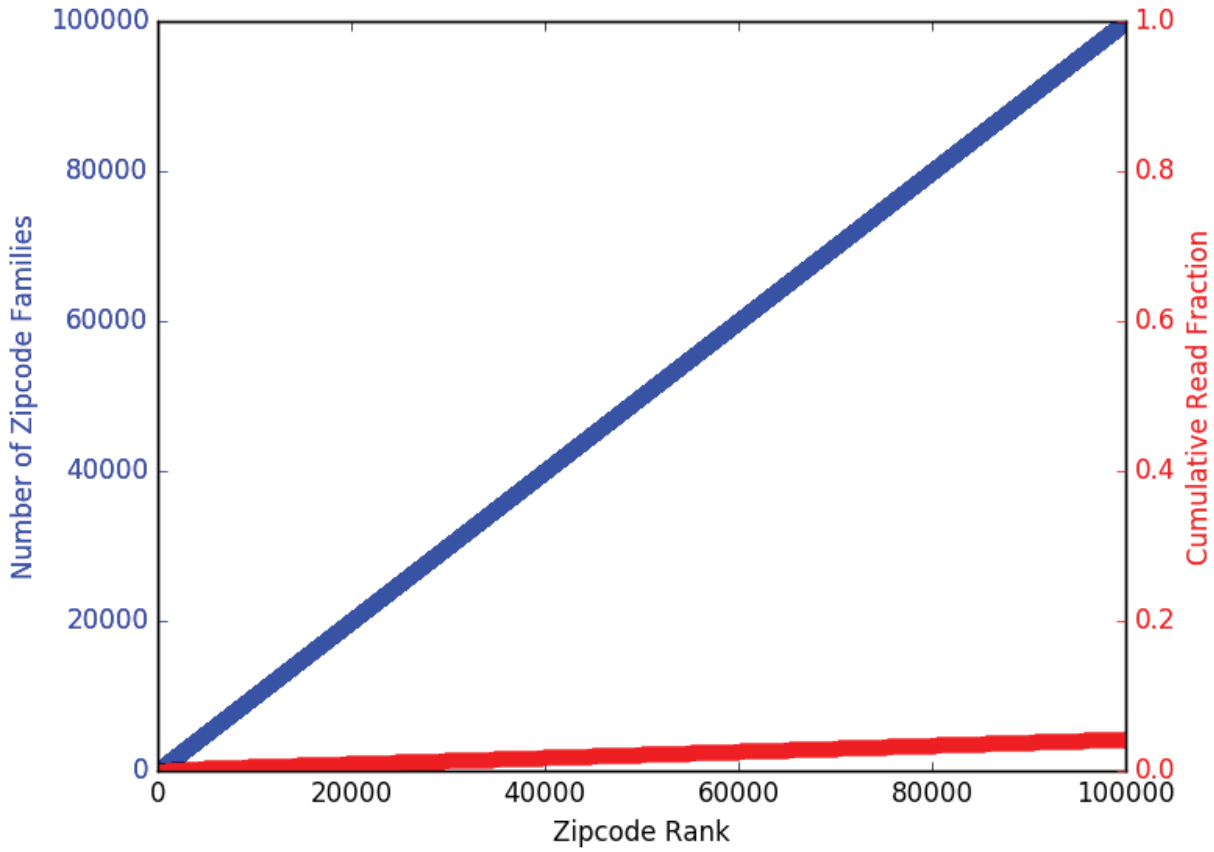


Figure S2. 1. Zip code complexity in Gibson assembly mix used to generate zip coded virion RNAs. A zip code amplicon was made from 1% of the Gibson assembly mix used in transfections to generate zip coded virus. The amplicon was high throughput sequenced and zip codes were clustered into zip code families. Of 6.23 million sequencing reads, the plot shows ~4% of the reads (right axis: red) contained 100,000 zip codes (left axis: blue). Zip code rank refers to the order of zip coudes, sorted by read abundance.

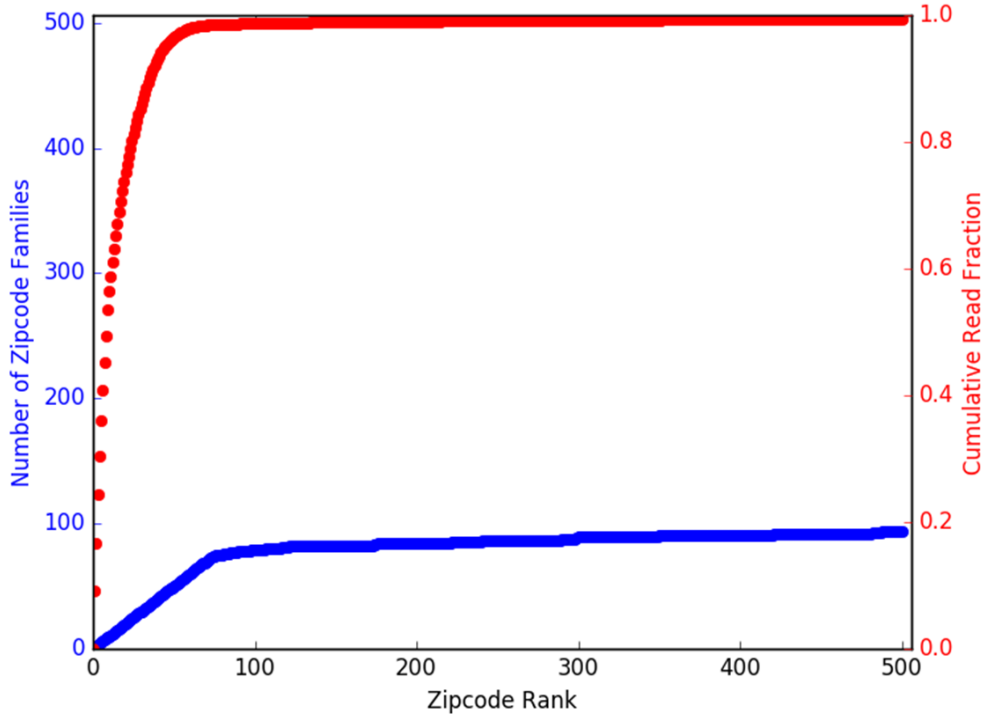
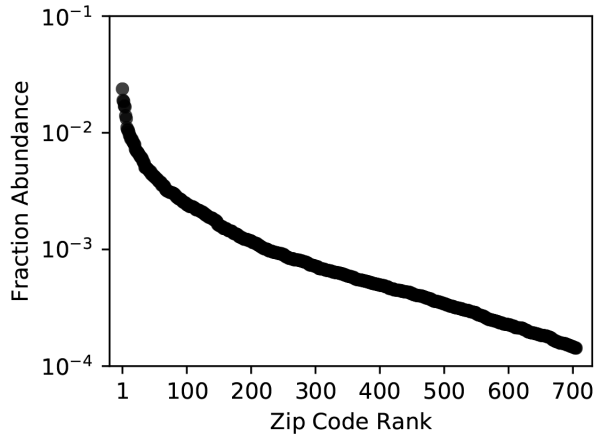
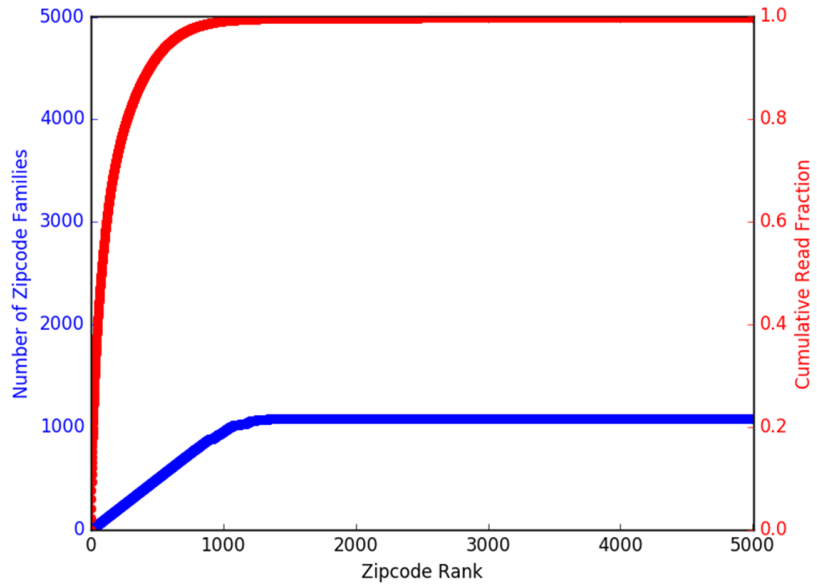


Figure S2. 2. Zip code family and read abundance for single cycle pilot experiment. The red line (right axis) shows the cumulative fraction of reads accounted for by each unique zip code. The blue line (left axis) shows the number of unique zip code families determined after clustering the indicated number of unique zip codes. The inflection point on the blue line indicated that the zip codes clustered into 74 families.

Figure S2. 3. Analysis of Jurkat cell pool high throughput sequencing reads and assignment of zip code families.



Zip code fractional abundance. Each of 706 zip code families identified in the Jurkat pool is depicted by a single point. The clones are arrayed left to right from the most abundant to the least abundant, with the fractional abundance of total reads assigned to that zip code on the Y axis.



Zip code rank and fractional abundance for Jurkat pool. The red line (right axis) show the cumulative fraction of reads accounted for by each unique zip code. The blue line (left axis) shows the number of zip code families determined by clustering the indicated number of unique zip codes.

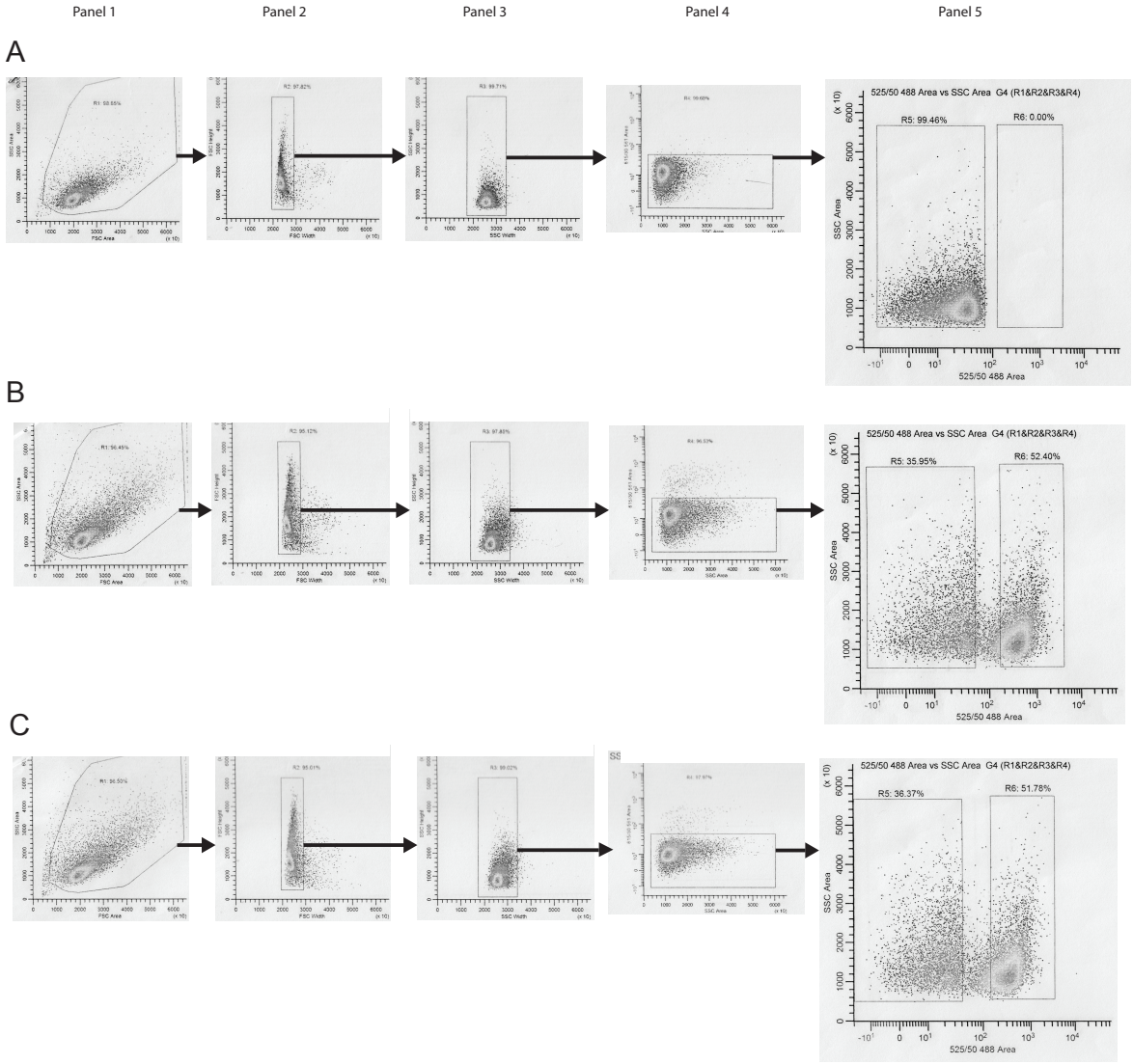


Figure S2. 4. Gating of GFP+ and GFP- subpopulations for sorting. Prior to sorting, cells were stained with propidium iodide. (A) Uninfected Jurkat cells were gated based on FSC-Area and SSC-A to gate out cellular debris (panel 1), and based on FSC and SSC widths and heights to exclude doublets (panels 2 and 3). Next, the propidium iodide positive cells were gated out in the PE channel to exclude dead cells (panel 4) and GFP- and GFP+ gates were drawn in the FITC channel as shown panel 5. These gates were then applied to (B) Pool 1, and (C) Pool 2 and sorted into GFP+ and GFP-.

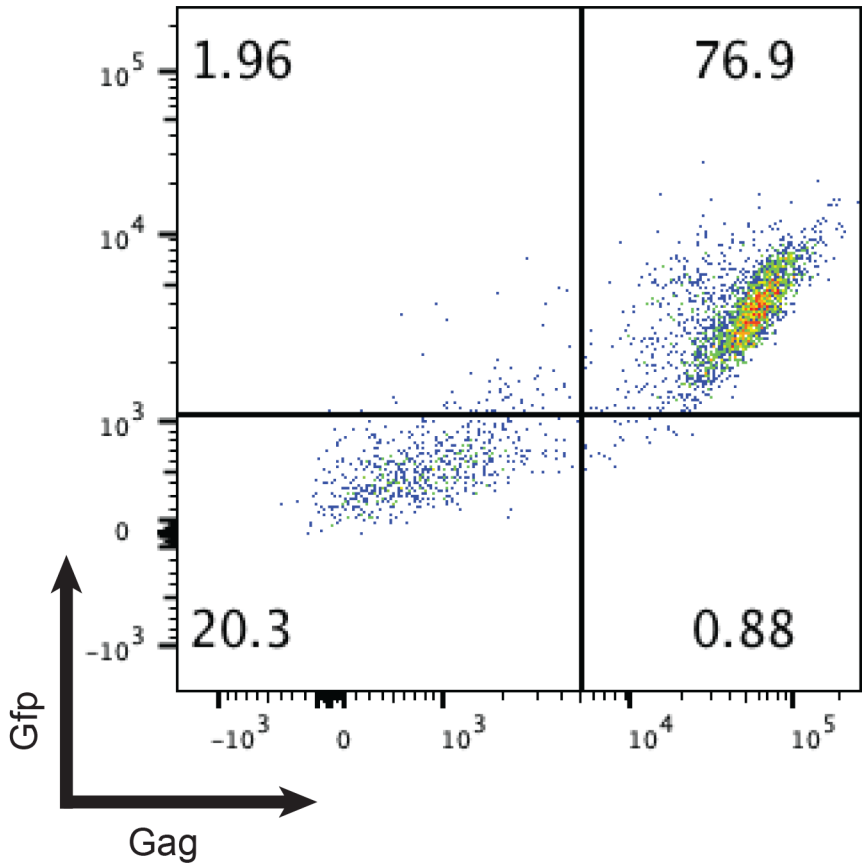


Figure S2. 5. ***Flow cytometric analysis for the co-occurrence of intracellular Gag staining and GFP***

Performed using Jurkat cells containing zip coded HIV GPV- library as described in Materials and Methods. Numbers in each quadrant indicate the proportion of total cells in that quadrant.

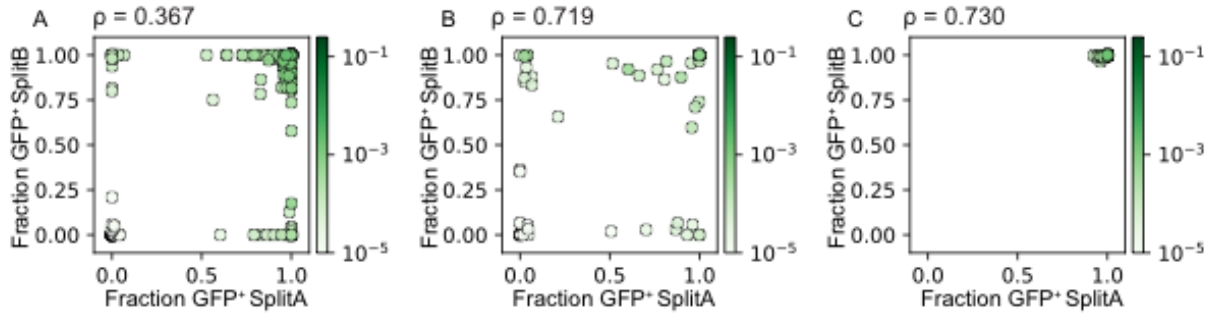


Figure S2. 6. ***GFP+ fractions in primary cells.*** *Scatter plots of GFP+ proportions for three primary cell experiments. The percent GFP+ for each zip code was calculated assuming 95% of cells were GFP+. Inclusion criteria were selected to identify zip codes sufficiently abundant following the limited cellular divisions in the passaged primary cells. For each zip code plotted, the abundance fraction determined in summed GFP+ and GFP- sorts was required to be >0.0001.* ***(A)*** *Scatter plot of GFP+ proportions for experiment 1. A total of 349 zip codes passed the inclusion criteria, and show a Spearman correlation of ρ=0.367 ($p=6.45x10^{-12}$) among SplitA and SplitB replicates.* ***(B)*** *GFP+ proportions from experiment 2. 73 zip codes passed the inclusion criteria, with a Spearman correlation of ρ=0.719 ($p=7.65x10^{-13}$).* ***(C)*** *GFP+ proportions from experiment 3. 90 zip codes passed the inclusion criteria, with a Spearman correlation of ρ=0.730 ($p=3.22x10^{-16}$). In each case, points are colored based on average fraction abundance in the GFP+ pools (green color bar). We note that sequencing libraries for experiments 2 and 3 were prepared at the same time and 6 additional PCR cycles were required due to low input of starting material. Some high abundance zip codes, representing potential contamination, were found in both experiment 2 and 3 libraries and were removed from all analyses. Note also that in primary cell experiments, many zip codes were detected in only one pool, likely representing unintegrated viral DNA or infected cells that had divided too few times to be sampled evenly. For example, aliquots of experiments 2 and 3 unsorted infected cell pools displayed totals of 43,525 and 33,114 zip codes, respectively. Of these, 35,686 (82%) or 28,842 (87%) were not observed in either GFP+ or GFP- fractions of SplitA or SplitB. This high rate of zip codes observed only in the unsorted pool illustrates the challenges in the primary cell analysis, which relies upon sufficient cellular divisions for zip codes to be detected in a reproducible manner among split pools.*

Experimental repetition	SplitA/SplitB %GFP+	# zip codes that pass inclusion criteria	Spearman correlations
A	0.4/0.41	349	0.367
B	0.2/0.22	73	0.719
C	1.0/0.9	90	0.730

Table S2. 1. ***Summarized data for the three primary cell experiments.*** *Note that the GFP- sorted cells included uninfected cells. Experimental information is provided above and in Materials and Methods.*

Table S2. 2. ***Randomized sequence tags in trial proviral clones***

Library	gccatcgatgNNNNNNNNNNNNNNNNNNtggatctaccac
Clone 1	gccatcgatgGGTACGGTGGTCATAGCATGtggatctaccac
Clone 2	gccatcgatgGCTCGCGACTGGGGGTGGTGtggatctaccac
Clone 3	gccatcgatgGGTTAGTTGGCGCAAGTGCTtggatctaccac
Clone 4	gccatcgatgTGTCTCATCGGACGGAGGATtggatctaccac
Clone 5	gccatcgatgACATCATGTGTCGCTCCTGCtggatctaccac
Clone 6	gccatcgatgAGCTCGGAGCGTGCGACGGCtggatctaccac
Clone 7	gccatcgatgTCAGCGATCGAAACATCGCGtggatctaccac
Clone 8	gccatcgatgAATTACCGCGGGGACCGGCTtggatctaccac
Clone 9	gccatcgatgGATGTGGAGTAAGCCAGTCGtggatctaccac

zipcode	ins Chrom	ins BP	ins Strand	num Shear Points (3nt Rule)	num Shear Points (unique)	total Reads
AAACGGTTACAGCGTGGCAC	chr3	18395421	rev	17	48	2050
AAAGCGCCTTACGCGCACGA	chr1	108595225	fwd	15	94	8901
AAATCCACTGTTGTATAATG	chr2	63820806	fwd	27	102	4584
AACACAGGCTGTCATAGCTG	chr17	74702540	rev	20	49	1582
AACAGTGTCTTTGAAAACCC	chrX	118686433	rev	3	3	3
AACATAAATGGATAGCCATG	chr16	14672056	fwd	9	32	977
AACTCGTTTACCCCTTCTAC	chr3	126502143	rev	17	24	732
AAGAATCGCCCCTTGCGCGA	chr12	123700550	fwd	27	73	2672
AAGAGTAAGCAGCAATATA	chr1	194511803	fwd	9	17	606
AAGCGTCTCAACCAATCGAA	chr15	83807195	rev	14	32	637
AAGGTAGTCTCCACGAATAT	chr21	33751694	rev	14	29	389
AATCGGCCTCCTTAACGATC	chr11	68293424	rev	8	13	631
AATTCTGTAAGTGTGTCGAA	chr2	32110713	rev	34	159	8104
AATTGTTCACCTTTACTGGA	chr3	125254009	fwd	14	34	832
ACACGAGAAGTTACTTGCTG	chr6	35893422	rev	27	99	4792
ACCCTTGCATAGTATCAGAC	chr4	83805060	rev	11	26	708
ACCGCGAATGACTTGGACTA	chr14	24691959	rev	15	37	1316
ACCGCGACAAGGGGCTCCTC	chr11	108580222	fwd	12	44	1544
ACGACGTCGAATCGTGATAT	chr10	95445139	rev	13	25	254
ACGCAACGGGGGACTGGGGC	chr20	35705414	rev	10	14	73
ACGCGTTACACATGTACGCA	chr2	69552981	fwd	13	19	328

ACGGAACGCCAATATGCGTG	chr1	59456184	fwd	13	26	678
ACTGAGTTCCACAGATAGCA	chr16	2299453	fwd	14	46	2190
ACTGCAAAGCGGGCCGACAG	chr12	57457811	fwd	8	10	176
AGAAATCCCTTAACAGAGAG	chr2	55749153	fwd	7	14	216
AGAAATCTGATAACTTTAGG	chr1	153869984	rev	4	4	5
AGAATGAAGGGGAACTGACT	chr17	29592319	rev	1	2	2
AGACCCTGCGGTCACATCAG	chr12	50917807	fwd	13	40	1693
AGACCTCCTTGGGGAGTCCG	chr11	134304579	fwd	5	11	92
AGACTGAGAGTCAACGCTTC	chr11	65634277	fwd	5	13	119
AGAGACAAATTGAAACCGAT	chr3	52295062	rev	21	43	2651
AGATGCCCAAGCAGAAAACT	chr9	88613604	rev	15	34	1496
AGATGTTCGACGAGCAACGT	chr17	4204916	rev	12	50	3375
AGCGACGTATCATAGAGAGG	chr20	43269953	fwd	15	29	826
AGCGTGACCCTAGTTCTATC	chr15	89804561	rev	21	44	1395
AGCTATTCATGAAGACCTGG	chr19	19234772	fwd	21	62	3387
AGCTTCTGCGTACCCCATAT	chr16	399531	rev	21	67	3725
AGGCCGGGACCGCATCGGGT	chr3	169864928	rev	20	55	2190
AGGGATTCGGTCGCCGATAT	chr3	15754061	rev	16	37	1599
AGGTTCTCATACGAGAACCT	chr11	65046054	rev	21	52	2336
AGTAACCCTTAGTTAGGGTA	chr17	17175509	rev	23	166	11202
AGTTGTCATACGTATGCAGT	chr21	47923541	rev	11	56	3298
ATAAACCCTTGGGAGTATCT	chr15	62283187	fwd	18	28	451
ATAATTGCTACTCACGGATG	chr1	185088174	fwd	12	26	1437

ATACCGATGCTATTTTTATT	chr1	149731511	rev	7	28	1188
ATACGAGCATCCCCCGGTAC	chr2	128861119	fwd	20	47	2132
ATATCTGGAAACTTCTGTTC	chr11	68337332	rev	15	33	801
ATATGAGCCGTTTAACAGAG	chr17	7214174	fwd	10	19	634
ATCATTCTTTCGCAAAAAAA	chr3	195787073	rev	5	6	6
ATCCATCTAGAATCGGGGAC	chr1	93603537	fwd	18	54	1383
ATCGGTATTGATGTGCAAGC	chr10	5757714	fwd	25	96	5688
ATGGCAACCCTAACCACTAT	chr1	116042595	fwd	18	55	1812
ATGTAGCATTGCATGTATGC	chr3	34152671	rev	10	15	530
ATTCTCACTTTCTGTCCCGC	chr19	42058337	fwd	26	71	2991
ATTCTGGTTCTTAGGGGACT	chr6	135738933	rev	12	24	1614
ATTGACGCCAAGTGCCAGCG	chr16	66849235	rev	17	49	1562
ATTGGCAGACAGCTAAATCA	chr12	62768615	fwd	21	87	3327
ATTGTAAGGCGCTCTTCCAC	chr22	41328631	rev	3	5	12
CAAACTAGGGTTTAGGATAA	chr19	13258242	fwd	1	2	2
CAACGGCCTGTGAGGATGGA	chr1	11730262	rev	21	69	669
CAACTCCTGCTCCCACCAAT	chr21	45454599	rev	8	13	610
CAAGCAGTGAAACCGATTAG	chr19	52072025	fwd	21	42	1106
CAATTAAGTGTGAATTGTCG	chr17	58008202	fwd	29	94	4486
CACAATTCTGATTTCCTATC	chr18	59800169	rev	20	68	2741
CACGGGTGATACGTCCTGAA	chr2	160590814	rev	1	1	1
CACTCTACATCATAAGTAGT	chr7	135073658	fwd	22	67	3739
CACTCTCTATGTCGTGTTAT	chr2	51538264	rev	5	7	347

CACTGCCCAAGCAATTTCAT	chr6	89600354	fwd	10	13	251
CACTGTGTAACCCTCGTCGC	chr16	57243954	fwd	19	58	4066
CACTTGCGGTGATCCCTGGG	chr20	52452675	rev	23	196	9865
CAGCACGATAAAATACTCTC	chr2	99326959	rev	2	2	2
CATATTGACACCCGCTGGCT	chr1	114972800	rev	4	7	39
CATCCAGACTCTTAGTAACT	chr16	67588223	fwd	20	48	1675
CATCGTAAAGTGAGCGTTAA	chr18	9183245	rev	12	22	720
CATCTGCGTCCTGCCTCGTC	chr7	43644182	fwd	13	24	1206
CATTAATTTACGAAGGTTTA	chr17	47727743	rev	2	3	3
CATTTGTTGACCAGCAGCGT	chr5	83921932	fwd	10	15	428
CCAACGGAATTGCCCTTTCC	chr3	125764667	rev	17	51	1992
CCACAGTTCACCCTTTGTAA	chr3	194120775	rev	1	1	1
CCCAATTATACTCAGGGCGA	chr20	21336740	fwd	10	17	264
CCCCACGGGTTCAGCTGATT	chr3	149585548	fwd	7	16	1131
CCCCATACCTGTACTCCGCC	chr2	26596975	rev	9	12	392
CCGGCAAAGTCACCACCACT	chr2	160590801	rev	18	102	4553
CCGGGCATGATCGATAATGT	chr17	60584939	fwd	6	13	503
CCGTAATCATGATCTAGTCC	chr4	57288570	fwd	18	39	1118
CCTCAAGCGGAAATTGGAGG	chr2	242230256	rev	12	19	435
CCTCCCCTTAAGTGGGGATG	chr2	128237892	fwd	24	65	2330
CCTCGATTCGAGAATTCGTG	chr1	213036048	fwd	19	85	3804
CGAGCATTTTTGAATGGTTG	chr5	138661871	fwd	22	45	1422
CGAGTGTATCATCACTCGAA	chr11	47082155	fwd	12	36	1634

CGATGGAAAAAATGACGTTA	chr12	101744563	rev	20	40	921
CGCAGTAGTATGAAGTATGA	chr20	34247250	rev	24	106	8913
CGCATGGGGCGGGAGCAGAG	chrX	70508662	rev	2	2	2
CGGAAAAAAAAGAGGCCGGG	chr5	132045324	rev	21	127	8668
CGGACCCACTTGCTTCACGC	chr1	31465909	fwd	17	28	850
CGGAGACAACTCTTCTGTTT	chr21	47905470	rev	10	18	220
CGGGACTCTGTTTCGCAAAC	chr21	23801760	rev	8	14	152
CGGGCCGGGCACCTACGCAG	chr12	49074757	fwd	10	27	834
CGGTTATCTTTCCGGCTCTA	chr17	32704009	rev	14	39	1187
CGGTTTGCGGGGTCGCCTAC	chr10	53467887	rev	13	24	699
CGTCGAGTGCACGAAGCACA	chr16	1696834	fwd	23	64	2896
CGTCTCACTGAGCAGCCAAG	chr20	43666534	fwd	10	23	444
CGTGATTTGTTGTAAGATGA	chr13	50649083	rev	7	13	424
CGTTATATGATCGATCATAT	chr4	40112573	fwd	17	58	2800
CTCATTTTAGCTATGTCCGT	chr5	31521263	rev	11	31	1709
CTCCCCCGCACAATAGTGTG	chr17	47405396	fwd	11	50	2261
CTCGTGATGCCCGTAGGAGC	chr17	78596696	fwd	1	1	1
CTGATATAGCTATAGAGTGA	chr15	91341982	fwd	4	8	197
CTGCCGAAGTATATCAGCAG	chr22	40863064	fwd	5	20	1421
CTGCGGATTAGCGTCTGTAA	chr1	114982359	rev	8	12	229
CTGCGGTGTGGTTCTTGGCA	chr15	77438425	rev	8	16	325
CTGCTCGTATAAAGTAGTAC	chr21	38872071	fwd	7	14	326
CTGGAAGAACTTTTGCAGCA	chr11	33958369	fwd	1	1	1

GAAAAGTTTGCCCACAATAC	chr13	60681242	rev	16	24	585
GAAACTCTGCACCTATTAAT	chr18	20598740	rev	15	25	980
GAACAGGCTAAGTACTCCCT	chr17	30682805	rev	19	67	3192
GAAGACACATTAACTTGTAC	chr14	106458885	rev	16	25	311
GACATACTAGCGCGGGCTAT	chr1	149850614	fwd	24	111	3660
GACCTATGCGTATTGCGGCA	chr9	131725938	fwd	21	59	1768
GACGACAACCACCGACCGAC	chr9	15725780	rev	19	52	1139
GACGCTCCAAGATTCTTAAC	chr3	150910355	rev	18	123	6230
GAGCGGCGAAATCGGACCTG	chr9	132775655	fwd	6	12	176
GAGTCGTAGGTATTAGCGAA	chr17	74082586	fwd	24	65	1830
GATAAAACGAGCATTAGAAA	chr5	176681791	fwd	10	21	828
GATAGCCCGTAGGCGACGTC	chr1	156314957	fwd	15	23	496
GATGTCTGAGGCCGTACAAT	chr17	4188572	rev	5	24	1462
GCACATAAAGAAGTGCTTTG	chr13	99864323	rev	6	20	457
GCACATCACGCGCAGCGGGC	chr7	23643901	rev	8	16	302
GCACCCCGGGAGTGAAAACC	chr12	48443290	rev	14	25	612
GCAGGGTAGCAACATAGGCC	chr17	27837380	fwd	17	200	16349
GCATCTCTCCCGTAACCGCT	chr17	64798415	rev	16	26	961
GCATTCGTCGCCTTTTTCGA	chr6	147631249	fwd	5	11	130
GCCACTTACATCTAGAAAGC	chr6	111025400	rev	5	5	15
GCCATGTGGCGCGATATACC	chr3	151072823	fwd	10	37	1260
GCCCATCCGGTCCGGAGATC	chr15	72838284	fwd	13	29	653
GCCCCAAAGGGACTGAATCC	chr4	48876058	rev	4	5	11

GCCCCGACCCGCGAGGGGGA	chr7	44689340	rev	16	46	3626
GCCTGACTCATATTCAGGGA	chr19	18431133	fwd	26	110	4835
GCCTTGATTGTAAGGTGGGG	chr2	105929621	rev	7	13	690
GCGACCCCCAAGAACCAGAT	chr7	72402060	fwd	16	33	855
GCGACGCTTGACAATGTCGA	chr4	83782207	fwd	26	91	3419
GCGTCAAGAAATAGAGAATG	chr8	86106668	rev	13	31	491
GCGTCATGGAGTTCCTCCAT	chr16	3027720	fwd	24	166	8948
GCGTCCATCAGCACATCAGG	chr19	50292416	fwd	11	26	760
GCGTGAACATTATACACTTT	chr1	152355191	fwd	27	103	7424
GCGTGCGGTTGCTACACAGA	chr1	168154114	rev	11	21	666
GCGTTAGTTAGCTCGGTTGG	chr9	131819215	rev	19	43	1967
GCTACAGATGTTCTTGCCGC	chr11	62400130	rev	2	2	2
GCTGTTTCTGGGGCGCTCA	chr1	51922203	fwd	9	17	427
GCTTAAATCGACAAGCGTAT	chr12	25484506	fwd	16	38	942
GGAGACCGCAATCCTACGCG	chr2	26026282	rev	5	9	306
GGCACCGGGGCAATCCGGCA	chr20	47890056	fwd	6	9	36
GGCCCCTGATGCTTGGATTT	chr4	184917808	rev	6	7	116
GGCCGTAAAGGGCTTAAATG	chr7	92200060	rev	17	38	1116
GGCTAGAGTAGCCGATCGAG	chr6	97712715	rev	11	22	722
GGGATTCTGGACACTAGCCG	chr2	61515437	fwd	5	6	6
GGGCACTAGGCCGATAGTTT	chr5	52074929	fwd	3	3	3
GGGCATTACATCTAATCAAA	chrY	23350488	fwd	17	38	788
GGGGAACCCCCGCGAGGGGA	chr9	96241196	fwd	2	2	2

GGGGCTTCAAATCACGCAAT	chr4	69197416	rev	19	48	1863
GGTAAAGTTACATCGGATAC	chr9	15725794	rev	1	1	1
GGTACATATCCCAAAAGGGA	chr12	102528184	fwd	14	62	4835
GGTGATTCGGTTATGTGCTA	chr1	186305474	rev	18	44	1527
GTACAAGCTTTGTAACCCGA	chr7	99079633	fwd	14	35	703
GTATTCTCGCACCACCCATC	chr19	10267306	rev	21	54	1956
GTCGGATATAAATCCGCTCT	chr10	27413350	rev	15	22	97
GTGGTAAAGTAAAACTTGGG	chr16	14583042	rev	11	23	482
GTGTCGTTAAGCCCGGCATA	chr12	57627963	rev	23	74	3910
GTTAGGTCACTAACAACGGA	chr7	24684959	fwd	6	8	250
GTTCGCTTCTGGTTCGTGT	chr10	116595096	fwd	17	26	275
GTTTATCTGGAAGCGTACTC	chr15	78338420	fwd	2	3	18
TAACAGTATTTTCTAGGAAC	chr14	74139316	rev	9	13	56
TAATGTTAGGTAATTCCGGA	chr2	20509207	rev	5	7	9
TAATTGCATAAATCTCGCCG	chr2	192268452	fwd	8	22	728
TACCTACATGCTTCGCCGTA	chr12	49816602	fwd	16	43	2075
TACGGACAGGTTTAGACCTA	chr11	108219643	fwd	16	32	557
TACGTTCCGCATAAACCGTC	chr17	7243712	rev	17	27	1074
TACTGAAGATCTAGTCATGA	chr5	32547909	fwd	27	97	4465
TAGAACCCCAATCCGGGGCC	chr3	196348422	rev	27	110	5514
TAGTGTGCTAGGTACTTCCC	chrX	13757756	rev	8	12	346
TATAATGTTTAGTCGCCTGG	chr6	37905815	fwd	27	109	3028
TATACAGCCCGAACACAATG	chr6	121609102	rev	6	8	48

TATCTCCCGCATTATGTCGC	chr2	9666113	rev	18	46	1224
TATCTTAAATAGTTTTTATA	chr7	7264648	rev	6	14	343
TATGGTAGGATGCTCAACAC	chr4	71573510	rev	15	65	2652
TCAGACCCCCGAAATGGGCC	chr16	69186487	rev	16	73	3935
TCATACTCAAGACAGCCGAT	chr17	65911250	fwd	25	131	6496
TCCATTGGAGCCCAAGTGTT	chr3	50091269	fwd	11	26	768
TCCGAGGCGACGCCTTTCGG	chr12	49069223	rev	27	140	10935
TCCGATCCAAGCATAAAAGA	chr1	1716276	rev	25	79	2336
TCCGTCCGGTATTTAGATTT	chr11	65870281	rev	11	20	151
TCGAAACAGAGTTTGTAGAG	chr3	182913103	fwd	20	129	8051
TCGGCAGCCCCCTTCGAATG	chr12	49429664	rev	13	40	1812
TCGGCGGGTACAACCAATGA	chr7	135273281	fwd	16	25	303
TCGTAGGACAGTCGAGTTTG	chr3	152006563	fwd	8	16	1008
TCGTGGTACCGAACGACTAG	chr17	3723709	rev	15	31	873
TCTCACCCCCTTGAATACAG	chr1	114440807	rev	29	110	4595
TCTCGGTACTAGGGGGCCTT	chr6	109436863	rev	17	111	7527
TCTTACACGGATTAGTGCAA	chr17	2026530	fwd	23	119	6859
TGAAACCTGACGAAACTGTG	chr15	57245476	rev	11	24	1237
TGAACCGCGGATTTTGGGCG	chrY	2825026	rev	12	24	586
TGACCTCCGATCACCCGGAG	chr20	62246791	fwd	12	36	1146
TGAGAGATCGCTACACTGTA	chr3	196291350	rev	29	171	9770
TGCAATATCATGGCGGTTGC	chr1	26287657	rev	11	42	1960
TGCCCTTGTGTGTCTGATAT	chr17	55342783	rev	12	35	1151

TGCCTAGAGCATACCTTAAG	chr17	79812870	fwd	4	7	30
TGCGCCCGATTTCTATCAAA	chr12	50989965	fwd	12	18	293
TGGAATCGCGCGGTATACCA	chr4	74067540	fwd	5	13	424
TGGACACTTGGTTCCTAGTG	chr22	41263568	fwd	9	12	62
TGGGATATCACTAACGATCC	chr7	102836180	fwd	15	38	1160
TGGTCTATCAGAGCGATATG	chr14	56106865	fwd	5	13	325
TGTATCAGCTTGATATCTTT	chr2	128252115	rev	10	17	767
TGTGAGCGGAGTCATCCGCG	chr1	16635538	rev	19	41	239
TTCAGAAATTCAACCGCATC	chr18	12959748	fwd	21	40	1145
TTCATGCAGCGGGGGTACAA	chr2	240209559	rev	28	98	4484
TTCGATAAACGCAGACTATT	chr4	154397977	rev	19	37	1137
TTGCACGGCATGTTACATTT	chr5	138682146	fwd	18	35	929
TTGCCTAGATCGCCGCTTCT	chr9	134275447	fwd	17	50	2081
TTGCTGTCCCGCTAGCGAGC	chr6	135481600	rev	15	45	920
TTGCTTACGAACTCAGTCTC	chr17	57165969	rev	21	108	5461
TTGGAAGAGGGTACTTGGAA	chr22	36237160	rev	10	12	198
TTTAGTGACGTGAGGATCAC	chr19	2148971	fwd	15	59	6052
TTTATTGCCCGACCACAATT	chr16	66837004	fwd	10	23	331
TTTCTCTAACCAACAATCCG	chr19	58337901	rev	9	18	392
TTTGAGGGGTACTACACCGC	chr6	42633342	fwd	6	9	175

Table S2. 3. ***Integration sites***

References

Alabert, C., T. K. Barth, N. Reveron-Gomez, S. Sidoli, A. Schmidt, O. N. Jensen, A. Imhof and A. Groth (2015). "Two distinct modes for propagation of histone PTMs across the cell cycle." Genes Dev **29**(6): 585-590.

Anderson, E. M. and F. Maldarelli (2018). "The role of integration and clonal expansion in HIV infection: live long and prosper." Retrovirology **15**(1): 71.

Archin, N. M., A. L. Liberty, A. D. Kashuba, S. K. Choudhary, J. D. Kuruc, A. M. Crooks, D. C. Parker, E. M. Anderson, M. F. Kearney, M. C. Strain, D. D. Richman, M. G. Hudgens, R. J. Bosch, J. M. Coffin, J. J. Eron, D. J. Hazuda and D. M. Margolis (2012). "Administration of vorinostat disrupts HIV-1 latency in patients on antiretroviral therapy." Nature **487**(7408): 482-485.

Battich, N., T. Stoeger and L. Pelkmans (2015). "Control of Transcript Variability in Single Mammalian Cells." Cell **163**(7): 1596-1610.

Battivelli, E., M. S. Dahabieh, M. Abdel-Mohsen, J. P. Svensson, I. Tojal Da Silva, L. B. Cohn, A. Gramatica, S. Deeks, W. C. Greene, S. K. Pillai and E. Verdin (2018). "Distinct chromatin functional states correlate with HIV latency reactivation in infected primary CD4(+) T cells." Elife **7**.

Bieniasz, P. and A. Telesnitsky (2018). "Multiple, Switchable Protein:RNA Interactions Regulate Human Immunodeficiency Virus Type 1 Assembly." Annu Rev Virol.

Bruner, K. M., A. J. Murray, R. A. Pollack, M. G. Soliman, S. B. Laskey, A. A. Capoferri, J. Lai, M. C. Strain, S. M. Lada, R. Hoh, Y. C. Ho, D. D. Richman, S. G. Deeks, J. D. Siliciano and R. F. Siliciano (2016). "Defective proviruses rapidly accumulate during acute HIV-1 infection." Nat Med **22**(9): 1043-1049.

Carter, C. C., A. Onafuwa-Nuga, L. A. McNamara, J. t. Riddell, D. Bixby, M. R. Savona and K. L. Collins (2010). "HIV-1 infects multipotent progenitor cells causing cell death and establishing latent cellular reservoirs." Nat Med **16**(4): 446-451.

Chavez, L., V. Calvanese and E. Verdin (2015). "HIV Latency Is Established Directly and Early in Both Resting and Activated Primary CD4 T Cells." PLoS Pathog **11**(6): e1004955.

Chen, H. C., J. P. Martinez, E. Zorita, A. Meyerhans and G. J. Filion (2017). "Position effects influence HIV latency reversal." Nat Struct Mol Biol **24**(1): 47-54.

Chomont, N., M. El-Far, P. Ancuta, L. Trautmann, F. A. Procopio, B. Yassine-Diab, G. Boucher, M. R. Boulassel, G. Ghattas, J. M. Brenchley, T. W. Schacker, B. J. Hill, D. C. Douek, J. P. Routy, E. K. Haddad and R. P. Sekaly (2009). "HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation." Nat Med **15**(8): 893-900.

Chun, T. W., L. Stuyver, S. B. Mizell, L. A. Ehler, J. A. Mican, M. Baseler, A. L. Lloyd, M. A. Nowak and A. S. Fauci (1997). "Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy." Proc Natl Acad Sci U S A **94**(24): 13193-13197.

Ciuffi, A., M. Llano, E. Poeschla, C. Hoffmann, J. Leipzig, P. Shinn, J. R. Ecker and F. Bushman (2005). "A role for LEDGF/p75 in targeting HIV DNA integration." Nat Med **11**(12): 1287-1289.

Consortium, E. P. (2012). "An integrated encyclopedia of DNA elements in the human genome." Nature **489**(7414): 57-74.

Corman, T. H., C. E. Leiserson, R. L. Rivest and C. Stein (2009). Introduction to Algorithms, Third Edition, The MIT Press.

Costin, J. M. (2007). "Cytopathic mechanisms of HIV-1." Virol J **4**: 100.

Coulon, A., C. C. Chow, R. H. Singer and D. R. Larson (2013). "Eukaryotic transcriptional dynamics: from single molecules to cell populations." Nat Rev Genet **14**(8): 572-584.

Dahabieh, M. S., E. Battivelli and E. Verdin (2015). "Understanding HIV latency: the road to an HIV cure." Annu Rev Med **66**: 407-421.

Dahabieh, M. S., M. Ooms, V. Simon and I. Sadowski (2013). "A doubly fluorescent HIV-1 reporter shows that the majority of integrated HIV-1 is latent shortly after infection." J Virol **87**(8): 4716-4727.

Deeks, S. G. (2012). "HIV: Shock and kill." Nature **487**(7408): 439-440.

Fennessey, C. M., M. Pinkevych, T. T. Immonen, A. Reynaldi, V. Venturi, P. Nadella, C. Reid, L. Newman, L. Lipkey, K. Oswald, W. J. Bosche, M. T. Trivett, C. Ohlen, D. E. Ott, J. D. Estes, G. Q. Del Prete, J. D. Lifson, M. P. Davenport and B. F. Keele (2017). "Genetically-barcoded SIV facilitates enumeration of rebound variants and estimation of reactivation rates in nonhuman primates following interruption of suppressive antiretroviral therapy." PLoS Pathog **13**(5): e1006359.

Finzi, D., M. Hermankova, T. Pierson, L. M. Carruth, C. Buck, R. E. Chaisson, T. C. Quinn, K. Chadwick, J. Margolick, R. Brookmeyer, J. Gallant, M. Markowitz, D. D. Ho, D. D. Richman and R. F. Siliciano (1997). "Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy." Science **278**(5341): 1295-1300.

Finzi, D., S. F. Plaeger and C. W. Dieffenbach (2006). "Defective virus drives human immunodeficiency virus infection, persistence, and pathogenesis." Clin Vaccine Immunol **13**(7): 715-721.

Gallastegui, E., G. Millan-Zambrano, J. M. Terme, S. Chavez and A. Jordan (2011). "Chromatin reassembly factors are involved in transcriptional interference promoting HIV latency." J Virol **85**(7): 3187-3202.

Hakre, S., L. Chavez, K. Shirakawa and E. Verdin (2012). "HIV latency: experimental systems and molecular models." FEMS Microbiol Rev **36**(3): 706-716.

Han, Y., Y. B. Lin, W. An, J. Xu, H. C. Yang, K. O'Connell, D. Dordai, J. D. Boeke, J. D. Siliciano and R. F. Siliciano (2008). "Orientation-dependent regulation of integrated HIV-1 expression by host gene transcriptional readthrough." Cell Host Microbe **4**(2): 134-146.

Hnisz, D., A. S. Weintraub, D. S. Day, A. L. Valton, R. O. Bak, C. H. Li, J. Goldmann, B. R. Lajoie, Z. P. Fan, A. A. Sigova, J. Reddy, D. Borges-Rivera, T. I. Lee, R. Jaenisch, M. H. Porteus, J. Dekker and R. A. Young (2016). "Activation of proto-oncogenes by disruption of chromosome neighborhoods." Science **351**(6280): 1454-1458.

Ho, Y. C., L. Shan, N. N. Hosmane, J. Wang, S. B. Laskey, D. I. Rosenbloom, J. Lai, J. N. Blankson, J. D. Siliciano and R. F. Siliciano (2013). "Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure." Cell **155**(3): 540-551.

Jordan, A., P. Defechereux and E. Verdin (2001). "The site of HIV-1 integration in the human genome determines basal transcriptional activity and response to Tat transactivation." EMBO J **20**(7): 1726-1738.

Kaczmarek, K., A. Morales and A. J. Henderson (2013). "T Cell Transcription Factors and Their Impact on HIV Expression." Virology (Auckl) **2013**(4): 41-47.

Kaern, M., T. C. Elston, W. J. Blake and J. J. Collins (2005). "Stochasticity in gene expression: from theories to phenotypes." Nat Rev Genet **6**(6): 451-464.

Keene, S. E., S. R. King and A. Telesnitsky (2010). "7SL RNA is retained in HIV-1 minimal virus-like particles as an S-domain fragment." Journal of virology **84**(18): 9070-9077.

Kharytonchyk, S., J. D. Brown, K. Stilger, S. Yasin, A. S. Iyer, J. Collins, M. F. Summers and A. Telesnitsky (2018). "Influence of gag and RRE Sequences on HIV-1 RNA Packaging Signal Structure and Function." J Mol Biol **430**(14): 2066-2079.

Kharytonchyk, S., S. R. King, C. B. Ndongmo, K. L. Stilger, W. An and A. Telesnitsky (2016). "Resolution of Specific Nucleotide Mismatches by Wild-Type and AZT-Resistant Reverse Transcriptases during HIV-1 Replication." Journal of molecular biology **428**(11): 2275-2288.

Kim, Y. H., L. Zhu, K. Pyaram, C. Lopez, R. G. Ohye, J. V. Garcia, G. E. Green and C. H. Chang (2018). "PLZF-expressing CD4 T cells show the characteristics of terminally differentiated effector memory CD4 T cells in humans." Eur J Immunol **48**(7): 1255-1257.

Kircher, M., S. Sawyer and M. Meyer (2012). "Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform." Nucleic Acids Res **40**(1): e3.

Kok, Y. L., S. Schmutz, A. Inderbitzin, K. Neumann, A. Kelley, L. Jorimann, M. Shilaih, V. Vongrad, R. D. Kouyos, H. F. Gunthard, C. Berens and K. J. Metzner (2018). "Spontaneous reactivation of latent HIV-1 promoters is linked to the cell cycle as revealed by a genetic-insulators-containing dual-fluorescence HIV-1-based vector." Sci Rep **8**(1): 10204.

Lewinski, M. K., D. Bisgrove, P. Shinn, H. Chen, C. Hoffmann, S. Hannenhalli, E. Verdin, C. C. Berry, J. R. Ecker and F. D. Bushman (2005). "Genome-wide analysis of chromosomal features repressing human immunodeficiency virus transcription." J Virol **79**(11): 6610-6619.

Lu, K., X. Heng, L. Garyu, S. Monti, E. L. Garcia, S. Kharytonchyk, B. Dorjsuren, G. Kulandaivel, S. Jones, A. Hiremath, S. S. Divakaruni, C. LaCotti, S. Barton, D. Tummillo, A. Hosic, K. Edme, S. Albrecht, A. Telesnitsky and M. F. Summers (2011). "NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging." Science **334**(6053): 242-245.

Ma, Y., K. Kanakousaki and L. Buttitta (2015). "How the cell cycle impacts chromatin architecture and influences cell fate." Front Genet **6**: 19.

Magoc, T. and S. L. Salzberg (2011). "FLASH: fast length adjustment of short reads to improve genome assemblies." Bioinformatics **27**(21): 2957-2963.

Maldarelli, F., X. Wu, L. Su, F. R. Simonetti, W. Shao, S. Hill, J. Spindler, A. L. Ferris, J. W. Mellors, M. F. Kearney, J. M. Coffin and S. H. Hughes (2014). "HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells." Science **345**(6193): 179-183.

Martins, L. J., P. Bonczkowski, A. M. Spivak, W. De Spiegelaere, C. L. Novis, A. B. DePaula-Silva, E. Malatinkova, W. Trypsteen, A. Bosque, L. Vanderkerckhove and V. Planelles (2016). "Modeling HIV-1 Latency in Primary T Cells Using a Replication-Competent Virus." AIDS Res Hum Retroviruses **32**(2): 187-193.

Mbonye, U. and J. Karn (2017). "The Molecular Basis for Human Immunodeficiency Virus Latency." Annu Rev Virol **4**(1): 261-285.

Mei, J. M., F. Nourbakhsh, C. W. Ford and D. W. Holden (1997). "Identification of Staphylococcus aureus virulence genes in a murine model of bacteraemia using signature-tagged mutagenesis." Mol Microbiol **26**(2): 399-407.

Menendez-Arias, L. (2009). "Mutation rates and intrinsic fidelity of retroviral reverse transcriptases." Viruses **1**(3): 1137-1165.

Mullins, J. I. and L. M. Frenkel (2017). "Clonal Expansion of Human Immunodeficiency Virus-Infected Cells and Human Immunodeficiency Virus Persistence During Antiretroviral Therapy." J Infect Dis **215**(suppl_3): S119-S127.

Ne, E., R. J. Palstra and T. Mahmoudi (2018). "Transcription: Insights From the HIV-1 Promoter." Int Rev Cell Mol Biol **335**: 191-243.

Nolan-Stevaux, O., D. Tedesco, S. Ragan, M. Makhanov, A. Chenchik, A. Ruefli-Brasse, K. Quon and P. D. Kassner (2013). "Measurement of Cancer Cell Growth Heterogeneity through Lentiviral Barcoding Identifies Clonal Dominance as a Characteristic of In Vivo Tumor Engraftment." PLoS One **8**(6): e67316.

Pace, M. J., L. Agosto, E. H. Graf and U. O'Doherty (2011). "HIV reservoirs and latency models." Virology **411**(2): 344-354.

Pace, M. J., E. H. Graf, L. M. Agosto, A. M. Mexas, F. Male, T. Brady, F. D. Bushman and U. O'Doherty (2012). "Directly infected resting CD4+T cells can produce HIV Gag without spreading infection in a model of HIV latency." PLoS Pathog **8**(7): e1002818.

Pinzone, M. R., D. J. VanBelzen, S. Weissman, M. P. Bertuccio, L. Cannon, E. Venanzi-Rullo, S. Migueles, R. B. Jones, T. Mota, S. B. Joseph, K. Groen, A. O. Pasternak, W. T. Hwang, B. Sherman, A. Vourekas, G. Nunnari and U. O'Doherty (2019). "Longitudinal HIV sequencing reveals reservoir expression leading to decay which is obscured by clonal expansion." Nat Commun **10**(1): 728.

Pocock, G. M., L. L. Zimdars, M. Yuan, K. W. Eliceiri, P. Ahlquist and N. M. Sherer (2017). "Diverse activities of viral cis-acting RNA regulatory elements revealed using multicolor, long-term, single-cell imaging." Mol Biol Cell **28**(3): 476-487.

Pollack, R. A., R. B. Jones, M. Pertea, K. M. Bruner, A. R. Martin, A. S. Thomas, A. A. Capoferri, S. A. Beg, S. H. Huang, S. Karandish, H. Hao, E. Halper-Stromberg, P. C. Yong, C. Kovacs, E. Benko, R. F. Siliciano and Y. C. Ho (2017). "Defective HIV-1 Proviruses Are Expressed and Can Be Recognized by Cytotoxic T Lymphocytes, which Shape the Proviral Landscape." Cell Host Microbe **21**(4): 494-506 e494.

Quail, M. A., I. Kozarewa, F. Smith, A. Scally, P. J. Stephens, R. Durbin, H. Swerdlow and D. J. Turner (2008). "A large genome center's improvements to the Illumina sequencing system." Nat Methods **5**(12): 1005-1010.

Rasmussen, T. A. and S. R. Lewin (2016). "Shocking HIV out of hiding: where are we with clinical trials of latency reversing agents?" Curr Opin HIV AIDS **11**(4): 394-401.

Re, F., D. Braaten, E. K. Franke and J. Luban (1995). "Human immunodeficiency virus type 1 Vpr arrests the cell cycle in G2 by inhibiting the activation of p34cdc2-cyclin B." J Virol **69**(11): 6859-6864.

Reeder, J. E., Y. T. Kwak, R. P. McNamara, C. V. Forst and I. D'Orso (2015). "HIV Tat controls RNA Polymerase II and the epigenetic landscape to transcriptionally reprogram target immune cells." Elife **4**.

Reveron-Gomez, N., C. Gonzalez-Aguilera, K. R. Stewart-Morgan, N. Petryk, V. Flury, S. Graziano, J. V. Johansen, J. S. Jakobsen, C. Alabert and A. Groth (2018). "Accurate Recycling of Parental Histones Reproduces the Histone Modification Landscape during DNA Replication." Mol Cell **72**(2): 239-249 e235.

Satou, Y., H. Katsuya, A. Fukuda, N. Misawa, J. Ito, Y. Uchiyama, P. Miyazato, S. Islam, A. Fassati, A. Melamed, C. R. M. Bangham, Y. Koyanagi and K. Sato (2017). "Dynamics and mechanisms of clonal expansion of HIV-1-infected cells in a humanized mouse model." Sci Rep **7**(1): 6913.

Schroder, A. R., P. Shinn, H. Chen, C. Berry, J. R. Ecker and F. Bushman (2002). "HIV-1 integration in the human genome favors active genes and local hotspots." Cell **110**(4): 521-529.

Serrao, E. and A. N. Engelman (2016). "Sites of retroviral DNA integration: From basic research to clinical applications." Crit Rev Biochem Mol Biol **51**(1): 26-42.

Sherrill-Mix, S., M. K. Lewinski, M. Famiglietti, A. Bosque, N. Malani, K. E. Ocwieja, C. C. Berry, D. Looney, L. Shan, L. M. Agosto, M. J. Pace, R. F. Siliciano, U. O'Doherty, J. Guatelli, V. Planelles and F. D. Bushman (2013). "HIV latency and integration site placement in five cell-based models." Retrovirology **10**: 90.

Singh, A., B. Razooky, C. D. Cox, M. L. Simpson and L. S. Weinberger (2010). "Transcriptional bursting from the HIV-1 promoter is a significant source of stochastic noise in HIV-1 gene expression." Biophys J **98**(8): L32-34.

Spivak, A. M. and V. Planelles (2018). "Novel Latency Reversal Agents for HIV-1 Cure." Annu Rev Med **69**: 421-436.

Sunshine, S., R. Kirchner, S. S. Amr, L. Mansur, R. Shakhbatyan, M. Kim, A. Bosque, R. F. Siliciano, V. Planelles, O. Hofmann, S. Ho Sui and J. Z. Li (2016). "HIV Integration Site Analysis of Cellular Models of HIV Latency with a Probe-Enriched Next-Generation Sequencing Assay." J Virol **90**(9): 4511-4519.

Swain, P. S., M. B. Elowitz and E. D. Siggia (2002). "Intrinsic and extrinsic contributions to stochasticity in gene expression." Proc Natl Acad Sci U S A **99**(20): 12795-12800.

Tyagi, M. and F. Romerio (2011). "Models of HIV-1 persistence in the CD4+ T cell compartment: past, present and future." Curr HIV Res **9**(8): 579-587.

Verdin, E., P. Paras, Jr. and C. Van Lint (1993). "Chromatin disruption in the promoter of human immunodeficiency virus type 1 during transcriptional activation." EMBO J **12**(8): 3249-3259.

Weinberger, L. S., J. C. Burnett, J. E. Toettcher, A. P. Arkin and D. V. Schaffer (2005). "Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity." Cell **122**(2): 169-182.

Wiegand, A., J. Spindler, F. F. Hong, W. Shao, J. C. Cyktor, A. R. Cillo, E. K. Halvas, J. M. Coffin, J. W. Mellors and M. F. Kearney (2017). "Single-cell analysis of HIV-1 transcriptional activity reveals expression of proviruses in expanded clones during ART." Proc Natl Acad Sci U S A **114**(18): E3659-E3668.

Wong, J. K., M. Hezareh, H. F. Gunthard, D. V. Havlir, C. C. Ignacio, C. A. Spina and D. D. Richman (1997). "Recovery of replication-competent HIV despite prolonged suppression of plasma viremia." Science **278**(5341): 1291-1295.

Wong, R. W., J. I. Mamede and T. J. Hope (2015). "Impact of Nucleoporin-Mediated Chromatin Localization and Nuclear Architecture on HIV Integration Site Selection." J Virol **89**(19): 9702-9705.

Yang, S., R. Delgado, S. R. King, C. Woffendin, C. S. Barker, Z. Y. Yang, L. Xu, G. P. Nolan and G. J. Nabel (1999). "Generation of retroviral vector for clinical studies using transient transfection." Hum Gene Ther **10**(1): 123-132.

Zentilin, L., G. Qin, S. Tafuro, M. C. Dinauer, C. Baum and M. Giacca (2000). "Variegation of retroviral vector gene expression in myeloid cells." Gene Ther **7**(2): 153-166.

Zhang, Y., T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li and X. S. Liu (2008). "Model-based analysis of ChIP-Seq (MACS)." Genome Biol **9**(9): R137.

Chapter 3 [2]Bimodal Expression Patterns, And Not Viral Burst Sizes, Predict the Effects of Vpr on HIV-1 Proviral Populations in Jurkat Cells

Abstract:

Integration site landscapes, clonal dynamics, and latency reversal with or without *vpr* were compared in HIV-1-infected Jurkat cell populations, and the properties of individual clones were defined. Clones differed in fractions of LTR-active daughter cells, with some clones containing few to no LTR-active cells while almost all cells were LTR-active for others. Clones varied over four orders of magnitude in virus release per active cell. Proviruses in largely LTR-active clones were closer to preexisting enhancers and promoters than low LTR-active clones. Unsurprisingly, major *vpr*+ clones contained fewer LTR-active cells than *vpr*- clones, and predominant *vpr*+ proviruses were farther from enhancers and promoters than those in *vpr*- pools. Distances to these marks among intact proviruses previously reported for ART-suppressed patients revealed that patient integration sites were more similar to those in the *vpr*+ pool than to *vpr*- integrants. Complementing *vpr*-defective proviruses with *vpr* led to rapid loss of high LTR-active clones, indicating that the effect of Vpr on proviral populations occurred post-integration. However, major clones in the complemented pool and its *vpr*- parent population did not differ in burst sizes. When the latency reactivation agents prostratin and JQ1 were applied separately or in combination, *vpr*+ and *vpr*- population-wide trends were similar, with dual treatment enhancement due in part to reactivated clones that did not respond to either drug applied separately. However, expression

[2] This chapter has been published as Atindaana, E., Kissi-Twum, A., Emery, S., Burnett, C., Pitcher, J., Visser, M., Kidd, J.M. and Telesnitsky, A., 2022. Bimodal Expression Patterns, and Not Viral Burst Sizes, Predict the Effects of Vpr on HIV-1 Proviral Populations in Jurkat Cells. *Mbio*, pp.e03748-21.

signatures of individual clones differed between populations. These observations highlight how Vpr, exerting selective pressure on proviral epigenetic variation, can shape integration site landscapes, proviral expression patterns, and reactivation properties.

Introduction

HIV-1 establishes stable reservoirs in patients treated with antiretroviral therapy (ART), which consist of cells containing replication competent proviruses that are not cleared by the immune system and that can rekindle spreading infection (Bailey, Blankson et al. 2004). It is generally assumed that the latent reservoir consists of proviruses that are transcriptionally silent, but what causes this is unclear. Reservoir establishment and maintenance is multifaceted and may involve intracellular depletion of transcription factors, viral regulatory protein deficiencies, provirus integration position effects and epigenetic variation (Verdin, Paras et al. 1993, Emiliani, Van Lint et al. 1996, Emiliani, Fischle et al. 1998, Kulkosky, Sullivan et al. 2004, Lassen, Han et al. 2004, Archin and Margolis 2006, Williams and Greene 2007, Han, Lin et al. 2008, Lenasi, Contreras et al. 2008, Blazkova, Trejbalova et al. 2009, Mbonye, Wang et al. 2018). The oligoclonal nature of ART-suppressed patients' proviruses suggests the reservoir's long-lived nature is at least partially due to infected cells' proliferation, which may be either homeostatic or driven by T cell receptor engagement (Wang, Gurule et al. 2018, Mendoza, Jackson et al. 2020, Simonetti, Zhang et al. 2021).

Fundamental aspects of latency in patients, such as the size of the latent reservoir, remain poorly defined. Method improvement for reservoir quantification is ongoing, as recent studies have demonstrated that quantitative viral outgrowth assays (QVOA) underestimate the latent reservoir while PCR-based quantification can overestimate it due to the predominance of defective proviruses (Eriksson, Graf et al. 2013, Bruner, Wang et al. 2019, Wonderlich, Subramanian et al. 2019). Experimental approaches for reactivation must sometimes be applied multiple times to achieve reactivation *ex vivo* (Ho, Shan et al. 2013). Moreover, reported genetic dissimilarities

between *ex vivo* outgrowth virus and reemergent viremia suggest existing reactivation approaches may not accurately detect rebound-competent virus (Cohen, Lorenzi et al. 2018, Lu, Pai et al. 2018). Large interclonal differences in virion release per T-cell and variability in virion release upon latency disruption may further complicate reservoir size determination (Bui, Mellors et al. 2016, Hataye, Casazza et al. 2019).

The rarity of latently infected cells in virally suppressed individuals makes the study of *in vitro* latency models necessary. Some tissue culture models for HIV-1 latency and reactivation, such as the widely used J-Lat clones, assess LTR promoter activity by reporter gene expression but lack genetic elements believed dispensable in HIV-1 latency (Jordan, Bisgrove et al. 2003, Pace, Agosto et al. 2011, Darcis, Kula et al. 2015). One such element is the *vpr* gene, whose product plays roles in viral infectivity *in vivo* (Eckstein, Sherman et al. 2001, Kino, Gragerov et al. 2002, Sato, Misawa et al. 2013, Zhang, Kang et al. 2021) but also causes cell-cycle arrest and can induce widespread changes in host gene expression (Yao, Mouland et al. 1998, Zhang and Bieniasz 2020, Bauby, Ward et al. 2021). Many latency models use *vpr*-defective proviruses which, cultured over time to allow proviral silencing, can be used for reactivation studies (Jordan, Defechereux et al. 2001, Chen, Martinez et al. 2017, Fennessey, Pinkevych et al. 2017, Jefferys, Burgos et al. 2021). However, in such systems, the extent to which proviral quiescence represents the silencing and outgrowth of a subset of integrant clones versus global proviral silencing is unknown.

Cell-based latency models have been used to pilot candidate cure strategies, including “shock and kill” and “block and lock” approaches. In contrast to ART, which prevents viral spread, the shock and kill method involves inducing virus expression with the intention that this will lead either to

cytopathic death of reactivated cells or to host immune recognition and infected cell clearance (Deeks 2012). Candidate latency reactivation approaches that perturb cellular pathways or complement intracellular deficiencies in experimental models include the use of prostratin, which stimulates T cells without inducing cellular proliferation and increases the level of NFkB (Korin, Brooks et al. 2002, Williams, Chen et al. 2004). Other latency-reversing agents (LRAs) include those that act to increase the level of pTEFb, including the BET bromodomain inhibitor JQ1 (Boehm, Calvanese et al. 2013) as well as treatments that modify the chromatin environment, such as histone deacetylase inhibitors like suberoyl anilide hydroxamic acid (SAHA) and entinostat (Archin, Espeseth et al. 2009, Contreras, Schweneker et al. 2009, Edelstein, Micheva-Viteva et al. 2009, Zhu, Gaiha et al. 2012, Jiang, Mendes et al. 2014). In contrast, the block and lock approach seeks to permanently silence proviruses to prevent their reactivation (Li, Mori et al. 2021).

Molecules that are effective at reversing latency in various tissue culture models have been identified, but evidence thus far suggests that these are either too toxic to be therapeutically useful or that they fail to reduce the size of the latent reservoir in patients (Archin, Espeseth et al. 2009, Yang, Xing et al. 2009, Xing, Bullen et al. 2011, Archin, Liberty et al. 2012, Spivak, Andrade et al. 2013). This discordance may be due in part to differences among cell-based latency models, as some use infectious virus while others use viral reporters, and some use clonal integrants while others use heterogeneous proviral populations (Pace, Agosto et al. 2011). Thus, the inconsistent results in LRA reactivation studies using these models may reflect that each model captures at best a subset of the expression properties that contribute to the latent reservoir (Spina, Anderson et al. 2013). And whereas effects on heterogeneous populations of integrants have been examined

extensively, the contributions of individual clones' behavior to aggregate population responses largely have not been explored.

In the current study, HIV-1 expression properties of hundreds of individual integrant clones were compared within polyclonal populations of barcoded proviruses. The influence of integration sites and the presence or absence of *vpr* on populations' clonal structures and on their reactivation dynamics were investigated. Together these findings suggest that Vpr's cytotoxic activity disproportionately affects a distinct subset of infected cells and plays critical roles both in shaping transcriptionally inactive proviral populations and in defining their reactivation potential.

Results

vpr- and vpr+ proviral pools differed in numbers of LTR-active cells

To study the effects of Vpr on the responsiveness of individual proviruses in the absence of virus spread, Jurkat cells were infected with distinguishable *vpr+* or *vpr-* versions of the NL4-3 derivative shown in Figure 3.1A (Read, Atindaana et al. 2019). Using the EF1α promoter to drive constitutive expression of the puromycin resistance gene allowed selection of infected cells independent of LTR expression. Each genomic RNA in an infecting virus contained a unique 20b randomized "barcode" inserted in U3, which was duplicated in both LTRs after reverse transcription and enabled tracking individual proviruses. We refer to the barcodes as "zip codes" because in the context of proviruses, they report the genetic neighborhoods of each integrant. Infected cell populations were passaged without cell cloning, thus generating polyclonal integrant populations within which transcriptionally active cells were identified using LTR-driven GFP expression or by progeny genomic RNA released in Env- virions.

It has long been recognized that at least some HIV-1 integrants establish clones in which a fraction of the total daughter cells possess active LTRs while proviruses are silenced in other sibling cells (Weinberger, Burnett et al. 2005). Previous work using zip coded *vpr*- derivatives of the vector used here has shown that for each integrant, the clone gives rise to a mixture of cellular progeny that includes some GFP+ cells and some GFP- cells, and that over time sibling cells switch between LTR-active (GFP+) and -inactive (GFP-) expression phenotypes while maintaining LTR-active and -inactive cells in stable equilibrium proportions (Read, Atindaana et al. 2019). To confirm that individual integrant clones in the current study also contained clone-specific mixtures of GFP+ and GFP- cells, single cells were isolated by limiting dilution from the *vpr*- pool and expanded, and cells from each clone were then subjected to flow cytometry (Figure 3.1B). Consistent with previously reported differences in bimodal expression patterns among proviral clones (Read, Atindaana et al. 2019), the results showed that each clonal pool was comprised of both GFP+ and GFP- cells, with LTR-active proportions discernable by GFP+ percentages (%GFP+) that differed among the clones: mostly GFP+ cells for some clones (high %GFP+ clones), and distinctly different GFP+ percentages for others (Figure 3.1B). To confirm that these cells were clonal and contained single proviruses, and thus that observed mixed cell phenotypes did not reflect the presence of more one than one clone, PCR amplicons from these clones were Sanger-sequenced without subcloning and determined to display unique sequences, indicating that at least a large majority of each clones' cells contained a unique zip code (See Materials and Methods).

In the current study, four independent polyclonal integrant pools were established by infecting roughly $5x10^6$ Jurkat T cells at a low multiplicity of infection (<0.0005) to ensure puromycin

resistant infected cells contained only one provirus per cell. Two of the pools contained *vpr*+ proviruses and two had integrants lacking *vpr*. Flow cytometry analysis after fourteen days of expansion showed significantly fewer GFP+ cells in the two *vpr*+ infected cell pools than in the two *vpr*- provirus pools ($p=0.042$, paired t-test; <5% vs. 73% GFP+, respectively (Figure 3.1C and 3.1D). To address the possibility that uninfected cell outgrowth might have contributed to GFP- cell populations, an uninfected Jurkat cell control was cultured in parallel and subjected to the same schedule of puromycin treatment and selection-free media exposure as the infected cells. After a total of 14 days, these uninfected control cultures were analyzed by flow cytometry to ensure no surviving uninfected cells were detectable that might confound observations. In an additional test of the possible presence of uninfected cells, dual aliquots of each pool were analyzed about 30 days after the pool establishment. One aliquot was cultured with puromycin and the other left untreated. When analyzed by propidium iodide staining for cell death, no difference was observed between treated and untreated pool controls.

Fourteen days post-infection *vpr*- and *vpr*+ cell pools were each sorted into GFP+ and GFP- subpools, and cellular DNA was harvested from an aliquot of each subpool immediately after sorting. Integrants' zip codes were amplified from the cellular DNA samples in at least two separate PCR reactions per cellular DNA sample, and the subpools' zip code contents were catalogued by high throughput sequencing. Correlation analysis for the fractional abundances of specific zip codes in replicate reactions showed reproducible zip code detection (Figure S3.1).

After analyzing roughly three million sequencing reads per pool, zip codes were ordered by read abundance. Determining how many unique barcodes were present in each GFP+ pool revealed that

similar *numbers* of zip codes (approximately 2000)—each indicative of a single integration event—were detected in the GFP+ sorted cells from both *vpr*+ and *vpr*- pools, even though the GFP+ cell fraction of the *vpr*+ pool was very small. Because all four pools had been infected at the same multiplicity of infection, this finding suggested that the number of GFP+ integrants initially established did not differ markedly between pools, which is consistent with previous work in dendritic cells that indicates that the extent of proviral integration does not differ depending on the presence or absence of Vpr (Miller, Akiyama et al. 2017).

Combining population-wide percentages of GFP+ cells in the unsorted cells, as determined by flow cytometry, with zip code read counts within sorted sub-pool libraries allowed calculating the percentage of LTR-active cells (%GFP+ values) within each cell clone (Figure 3.2A). Clones' %GFP+ values were calculated using zip code abundances in FACS sorted GFP+ and GFP- cells' DNA and normalizing to population-wide GFP positivity levels at sort time, as described in Material and Methods. Notably, for the top 500 most abundant clones in the unfractionated *vpr*+ and *vpr*- pools, median %GFP+ was significantly lower in the *vpr*+ pool ($p=0.0001$, Mann-Whitney U two-tailed test), indicating that most cells in the *vpr*+ pool were members of low LTR-active clones (Figure 3.2A). When GFP+ sorted cells from *vpr*+ and *vpr*- pools were cultured further, no viable *vpr*+ GFP+ cells were detected after three days, and thus although integrated zip codes in the freshly sorted cells were determined, the expression properties and clonal structures of *vpr*+ GFP+ cells could not be analyzed further. These results indicated that integration events that resulted in GFP expression were equally likely in *vpr*+ and *vpr*- proviral pools when the cells were examined early post-infection. However, the depletion of the *vpr*+ GFP+ cell sub-populations

suggested that whenever cells with *vpr*+ proviruses switched phenotypes from LTR-inactive to LTR-active, the resulting GFP+ cells died upon further cell culturing.

Vpr shaped the clonal structure of infected pools

Observing fewer GFP+ cells in the *vpr*+ pool than in cells with *vpr*- proviruses was not surprising due to Vpr's well-known cytotoxic effects. Along with *env, tat, nef,* and *vpu, vpr* is one of the HIV-1 genes reported to be cytotoxic in at least some context (Bartz and Emerman 1999, Ohagen, Ghosh et al. 1999, Xu, Laffert et al. 1999). However, despite Vpr's cytotoxicity, a small fraction of *vpr*+ GFP+ cells were observed among unsorted *vpr*+ cells, even though sorted *vpr*+ GFP+ cells did not survive three days of culture.

One plausible reason unsorted *vpr*+ populations contained rare GFP+ cells was that these cells arose via recent phenotypic switches from clones that were largely inactive. To test this possibility, zip codes in GFP+ and GFP- sorted subpools were compared for both *vpr*+ and *vpr*- integrants (Figure 3.2B). Ordering zip codes by their abundance in the unsorted pools and analyzing those that comprised the top 500 revealed that among *vpr*- cells, about 73.2% of all the zip codes were found in both subpools, while 4% were observed only in GFP- cells and 22.8% only in the GFP+ subpool. In contrast, 86.2% of the *vpr*+ cells' zip codes were found in both subpools, with 13.8% observed only in the GFP- subpool, and none of the zip codes found exclusively in the GFP+ subpool. This suggested that the small fraction of GFP+ cells in the *vpr*+ pool (Figure 3.1C, D; *vpr*+) resulted from the recent acquisition of LTR expression by cells from the larger GFP- cell pool. If the *vpr*- pool is assumed to be relatively representative of a population that results when

all clones are equally viable, this suggests that when initially integrated, proviruses whose daughter cells were largely or always GFP+ were the dominant class of clones.

Clones differed in LTR-active cell burst sizes

Because the proviral vectors used in this study contained most HIV-1 genes and expressed packageable RNAs as well as a GFP reporter, expression properties could be measured both by GFP expression and by virion release. To determine the amount of virus released per LTR-active cell in each clone, total virus release from the *vpr*- GFP+ fractionated cell pool was quantified by p24 release; the limited viability of the *vpr*+ GFP+ cell fraction prevented this analysis for the *vpr*+ pool. Relative zip code amounts in virion RNA and in genomic DNA extracted from the GFP+ cells were then determined and used to calculate the amount of virus released per LTR-active cell for each clone. These results are indicated in Figure 3.2C, with each clone's release per LTR-active cell presented on the Y axis and clones ordered by their member cells' %GFP+ on the X axis. All clones in Figure 3.2C are represented in light grey; regions that appear to display darker shading indicate the presence of multiple clones at the same coordinates. We and others have previously reported significant differences among HIV-1 infected cell clones in the amount of virus released per activate cell (Bui, Mellors et al. 2016, Read, Atindaana et al. 2019), and the results here indicated that burst sizes for the integrant clones in the current study ranged over four orders of magnitude.

Notably, burst sizes were slightly negatively correlated with %GFP+ (Pearson's R = -0.28) (Figure 3.2C), indicating that clones that displayed higher %GFP+ tended to produce fewer virions per LTR-active cell than low %GFP+ clones. Because the intactness of these proviruses was not

addressed directly, it is possible that some of the observed variation in virus release reflected provirus defects. However, because this work quantified encapsidated viral RNAs, even low virus-expressing proviruses must have remained largely intact, as proviruses with large internal deletions would have lost the ability to assemble viral particles. Another possible contributor to these differences may be that large burst sizes conferred some selection against high LTR-active clones during the weeks of infected cell passage that preceded these measurements. Although Vpr, which was not present in these cells, is the HIV-1 protein best recognized as cytotoxic, expression of the retained HIV-1 proteins or other elements may attenuate high %GFP+ clones in more subtle ways.

High LTR-active clones' proviruses were in closer proximity to genome marks associated with active gene expression

The finding that LTR-active *vpr*+ cells were rapidly lost suggested that proviruses that had integrated into more active regions of the host genome might be selected against within polyclonal *vpr*+ populations. To examine this notion, we compared the proximities of *vpr*+ and *vpr*- integration sites to genomic features associated with active gene expression. First, zip code integration sites were determined using cellular DNA harvested ten days after establishment of each pool and mapped to 1171 and 1121 unique sites in the human genome for *vpr*+ and *vpr*- pool members, respectively (see Methods). Then, the proportion of zip codes located in genes versus those not in genes (as defined by ENCODE for Jurkat cells (ENCODE)) were compared (Figure 3.3A, S3.2 Figure For first and duplicate pools respectively). Consistent with previous reports (Poon and Chen 2003, Sherrill-Mix, Lewinski et al. 2013, Vansant, Chen et al. 2020), the results indicated that similar majorities of integrants were established within genes regardless of whether or not *vpr* was present. Next, *vpr*+ and *vpr*- integrants were compared for their proximities to

specific genome marks associated with active gene expression that have been reported to pre-exist in Jurkat cells (Consortium 2012). No differences were found in distances to closest DNase I sensitivity sites, which are associated with open chromatin (Mann-Whitney U test, p=0.1854) (Figure 3.3B). However, proximities to H3K27ac (associated with enhancers) (Mann-Whitney U two-tailed test, p<0.0001) and H3K4me3 (associated with active promoters) (Mann-Whitney U two-tailed test, p<0.0001) marks differed significantly, with proviruses from *vpr*+ pools somewhat farther from these marks (Figure 3.3C-D).

To correlate clones' LTR-active proportions with their proximities to active chromatin marks, distances from these marks were compared for high- and low-LTR active clones. Few clones displayed intermediate %GFP+ levels (Fig 3.2A), and all major clones in the *vpr*+ pool had clonal %GFP+ proportions of 30% or less. Therefore, only integrants from the *vpr*- pool were examined, and low LTR-active clones (%GFP+ < 30%) were compared to high LTR-active clones (%GFP+ >= 60%) (Figure 3.4A). This analysis showed that distances to both H3K27ac and H3K4me3 marks were shorter among high LTR-active clones than low LTR-active clones, albeit not dramatically, suggesting that the difference in integration site proximities observed between *vpr*+ and *vpr*- proviruses may be due to the survival of high LTR-active clones in the absence of Vpr.

Because the *vpr*+ pool was devoid of high LTR-active clones, its high- and low LTR-active integrants could not readily be compared. However, with the reasoning that abundances of residual high LTR-active clones would gradually diminish over time, the dynamics of the v*pr*+ pool were examined by comparing samples harvested two weeks apart. Examination of the 100 most abundant zip codes in the unsorted *vpr*+ cell pool showed that only 68 of the top 100 zip codes

present on day 10 post-infection (Time Point 1) were observed among the 100 most abundant zip codes on day 24. When integration site proximities to H3K4me3 and H3K27ac marks were compared, the 32 zip codes that were absent from the top-100 on day 24 (Time Point 2) (Figure 3.4B) were significantly closer to H3K4me3 (p=0.0193, Mann-Whitney U two-tailed test), and tended to be closer to H3K27ac marks (although not with statistical significance: p=0.1581, Mann-Whitney U two-tailed test) than the 68 Time Point 1 clones that remained within the top-100 most abundant clones at Time Point 2.

The spectra of LRA responses were similar for vpr+ and vpr- populations, but clonal behaviors differed

To address possible functional consequences of *vpr*, the reactivation properties of LTR-inactive *vpr*- and *vpr*+ subpools were compared. The LRAs prostratin, a protein kinase C agonist, and JQ1, a bromodomain inhibitor, were applied separately or in combination to the GFP- sub-populations of each pool for 24 hours. Reactivation for each treated GFP- population was monitored both by changes in the frequency of GFP+ cells, as measured using flow cytometry, and by quantifying virus release (Figure 3.5A-B; S3.3A-B Figure). The results indicated that compared to single prostratin and JQ1 treatments, dual treatment resulted in additive levels of reactivation in both *vpr*+ and *vpr*- populations by both criteria (Figure 3.5A shows reactivation monitored by GFP+ cells and Figure 3.5B shows virus release; left panel indicates reactivation for *vpr*+ and right panel for *vpr*-). In dual LRA treatments for both pools, there was an approximately 4-fold increase in GFP+ cells, while virus release increased by approximately 30-fold relative to DMSO control. The most significant difference between the *vpr*- and *vpr*+ pools was that the absolute amount of virus release upon treatment was 3-fold higher in *vpr*- pools, and the responsiveness to JQ1 was lower

in the *vpr*+ pools. These differences between pools in their extents of reactivation were not due to differences in cell viability (Figure S3.4C). The observation that reactivation was enhanced by dual prostratin and JQ1 treatment is consistent with previous works by Boehm *et al.,* and Darcis *et al.,* using the same drugs in cell culture models of latency and *ex vivo* treated cells from HIV-1 patients respectively (Boehm, Calvanese et al. 2013, Darcis, Kula et al. 2015).

Next, the behaviors of individual proviral clones within the populations were determined. Virus release from the treated cells was quantified by p24 equivalent and cDNA was generated using virion genomic RNA upon LRA treatment. Zip codes were amplified from the viral cDNA and also from an untreated aliquot of the GFP- cells' DNA, and zip code libraries were high through-put sequenced. The results were normalized to calculate average virus release per treated cell for each clone (Figure 3.5C; Figure S3.3D).

This analysis revealed that many clones were not detectably reactivated. For example, of the most abundant five hundred clones in the *vpr*- pool, sequencing virion RNA after reactivation treatment revealed no evidence of virus production for 102 of these abundant clones in the dual prostratin and JQ1-induced virion cDNA pools. When cells were treated with both drugs, some clones in both the *vpr*+ and *vpr*- pools displayed enhanced virus release per treated cell compared to single treatment conditions. Interestingly, both proviral pools included a subset of clones that were not detectably reactivated by either prostratin or JQ1 when the LRAs were applied alone, but that were reactivated upon dual LRA treatment.

Surprisingly, and only in the *vpr*- pool, reactivation levels observed under dual LRA conditions were lower than those observed with single LRA use for a subset of clones (Figure 3.5C: right panel. Compare, for example, the top dozen rows in the *vpr*- columns to those for the *vpr*+ clones). These same patterns of reactivation were observed when experiments were repeated using a second set of independently established *vpr*- and *vpr*+ pools (Figure S3.3D). A possibility suggested by this pattern of reactivation is that Vpr degrades a negative regulator of HIV-1 gene expression that is induced in Jurkat cells by prostratin treatment. And indeed, Vpr has been reported to cause depletion or mislocalization of several factors that can repress HIV-1 expression, including Class I histone deadenylases, the transcription factor ZBTB2, and the negative regulator of pTEFb, CTIP2 (Le Douce, Herbein et al. 2010, Cherrier, Le Douce et al. 2013, Romani, Kamali Jamil et al. 2016, Bruce, Bracken et al. 2021).

Complementation of the vpr-defective pool led to depletion of high LTR-active clones but did not discriminate against high burst size clones

As an additional test of the effects of *vpr* on polyclonal population composition, a functional copy of the *vpr* gene was added to cells harboring *vpr*- proviruses 16 days post-infection. This allowed comparing the behaviors of each of several hundred specific individual integrants in the presence or absence of *vpr*. This was achieved by transducing the *vpr*- pool with a *tat*-deficient lentiviral vector containing LTR driven *vpr,* which also contained the fluorescent marker *mKO* expressed from the constitutive SV40 promoter. This resulted in expression of Vpr only when a pre-existing *vpr*- provirus was transcriptionally active. mKO expression was used as a proxy for *vpr* vector transduction, and enabled cells that were successfully transduced to be sorted in the phycoerythrin (PE) channel (Figure 3.6A). 48 hours post-transduction, an unsorted aliquot was saved while the

remaining cells were sorted to identify GFP- PE+ cells (Figure S3.4A-B) for work described below.

Comparing the proportions of GFP+ cells in the parental *vpr*- pool to the proportions in the same pool "reverted" *to vpr*+ by introduction of the *vpr* vector (hereafter referred to as the vpr^{rev} pool) showed a marked decrease in GFP+ cells between day-2 and day-14 post transduction (Figure 3.6B). Although it cannot be ruled out that some feature of the transduction vector other than *vpr* influenced the cells' properties, the selective loss of GFP+ transduced cells was consistent with predictions based on observations with the *vpr*+ pool, which showed that LTR-active cells were depleted in the presence of Vpr.

Proviral zip code abundances in unsorted *vpr*- and vpr^{rev} pools were then compared. Of the 500 most abundant clones in the *vpr*- pool, zip codes were split into two groups based on observed fold-change in relative abundance: those that were reduced by 10-fold or more in vpr^{rev} (reduced clones), and those with no observed reduction in their relative abundance upon *vpr* addition (not reduced). A comparison between the "reduced" and "not reduced" groups revealed that nearly all *vpr*- pool clones that displayed high %GFP+ --that is, clones in which most member cells displayed LTR activity-- were reduced in the vpr^{rev} pool (Figure 3.6C).

In contrast, when the burst sizes of individual clones present in the vpr^{rev} population, as determined above for the parental *vpr*- pool (Figure 3.6D), were compared to the clones that were reduced versus for those that were not reduced upon complementation, no differences in the amount of virus released per active cell were detectable between groups.

The *vpr*^{rev} pool was then used to address whether *vpr*- pools rendered *vpr*+ by complementation displayed reactivation patterns similar to those of the original *vpr*+ pools. Transduced GFP- *vpr*^{rev} subpools were subjected to prostratin, JQ1, or dual treatment and analyzed by flow cytometry and virus release (Figure 3.6E). In general, the magnitude of reactivation as measured by flow cytometry did not differ significantly compared to no-drug control in the transduced cells (left panel). However, despite significant increase in virus release for dual treated samples (right panel), the magnitude was diminished relative to parental *vpr*- GFP- cells, with cells responding less to JQ1 treatment than they did before Vpr addition. Consistent with expectations if the phenotypes above reflected Vpr, transduced cells showed enhanced reactivation upon dual LRA treatment (Figure 3.6E; Figure S3.4C). Notably, clones (such as clones numbered 53, 133, 161, 293, 420, etc.in Figure 3.6F) that were highly responsive to JQ1 treatment in the parental untransduced *vpr*- pool, and that differentiated the LRA responsiveness of the *vpr*- pool from the *vpr*+ pool, were highly represented among those lost upon *vpr* addition.

Genomic features of persistent intact proviruses in patients on ART are more similar to vpr+ than vpr- pool members

Although patients' *vpr* genes are not routinely sequenced, it seems likely that most proviruses in HIV-1 patients on ART contain intact Vpr and have survived its selective pressures. Having observed differences between *vpr*+ and *vpr*- viruses in cultured cells in terms of their reactivation patterns and proximity to genome features, we sought to determine which class of these proviruses more closely resembled those in persistent clinical isolates.

To do this, previously published data on patients' intact provirus integration sites (Einkauf, Lee et al. 2019) were analyzed for their proximity to H3K4me3, H3K27ac, and DNase I hypersensitivity sites that have been reported for Jurkat T cells and primary CD4+ T cells (see Methods) (Consortium 2012). When these proximities were compared to those in the *vpr*- and *vpr*+ pools established here, prominent proviruses in the *vpr*- pool, and not those in the *vpr*+ pool, were found to significantly differ from the patient proviruses in their proximities to H3K4me3 and H3K27ac marks (Figure 3.7A-B; Figure S3.5), but not DNase I hypersensitivity sites (Figure 3.7C). These results suggested that the integration site distribution of persistent proviruses in patients across the tested genome marks was similar to those observed in *in vitro* established *vpr*+ proviral populations but significantly different from what was observed within *vpr*- proviral pools.

Discussion

Here we describe how Vpr disrupts polyclonal provirus population structures and alters the expression properties and latency reversing agent responsiveness of residual proviral populations in cultured cells. The results illustrate the functional significance of HIV-1's bimodal expression phenotypes in shaping proviral populations and show that whereas Vpr's cytotoxicity will lead to the rapid depletion of clones constitutively or frequently expressing HIV-1 genes, *vpr*+ proviruses that are capable of supporting brief robust bursts of virion production can readily persist in proliferating infected cell populations *in vitro*.

Cell-based models are critical to HIV-1 persistence work, but they differ by cell type, form of HIV-1 used, whether they use clones or polyclonal populations, and other parameters. Many experimental latency systems include reporter genes and/or delete viral genes believed

unnecessary for silencing and reactivation (Fennessey, Pinkevych et al. 2017, Marsden, Zhang et al. 2020, Jefferys, Burgos et al. 2021). For example, two prominent studies that used barcoded proviruses to track virus dissemination in animals used *vpr*- proviruses (Fennessey, Pinkevych et al. 2017, Marsden, Zhang et al. 2020). To address the consequences of such exclusions, the current study compared proviral populations with and without *vpr.* Note, however, that all vectors here also lacked *env* and *nef,* which are cytotoxic in some conditions and affect cell physiology in many ways, and thus these omissions may have biased outcomes (Fujii, Otake et al. 1996).

We have previously studied interclonal variation in HIV-1 gene expression using barcoded proviruses (Read, Atindaana et al. 2019). That work revealed a broad range of expression variation among integrant clones. When virions produced by barcoded integrants were pseudotyped and polyclonal progeny generated, expression properties associated with specific barcodes in the first-generation were lost, demonstrating that expression variation was non-deterministic and may be influenced by position effects. If similar variation exists *in vivo*, proviruses during suppressive ART may exist in epigenetic quasispecies on which selection can act (Domingo, Sheldon et al. 2012). Vpr was used as a source of selective pressure on infected cell populations here.

In our previous study, we compared expression patterns of barcoded *vpr*- proviruses in primary CD4+ T cells to those in Jurkat T cells, and observed indistinguishable spectra of expression patterns (Read, Atindaana et al. 2019). Although primary CD4+ T cells are a closer system to natural infection, these cells only survive 2-4 weeks *ex vivo* unless treated with antiapoptotic agents (Xing, Bullen et al. 2011, Kim, Hosmane et al. 2014) and do not proliferate unless stimulated in ways that can affect proviral expression patterns. Thus, because extended cell passaging was

required and retaining expression patterns associated with initial integration events was desirable, we chose to work with Jurkat cells here, which limits the physiologic relevance of our findings. Note, however, that Vpr reportedly exerts similar effects in primary CD4+ and Jurkat T cells (Gummuluru and Emerman 1999).

Our and others' work has shown that individual proviral clones can contain mixtures of LTR-active and -inactive cells (Weinberger and Weinberger 2013, Battivelli, Dahabieh et al. 2018, Read, Atindaana et al. 2019, Telwatte, Morón-López et al. 2019). Furthermore, the clonal progeny of individual infected cells can shift from being silenced to expressing their proviruses, while maintaining overall proportions of cells with active proviruses (Weinberger and Weinberger 2013, Battivelli, Dahabieh et al. 2018, Read, Atindaana et al. 2019). For example, in polyclonal *vpr*- populations, individual member cells within each clone either express HIV-1 genes, as monitored by GFP reporter, p24 staining, and virion release, or they do not. Although some diminution of LTR expression is observed over time, for the most part each clone adopts a stable, heritable pattern of bimodal gene expression (Read, Atindaana et al. 2019).

Earlier reports suggest that HIV-1's bifurcating expression profiles exist in a minority of clones (Weinberger, Burnett et al. 2005), and that stochastic fluctuations in gene expression enable probabilistic LTR on/off fate decisions that are initially unstable but become stabilized by post-transcriptional feedback mechanisms (Hansen, Wen et al. 2018). Our observations of bimodal HIV-1 expression patterns likely describe a similar phenomenon, although the majority of our clones display mixed expression phenotypes and these phenotypes interchange over time in

replicating cell populations, suggesting that they may exist in an oscillating circuit (Weinberger, Burnett et al. 2005, Rosier and De Greef 2015, Read, Atindaana et al. 2019).

The zip coded provirus approach used here enabled comparing *vpr*+ and *vpr*- provirus-containing cell pools. Not surprisingly, *vpr*+ pools contained significantly fewer GFP+ cells (indicative of transcriptionally active proviruses) than did *vpr*- pools. Dominant *vpr*+ clones included very few member cells that expressed HIV-1 genes (ie: they were low LTR-active clones) whereas high LTR-active clones dominated the *vpr*- pool.

These findings may help explain inconsistences in previous estimates of the fractions of HIV-1 infected cells that are transcriptionally active. Using polyclonal *vpr*+ proviruses *in vitro*, Dahabieh *et al*. suggested that most integrated proviruses are silent, even though it has been estimated that only a small fraction of infected patient cells are latent (Stewart, Poon et al. 1997, Yao, Mouland et al. 1998, Dahabieh, Ooms et al. 2013). Consistent with Dahabieh *et al.,* we found that HIV-1 is transcriptionally inactive in most cells that persisted in *vpr*+ populations, with this reflecting the selective survival and relative amplification of low LTR-active clones. However, the results showed that low LTR-active clones with large burst sizes were not appreciably depleted. This finding is striking considering that GFP+ cells--cells that in a binary sense would score LTR-active-- displayed inter-clonal variation in burst size that spanned four orders of magnitude.

These observations support the following model (Figure8): initial integration site distribution will be Vpr-agnostic, as previously demonstrated (Miller, Akiyama et al. 2017). As infected cells proliferate to form clones (Wang, Gurule et al. 2018, Mendoza, Jackson et al. 2020, Simonetti,

Zhang et al. 2021), some daughter cells will be LTR-active while other cells will be LTR-inactive (Weinberger, Burnett et al. 2005, Rosier and De Greef 2015, Read, Atindaana et al. 2019). Each clone will maintain a clone-specific equilibrium population of LTR-active vs -inactive cells over time, which is determined at least in part by integration site features. However, within clones, individual member cells' phenotypes are transitory, with cells alternating between LTR on and off states (Weinberger, Burnett et al. 2005, Rosier and De Greef 2015, Read, Atindaana et al. 2019). For high LTR-active clones, most cells will be LTR active and subject to Vpr's cytopathic effects. In contrast, for low LTR-active clones--even clones that display very high burst sizes when cells are LTR-active-- clonal proliferation will proceed largely unimpeded.

The *vpr*+ and *vpr*- proviral populations and their constituent clones here were compared for latency reactivation agent responsiveness, using LRAs selected based on their reported ability to induce HIV-1 to levels comparable to maximum reactivation by T cell activation (Laird, Bullen et al. 2015). Overall pool trends for the *vpr*+ and *vpr*- proviral populations were similar, although responses differed in magnitude. In contrast, zip code analysis revealed that the behavior of individual clones within *vpr*+ and *vpr*- populations differed. What caused observed reduced levels of reactivation in the *vpr*+ pools was not determined. The widespread transcriptomic changes Vpr induces in CD4 T cells may have played a role (Bauby, Ward et al. 2021), and surviving LTR-inactive clones may have included defective proviruses. Another possibility is that selective pressure exerted by Vpr may have depleted clones prone to high levels of reactivation. Consistent with this notion, proviruses farther from genome marks associated with active gene expression-- a feature associated with *vpr*+ clones in the present study-- are reportedly relatively resistant to

reactivation (Einkauf, Osborn et al. 2022). The adoption by the *vpr*^{rev} pools of the *vpr*+ pool's patterns of LRA responsiveness further corroborated the impact of Vpr on reactivation properties.

Comparing integrants that dominated passaged *vpr*+ and *vpr*- pools suggested Vpr selects against proviruses proximal to genome marks associated with active gene expression. To address the relationship of these observations to persistent populations *in vivo*, we mapped distances to pre-existing marks for previously published integration sites of intact, inferred to be intact, or large proviruses in three patients on antiretroviral therapy (ART) (Einkauf, Lee et al. 2019). This analysis indicated that the proximities observed in patients were significantly more similar to those of *vpr*+ proviruses than to *vpr*- pool members.

While proviral expression and latency have traditionally been viewed as mutually exclusive states, recent evidence, such as the selective depletion of patients' intact proviruses over time and detectable HIV-1 RNAs in unstimulated latent cells, suggests that both basal expression and proviral inducibility may be heterogenous, even within individual cell types (Neidleman, Luo et al. 2020, Peluso, Bacchetti et al. 2020). Based on their study of replication competent non-induced proviruses, Ho and colleagues noted that certain proviruses' expression was infrequent despite T cell activation, suggesting induction of latent proviruses might be a stochastic process (Ho, Shan et al. 2013). Stochastic modeling coupled with laboratory experimentation in a separate study provided evidence that the transition from HIV-1 latency to viral outgrowth is a stochastic property (Hataye, Casazza et al. 2019). These examples of unanticipated expression patterns *ex vivo*, and their conclusions of HIV-1 expression stochasticity, were based in part on a failure to establish causal links between conditions believed to promote activation and its occurrence.

A possibly related but little-discussed problem observed with transcriptionally silenced *in vitro* latency models is that they tend to "spontaneously revert" during passage and generate small proportions of cells with transcriptionally active proviruses (McNamara, Ganesh et al. 2012, Battivelli, Dahabieh et al. 2018). Our previous work and findings here suggest an explanation slightly different from "spontaneous reversion". Previous quantification of phenotypic switching within expression-sorted polyclonal populations demonstrated that the phenotypes of active- and inactive- subpopulations of proviral clones are not static (Read, Atindaana et al. 2019). Instead, these phenotypes equilibrate, such that most LTR-inactive cells are not stably silenced but rather represent the transient off-state of an on-off continuum. Thus, it seems conceivable that "spontaneous reversion" by a small subset of latency model cells, like the low proportion of GFP+ cells in the current study's *vpr*+ pool, might more accurately be viewed as the programmed equilibration of a low LTR-active clone than the outlier behavior in a subset of silenced cells.

Burst size variation similar in magnitude to that described here is observed among patient samples reactivated *ex vivo (Bui, Mellors et al. 2016),* and it seems likely that at least vestiges of bimodal expression are retained during natural infection. Weinberger and colleagues have proposed that HIV-1's bifurcating expression phenotypes may represent a bet-hedging strategy, wherein some daughter cells support robust virus production while others remain quiescent, thus allowing clonal persistence whether or not the environment is conducive for viral expression (Weinberger, Burnett et al. 2005). The selective pressures exerted by Vpr that were revealed in the present work may constitute one such determinant that specifies which proviruses will survive and contribute to persistent HIV-1 populations.

Materials and Methods

Tissue culture and cell lines

Two immortalized mammalian cell lines were used in this study. Human Embryonic Kidney (HEK) 293T and Jurkat cells were purchased from ATCC (Cat No. CRL3216 and TIB-152 respectively). Both cell lines were preserved as frozen stocks. 293T cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% Fetal Bovine Serum (FBS) (Gemini) and 125 μM gentamycin. Jurkat cells were cultured in RPMI media supplemented with 10% FBS, 100 U/ml penicillin, 100 μg/ml streptomycin, 2 mM glutamine and 55 μM β-mercaptoethanol. To propagate cells, frozen 1 ml aliquots of each cell line were thawed rapidly in 37° C water bath and added to 9 ml of their respective pre-warmed media. After mixing gently, cells were centrifuged at 2500 rpm for 5 minutes. Supernatant was discarded and the cell pellet was resuspended gently in pre-warmed media and plated in an appropriately sized tissue culture plate. HEK 293T cells were sub-cultured when confluency was between 75-100%. Jurkat cells were passaged 1/10 when the cell concentration reached 1 x 10^6 cells/ml.

Zip coded vector and virus production

Zip coded vector DNA templates were generated by digestion of previously described modified "inside-out" forms of GKO, which here are called GPV+ and GPV$^-$ (Battivelli, Dahabieh et al. 2018, Read, Atindaana et al. 2019), with Cla I and Mlu I. The resulting 11.4 kb DNA fragment, which was devoid of plasmid backbone, was gel purified using QIAquick Gel Extraction Kit (Cat No./ID: 28706 Qiagen, Germantown, MD). A 304 bp insert fragment was generated by PCR with GPV$^+$ or GPV$^-$ as template for *vpr*+ and *vpr*- respectively, Phusion High-Fidelity DNA

Polymerase (New England Biolabs, Inc., Ipswich, MA), and primers 5'-GACAAGATATCCTTGATCTGNNN NNNNNNNNNNNNNNNNGCCATCGATGTGGATCTACCACACACAAGGC-3' and 5'-CGGTGCCTGATTAATTAAACGCGTGCTCGAGACCTGGAAAAAC-3'. The 11.4 kb and 304 bp degenerate barcode containing fragments were joined by Gibson assembly using HiFi DNA assembly mix (New England Biolabs) in a molar ratio of 1:5. 3 μg of the resulting covalently closed circular DNA was directly co-transfected with 330 ng of pHEF-VSV-G (Addgene Plasmid #22501) into a 70% confluent monolayer of 293T cells in a 10 cm dish using polyethylenimine (Polysciences Inc, Warrington, PA.) at 4X total transfected DNA in 800 μl 150 mM NaCl. DMEM was replaced with 4 ml RPMI1640 medium with 10% FBS and 1% Pen/strep 24 hours post transfection and culture media was harvested by filtering through a 0.22 μm filter (Fisher Scientific. Cat. No. 09-720-511) 48 hours post-transfection.

Generation of zip coded proviral pools

1000 μl virus-containing media was mixed with polybrene at a final concentration of 0.5 μg/ml and brought to a total volume of 2000 μl by addition of RPMI. This infection mixture was added to 5 x 10^6 Jurkat cells and incubated in two wells of a 12-well plate at 37˚C with 5% CO_2 for 5 hours. Infected cells were then transferred to 10 ml Falcon tubes and centrifuged for 5 minutes at 2500 rpm at 4℃. Following centrifugation, supernatants were replaced with fresh media and cell pellets were resuspended and cultured in two wells of a 6-well plate at 37℃ with 5% CO_2. At 24 hours post-infection, puromycin was added to a final concentration of 0.5 μg/ml. The infected cells were expanded into 6 cm culture plates without puromycin on day 5. Ten days post-infection, the

culture supernatant was replaced with fresh media and the cultures were divided into aliquots, to be either frozen, prepared for integration-site sequencing or further expanded for four additional days and sorted into GFP+ and GFP- subpools for subsequent experiments. Puromycin resistant colony forming titers of pseudotyped barcoded virus stocks were determined by infection of 293T cells, and equivalent infectious units were used to generate *vpr*+ and *vpr*- Jurkat cell pools.

Isolation of clonal populations

Single cells were isolated from the above infected Jurkat cell libraries by limiting dilution and expanded to generate clonal populations. The purity of clones was verified by zip code. Briefly, a 555bp U3-R PCR product was amplified from cellular DNA extracted from each clonal population using the primers 5'-ACGAAGACAAGATATCCTTGATC-3' and 5'-GCACTCAAGGCAAGCTT-3', which flank the zip coded region. PCR reactions used Q5 Hot Start High-Fidelity 2X Master Mix (New England Biolabs) following the manufacturer's protocol and for 30 cycles with a 1-minute extension at 72°C and a 60°C annealing temperature. PCR amplicons were gel purified using QIAquick Gel Extraction Kit (Cat No./ID: 28706 Qiagen, Germantown, MD) and submitted to GENEWIZ (South Plainfield, NJ) for Sanger sequencing with 5'-GCACTCAAGGCAAGCTT-3'. The results revealed that clones 1, 2, and 3 (Fig 1B) contained the zip codes AATACAAGTCGGACCACCTG, GTAACCTTGGCGTCAGGAG, and GTGATGGTAGCGACAGCGTG respectively.

Construction of vpr+ lentiviral vector and its use

A 2933 bp in vitro DNA synthesized *IRES-vpr mKO* fragment was ordered from Genewiz (South Plainfield, NJ) using an IRES sequence from pTRIPZ-hDDX5/7 (Addgene plasmid #71307) (Yang, Zhao et al. 2015) and *vpr* and *mKO* from GPV^{+}. An HIV-1 lentiviral vector fragment was generated by Xba I and Mfe I digestion of pWA18puro (Kharytonchyk, King et al. 2016) to remove its puromycin resistance cassette, and was gel purified using QIAquick Gel Extraction Kit (Cat No./ID: 28706 Qiagen, Germantown, MD). The resulting DNA fragments were joined by Gibson assembly using HiFi DNA assembly mix (New England Biolabs) following the manufacturer's protocol to generate plasmid pEA216-1. 10 μg of pEA216-1 was then co-transfected with 5 μg of the pCMVΔR8.2 helper plasmid (Addgene plasmid #122263) and 1 μg of pHEF-VSV-G into 70% confluent monolayers of 293T cells in a 10 cm dish using polyethylenimine (Polysciences Inc, Warrington, PA.) at 4X total transfected DNA in 800 μl 150 mM NaCl. DMEM was replaced with 4 ml RPMI1640 medium with 10% FBS and 1% pen/strep 24 hours post transfection and culture media was harvested by filtering through a 0.22 μm filter (Fisher Scientific. Cat. No. 09-720-511) 48 hours post-transfection. Parental *vpr*- pools were infected with this filtered media and sorted for mKO and GFP expression 48 hours post-transduction to generate *vpr-*^{rev} pools. Vpr expression was evaluated by cell death. All mKO+ cells were assumed to contain the Tat-inducible *vpr* vector. Therefore, survival of mKO+ cells (PE) compared to dual positive cells (S4 A and B Fig; left panels) suggested that in the absence of Tat from a pre-existing provirus, uninduced Vpr expression was not pronounced enough to kill at least most transduced cells. However, the possibility that low levels of uninduced Vpr had led to the depletion of some single positive cells cannot be ruled out.

In determining the effect of Vpr addition on parental *vpr*- pools, zip code abundances in sequences amplified from genomic DNA of parental *vpr*- unsorted cells and from Vpr expression vector-transduced vpr^{rev} mKO+ cells were compared. Two categories of zip codes were defined: reduced zip codes were those that were reduced 10-fold or more after Vpr transduction, whereas those with a change of 1.5-fold or less in their relative abundances were regarded as not affected by *vpr* addition (Not reduced).

Flow cytometry and cell sorting

For Fluorescence Activated Cell Sorting (FACS) analysis by flow cytometry, Jurkat cells were suspended in phosphate buffered saline (PBS) containing 1% FBS (FACS buffer). Dead cells were excluded from all analyses and sorting experiments using propidium iodide (PI). Acquisition was carried out on the FITC channel for GFP and PE channel for PI. Cell fluorescence was assessed using BD LRS Fortessa (BD Biosciences) and data were analyzed using FlowJo software, version 10.6 (FlowJo, LLC., Ashland, Oregon). Infected cells were sorted into GFP+ and GFP- sub-populations by flow cytometry using FACS Aria II (BD Biosciences, Franklin Lakes, NJ) or iCyt Synergy SY3200 (Sony Biotechnology, San Jose, CA) cell sorters at the flow cytometry core of the University of Michigan.

Latency reversing agents and reactivation

JQ1 and prostratin were purchased from Sigma Aldrich. Each LRA was dissolved in DMSO (Thermofisher) to produce stocks. For each experiment, stocks were added to culture medium to achieve 2 μM JQ1 and 10 μM prostratin final concentrations. Dual LRA treatment was performed by adding the two LRAs to the same culture medium to achieve their respective single LRA

concentrations. For reactivation experiments, GFP- cells sorted on day 14 post-infection were co-cultured with appropriate LRA for 24 hours. Cells were then centrifuged at 2000rpm, 4ºC for 5 mins, cells pellets were washed twice with ice cold FACS buffer after being stained with PI for 5 mins at room temperature. The resulting cells were then washed and assessed by flow cytometry and the p24 concentration of culture supernatants was determined by a reverse transcriptase assay.

Zip code sequencing libraries

Zip codes were amplified from the genomic DNA of infected cells as well as from the RNA of virus released into cell media. Generation of zip coded sequencing libraries from infected cells was initiated by harvesting DNA from an aliquot of 2×10^6 infected cells. Genomic DNA extraction was carried out using the Qiagen DNeasy Blood & Tissue Kit (Qiagen, Germantown, MD). Zip codes were then amplified by PCR from 200 ng of DNA template using Phusion High-Fidelity DNA Polymerase (New England Biolabs) in HF Buffer. Primers were designed to flank the zip code region (primer sequences: 5'-NNACGAAGACAAGATATCCTTGATC-3' and 5'-NNTGTGTGGTAGATCCACATCG-3'). Multiple copies of these primers were created, each with a unique pair of known, randomized nucleotides at the 5' end to confirm that no cross-contamination had occurred between samples. Reactions were cycled 29 times with a 30 second extension at 72°C and a 59°C annealing temperature. Zip code amplicons were purified using the DNA Clean and Concentrator-5 kit (Zymo Research, CA. Cat. No. D4013) and eluted in 15 μL of Milli-Q ® H_2O. To amplify zip codes from virus, a tissue culture plate of infected cells was decanted into a conical tube and centrifuged at 2500 rpm for 5 minutes. The virus-containing media was separated from the cell pellet and passed through a 0.22 μm filter. To concentrate virus, media was subjected to ultra-

centrifugation (25,000 rpm) for 120 minutes through a 20% sucrose cushion. Viral pellets were then resuspended in 200 µl PBS and viral RNA was extracted using Quick-RNA Viral Kit (Zymo Research, CA. Cat. No. R1034 & R1035) following the manufacturer's protocol and eluted in 10 µl RNase-free water. cDNA was synthesized using 5 µl of the eluent as template using the U3 antisense primer 5' TGTGTGGTAGATCCACATCG-3' and M-MLV RT RNase [H-] (Promega, WI. Cat. No. MR3681) following the manufacturer's protocol. Zip codes were amplified from this cDNA using conditions above. The zip code amplicons were then used to generate MiSeq libraries for sequencing as described previously (Read, Atindaana et al. 2019).

Integration site determination

Genomic DNA was extracted from *vpr*+ and *vpr*- cells 10 days post-infection using Qiagen DNeasy Blood & Tissue kit (Qiagen) and 200 ng of DNA was sheared to 1 kb fragments using Covaris M220 and micro- TUBE according to the manufacturer's recommended settings (Covaris, Woburn, MA). HIV-1 insertion site libraries were prepared and sequenced using methods described previously (Read, Atindaana et al. 2019). All sequence data has been deposited to the Sequence Read Archive (SRA) SUB10130489.

Quantification of virus release

Virion production was quantified using a real-time reverse transcription PCR assay developed by Pizzato as modified by Kharytonchyk et al (Pizzato, Erlwein et al. , Kharytonchyk, King et al. 2016). Briefly, viral lysates were prepared by adding 5 µl of culture supernatant to 5 µl of lysis buffer. Using MS2 RNA as template, MS2 cDNA was synthesized with viral lysates and quantified

by real-time PCR in one reaction. Released virus was quantified and normalized for p24 level based on values determined in parallel for reference samples.

Zip code quantification and analysis

Zip codes were identified and quantified from Illumina sequencing reads using a previously described custom suite of tools implemented in Python (https://github.com/KiddLab/hiv-zipcode-tools). Briefly, 2x75 bp paired reads were merged together using *flash* v1.2.11 [73]. Zip codes were identified by searching for known flanking sequence (with up to 1 mismatch). Only candidate zip codes with a length of 17–23 nucleotides were considered and the read count for each unique zip code was tabulated. To identify sets of zip codes for further analysis, zip code families, which account for PCR and sequencing errors, were determined by clustering together the observed unique zip codes. Abundances for the zip codes were then determined by assigning unique zip codes to the most abundant family whose sequence was within 2 mismatches and summing their associated read counts. Only zip code families with corresponding data in the integration site data were selected for further analysis.

For each latency reversal treatment condition, clonal virus release was determined by multiplying the fractional abundances of zip codes from the cDNA sequencing libraries of each treatment by their corresponding sample's pool p24 concentration, as measured 24 hours post LRA treatment. The resulting clonal p24 values for each zip code in prostratin, JQ1, and combination prostratin-JQ1 treated samples were divided by clonal p24 values defined for cell samples exposed only to DMSO to determine fold change after LRA treatment.

The burst sizes of clones were determined by multiplying the fractional abundances of zip codes from cDNA library of unstimulated sample's pool p24 concentration (basal viral release). The resulting p24 values for each zip code were then divided by their corresponding fraction of abundances determined in gDNA sequencing library of GFP+ sorted cells from the same unstimulated sample cDNA library was made.

The %GFP+ for each zip code was determined as $F_i = ((G_i * P)/ (G_i * P + W_i * Q))*100$ (Read, Atindaana et al. 2019) where F_i is the GFP+ percent of zip code i, G_i is the fraction abundance of zip code i in the GFP+ sorted pool, W_i is the fraction abundance of zip code i in the GFP- sorted pool, P is the fraction of cells that sorted into the GFP+ pool and Q is the fraction of cells that sorted into the GFP- pool.

Determination of chromatin marks

H3K27ac marks annotated for the Jurkat E-6-1 clone were sourced from ChIP-Atlas ((http://chip-atlas.org/view?id=SRX1041803), H3K4me3 and DNase I sensitivity site data sets were downloaded from the ENCODE Project (https://www.encodeproject.org/) with Sequence Read Archive accession no. SRX1041803 and ENCODE Project identifiers ENCFF304GVP (https://www.encodeproject.org/experiments/ENCSR000EOS/) for genome marks preexisting in Jurkat cells, ENCFF341XUX (https://www.encodeproject.org/experiments/ENCSR807WEO/), and ENCFF053LHH (https://www.encodeproject.org/experiments/ENCSR724GUS/) (Consortium 2012). Bedtools was then used to map the distance to the closest known annotated

marks. Analysis of proximity to these genome marks was done using matplotlib and scipy.stats packages in python and the results were exported into Graphpad Prism version 9.1.2 to plot graphs.

Acknowledgements

We thank Rob Gifford for invaluable discussion and Stephen Goff, Akira Ono and Joel Swanson for helpful comments on the manuscript. This work was supported by NIH/NIAID grant number R33 AI116190 to AT and JMK and by a pilot award to AT from the Rogel Cancer Center, NIH/NCI grant number P30CA046592.

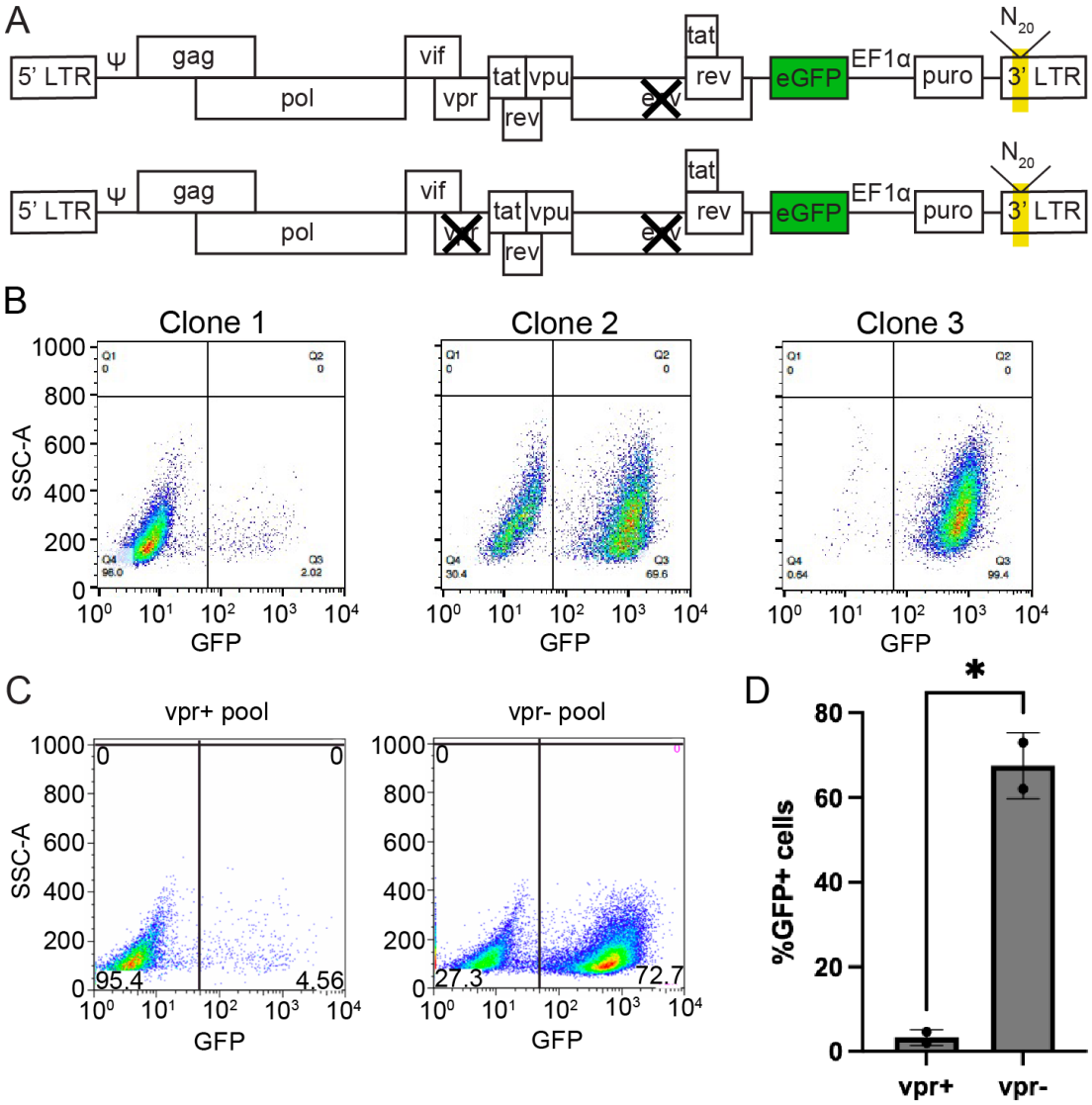


Figure 3. 1. ***Generation of zip coded HIV-1 proviral pools.*** *Polyclonal cell pools containing vpr+ and vpr- proviruses were established in Jurkat T cells and compared.* ***A.*** *Schematic of vectors used to produce zip coded virus (not to scale). The yellow shade in the 3' LTR indicates position of 20bp randomized barcode in U3.* ***B.*** *vpr- clone expression properties. Three clones were isolated by limiting dilution of vpr- pool cells, expanded for 14 days and subjected to flow cytometry. The flow plots show clonal pools 1-3 have different percents GFP+ cells. X axis indicates GFP signal detected in the FITC channel; numbers in each gate indicate percent of cells gated GFP- or GFP+.* ***C.*** *Representative flow cytometry plots of polyclonal Jurkat T cells generated by infection with barcoded vpr+ (left panel) or vpr- (right panel) viruses, that were puromycin-selected for four days and expanded for an additional ten days. D. Bar graph comparing the percent GFP+ puroR cells in polyclonal vpr+ and vpr- populations. Data represent two vpr+ pools and two vpr- pools (p=* indicates p<0.05; paired t-test).*

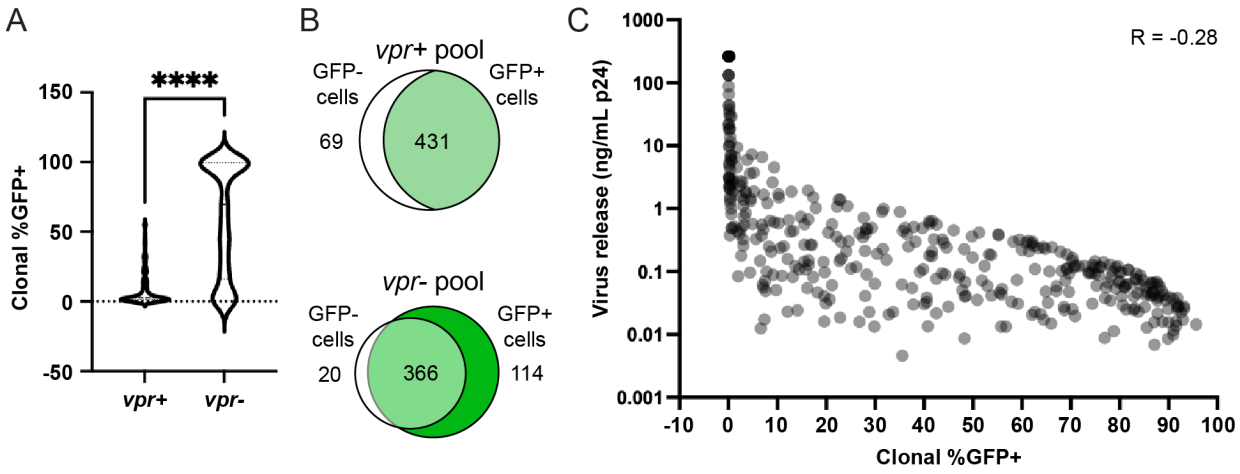


Figure 3. 2. ***Expression properties of individual clones within the vpr+ and vpr- pools****. A. Comparison of the numbers of high and low LTR-active clones in vpr+ and vpr- pools. The fractions of total cells in each clone that sorted GFP+ (LTR-active) were calculated for the most abundant 500 clones in each pool by quantifying the fraction of each zip code (integrated barcode) in the GFP+ sorted cells' genomic DNA to its quantity in the unsorted pool by high throughput sequencing and normalizing for the percent of the unsorted pool that was GFP+ (see Materials and Methods). Violin plots compare the %GFP+ (indicative of clones' fraction of LTR-active daughter cells) for the top 500 most abundant zip codes in each pool. (p= **** indicated shows p<0.0001 Mann-Whitney U two-tailed test.). B. Venn diagrams of the number of independent zip codes in GFP+ and GFP- sorted sub pools for vpr+ and vpr- pools. The number in the white region indicates the number of different zip codes observed only in GFP- sorted cells, that in the light-green region of intersection represents the number of zip codes that were present in both GFP+ and GFP- sorted subpools, and the number in the dark green region represents zip codes present in only the green sorted cells. Reported 500 zip codes in each pool represent those present in the top 78% and 88% of reads for vpr+ and vpr- pools respectively, when clones were ordered by read abundance. C. A scatter plot displaying each clone's %GFP+ value on the X-axis and virus release per GFP+ cell on the Y-axis (Pearson's R = -0.28).*

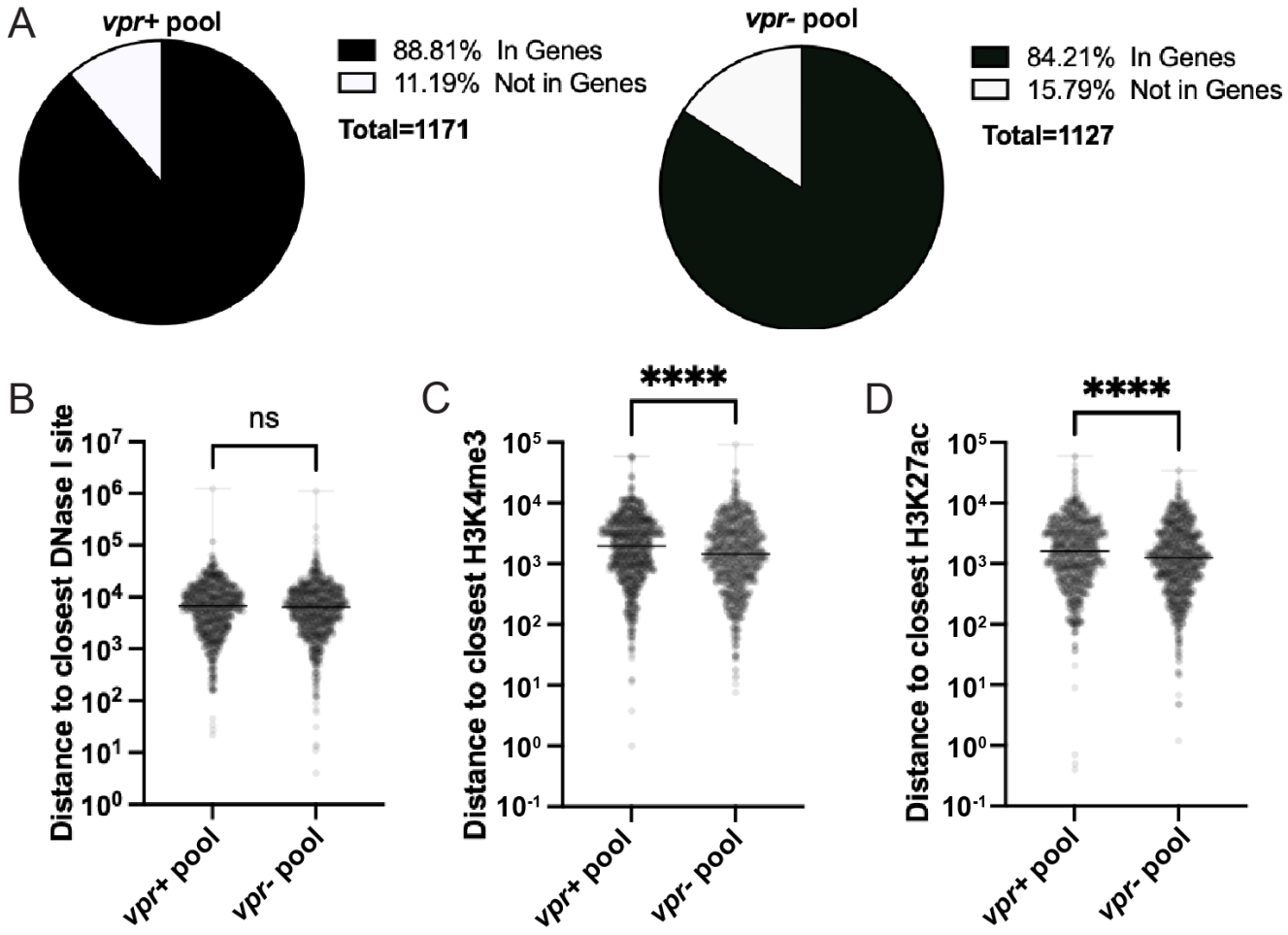


Figure 3. 3. ***Comparison of integration sites in vpr+ and vpr- pools.*** *Integration sites for both vpr+ and vpr- pool cells harvested 10 days post-infection were determined and distances to the nearest indicated genome features were mapped and compared. A. Pie charts showing the percentage of integration sites found within genes versus those not found in genes for vpr+ and vpr- pools. B.-D. Box plots showing pairwise comparisons between vpr+ and vpr- integration sites' distances to the closest: B. DNase I hypersensitivity site (Mean=10687 and 10236 respectively), C. H3K4me3 (Mean=32989, 27047 respectively), and D. H3K27ac (Mean = 28485 and 22228 respectively). Mann-Whitney U two-tailed test was conducted for pairwise comparisons for B-D (p= ns, and **** indicated shows $p>0.05$ and $p<0.0001$ respectively, n=1171 and 1127 for vpr+ and vpr- respectively).*

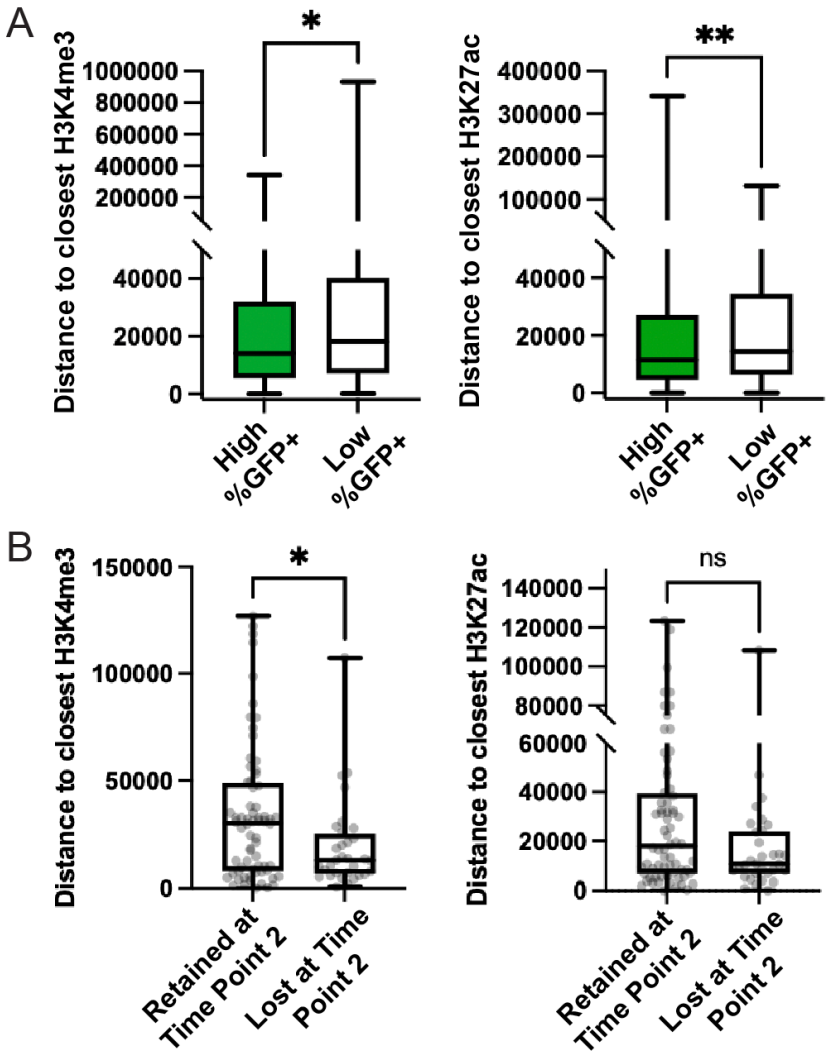


Figure 3. 4. ***Integration site proximities to chromatin marks, and their changes over time in vpr+ pools****. A. Distances to the indicated marks for high and low %GFP+ clones. The %GFP+ values for each zip code from the vpr- pool was determined as described in Materials and Methods. Zip codes with %GFP+ >= 60 were binned as "High LTR-active clones" and zip codes (595 zip codes) with less than 30%GFP+ were binned as "Low LTR-active" (222 zip codes). Pairwise comparison of the distance to the closest H3K4me3 (left) and H3K27ac (right) marks are shown. B. Changes in predominant integration site features for the vpr+ pools over time. Among the top 100 most abundant zip codes in the unsorted vpr+ pool at Time Point 1 (10 days post-infection), 68 were retained among the top 100 at Time Point 2 (24 days post-infection), while the remaining 32 Time Point 1 top 100 clones were lost at Time Point 2 and supplanted by other, formerly less-abundant zip codes. Box plots comparing the closest H3K4me3 or H3K27ac mark for the 68 Time Point 1 zip codes that were retained in the top 100 at Time Point 2 to the 32 Time Point 1 zip codes that had been lost at Time Point 2 (p=ns, indicated shows p>0.05; Mann-Whitney U two-tailed test).*

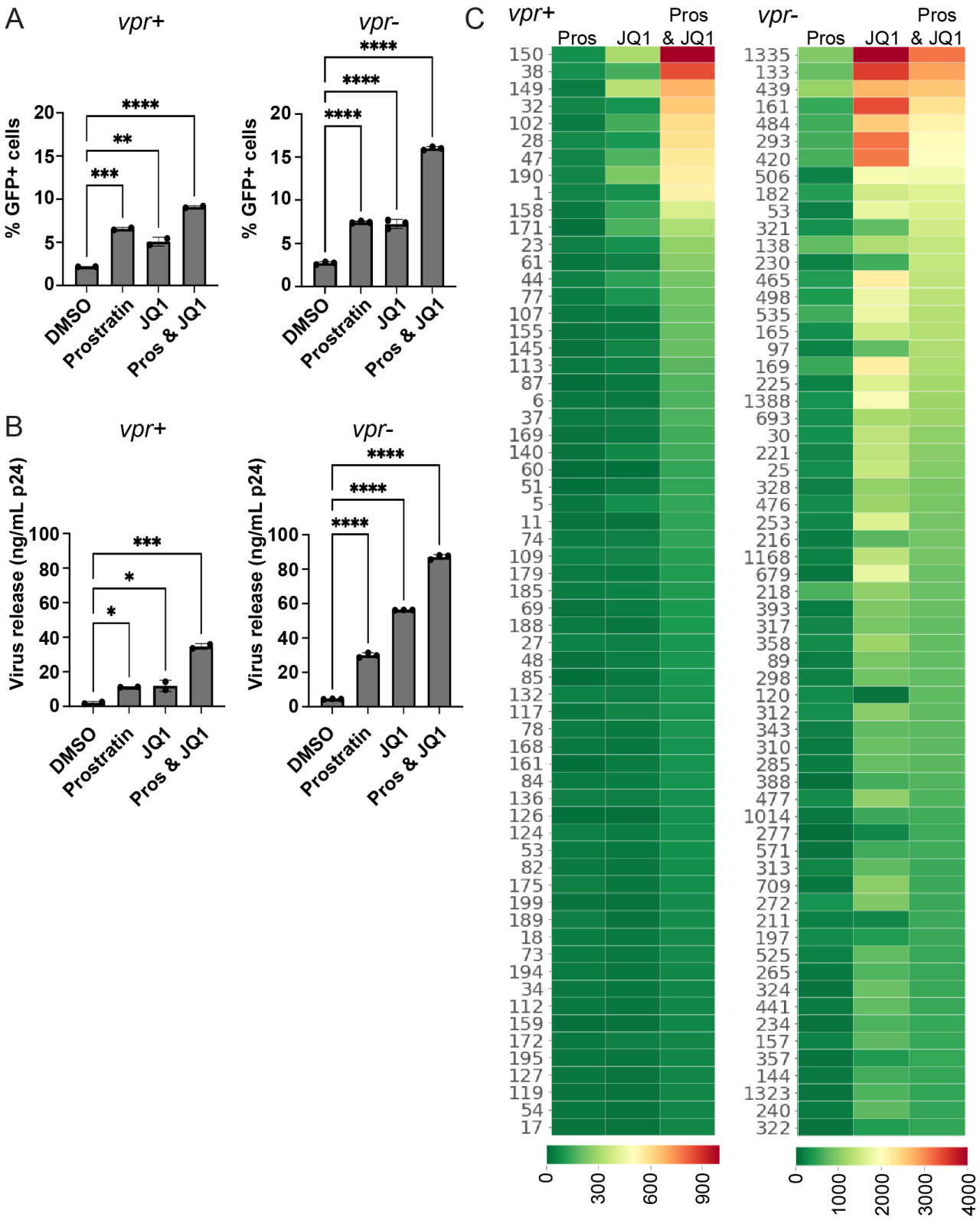


Figure 3. 5. ***Effects of LRAs on zip coded pools and clones.*** *vpr+ and vpr- GFP- cell fractions were exposed to 0.1% DMSO, 2μM JQ1, 2μM prostratin, or to a combination of prostratin and JQ1. Reactivation to GFP+ was measured at 24 hours post-infection by flow cytometry, and virus release was quantified using a reverse*

transcriptase assay and normalized to define p24 levels. A. Bar graphs showing the frequency of GFP+ cells post LRA treatment (left panel for vpr- and right panel for vpr+) and B. the amount of virus released into culture supernatant (left panel for vpr- and right panel for vpr+) for the indicated polyclonal pools. The error bars show the mean and standard deviation from two experimental treatment repetitions. C. Heatmaps of the clonal (zip code) virus release per treated GFP- cell (left: vpr+ pool and right: vpr- pool). Numbers at the left of each panel are clone identifiers generated by ordering proviral zip codes in decreasing relative abundance, as determined for the unsorted pools. Every row represents a unique cell clone's response. The clones were ordered from top to bottom by diminishing virus release per treated cell upon dual LRA treatment. The color bar indicates the extent of release per treated cell based on p24 values in arbitrary units. Note that the same unit values were used for both vpr+ and vpr- pools, but that the color scales, as presented at the bottom of the panels, differ between pools.

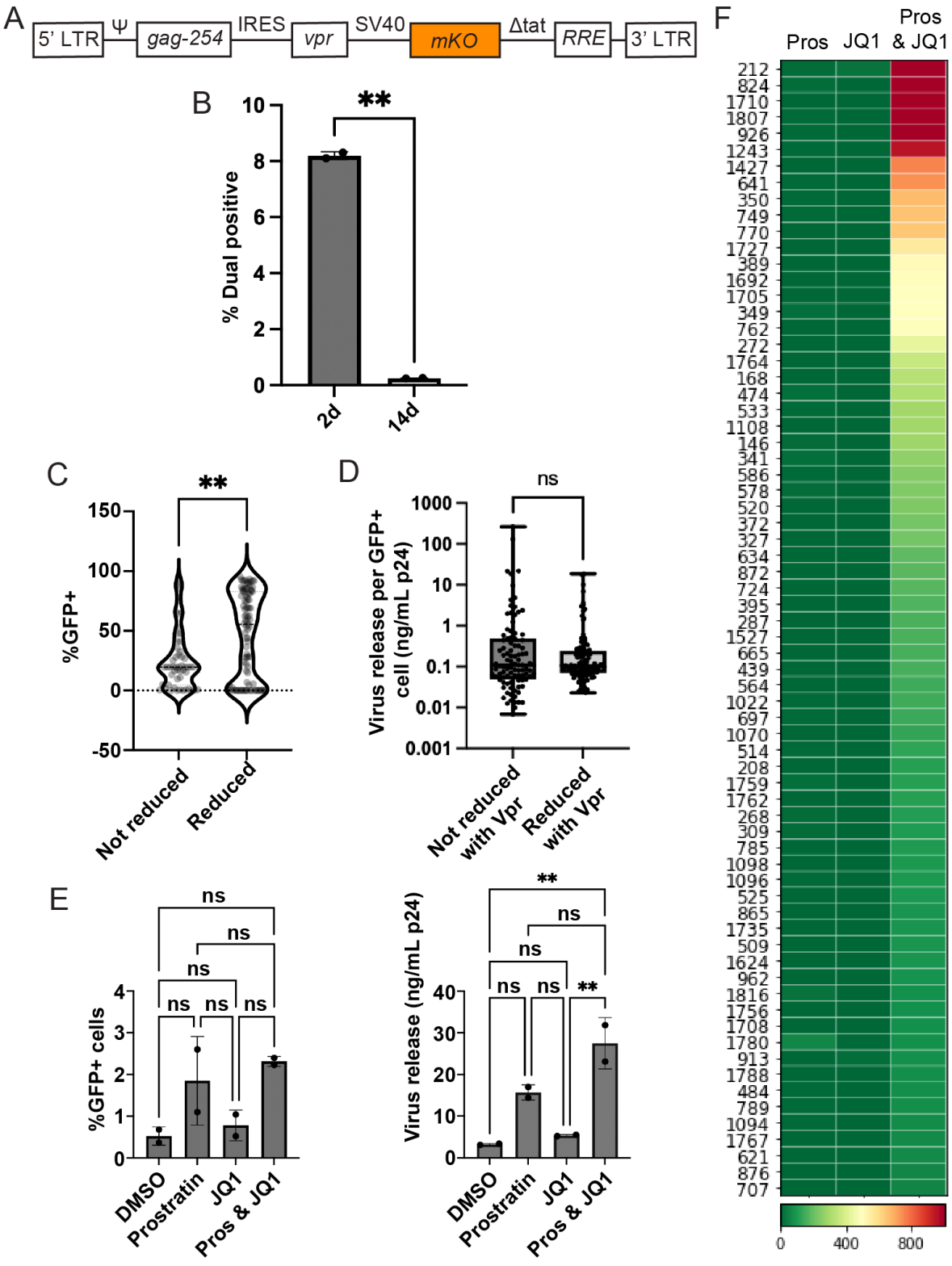


Figure 3. 6. ***Complementation of vpr- pools with Vpr expression vector.*** *A. Schematic diagram of the Tat-deficient Vpr expression vector (not to scale). This vector was used to add vpr to vpr- pools and effects were determined by high throughput analysis of the indicated virus and cell samples. The vector contains IRES-vpr under the HIV-1 LTR promoter and an SV40 promoter-driven mKO fluorescent reporter. B. Population changes after vpr addition.*

Bar graph of the percent GFP+ cells for two polyclonal vpr^{rev} pools established in parallel at day-2 (2d) and day-14 (14d) post-transduction (p=0.0076, significant pairwise comparison by paired t-test). C. Changes in clone sizes. Pairwise comparisons of %GFP+ between the clones that were reduced (10-fold or more reduction in relative abundance) and those that were not reduced (less than 1.5-fold reduction in relative abundance) after vpr addition (significant pairwise comparison by Mann-Whitney U two-tailed test). D. Changes in clone sizes is not affected by viral burst sizes. Pairwise comparisons of virus release per GFP+ cell between the clones that were reduced (10-fold or more reduction in relative abundance) and those that were not reduced (less than 1.5-fold reduction in relative abundance) after vpr addition (significant pairwise comparison by Mann-Whitney U two-tailed test). E. Bar graphs showing the percent of GFP+ cells 24 hours post reactivation with JQ1, prostratin, and dual prostratin-JQ1 for the vpr^{rev} pool (left) and vpr- pool (right). F. Heatmap of the clonal (zip code) virus release per treated GFP- cell of vpr^{rev} pool. Numbers at the left of each panel are clone identifiers generated by ordering proviral zip codes in decreasing relative abundance, as determined for the unsorted pools. Every row represents a unique cell clone's response. The clones were ordered from top to bottom by diminishing virus release per treated cell upon dual LRA treatment. The color bar indicates the extent of release per treated cell in arbitrary units. Cell density and other culture and assay conditions were as in Fig 5.

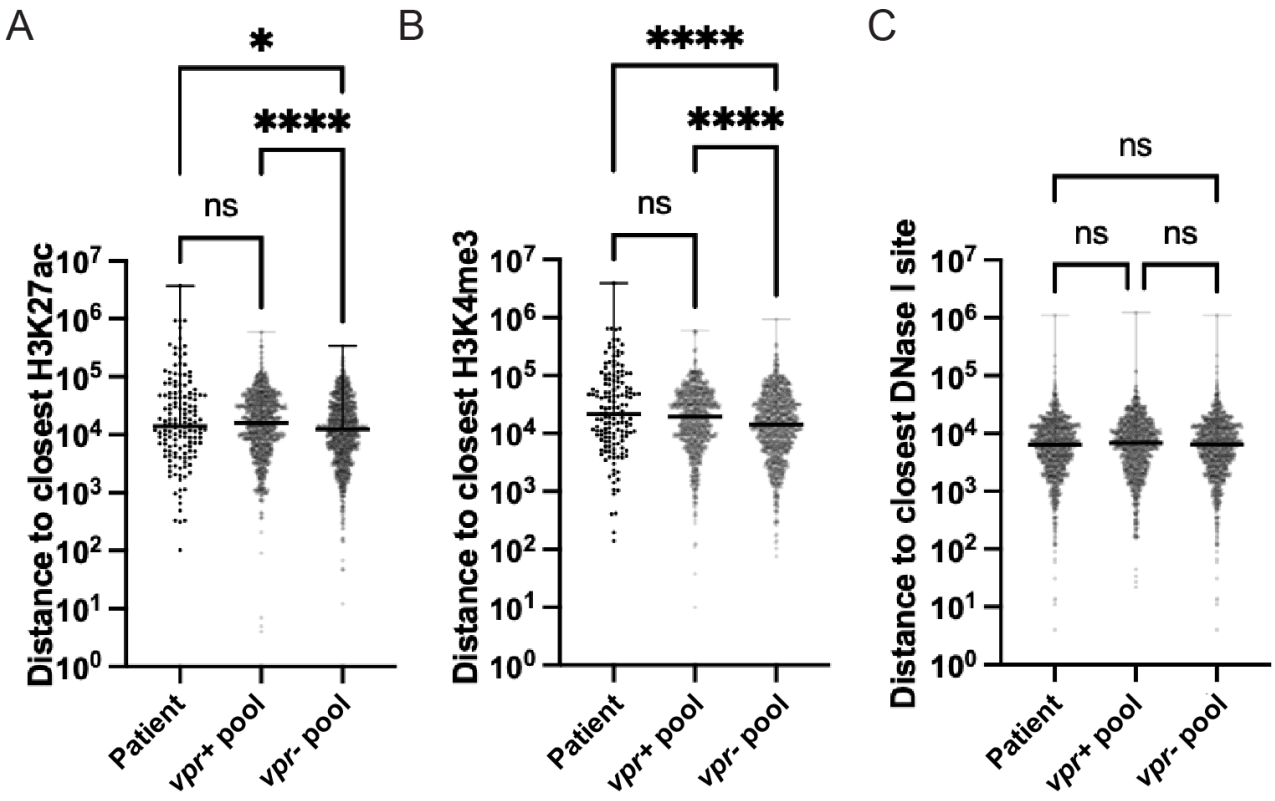


Figure 3. 7. ***Comparison of patients' integration site features with vpr+ and vpr- pools.*** *Previously published HIV integration sites from three patients were used to determine base distances to closest H3K27ac, H3K4me3, and DNase I sensitivity sites. Distances were compared to those of vpr+ and vpr- pools. Box plots comparing base distances of vpr+, vpr-, and patients' proviral integration sites to: A. H3K27ac (Mean = 28485, and 22228 and 89771 respectively), B. H3K4me3 (Mean=32989, 27047, and 97863 respectively), and C. DNase I sensitivity sites (Mean=10687, 10236, and 10191 respectively). (Kruskal-Wallis test. p=ns, *, **, ***, **** indicate p>0.05, < 0.05, <0.01, 0.001, and 0.0001 respectively).*

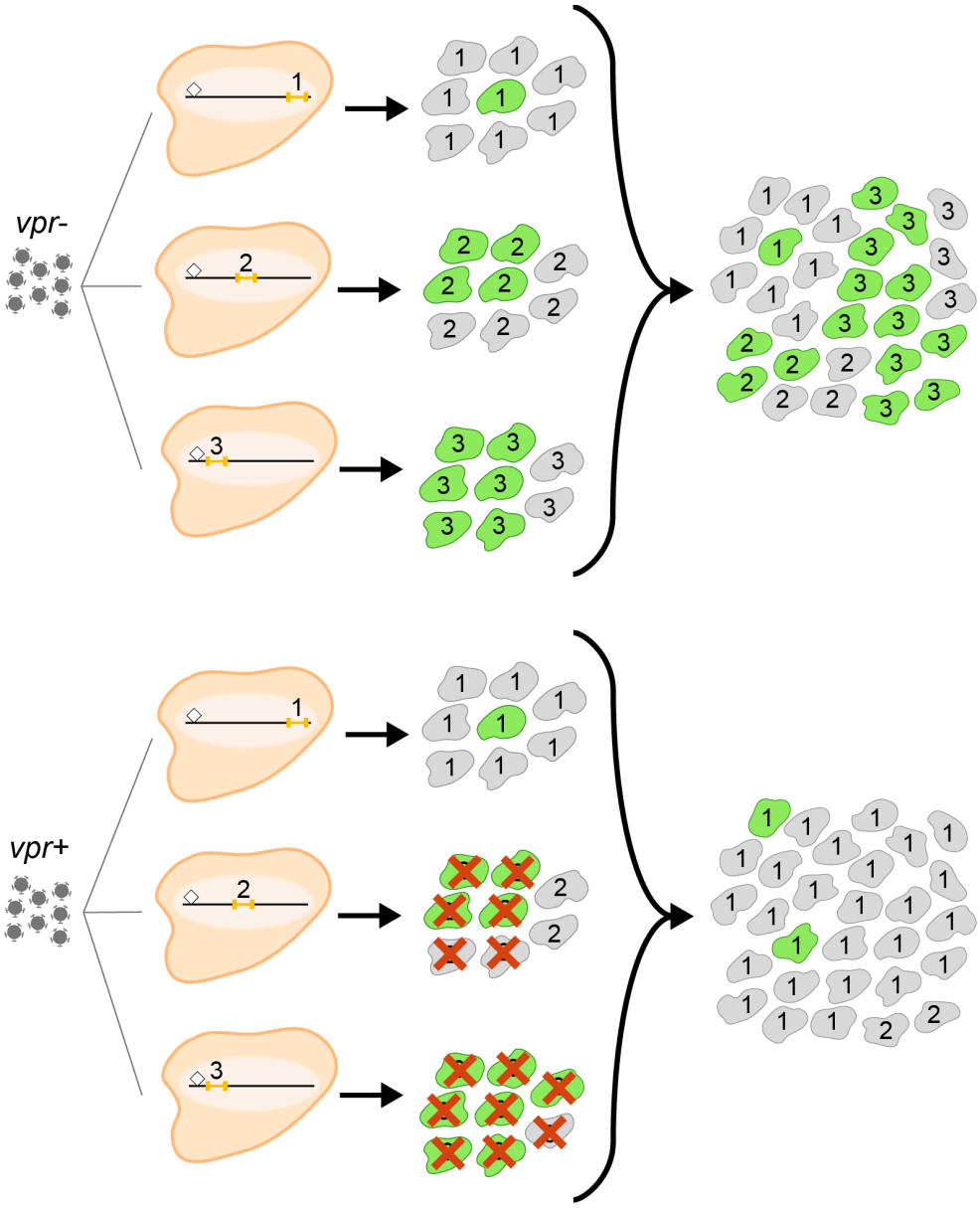


Figure 3. 8. ***Schematic model of the effect of Vpr on integrant populations' proviral landscapes.*** *From left: proviruses integrate at indistinguishable genomic positions (designated 1, 2, and 3 above the small orange bars, which indicate proviruses) regardless of whether the infecting virus was Vpr+ or Vpr-. Note that integrants 1, 2, and 3 differ in their distances to the closest chromatin mark, which is indicated by a small diamond, which may affect their expression characteristics and responses to LRAs. Next, infected cells divide to form cell clones. For clone 1, cells that express HIV genes (indicated by green cell) are rare (low LTR-active clone) and for other clones, more cellular members of each clone express HIV (eg: clone 3: high LTR-active clone). This pattern of bimodal gene expression is an intrinsic property of each clone (Read, Atindaana et al. 2019). The red Xs indicate that high LTR-active clone members are selectively depleted by Vpr. As a result, in the polyclonal populations shown at right, low LTR-active clones are enriched in the vpr+ pool.*

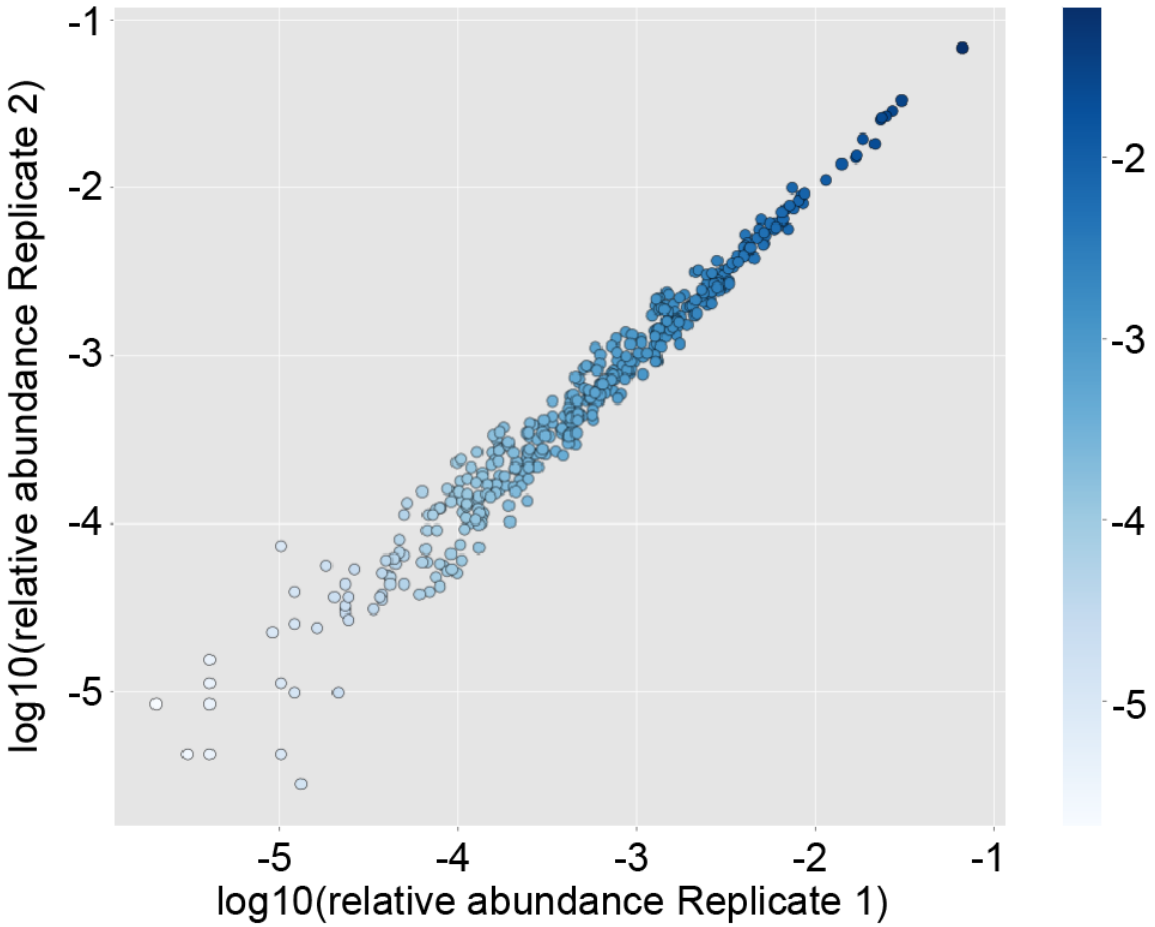


Figure S3. 1: ***Reproducible amplification and high throughput sequencing of zip codes.*** *Two independent PCR reactions were setup to generate amplicons zip codes from gDNA sample and high throughput sequenced. The relative abundance of zip codes were quantified and the log10 (relative abundance) of PCR amplicon 1 (replicate 1) and PCR amplicon 2 (replicate 2) were plotted on the X and Y axes respectively. The color bar indicates the log10 relative abundance of zip codes in replicate 2. The darker the blue scale, the more abundant the zip code.*

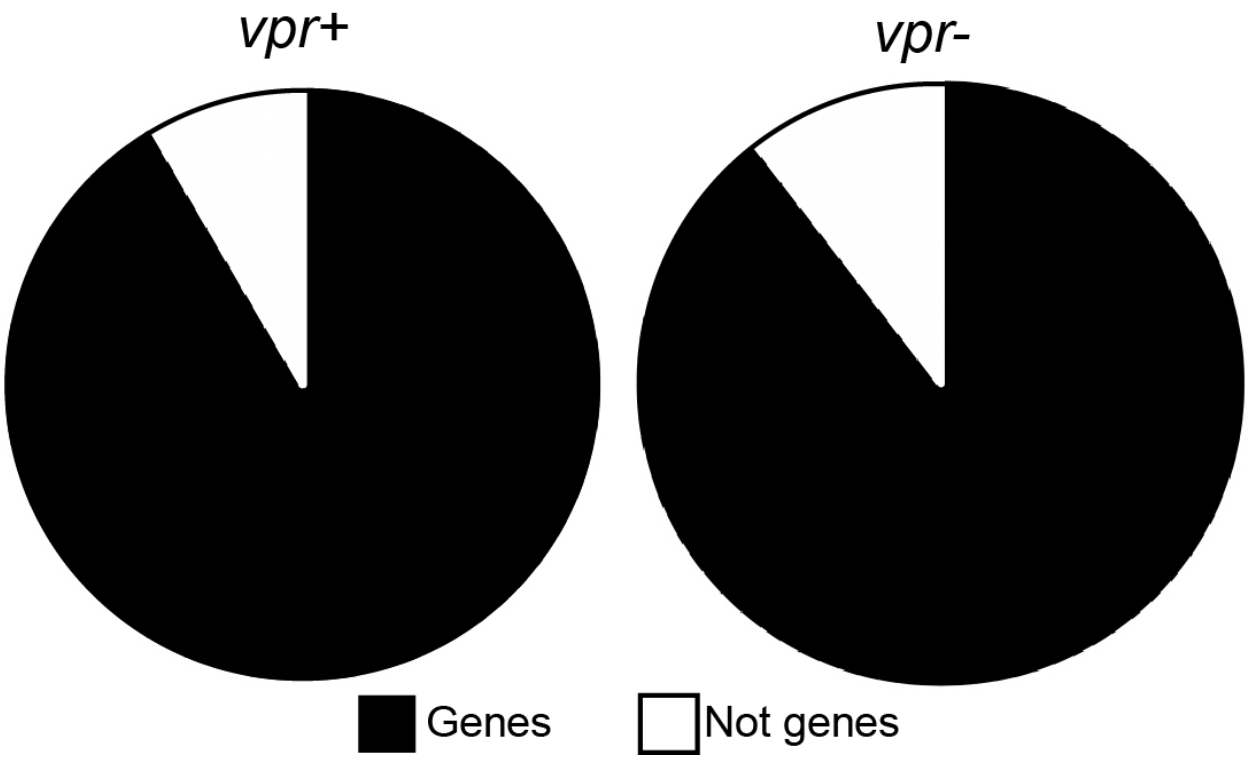


Figure S3. 2. ***Comparison of integration sites and %GFP+ between duplicate vpr- and vpr+ pools.*** *A. Integration sites were sequenced and mapped to genes in Jurkat T cells. 92% of integration sites were found in genes and 8% were not in genes for vpr+ pool 2 (n=166) compared to 90% within genes and 10% not in genes for vpr- pool 2 (n=67).*

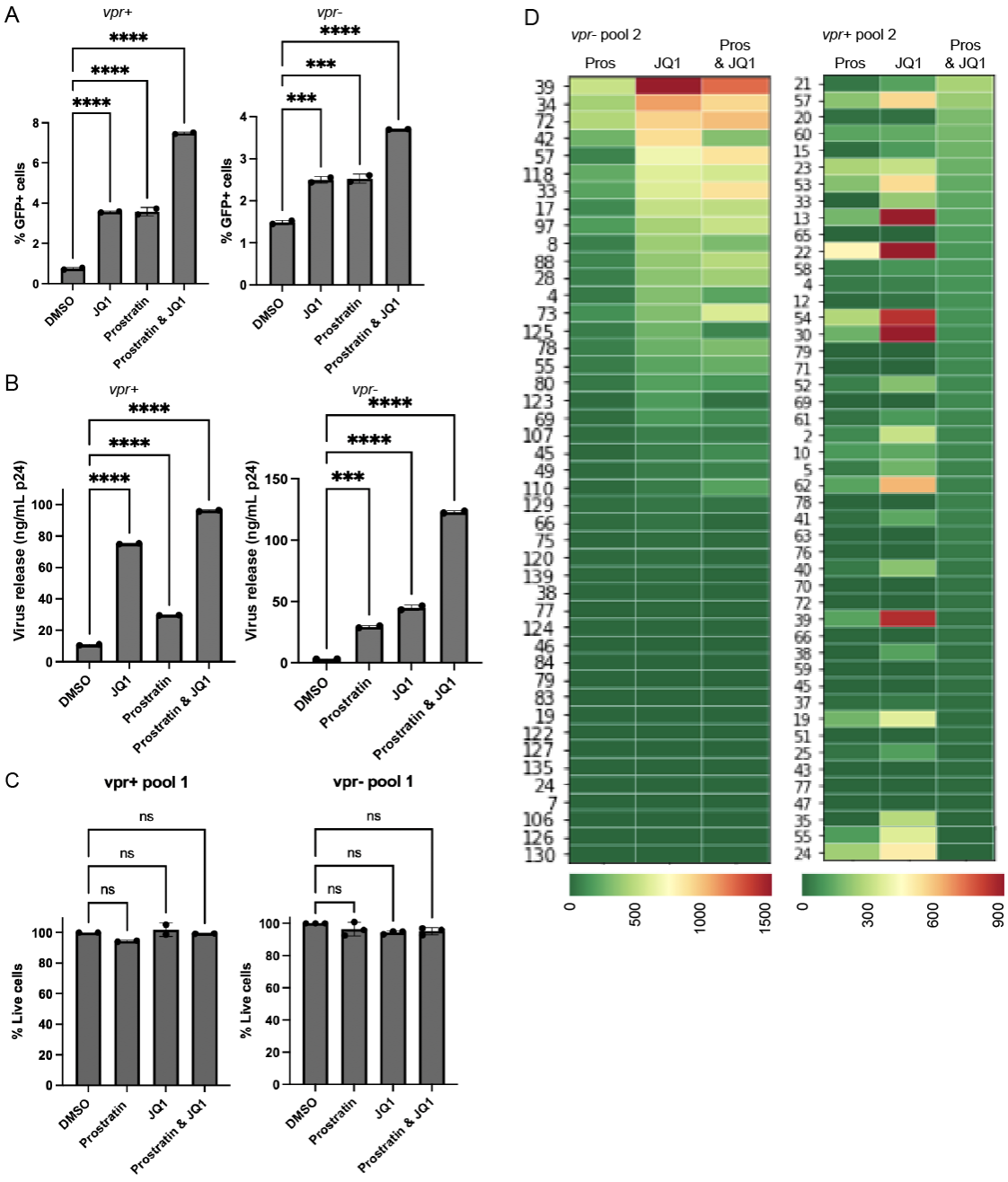


Figure S3. 3. ***Cell viability and reactivation of additional vpr- and vpr+ pools.*** *vpr+ and vpr- GFP- cells were exposed to 0.1% DMSO, 2μM JQ1, 2μM prostratin, and a combination of prostratin and JQ1. Reactivation was measured at 24 hours post-infection by flow cytometry, and virus release was quantified by reverse transcriptase assay. (A) Bar graphs showing the frequency of GFP+ cells post LRA treatment (left panel for vpr+ and right panel for vpr-). (B) The amount of virus released into culture supernatant (left panel for vpr+ and right panel for vpr-) for the indicated polyclonal pools. The error bars show the mean and standard deviation from two experimental replicates. (Ordinary One-Way ANOVA; p=ns, ***, and **** indicate p>0.05, p< 0.001, and 0.0001 respectively). The error bars show mean and standard deviation. (C) Viability was determined by propidium iodide staining and the percent viability was measured relative to the viability of cells in DMSO control. (D) A heatmap of virus release per cell for the indicated clones (left: vpr+ pool 2 and right: vpr- pool 2). Every column represents a unique cell clone's response. The color bar indicates the extent of release in arbitrary units.*

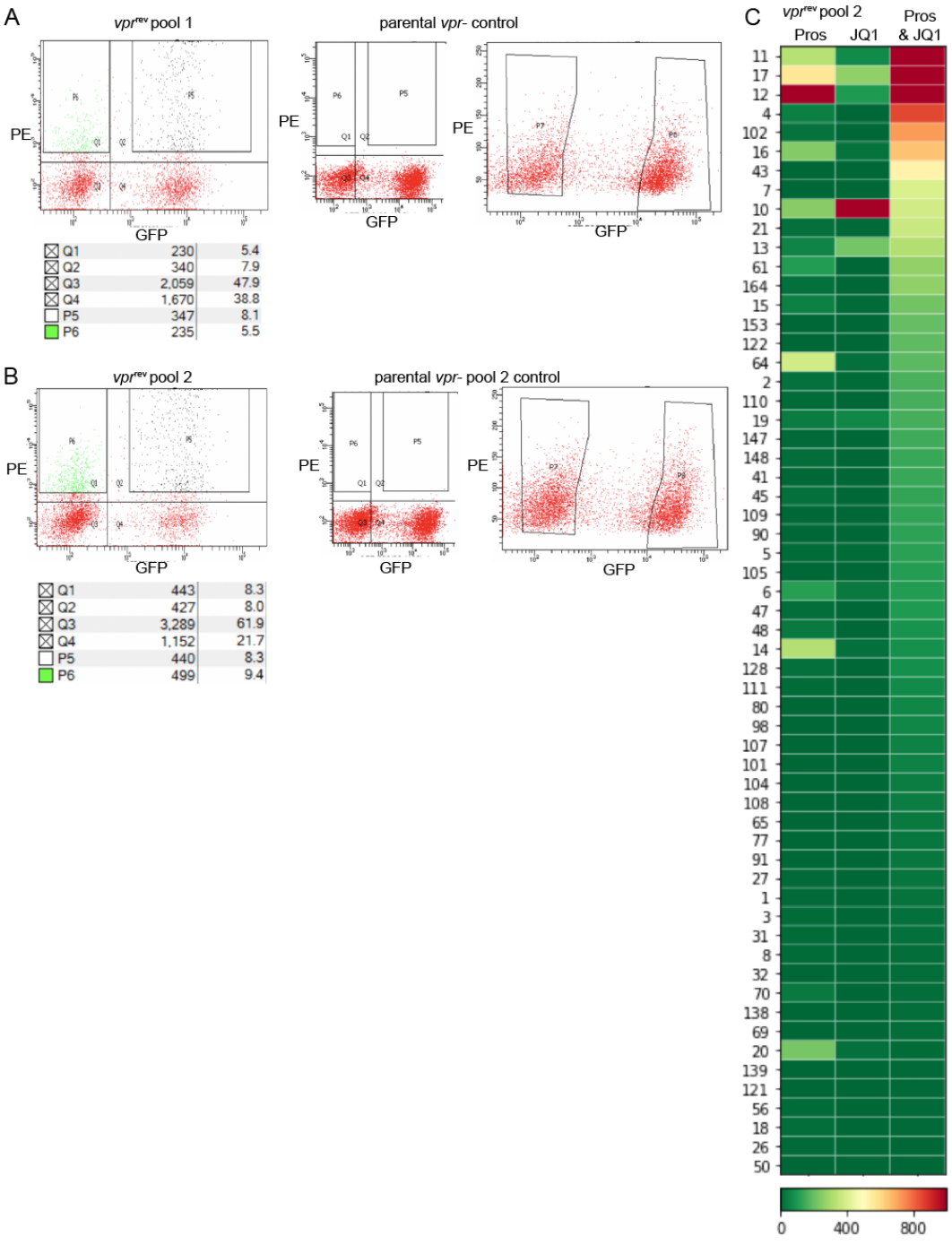


Figure S3. 4. ***Sort gates for vpr-tranduced vpr- pools.*** *Parental vpr- pools were transduced with Vpr expression vector. 48 hours post-transduction, cells were sorted into PE+ GFP- and PE+ GFP+ cells. Flow plots show the gates for sorting of A. Vpr vector transduced vpr- pool 1 (left), parental vpr- pool 1 (right), B. Vpr vector transduced vpr- pool 2 (left), and parental vpr- pool 2 (right). Quadrants P6 and P5 were collected for further analysis. B. Pairwise comparison of the distance to the closest H3K4me3 (left) and H3K27ac (right) marks of the zip codes that were similarly abundant at both time points (TP1 &TP2) and those that were only abundant at the first time point (TP1). Mann-Whitney two-tailed test p=ns and * indicate p>0.05 and p<0.05 respectively. C. A heatmap of virus release per treated cell for vpr^{rev}. The indicated clones in the first column are arranged from left to right by diminishing virus release per dual treatment. Pros indicates prostratin.*

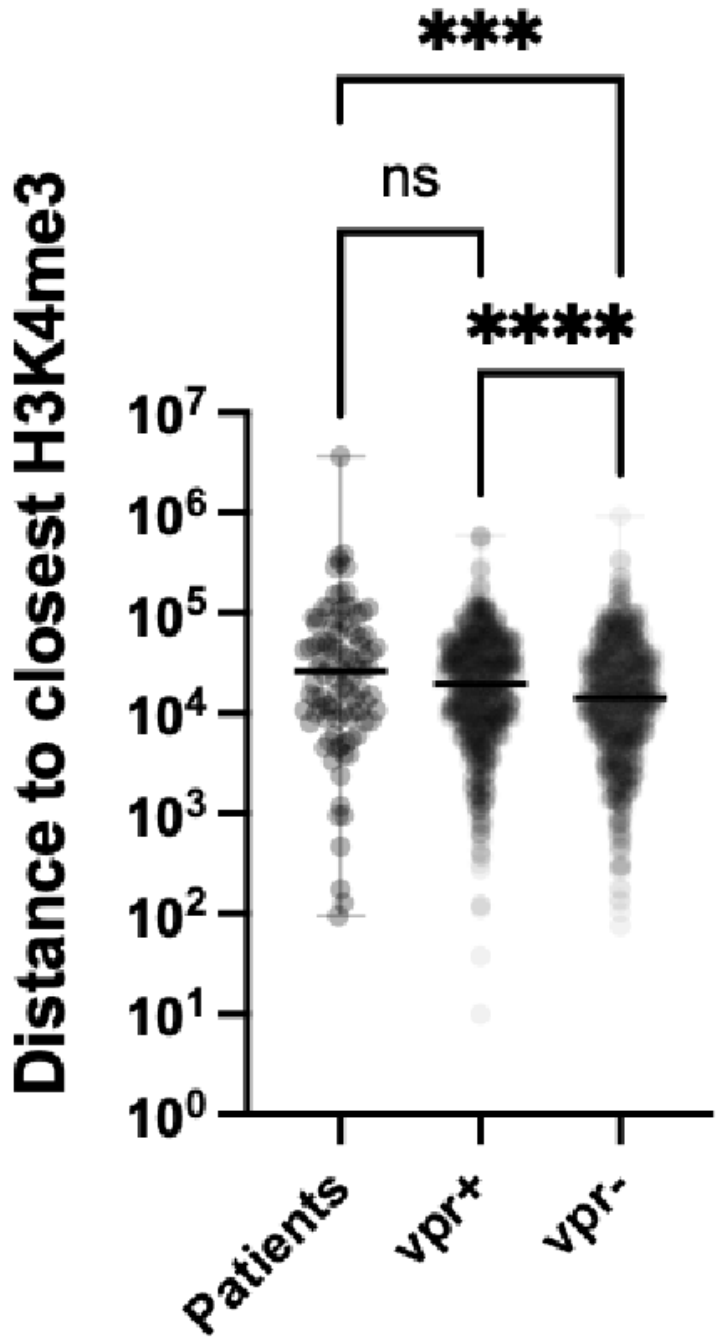


Figure S3. 5: ***Comparison of patient integration site distance to the nearest H3K4me3 mark to vpr+ and vpr- pools.*** *Previously reported integration sites of intact proviruses from three HIV patients were mapped to the nearest pre-existing H3K4me3 in memory CD4+ T cells. Significant comparisons by Kruskal-Wallis test. p=ns, *, **, ***, ****

Table S3. 1 Table of zip codes, these clones' fractions of abundance, and integration sites of vpr- and vpr+ pools

References

Archin, N. M., A. Espeseth, D. Parker, M. Cheema, D. Hazuda and D. M. Margolis (2009). "Expression of latent HIV induced by the potent HDAC inhibitor suberoylanilide hydroxamic acid." AIDS research and human retroviruses **25**(2): 207-212.

Archin, N. M., A. Liberty, A. D. Kashuba, S. K. Choudhary, J. Kuruc, A. Crooks, D. Parker, E. Anderson, M. Kearney and M. Strain (2012). "Administration of vorinostat disrupts HIV-1 latency in patients on antiretroviral therapy." Nature **487**(7408): 482.

Archin, N. M. and D. M. Margolis (2006). "Attacking latent HIV provirus: from mechanism to therapeutic strategies." Current Opinion in HIV and AIDS **1**(2): 134-140.

Bailey, J., J. N. Blankson, M. Wind-Rotolo and R. F. Siliciano (2004). "Mechanisms of HIV-1 escape from immune responses and antiretroviral drugs." Current opinion in immunology **16**(4): 470-476.

Bartz, S. R. and M. Emerman (1999). "Human immunodeficiency virus type 1 Tat induces apoptosis and increases sensitivity to apoptotic signals by up-regulating FLICE/caspase-8." Journal of virology **73**(3): 1956-1963.

Battivelli, E., M. S. Dahabieh, M. Abdel-Mohsen, J. P. Svensson, I. T. Da Silva, L. B. Cohn, A. Gramatica, S. Deeks, W. C. Greene and S. K. Pillai (2018). "Chromatin Functional States Correlate with HIV Latency Reversal in Infected Primary CD4+ T Cells." bioRxiv: 242958.

Battivelli, E., M. S. Dahabieh, M. Abdel-Mohsen, J. P. Svensson, I. T. Da Silva, L. B. Cohn, A. Gramatica, S. Deeks, W. C. Greene and S. K. Pillai (2018). "Distinct chromatin functional states correlate with HIV latency reactivation in infected primary CD4+ T cells." Elife **7**: e34655.

Bauby, H., C. C. Ward, R. Hugh-White, C. M. Swanson, R. Schulz, C. Goujon and M. H. Malim (2021). "HIV-1 Vpr Induces Widespread Transcriptomic Changes in CD4+ T Cells Early Postinfection." mBio **12**(3): e01369-01321.

Blazkova, J., K. Trejbalova, F. Gondois-Rey, P. Halfon, P. Philibert, A. Guiguen, E. Verdin, D. Olive, C. Van Lint and J. Hejnar (2009). "CpG methylation controls reactivation of HIV from latency." PLoS pathogens **5**(8): e1000554.

Boehm, D., V. Calvanese, R. D. Dar, S. Xing, S. Schroeder, L. Martins, K. Aull, P.-C. Li, V. Planelles and J. E. Bradner (2013). "BET bromodomain-targeting compounds reactivate HIV from latency via a Tat-independent mechanism." Cell cycle **12**(3): 452-462.

Bruce, J. W., M. Bracken, E. Evans, N. Sherer and P. Ahlquist (2021). "ZBTB2 represses HIV-1 transcription and is regulated by HIV-1 Vpr and cellular DNA damage responses." PLoS Pathogens **17**(2): e1009364.

Bruner, K. M., Z. Wang, F. R. Simonetti, A. M. Bender, K. J. Kwon, S. Sengupta, E. J. Fray, S. A. Beg, A. A. Antar and K. M. Jenike (2019). "A quantitative approach for measuring the reservoir of latent HIV-1 proviruses." Nature **566**(7742): 120-125.

Bui, J. K., J. W. Mellors and A. R. Cillo (2016). "HIV-1 virion production from single inducible proviruses following T-cell activation ex vivo." Journal of virology **90**(3): 1673-1676.

Chen, H.-C., J. P. Martinez, E. Zorita, A. Meyerhans and G. J. Filion (2017). "Position effects influence HIV latency reversal." Nature Structural and Molecular Biology **24**(1): 47.

Cherrier, T., V. Le Douce, S. Eilebrecht, R. Riclet, C. Marban, F. Dequiedt, Y. Goumon, J.-C. Paillart, M. Mericskay and A. Parlakian (2013). "CTIP2 is a negative regulator of P-TEFb." Proceedings of the National Academy of Sciences **110**(31): 12655-12660.

Cohen, Y. Z., J. C. Lorenzi, L. Krassnig, J. P. Barton, L. Burke, J. Pai, C.-L. Lu, P. Mendoza, T. Y. Oliveira and C. Sleckman (2018). "Relationship between latent and rebound viruses in a clinical trial of anti–HIV-1 antibody 3BNC117." Journal of Experimental Medicine **215**(9): 2311-2324.

Consortium, E. P. (2012). "An integrated encyclopedia of DNA elements in the human genome." Nature **489**(7414): 57-74.

Contreras, X., M. Schweneker, C.-S. Chen, J. M. McCune, S. G. Deeks, J. Martin and B. M. Peterlin (2009). "Suberoylanilide hydroxamic acid reactivates HIV from latently infected cells." Journal of Biological Chemistry **284**(11): 6782-6789.

Dahabieh, M. S., M. Ooms, V. Simon and I. Sadowski (2013). "A doubly fluorescent HIV-1 reporter shows that the majority of integrated HIV-1 is latent shortly after infection." Journal of virology **87**(8): 4716-4727.

Darcis, G., A. Kula, S. Bouchat, K. Fujinaga, F. Corazza, A. Ait-Ammar, N. Delacourt, A. Melard, K. Kabeya and C. Vanhulle (2015). "An in-depth comparison of latency-reversing agent combinations in various in vitro and ex vivo HIV-1 latency models identified bryostatin-1+ JQ1 and ingenol-B+ JQ1 to potently reactivate viral gene expression." PLoS pathogens **11**(7): e1005063.

Deeks, S. G. (2012). "HIV: Shock and kill." Nature **487**(7408): 439.

Domingo, E., J. Sheldon and C. Perales (2012). "Viral quasispecies evolution." Microbiology and Molecular Biology Reviews **76**(2): 159-216.

Eckstein, D. A., M. P. Sherman, M. L. Penn, P. S. Chin, C. M. De Noronha, W. C. Greene and M. A. Goldsmith (2001). "HIV-1 Vpr enhances viral burden by facilitating infection of tissue macrophages but not nondividing CD4+ T cells." The Journal of experimental medicine **194**(10): 1407-1419.

Edelstein, L. C., S. Micheva-Viteva, B. D. Phelan and J. P. Dougherty (2009). "Activation of latent HIV type 1 gene expression by suberoylanilide hydroxamic acid (SAHA), an HDAC

inhibitor approved for use to treat cutaneous T cell lymphoma." AIDS research and human retroviruses **25**(9): 883-887.

Einkauf, K. B., G. Q. Lee, C. Gao, R. Sharaf, X. Sun, S. Hua, S. M. Chen, C. Jiang, X. Lian and F. Z. Chowdhury (2019). "Intact HIV-1 proviruses accumulate at distinct chromosomal positions during prolonged antiretroviral therapy." The Journal of clinical investigation **129**(3): 988-998.

Einkauf, K. B., M. R. Osborn, C. Gao, W. Sun, X. Sun, X. Lian, E. M. Parsons, G. T. Gladkov, K. W. Seiger and J. E. Blackmer (2022). "Parallel analysis of transcription, integration, and sequence of single HIV-1 proviruses." Cell.

Emiliani, S., W. Fischle, M. Ott, C. Van Lint, C. A. Amella and E. Verdin (1998). "Mutations in the tat gene are responsible for human immunodeficiency virus type 1 postintegration latency in the U1 cell line." Journal of virology **72**(2): 1666-1670.

Emiliani, S., C. Van Lint, W. Fischle, P. Paras, M. Ott, J. Brady and E. Verdin (1996). "A point mutation in the HIV-1 Tat responsive element is associated with postintegration latency." Proceedings of the National Academy of Sciences **93**(13): 6377-6381.

ENCODE. from https://www.encodeproject.org/experiments/ENCSR000BXX/.

Eriksson, S., E. H. Graf, V. Dahl, M. C. Strain, S. A. Yukl, E. S. Lysenko, R. J. Bosch, J. Lai, S. Chioma and F. Emad (2013). "Comparative analysis of measures of viral reservoirs in HIV-1 eradication studies." PLoS pathogens **9**(2): e1003174.

Fennessey, C. M., M. Pinkevych, T. T. Immonen, A. Reynaldi, V. Venturi, P. Nadella, C. Reid, L. Newman, L. Lipkey and K. Oswald (2017). "Genetically-barcoded SIV facilitates enumeration of rebound variants and estimation of reactivation rates in nonhuman primates following interruption of suppressive antiretroviral therapy." PLoS pathogens **13**(5): e1006359.

Fujii, Y., K. Otake, M. Tashiro and A. Adachi (1996). "Soluble Nef antigen of HIV-1 is cytotoxic for human CD4+ T cells." FEBS letters **393**(1): 93-96.

Gummuluru, S. and M. Emerman (1999). "Cell cycle-and Vpr-mediated regulation of human immunodeficiency virus type 1 expression in primary and transformed T-cell lines." Journal of virology **73**(7): 5422-5430.

Han, Y., Y. B. Lin, W. An, J. Xu, H.-C. Yang, K. O'Connell, D. Dordai, J. D. Boeke, J. D. Siliciano and R. F. Siliciano (2008). "Orientation-dependent regulation of integrated HIV-1 expression by host gene transcriptional readthrough." Cell host & microbe **4**(2): 134-146.

Hansen, M. M., W. Y. Wen, E. Ingerman, B. S. Razooky, C. E. Thompson, R. D. Dar, C. W. Chin, M. L. Simpson and L. S. Weinberger (2018). "A post-transcriptional feedback mechanism for noise suppression and fate stabilization." Cell **173**(7): 1609-1621. e1615.

Hataye, J. M., J. P. Casazza, K. Best, C. J. Liang, T. T. Immonen, D. R. Ambrozak, S. Darko, A. R. Henry, F. Laboune and F. Maldarelli (2019). "Principles governing establishment versus collapse of HIV-1 cellular spread." Cell host & microbe **26**(6): 748-763. e720.

Ho, Y.-C., L. Shan, N. N. Hosmane, J. Wang, S. B. Laskey, D. I. Rosenbloom, J. Lai, J. N. Blankson, J. D. Siliciano and R. F. Siliciano (2013). "Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure." Cell **155**(3): 540-551.

Jefferys, S. R., S. D. Burgos, J. J. Peterson, S. R. Selitsky, A.-M. W. Turner, L. I. James, Y.-H. Tsai, A. R. Coffey, D. M. Margolis and J. Parker (2021). "Epigenomic characterization of latent HIV infection identifies latency regulating transcription factors." PLoS pathogens **17**(2): e1009346.

Jiang, G., E. A. Mendes, P. Kaiser, S. Sankaran-Walters, Y. Tang, M. G. Weber, G. P. Melcher, G. R. Thompson III, A. Tanuri and L. F. Pianowski (2014). "Reactivation of HIV latency by a newly modified Ingenol derivative via protein kinase Cδ–NF-κB signaling." AIDS (London, England) **28**(11): 1555.

Jordan, A., D. Bisgrove and E. Verdin (2003). "HIV reproducibly establishes a latent infection after acute infection of T cells in vitro." The EMBO journal **22**(8): 1868-1877.

Jordan, A., P. Defechereux and E. Verdin (2001). "The site of HIV-1 integration in the human genome determines basal transcriptional activity and response to Tat transactivation." The EMBO journal **20**(7): 1726-1738.

Kharytonchyk, S., S. R. King, C. B. Ndongmo, K. L. Stilger, W. An and A. Telesnitsky (2016). "Resolution of Specific Nucleotide Mismatches by Wild-Type and AZT-Resistant Reverse Transcriptases during HIV-1 Replication." Journal of molecular biology **428**(11): 2275-2288.

Kim, M., N. N. Hosmane, C. K. Bullen, A. Capoferri, H.-C. Yang, J. D. Siliciano and R. F. Siliciano (2014). "A primary CD4+ T cell model of HIV-1 latency established after activation through the T cell receptor and subsequent return to quiescence." Nature protocols **9**(12): 2755-2770.

Kino, T., A. Gragerov, O. Slobodskaya, M. Tsopanomichalou, G. P. Chrousos and G. N. Pavlakis (2002). "Human immunodeficiency virus type 1 (HIV-1) accessory protein Vpr induces transcription of the HIV-1 and glucocorticoid-responsive promoters by binding directly to p300/CBP coactivators." Journal of virology **76**(19): 9724-9734.

Korin, Y. D., D. G. Brooks, S. Brown, A. Korotzer and J. A. Zack (2002). "Effects of prostratin on T-cell activation and human immunodeficiency virus latency." Journal of virology **76**(16): 8118-8123.

Kulkosky, J., J. Sullivan, Y. Xu, E. Souder, D. H. Hamer and R. J. Pomerantz (2004). "Expression of latent HAART-persistent HIV type 1 induced by novel cellular activating agents." AIDS research and human retroviruses **20**(5): 497-505.

Laird, G. M., C. K. Bullen, D. I. Rosenbloom, A. R. Martin, A. L. Hill, C. M. Durand, J. D. Siliciano and R. F. Siliciano (2015). "Ex vivo analysis identifies effective HIV-1 latency–reversing drug combinations." The Journal of clinical investigation **125**(5): 1901-1912.

Lassen, K., Y. Han, Y. Zhou, J. Siliciano and R. F. Siliciano (2004). "The multifactorial nature of HIV-1 latency." Trends in molecular medicine **10**(11): 525-531.

Le Douce, V., G. Herbein, O. Rohr and C. Schwartz (2010). "Molecular mechanisms of HIV-1 persistence in the monocyte-macrophage lineage." Retrovirology **7**(1): 1-16.

Lenasi, T., X. Contreras and B. M. Peterlin (2008). "Transcriptional interference antagonizes proviral gene expression to promote HIV latency." Cell host & microbe **4**(2): 123-133.

Li, C., L. Mori and S. T. Valente (2021). "The block-and-lock strategy for human immunodeficiency virus cure: lessons learned from didehydro–cortistatin A." The Journal of Infectious Diseases **223**(Supplement_1): S46-S53.

Lu, C.-L., J. A. Pai, L. Nogueira, P. Mendoza, H. Gruell, T. Y. Oliveira, J. Barton, J. C. Lorenzi, Y. Z. Cohen and L. B. Cohn (2018). "Relationship between intact HIV-1 proviruses in circulating CD4+ T cells and rebound viruses emerging during treatment interruption." Proceedings of the National Academy of Sciences **115**(48): E11341-E11348.

Marsden, M. D., T.-h. Zhang, Y. Du, M. Dimapasoc, M. S. Soliman, X. Wu, J. T. Kim, A. Shimizu, A. Schrier and P. A. Wender (2020). "Tracking HIV Rebound following Latency Reversal Using Barcoded HIV." Cell Reports Medicine **1**(9): 100162.

Mbonye, U., B. Wang, G. Gokulrangan, W. Shi, S. Yang and J. Karn (2018). "Cyclin-dependent kinase 7 (CDK7)-mediated phosphorylation of the CDK9 activation loop promotes P-TEFb assembly with Tat and proviral HIV reactivation." Journal of Biological Chemistry **293**(26): 10009-10025.

McNamara, L. A., J. A. Ganesh and K. L. Collins (2012). "Latent HIV-1 infection occurs in multiple subsets of hematopoietic progenitor cells and is reversed by NF-κB activation." Journal of virology **86**(17): 9337-9350.

Mendoza, P., J. R. Jackson, T. Y. Oliveira, C. Gaebler, V. Ramos, M. Caskey, M. Jankovic, M. C. Nussenzweig and L. B. Cohn (2020). "Antigen-responsive CD4+ T cell clones contribute to the HIV-1 latent reservoir." Journal of Experimental Medicine **217**(7).

Miller, C. M., H. Akiyama, L. M. Agosto, A. Emery, C. R. Ettinger, R. I. Swanstrom, A. J. Henderson and S. Gummuluru (2017). "Virion-associated Vpr alleviates a postintegration block to HIV-1 infection of dendritic cells." Journal of virology **91**(13): e00051-00017.

Neidleman, J., X. Luo, J. Frouard, G. Xie, F. Hsiao, T. Ma, V. Morcilla, A. Lee, S. Telwatte and R. Thomas (2020). "Phenotypic analysis of the unstimulated in vivo HIV CD4 T cell reservoir." Elife **9**: e60933.

Ohagen, A., S. Ghosh, J. He, K. Huang, Y. Chen, M. Yuan, R. Osathanondh, S. Gartner, B. Shi and G. Shaw (1999). "Apoptosis induced by infection of primary brain cultures with diverse human immunodeficiency virus type 1 isolates: evidence for a role of the envelope." Journal of virology **73**(2): 897-906.

Pace, M. J., L. Agosto, E. H. Graf and U. O'Doherty (2011). "HIV reservoirs and latency models." Virology **411**(2): 344-354.

Peluso, M. J., P. Bacchetti, K. D. Ritter, S. Beg, J. Lai, J. N. Martin, P. W. Hunt, T. J. Henrich, J. D. Siliciano and R. F. Siliciano (2020). "Differential decay of intact and defective proviral DNA in HIV-1–infected individuals on suppressive antiretroviral therapy." JCI insight **5**(4).

Pizzato, M., O. Erlwein, D. Bonsall, S. Kaye, D. Muir and M. O. McClure (2009). "A one-step SYBR Green I-based product-enhanced reverse transcriptase assay for the quantitation of retroviruses in cell culture supernatants." Journal of virological methods **156**(1-2): 1-7.

Poon, B. and I. S. Chen (2003). "Human immunodeficiency virus type 1 (HIV-1) Vpr enhances expression from unintegrated HIV-1 DNA." Journal of virology **77**(7): 3962-3972.

Read, D. F., E. Atindaana, K. Pyaram, F. Yang, S. Emery, A. Cheong, K. R. Nakama, C. Burnett, E. T. Larragoite, E. Battivelli, E. Verdin, V. Planelles, C.-H. Chang, A. Telesnitsky and J. M. Kidd (2019). "Stable integrant-specific differences in bimodal HIV-1 expression patterns revealed by high-throughput analysis." PLOS Pathogens **15**(10): e1007903.

Romani, B., R. Kamali Jamil, M. Hamidi-Fard, P. Rahimi, S. B. Momen, M. R. Aghasadeghi and E. Allahbakhshi (2016). "HIV-1 Vpr reactivates latent HIV-1 provirus by inducing depletion of class I HDACs on chromatin." Sci Rep **6**: 31924.

Rosier, B. J. and T. F. De Greef (2015). "Synthetic Biology: How to make an oscillator." Elife **4**: e12260.

Sato, K., N. Misawa, S. Iwami, Y. Satou, M. Matsuoka, Y. Ishizaka, M. Ito, K. Aihara, D. S. An and Y. Koyanagi (2013). "HIV-1 Vpr accelerates viral replication during acute infection by exploitation of proliferating CD4+ T cells in vivo." PLoS pathogens **9**(12): e1003812.

Sherrill-Mix, S., M. K. Lewinski, M. Famiglietti, A. Bosque, N. Malani, K. E. Ocwieja, C. C. Berry, D. Looney, L. Shan and L. M. Agosto (2013). "HIV latency and integration site placement in five cell-based models." Retrovirology **10**(1): 90.

Simonetti, F. R., H. Zhang, G. P. Soroosh, J. Duan, K. Rhodehouse, A. L. Hill, S. A. Beg, K. McCormick, H. E. Raymond and C. L. Nobles (2021). "Antigen-driven clonal selection shapes the persistence of HIV-1–infected CD4+ T cells in vivo." The Journal of clinical investigation **131**(3).

Spina, C. A., J. Anderson, N. M. Archin, A. Bosque, J. Chan, M. Famiglietti, W. C. Greene, A. Kashuba, S. R. Lewin and D. M. Margolis (2013). "An in-depth comparison of latent HIV-1 reactivation in multiple cell model systems and resting CD4+ T cells from aviremic patients." PLoS pathogens **9**(12): e1003834.

Spivak, A. M., A. Andrade, E. Eisele, R. Hoh, P. Bacchetti, N. N. Bumpus, F. Emad, R. Buckheit III, E. F. McCance-Katz and J. Lai (2013). "A pilot study assessing the safety and latency-reversing activity of disulfiram in HIV-1–infected adults on antiretroviral therapy." Clinical infectious diseases **58**(6): 883-890.

Stewart, S. A., B. Poon, J. Jowett and I. Chen (1997). "Human immunodeficiency virus type 1 Vpr induces apoptosis following cell cycle arrest." Journal of Virology **71**(7): 5579-5592.

Telwatte, S., S. Morón-López, D. Aran, P. Kim, C. Hsieh, S. Joshi, M. Montano, W. C. Greene, A. J. Butte and J. K. Wong (2019). "Heterogeneity in HIV and cellular transcription profiles in cell line models of latent and productive infection: implications for HIV latency." Retrovirology **16**(1): 1-23.

Vansant, G., H.-C. Chen, E. Zorita, K. Trejbalová, D. Miklík, G. Filion and Z. Debyser (2020). "The chromatin landscape at the HIV-1 provirus integration site determines viral expression." Nucleic acids research **48**(14): 7801-7817.

Verdin, E., P. Paras and C. Van Lint (1993). "Chromatin disruption in the promoter of human immunodeficiency virus type 1 during transcriptional activation." The EMBO journal **12**(8): 3249-3259.

Wang, Z., E. E. Gurule, T. P. Brennan, J. M. Gerold, K. J. Kwon, N. N. Hosmane, M. R. Kumar, S. A. Beg, A. A. Capoferri and S. C. Ray (2018). "Expanded cellular clones carrying replication-competent HIV-1 persist, wax, and wane." Proceedings of the National Academy of Sciences **115**(11): E2575-E2584.

Weinberger, A. D. and L. S. Weinberger (2013). "Stochastic fate selection in HIV-infected patients." Cell **155**(3): 497-499.

Weinberger, L. S., J. C. Burnett, J. E. Toettcher, A. P. Arkin and D. V. Schaffer (2005). "Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity." Cell **122**(2): 169-182.

Williams, S. A., L.-F. Chen, H. Kwon, D. Fenard, D. Bisgrove, E. Verdin and W. C. Greene (2004). "Prostratin antagonizes HIV latency by activating NF-κB." Journal of Biological Chemistry **279**(40): 42008-42017.

Williams, S. A. and W. C. Greene (2007). "Regulation of HIV-1 latency by T-cell activation." Cytokine **39**(1): 63-74.

Wonderlich, E. R., K. Subramanian, B. Cox, A. Wiegand, C. Lackman-Smith, M. J. Bale, M. Stone, R. Hoh, M. F. Kearney, F. Maldarelli, S. G. Deeks, M. P. Busch, R. G. Ptak and D. A. Kulpa (2019). "Effector memory differentiation increases detection of replication-competent HIV-l in resting CD4+ T cells from virally suppressed individuals." PLoS pathogens **15**(10): e1008074.

Xing, S., C. K. Bullen, N. S. Shroff, L. Shan, H.-C. Yang, J. L. Manucci, S. Bhat, H. Zhang, J. B. Margolick and T. C. Quinn (2011). "Disulfiram reactivates latent HIV-1 in a Bcl-2-transduced primary CD4+ T cell model without inducing global T cell activation." Journal of virology **85**(12): 6060-6064.

Xu, X.-N., B. Laffert, G. R. Screaton, M. Kraft, D. Wolf, W. Kolanus, J. Mongkolsapay, A. J. McMichael and A. S. Baur (1999). "Induction of Fas ligand expression by HIV involves the

interaction of Nef with the T cell receptor ζ chain." Journal of Experimental Medicine **189**(9): 1489-1496.

Yang, H.-C., S. Xing, L. Shan, K. O'Connell, J. Dinoso, A. Shen, Y. Zhou, C. K. Shrum, Y. Han and J. O. Liu (2009). "Small-molecule screening using a human primary cell model of HIV latency identifies compounds that reverse latency without cellular activation." The Journal of clinical investigation **119**(11): 3473-3486.

Yang, J., Y. Zhao, M. Kalita, X. Li, M. Jamaluddin, B. Tian, C. B. Edeh, J. E. Wiktorowicz, A. Kudlicki and A. R. Brasier (2015). "Systematic determination of human cyclin dependent kinase (CDK)-9 interactome identifies novel functions in RNA splicing mediated by the DEAD Box (DDX)-5/17 RNA helicases." Molecular & Cellular Proteomics **14**(10): 2701-2721.

Yao, X.-J., A. J. Mouland, R. A. Subbramanian, J. Forget, N. Rougeau, D. Bergeron and E. A. Cohen (1998). "Vpr stimulates viral expression and induces cell killing in human immunodeficiency virus type 1-infected dividing Jurkat T cells." Journal of virology **72**(6): 4686-4693.

Zhang, F. and P. D. Bieniasz (2020). "HIV-1 Vpr induces cell cycle arrest and enhances viral gene expression by depleting CCDC137." Elife **9**: e55806.

Zhang, Q., Y. Kang, S. Wang, G. M. Gonzalez, W. Li, H. Hui, Y. Wang and T. M. Rana (2021). "HIV reprograms host m6Am RNA methylome by viral Vpr protein-mediated degradation of PCIF1." Nature Communications **12**(1): 1-13.

Zhu, J., G. D. Gaiha, S. P. John, T. Pertel, C. R. Chin, G. Gao, H. Qu, B. D. Walker, S. J. Elledge and A. L. Brass (2012). "Reactivation of latent HIV-1 by inhibition of BRD4." Cell reports **2**(4): 807-816.

Chapter 4 Discussion

4.1 ***Preliminary results, summary of dissertation, and discussion of results***

This chapter presents a discussion of my overall findings, preliminary experiments that further explore correlates of virion burst size variation among clones, the implications of these findings, and some suggestions for follow-up experiments.

4.2 Summary and discussion of results

I developed a high-throughput system to simultaneously study the expression properties of thousands of individual proviruses in a polyclonal pool of HIV-1 infected cells. This allowed for the analysis of the total pool's behavior while defining the contribution of individual member's behavior to the pool effect. Individual viruses were marked by a unique 20-base nucleotide sequence "barcode" that was incorporated into the U3 region and was duplicated into the long terminal repeats (LTRs) upon reverse transcription. Initial experimental results showed a similar population-wide trend to the unmodified version of the lentiviral vector and indicated that the incorporation of barcodes did not significantly alter the vector biology (Battivelli, Dahabieh et al. 2018).

Existing tissue culture models of HIV-1 latent infection are either composed of monoclonal or polyclonal populations. In studies where a monoclonal population is used, the results may not be

generalizable because HIV semi-randomly integrates into the host cells' DNA, and proviruses in different locations in the host cells' genome may be subject to locus-specific regulation. Therefore, the *in vitro* system described in this dissertation helps resolves some of the limitations of previous experimental systems.

Contrary to previous estimates that a large proportion of residual viruses in HIV-1 patients are defective (Chun, Stuyver et al. 1997, Finzi, Hermankova et al. 1997, Wong, Hezareh et al. 1997), in Chapter 2, the results show that mutational inactivation was rare in a single round of replication. However, tissue culture lacks the antiviral activities of the host that might differentially eliminate intact virus-containing cells. Paired with recent findings that defective proviruses decay slower than intact ones (Peluso, Bacchetti et al. 2020) our results suggest that the large proportion of defective proviruses in patients may be due more to selection against the persistence of infectious proviruses rather than to defective provirus formation.

While the results in Chapter 2 clearly show how rare these inactivation events occur, the findings are limited in sample size, by cell-type, and in not providing insight into the nature of mutations that inactivate proviruses. This experiment was performed in the HEK 293T transformed kidney cell line, which is not the natural host of HIV-1. Moreover, because HIV-1 infects less metabolically-active CD4+ expressing cell-types such as macrophages and microglia in addition to T cells (Epstein, Pantaleo et al. 1993, Burger and Poles 2003, Alexaki and Wigdahl 2008), it is possible that the rate of mutational inactivation may differ in other cell types, such as in macrophages where reduced dNTP concentrations lead to increases in mutations during reverse transcription by such mechanisms as increased misincorporation of ribonucleotides (rNTPs) by

RT (Kennedy, Amie et al. 2012). Therefore, further experiments are needed to ascertain the inactivation rate in these CD4+ expressing cell types. The nature of the mutations can also be tested by employing the use of nanopore sequencing technology to sequence whole proviral DNA.

Previous studies using a monoclonal population of provirus-containing cells showed that the active and silent/inactive binary expression status is stochastically determined and can be stabilized by a positive feedback loop (Weinberger, Burnett et al. 2005, Weinberger and Weinberger 2013). The work in Chapter 2 provides more insight by demonstrating that, in a pool of hundreds of clones whose proviruses were keyed to distinct integration sites, each clone's daughter cells established its own pattern of expression. This pattern of expression consisted of clonal cell populations that were phenotypically mixed with regard to HIV-1 expression, with fractions of LTR-active daughter cells that ranged from 0 to 100%. While intra-clonal variation was consistent with Weinberger et. al. (Weinberger and Weinberger 2013), we also observed inter-clonal variation. If the on or off transcriptional fate determination were entirely stochastic, we would not expect the significant differences in the percent of LTR-active daughter cells among clones that we did observe. Therefore, the results suggest that integration sites may contribute to the fate determination of clonal populations. Not only did the percentage of LTR-active cells differ among clones, but each clone maintained the same percentage of LTR-active cells over the course of nine days. This was despite the observation that daughter cell members of clones switched between the on and off transcriptional states within the same timeframe.

Stochastic gene expression has been proposed to be a bet hedging strategy that the virus uses to generate a wide range of expression phenotypes (Weinberger and Weinberger 2013). Therefore, it

may guarantee the survival of distinct subpopulations of daughter cells under different selective pressures. This stochastic fate determination strategy presents the following implications: 1) that independent of their integration sites, clones can generate daughter cells with a wide range of expression phenotypes, and that daughter cells from distinct clones will display the same spectra of expression profiles, 2) expression phenotypes among clones won't differ significantly, and 3) under selective pressures, all clones will have an equal chance of surviving. However, growing evidence suggests that persistent proviruses that are intact tend to accumulate in distinct chromosomal positions (Einkauf, Lee et al. 2019) and that not all integrated proviruses can persist in the face of selective pressures.

The results in Chapter 2 suggest a model that is supported by several studies. First, the observed differing percentages of LTR-activity that are stable and heritable among clones implies that proviral clones might decay differentially (display different half-lives), and such differences have been reported (Fischer, Joos et al. 2008, Peluso, Bacchetti et al. 2020). Clones in which the majority of cells are LTR-active may stay on long enough to either reveal themselves for immune clearance or accumulate viral cytotoxic proteins to levels that are high enough to kill those cells, and thus such clones may be rapidly lost. Meanwhile, clones with very few LTR-active cells at any given time will decay much more slowly since the majority of the clone's member cells are indistinguishable from uninfected cells *in vivo* and invisible to detection by the immune system.

LTR-inactive cells may not be completely silenced. Our data also showed that daughter cells of clones switched from on to off phenotypes and vice versa, thus suggesting that the clones' expression might be in an oscillating cycle. This appears to correlate with studies that show blips

of viremia in virally suppressed patients (Lee, Kieffer et al. 2006, Sörstedt, Nilsson et al. 2016), and suggests the possibility that the biomodal expression property described here may contribute to this phenomenon. Previous explanations for the observed blips of viremia in patients include that described by stochastic population switch models (Wang and Rong 2014), and the activation of resting CD4+ T cells that harbor latent proviruses (Jones and Perelson 2007).

HIV-1 latency is one of the main hurdles to curing the disease partly because of the oligoclonal nature of patients' reservoir. Chapter 3 explores how integrants' basal expression properties might help define which proviral clones contribute to HIV-1 persistence and the outcomes of latency reversal. Chapter 3 also compares the effect of Vpr (which was not present in our initial studies in Chapter 2) on the proviral landscape and latency reversal. We found that low LTR-active clones are capable of persisting in the presence of the cytotoxic protein Vpr and that these clones support higher burst sizes than high LTR-active clones. Furthermore, proviruses that are *vpr*+ are more resistant to reactivation by the tested LRAs and have clonal patterns of reactivation that differ from *vpr*- proviruses.

Some estimates suggest that the majority of infected cells are latent *in vitro* while in other studies, a small fraction is latent both *in vitro* and in patients (Jordan, Bisgrove et al. 2003, Eriksson, Graf et al. 2013). By comparing *vpr*+ and *vpr*- infected pools, we found that most infected cells in the *vpr*- pools were LTR-active, which is consistent with estimates that latent infections are infrequent (Eriksson, Graf et al. 2013). However, most infected cells in the *vpr*+ pools were LTR-inactive, possibly because most active cells died and did not contribute to the population, consistent with previous estimates of infected cells' half-life of about two days due to viral cytotoxic effects

(Perelson, Neumann et al. 1996, Stewart, Poon et al. 1997, Andersen, Le Rouzic et al. 2008). Because Vpr was not previously considered important to latency and inclusion of *vpr* has been inconsistent, the results in Chapter 2 contribute to explaining the discrepancies among previous studies.

Contrary to observations with *vpr*- pools, which included clonal populations whose percentages of LTR-activity ranged from 0-100%, *vpr*+ pool clones' LTR-activity only ranged from 0-60%. Mapping integration sites and examining their genomic loci revealed that high LTR active clones were closer in linear distance to the nearest H3K27ac and H3K4me3 marks of the host cells' chromatin than were low LTR-active clones. Because the proviral integration sites of high LTR-active clones were more proximal to these genome marks, the selective depletion of high LTR-active clones resulted in differences in the proviral integration landscape between *vpr*- and *vpr*+ pools. These effects of Vpr were evidently largely post-integration, because when *vpr*- pools were complemented with a Vpr expression vector, the complemented pool acquired integration landscape properties observed for the *vpr*+ pool. Interestingly high LTR-active clones were lost in the complemented pool regardless of viral burst sizes.

Some prominent *in vitro* culture and small animal models for studying HIV use Vpr-defective vectors (Jordan, Bisgrove et al. 2003, Pace, Agosto et al. 2011). Without Vpr, most high LTR-active clones survive and are often chosen for further studies. Our data shows that while these models contribute to our understanding of the basic mechanisms involved in HIV-1 latency establishment and reactivation, the residual reservoir in patients may consist of oligoclonal populations whose expression properties differ from those that are frequently chosen for studies in

in vitro models. Therefore, this study highlights the importance of including *vpr* in both tissue culture and animal models of HIV-1 infection and latency reversal studies.

Chapter 3 further demonstrates that the effect of Vpr in shaping the proviral landscape may be occurring *in vivo* since integration site proximities to H3K27ac and H3K4me3 marks in HIV-1 patients were more similar to those in the *vpr*+ pools than in the *vpr*- pools. These findings agree with previous data that the integration sites of persistent proviruses in patients on prolonged ART are different than those in acute infection (Einkauf, Lee et al. 2019). Our work defines an effect of Vpr in shaping the proviral population structure that was previously not described.

The expression properties of proviruses were further examined by measuring the amount of virus release once the proviruses were LTR-active. Interestingly, the clones' percentage of LTR-active cells was slightly negatively correlated with virus release. This suggested that low LTR-active clones had higher burst sizes than high LTR-active clones. Therefore, the few transcriptionally active cells of low LTR-active clones could support robust virus release and may help explain the blips of viremia in ART treated patients.

Just like all other viruses, the goal of HIV-1 is to persist and replicate. This makes HIV-1 gene expression a very critical component in the selection of persistent viruses under different anatomical conditions that might exert differing selective pressures. For example, in the central nervous system, lymphoid organs, and in macrophages, where higher drug doses are required to reach effective concentrations, the selective pressures are different from those in blood. Therefore, the expression properties of HIV-1 proviral clones that persist in these anatomical sites may differ.

Expanded clones can mask the dynamics of clonal reactivation by dominating the total pools' response. Clonal expansion by T cell receptor (TCR) engagement and by integration sites have been reported to contribute to HIV-1 persistent *in vivo* (Liu, Simonetti et al. 2020, Simonetti, Zhang et al. 2021, Yeh, Yang et al. 2021). We observed several examples of clonal expansion in the pools used in the studies in this dissertation. For instance, in some pools, a single clone could account for more than 30% of the total sequencing read counts after a few weeks of passage. In such instances, the total pool's behavior reflected the expression properties of the most dominant clone(s) and highlighted the importance of defining how each clone contributes to the total pool's behavior.

The reasons for clonal expansion in our system are still unclear and may require further investigation. It is unlikely that the expansion was driven by antigen-specific TCR engagement because all the clones had the same TCR specificity (a Jurkat clonal cell line was used). Speculatively, and in agreement with studies by Maldarelli et. al., (Maldarelli, Wu et al. 2014), the dominance of these clone(s) may be driven by integration sites. However, further experiments will be needed to ascertain the mechanism(s) of clonal expansion observed here.

Having observed that the virus release per LTR-active cell varied by about three orders of magnitude, and that the association between integration sites and expression properties was relatively weak, the differences in viral burst sizes among clones was further investigated. The preliminary results of an unfinished manuscript on these findings are described here:

To address how closely virus release was associated with GFP expression, culture supernatant was harvested from GFP+ and GFP- sorted subpools of the *vpr*- pool in Chapter 3, Figure. 3.1C (left panel). p24 concentrations in the culture media were then defined, and the results showed that GFP+ pool cells released about 1800-fold more virus than GFP- cells after subtraction of background values for uninfected controls (Figure. 4.1A).

To further characterize the burst size per cell for each clone, zip codes from pooled cells' DNA and viral cDNA were subjected to high throughput sequencing. Zip code abundance in unsorted cellular DNA and viral cDNA showed a positive correlation (Figure. 4.1B), However, when the viral cDNA abundance was normalized to zip code abundances in GFP+ cellular DNA, a weaker correlation was observed. These results suggested that both the size of a clone and the percentage of its total cells that were GFP positive contributed to the total clonal virus release. This is consistent with Chapter 3 data showing virus release spanned three orders of magnitude and was negatively correlated with %GFP+ (Chapter 3 Figure. 3.2C).

Recent evidence suggests that variation in HIV-1 promoter sequences can correlate with expression phenotypes (Qu, Li et al. 2016, Ali, Bhange et al. 2021). And indeed, some variation in HIV-1 gene expression is due to differences in the virus's promoter region. For example, HIV-1 subtype C has been reported to have low expression due to an additional copy of the retinoblastoma family variant 2 transcription factor (RBF-2) binding site (Ali, Bhange et al. 2021). This and other studies provide evidence that mutations in HIV-1 promoter elements can contribute to the basal expression levels, transcriptional noise, and other differences in expression (van Opijnen, Boerlijst et al. 2006, Qu, Li et al. 2016).

Although all proviral clones were derived from a single construct here, errors during reverse transcription could have generated sequence heterogeneity. Therefore, to test whether or not the clonal burst size variations observed correlated with proviral promoter variation, PCR-amplicons spanning U3-R were generated (Figure. 4.2A). Sequencing results showed no mutations within the SP1 and NFkB binding sites, and that all clones contained parental transcription start sites (Figure. 4.2B), suggesting burst size variation is not due to promoter sequences.

HIV-1 gene expression relies on a fine balance of unspliced and spliced viral RNAs, as virion assembly requires proteins translated from mRNA transcripts across the various splice classes, as well as unspliced transcripts to either serve as the genome packaged into virions or as Gag and Gag-Pol polyprotein mRNAs (Ocwieja, Sherrill-Mix et al. 2012). Having determined that burst size differences observed here were not due to variation in promoter sequence, the ratio of unspliced to spliced HIV-1 cellular RNA among clones was measured. To achieve this, cellular RNA extracted from the GFP+ subpool was subjected to polyA purification, and unspliced RNA was extracted due to its binding to a biotinylated oligonucleotide complementary to intronic sequences (Figure. 4.3A). qPCR quantification of various contigs within the cDNA from total cell, polyA+, and unspliced RNA confirmed successful enrichment of the unspliced RNA. Analysis revealed that the ratios of unspliced to spliced HIV-1 RNA per GFP+ cell differed among clones (Figure. 4.3B).

Based on these ratios, clones were grouped into “high unspliced” and into “low unspliced” clones. Burst sizes, or virus release per GFP+ cell, were then compared between the two groups (Figure.

4.3C) (p-value=0.001, student t-test). The results revealed that clones with higher unspliced:spliced HIV-1 RNA ratios released significantly more virus than “low unspliced” clones. These results suggest that HIV-1 RNA splicing differs among integrants and contributes to burst size variation among clones.

These preliminary results suggest that transcripts from HIV-1 proviruses integrated at distinct genomic locations are differentially spliced. By measuring the unspliced and spliced RNA per GFP+ cell, we show that virion production diminishes as the balance of RNAs favors spliced variants. This suggests that an over-splicing phenotype may be detrimental to the spread of the virus during ongoing replication.

The data here showed that the amount of HIV-1 RNA produced by each clonal population was positively correlated with that clone’s relative abundance in the pools’ cell DNA. However, the correlation between the RNA per GFP+ cell and proviral DNA abundance was much lower. This is consistent with Chapter 3 results, which showed that clones that rarely express HIV genes can support high bursts in virion release during brief proviral activity and indicated that the clone size was not a good indicator of the level of HIV expression. Whether or not this also is true *in vivo* is unknown. Clonal expansion of HIV-1 infected cells is an important feature for persistence in patients (Wang, Gurule et al. 2018, Liu, Simonetti et al. 2020). Recent evidence suggests that the reservoir may not be completely silenced and that cells with ongoing low frequency of expression persist (Atindaana, Kissi-Twum et al. , Einkauf, Osborn et al. 2022), and even low-level expressing cells can contribute to the rekindling of viral replication when ART is interrupted. Our preliminary data shows that low-level expression can be achieved by regulating the balance of spliced and

unspliced HIV-1 RNA variants. Therefore, even when HIV-1 DNA is integrated into regions of high transcriptional activity, differences in splicing may represent another layer of regulation of gene expression that the virus utilizes to achieve a repertoire of expression phenotypes.

Together, these results showed that integrant clones' GFP+ cells, which were those cells in which the HIV-1 promoter was active, released clone-specific amounts of virus. Differences among clones were not due the nucleotide sequence of the HIV-1 promoter. However, when the ratios of unspliced to total RNAs were compared, the ratios were observed to span about one order of magnitude, and clones with high virus release had a significantly higher amount of unspliced RNA than clones with low virus release.

In summary, after stable integration of HIV-1 DNA into the host chromosomes, proviruses established their own patterns of expression that were comprised of waves of LTR activity and inactivity. The majority of daughter cells from clones with proviruses integrated close to the histone marks H3K4me3 and H3K27ac were LTR-active (high LTR-active clones). On the other hand, only a small fraction of daughter cells from clones whose proviruses were integrated farther away from these marks were LTR-active (low LTR-active clones). These patterns of LTR-activity differed among clones and once established were transmitted across cell generations. Low LTR-active clones supported higher viral burst sizes than high LTR-active clones, possibly due in part to higher unspliced to spliced HIV RNA ratios in low LTR-active clones than in high LTR-active clones. In the presence of Vpr, high LTR-active clones died rapidly but low LTR-active clones were favored to persist. A schematic representation of this model is shown in Figure. 4.4.

4.3 ***Future directions***

My dissertation contributes to understanding of the variation in HIV-1 gene expression patterns, how the virus's intrinsic cytotoxicity causes the selection of a subset of clonal populations to persist in dividing cell populations, and parameters that influence the response of persistent proviruses to latency reversing agents. However, several open questions emerged from these findings that remain to be addressed. The following paragraphs outline some of the hypotheses that were generated during my studies and how they may be addressed.

4.4 ***The role of the viral promoter in the stable heritable pattern of gene expression.***

One of the major findings in Chapter 2 was that clonal populations had different percentages of LTR-active cells. Over a period of 9 days, the proportion of clonal LTR-active cells remained relatively the same despite the observation that cells within the LTR-active (GFP+ cells) and the LTR-inactive (GFP-) subpopulations switched between the two LTR-activity states (a phenomenon we referred to as "flickering"). Whether flickering is a function of the HIV-1 promoter is not known. This can be addressed by testing whether this heritable pattern of bimodal expression can be observed with other cellular promoters. To achieve this, a barcoded HIV-1 based self-inactivating (SIN) lentiviral vector can be designed to incorporate a test promoter upstream of the major 5' splice site and the Gag ORF. Experiments in Chapter 2 will need to be repeated with the SIN vector to ascertain if flickering is a function the HIV-1 promoter or is also observed for test promoters.

4.5 ***What is the timing of flickering among clones?***

The work presented in this dissertation suggests that latent HIV-1 expression is not completely silenced, and clonal populations may exist in an oscillation cycle that consists of waves of LTR activity and inactivity (flickering). How flickering is achieved and whether it differs among clones is still not known. A possibility is that the asymmetric inheritance of epigenetic marks and or host cellular factors can result in the temporal on/off phenotypic switches. Therefore, the on and off switches will only be observed once a cell divides. This hypothesis would be consistent with a previous study that spontaneous reactivation is linked to the cell cycle (Kok, Schmutz et al. 2018). To address this, experiments will be repeated in an adherent cell line such as 293T cells to allow for imaging by microscopy. Images of individual cells can then be taken during incubation for several cell divisions in an incucyte (Lanigan, Rasmussen et al. 2020). The incucyte technology allows for real time imaging of phase contrast and fluorescent samples in a culture medium. By observing different integrant clones and comparing the frequency of switching, how flickering differs among clones can be addressed.

4.6 ***Epigenetic silencing of HIV-1 by HDACs may not be limited to class I***

Experiments using J89GFP cells showed that class I HDACs are the most relevant in suppressing HIV-1 expression (Keedy, Archin et al. 2009). Hence, this led to suggestions that class-specific HDAC inhibitors may be promising candidates for reactivation of latent HIV-1 (Matalon, Rasmussen et al. 2011). Here, whether HIV-1 proviral silencing is limited to class I HDACs *in vitro* was investigated by comparing the responsiveness of a polyclonal population of latent cells to the pan-HDAC inhibitor SAHA, and to the class I-specific inhibitor entinostat. GFP- subpools of cells harboring zip coded proviruses were subjected to SAHA and entinostat treatments for 48

hours, at which point a significant increase in reactivation as measured by GFP+, relative to a DMSO only control, was observed (Figure. 4.5A). Unexpectedly, SAHA showed a significantly higher frequency of GFP+ cells than entinostat did, but no significant difference in the total amount of virus released (Figure. 4.5B) possibly due to relatively fewer viable cells in the SAHA treatment at 48 hours (data not shown) and/or the preservation of factors required for maximal HIV-1 expression by the class I-specific HDAC inhibitor; entinostat (Zaikos, Painter et al. 2018). However, the lower frequency of GFP+ cells relative to that observed after treatment with SAHA suggests that there may be clone(s) preferentially targeted by SAHA that are not entinostat responsive. To further explore if there were differences in clone targeting, zip code libraries generated from viral cDNA were compared among the treatment conditions. The results revealed that amongst the 10 most abundant zip codes, SAHA reactivated proviral clones 1, 5, and 7 while entinostat had little to no effect on these (Figure. 4.5C). Although the effect of SAHA on other cellular non-histone factors that affect the reactivation of HIV-1 cannot be ruled out, the results suggest that HIV-1 epigenetic silencing might not be limited to class I HDACs.

The promoter of HIV-1 is wrapped around three distinct nucleosomes, Nuc0, Nuc1 and Nuc2, that are positioned at -415/-255, +10/+155, and +256/+412 respectively. These nucleosomes can be modified by histone modifying enzymes that may affect viral gene expression. To further examine whether non-class I HDACs may be involved in the silencing HIV-1 expression, chromatin immunoprecipitation experiments using anti HDAC-specific antibodies should be performed on untreated and on entinostat and SAHA treated samples to determine which clones (zip codes) are associated with the specific HDACs.

4.7 The role of EZH2 in flickering

Enhancer of zeste homolog 2 (EZH2) is the enzymatic catalytic subunit of the polycomb repressive complex 2 (PRC2) that can alter downstream target genes expression by trimethylation of Lys-27 in histone 3 (H3K27me3). EZH2 can also regulate gene expression in ways besides H3K27me3 (He, Shen et al. 2012, Kim, Lee et al. 2018). In a preliminary experiment to test whether the on to off transition can be blocked by the EZH2 inhibitor GSK343, a GFP+ sorted *vpr*- sub-pool was treated with 2 μM GSK343 and compared to a DMSO control after 72 hours. The results revealed a more than 2-fold decrease in the on to off transition of GFP+ cells in the inhibitor treated cells (Figure. 4.6 A and B). This finding suggests that the on state can be stabilized by the inhibition of EZH2. This experiment needs to be repeated to ensure reproducibility. Also, EZH2 knockdown experiments should be conducted to ascertain if the observed effect is due to off-target effects of GSK343 treatment.

4.8 Conclusions

To summarize this thesis, my work established a high throughput approach to track and analyze the expression properties of thousands of individual proviral clones within polyclonal populations, defined expression properties and the integration site landscape of the subpopulation of proviruses that is selected to persist in the presence of Vpr, and demonstrated differences in the spectra of clonal responses to the latency reversing agents Prostratin and JQ1 when they were applied both separately and in combination. Importantly, my work also showed that HIV-1 gene expression displays bimodal on/off patterns that differ among clones and may exist in an oscillation cycle. The waves of on/off oscillations could be a significant contributor to blips of viremia in some virally suppressed patients, and therefore serve as a bet-hedging strategy for the virus. The future

work that I have proposed should continue to advance our understanding of HIV-1 expression variability and inform studies of HIV-1 persistence.

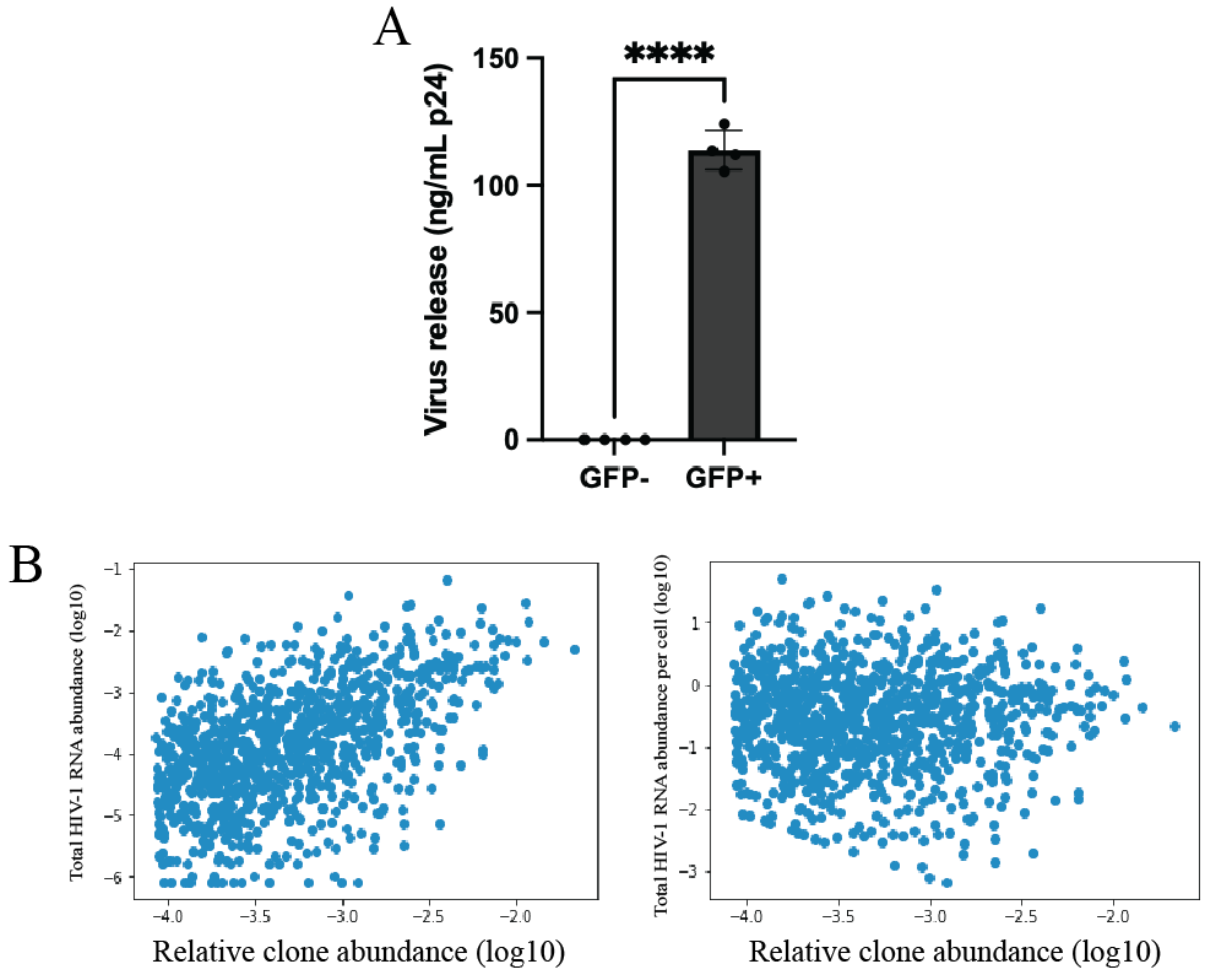


Figure. 4. 1: Correlates of vpr- proviral pool's expression. Polyclonal proviral populations of vpr- was established in Jurkat cells and sorted. A) A bar graph of the virus release that was measured in culture supernatant 24 hours after about 2 million cells each of the GFP- and GFP+ sorted subpools were cultured. B) Scatterplot of the relative abundance (log10) of clones as determined by the zip code abundance cell pool's DNA on the x axis versus the relative abundance (log10) of the zip codes in the total cellular RNA y axis (left panel), and cellular RNA per GFP+ cell y axis (right panel).

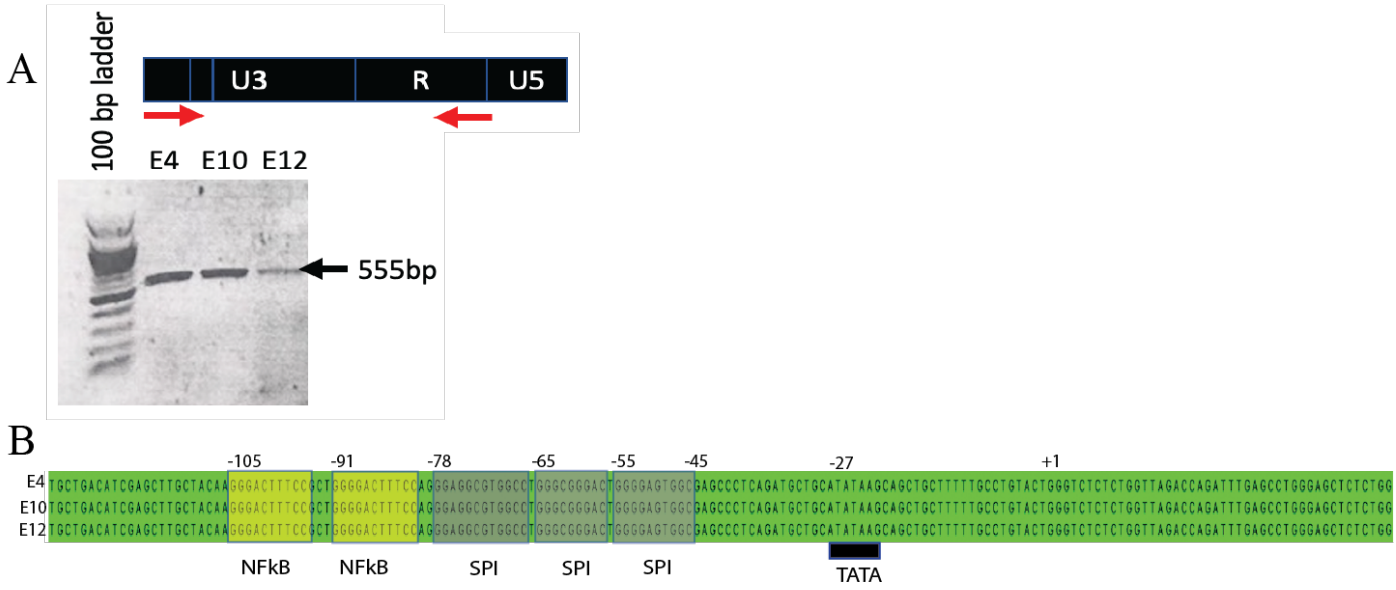


Figure. 4. 2: The intactness of proviral promoters. The promoter region of individual clones was amplified, cloned and the PCR inserts were subjected to Sanger sequencing. A) A 1% agarose gel image showing the bands corresponding to the 555 bp expected size of the PCR amplicon. The first lane was loaded with a 100 bp DNA ladder, and lanes 2-4 were loaded with PCR reactions of E4, 10, and E12 respectively. C) The nucleotide sequences of clones E4, E10, and E12. The yellow rectangles indicate the NFkB binding sites, the grey rectangles indicate the SP1 binding sites, and the black rectangle indicates the TATA sequence. The numbers on the top of the panel indicates the nucleotide position relative to the transcription start site (+1).

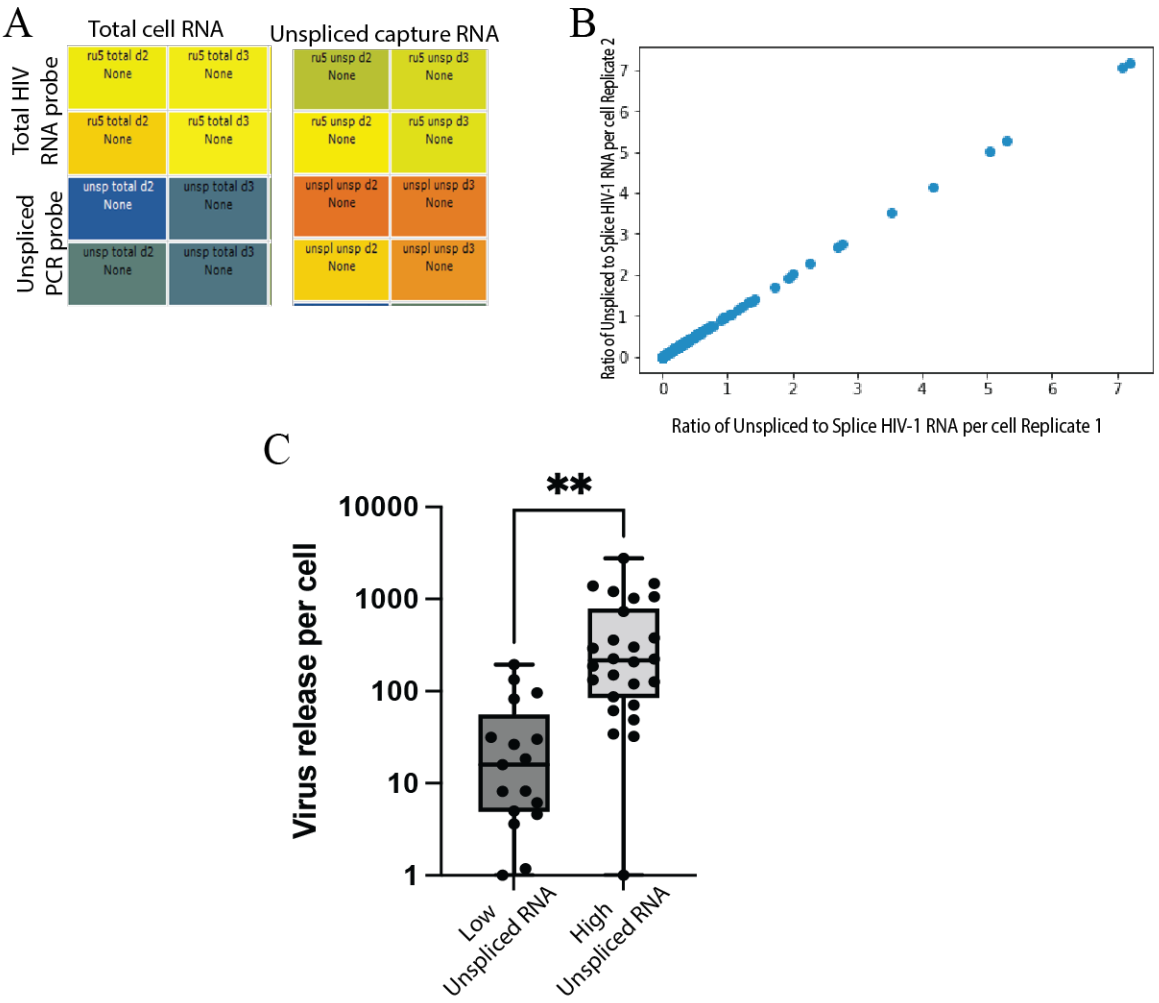


*Figure. 4. 3: HIV-1 RNA splicing among integrant clones. Cellular RNA was isolated from the polyclonal pool and split into two aliquots. One aliquot was subjected to RNA capture to isolate unspliced HIV-1 RNA, and the other aliquot was processed as total RNA (unspliced + spliced). A) Heat map of qPCR of the total RNA (left panel) and unspliced HIV-1 RNA capture (right panel) samples. The top two squares of both the left and right panels (from top to bottom: 5-fold dilution) shows the relative quantities the total HIV-1 RNA transcripts using a R-U5 primer pair that amplifies all HIV-1 RNA variants, and the bottom two squares (from top to bottom: 5-fold dilution) of both the right and left panels shows the relative quantity of unspliced HIV-1 RNA that was amplified using a primer pair that only amplifies unspliced HIV-1 RNA. B) A scatterplot of the ratio of unspliced to spliced RNA among clones. The x axis indicates the ratios of replicate 1 and the y axis indicates the ratios of replicate 2. C) A pairwise comparison of the clones with high unspliced HIV-1 RNA (rations >=1) and those with low unspliced (ratios <1) (significance test was by Unpaired t-test: **=p-value < 0.0001).*

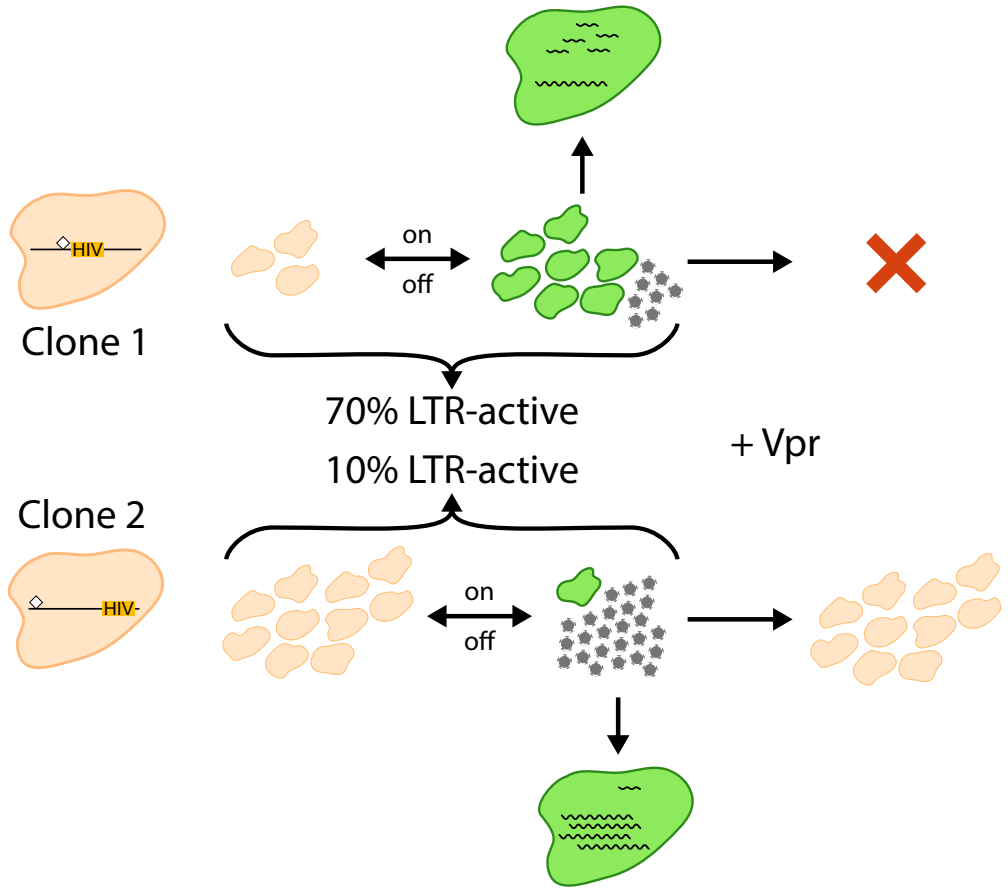


Figure. 4. 4: ***Model for HIV-1 persistence in vitro****. From left to right, after stable integration into the host cells' DNA, integrant clones (Clone 1 and Clone 2) divide into daughter cells. The provirus in Clone 1 is integrated clones to the H3K4me3 and H3K27ac marks (indicated by the diamond shape) and most of the daughter cells are transcriptionally active (70% LTR-active). Virus release is low among daughter cells of clone 1 due to its low unspliced to spliced intracellular RNA ratio. The provirus in clone 2 is farther away from the H3K4me3 and H3K27ac marks and most of its daughter cells are transcriptionally inactive (10% LTR-active). The few transcriptionally active cells of clone 1 releases high amount of virus due to the high unspliced to spliced intracellular viral RNAs levels. In the presence of Vpr, daughter cells of clone 1 die over time, however, clone 2 persists.*

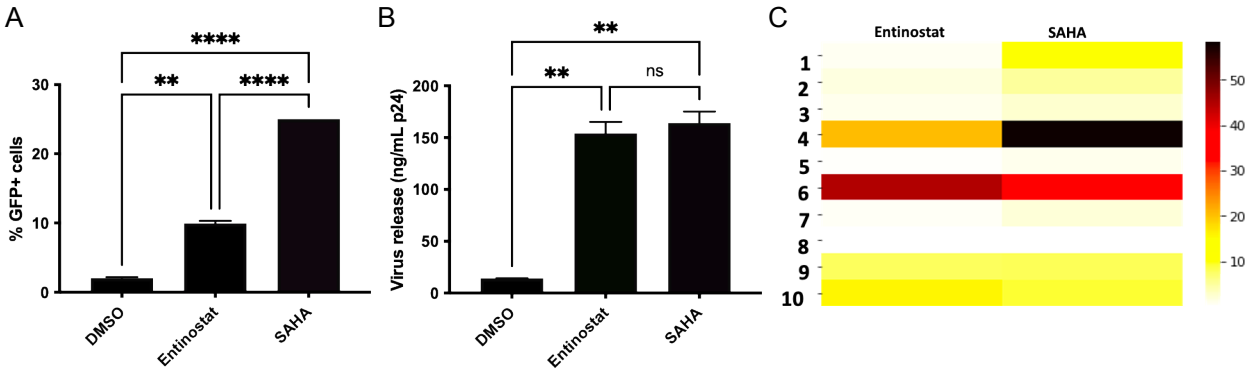


Figure. 4. 5: ***SAHA reactivates proviral-clones that are minimally affected by Entinostat.*** *vpr- pool was exposed to entinostat 500nM and 500nM SAHA for 48 hours post-infection by flow cytometry, and virus release was quantified using a reverse transcriptase assay and normalized to define p24 levels. (A) Bar graphs showing the frequency of GFP+ cells post LRA treatment. (B) The amount of virus released into culture supernatants (C) A heatmap of the clonal (zip code) virus release per treated GFP- cell. Numbers at the left of the panel are clone identifiers generated by ordering proviral zip codes in decreasing relative abundance, as determined for the GFP- subpool. Every row represents a unique cell clone's response. Only the top 10 most abundant clones are represented in the heatmap. The color bar indicates the extent of release per clone based on p24.*

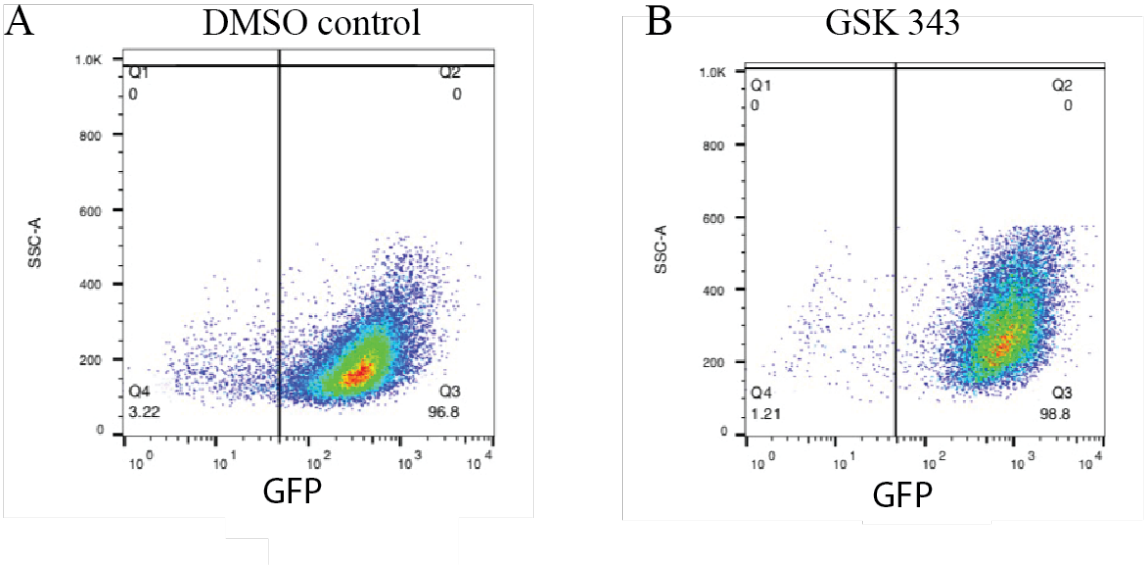


Figure. 4. 6: ***The EZH2 inhibitor GSK343 blocks the on to off transition.*** *GFP+ sorted vpr- subpool was split into aliquots. One was subjected to GSK343 treatment and compared to a DMSO control 72 hours post-treatment. The cells were then analyzed by flow cytometry. Flow cytometry plots showing (A) DMSO treated cells and (B) 2 µM GSK343 treatment.*

References

Alexaki, A. and B. Wigdahl (2008). "HIV-1 infection of bone marrow hematopoietic progenitor cells and their role in trafficking and viral dissemination." PLoS pathogens **4**(12): e1000215.

Ali, H., D. Bhange, K. Mehta, Y. Gohil, H. K. Prajapati, S. N. Byrareddy, S. Buch and U. Ranga (2021). "An emerging and variant viral promoter of HIV-1 subtype C exhibits low-level gene expression noise." Retrovirology **18**(1): 1-12.

Andersen, J. L., E. Le Rouzic and V. Planelles (2008). "HIV-1 Vpr: mechanisms of G2 arrest and apoptosis." Experimental and molecular pathology **85**(1): 2-10.

Atindaana, E., A. Kissi-Twum, S. Emery, C. Burnett, J. Pitcher, M. Visser, J. M. Kidd, A. Telesnitsky and M. J. Roth "Bimodal Expression Patterns, and Not Viral Burst Sizes, Predict the Effects of Vpr on HIV-1 Proviral Populations in Jurkat Cells." mBio **0**(0): e03748-03721.

Battivelli, E., M. S. Dahabieh, M. Abdel-Mohsen, J. P. Svensson, I. T. Da Silva, L. B. Cohn, A. Gramatica, S. Deeks, W. C. Greene and S. K. Pillai (2018). "Distinct chromatin functional states correlate with HIV latency reactivation in infected primary CD4+ T cells." Elife **7**: e34655.

Burger, S. and M. A. Poles (2003). Natural history and pathogenesis of human immunodeficiency virus infection. Seminars in liver disease, Copyright© 2002 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New ….

Chun, T.-W., L. Stuyver, S. B. Mizell, L. A. Ehler, J. A. M. Mican, M. Baseler, A. L. Lloyd, M. A. Nowak and A. S. Fauci (1997). "Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy." Proceedings of the National Academy of Sciences **94**(24): 13193-13197.

Einkauf, K. B., G. Q. Lee, C. Gao, R. Sharaf, X. Sun, S. Hua, S. M. Chen, C. Jiang, X. Lian and F. Z. Chowdhury (2019). "Intact HIV-1 proviruses accumulate at distinct chromosomal positions during prolonged antiretroviral therapy." The Journal of clinical investigation **129**(3): 988-998.

Einkauf, K. B., M. R. Osborn, C. Gao, W. Sun, X. Sun, X. Lian, E. M. Parsons, G. T. Gladkov, K. W. Seiger and J. E. Blackmer (2022). "Parallel analysis of transcription, integration, and sequence of single HIV-1 proviruses." Cell.

Epstein, F., G. Pantaleo, C. Graziosi and A. Fauci (1993). "The immunopathogenesis of human immunodeficiency virus infection." N. Engl. J. Med **328**(5): 327-335.

Eriksson, S., E. H. Graf, V. Dahl, M. C. Strain, S. A. Yukl, E. S. Lysenko, R. J. Bosch, J. Lai, S. Chioma and F. Emad (2013). "Comparative analysis of measures of viral reservoirs in HIV-1 eradication studies." PLoS pathogens **9**(2): e1003174.

Finzi, D., M. Hermankova, T. Pierson, L. M. Carruth, C. Buck, R. E. Chaisson, T. C. Quinn, K. Chadwick, J. Margolick and R. Brookmeyer (1997). "Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy." Science **278**(5341): 1295-1300.

Fischer, M., B. Joos, B. Niederöst, P. Kaiser, R. Hafner, V. von Wyl, M. Ackermann, R. Weber and H. F. Günthard (2008). "Biphasic decay kinetics suggest progressive slowing in turnover of latently HIV-1 infected cells during antiretroviral therapy." Retrovirology **5**(1): 1-16.

He, A., X. Shen, Q. Ma, J. Cao, A. von Gise, P. Zhou, G. Wang, V. E. Marquez, S. H. Orkin and W. T. Pu (2012). "PRC2 directly methylates GATA4 and represses its transcriptional activity." Genes & development **26**(1): 37-42.

Jones, L. E. and A. S. Perelson (2007). "Transient viremia, plasma viral load, and reservoir replenishment in HIV-infected patients on antiretroviral therapy." Journal of acquired immune deficiency syndromes (1999) **45**(5): 483.

Jordan, A., D. Bisgrove and E. Verdin (2003). "HIV reproducibly establishes a latent infection after acute infection of T cells in vitro." The EMBO journal **22**(8): 1868-1877.

Keedy, K. S., N. M. Archin, A. T. Gates, A. Espeseth, D. J. Hazuda and D. M. Margolis (2009). "A limited group of class I histone deacetylases acts to repress human immunodeficiency virus type 1 expression." Journal of virology **83**(10): 4749-4756.

Kennedy, E. M., S. M. Amie, R. A. Bambara and B. Kim (2012). "Frequent incorporation of ribonucleotides during HIV-1 reverse transcription and their attenuated repair in macrophages." Journal of Biological Chemistry **287**(17): 14280-14288.

Kim, J., Y. Lee, X. Lu, B. Song, K.-W. Fong, Q. Cao, J. D. Licht, J. C. Zhao and J. Yu (2018). "Polycomb-and methylation-independent roles of EZH2 as a transcription activator." Cell reports **25**(10): 2808-2820. e2804.

Kok, Y. L., S. Schmutz, A. Inderbitzin, K. Neumann, A. Kelley, L. Jörimann, M. Shilaih, V. Vongrad, R. D. Kouyos and H. F. Günthard (2018). "Spontaneous reactivation of latent HIV-1 promoters is linked to the cell cycle as revealed by a genetic-insulators-containing dual-fluorescence HIV-1-based vector." Scientific reports **8**(1): 1-14.

Lanigan, T. M., S. M. Rasmussen, D. P. Weber, K. S. Athukorala, P. L. Campbell, D. A. Fox and J. H. Ruth (2020). "Real time visualization of cancer cell death, survival and proliferation using fluorochrome-transfected cells in an IncuCyte® imaging system." Journal of Biological Methods **7**(2).

Lee, P. K., T. L. Kieffer, R. F. Siliciano and R. E. Nettles (2006). "HIV-1 viral load blips are of limited clinical significance." Journal of Antimicrobial Chemotherapy **57**(5): 803-805.

Liu, R., F. R. Simonetti and Y.-C. Ho (2020). "The forces driving clonal expansion of the HIV-1 latent reservoir." Virology Journal **17**(1): 1-13.

Maldarelli, F., X. Wu, L. Su, F. Simonetti, W. Shao, S. Hill, J. Spindler, A. Ferris, J. Mellors and M. Kearney (2014). "Specific HIV integration sites are linked to clonal expansion and persistence of infected cells." Science **345**(6193): 179-183.

Matalon, S., T. A. Rasmussen and C. A. Dinarello (2011). "Histone Deacetylase Inhibitors for Purging HIV-1 from the Latent Reservoir." Molecular Medicine **17**(5): 466-472.

Ocwieja, K. E., S. Sherrill-Mix, R. Mukherjee, R. Custers-Allen, P. David, M. Brown, S. Wang, D. R. Link, J. Olson and K. Travers (2012). "Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing." Nucleic acids research **40**(20): 10345-10355.

Pace, M. J., L. Agosto, E. H. Graf and U. O'Doherty (2011). "HIV reservoirs and latency models." Virology **411**(2): 344-354.

Peluso, M. J., P. Bacchetti, K. D. Ritter, S. Beg, J. Lai, J. N. Martin, P. W. Hunt, T. J. Henrich, J. D. Siliciano and R. F. Siliciano (2020). "Differential decay of intact and defective proviral DNA in HIV-1–infected individuals on suppressive antiretroviral therapy." JCI insight **5**(4).

Perelson, A. S., A. U. Neumann, M. Markowitz, J. M. Leonard and D. D. Ho (1996). "HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time." Science **271**(5255): 1582-1586.

Qu, D., C. Li, F. Sang, Q. Li, Z.-Q. Jiang, L.-R. Xu, H.-J. Guo, C. Zhang and J.-H. Wang (2016). "The variances of Sp1 and NF-κB elements correlate with the greater capacity of Chinese HIV-1 B′-LTR for driving gene expression." Scientific reports **6**(1): 1-11.

Simonetti, F. R., H. Zhang, G. P. Soroosh, J. Duan, K. Rhodehouse, A. L. Hill, S. A. Beg, K. McCormick, H. E. Raymond and C. L. Nobles (2021). "Antigen-driven clonal selection shapes the persistence of HIV-1–infected CD4+ T cells in vivo." The Journal of clinical investigation **131**(3).

Sörstedt, E., S. Nilsson, A. Blaxhult, M. Gisslén, L. Flamholc, A. Sönnerborg and A. Yilmaz (2016). "Viral blips during suppressive antiretroviral treatment are associated with high baseline HIV-1 RNA levels." BMC infectious diseases **16**(1): 1-9.

Stewart, S. A., B. Poon, J. Jowett and I. Chen (1997). "Human immunodeficiency virus type 1 Vpr induces apoptosis following cell cycle arrest." Journal of Virology **71**(7): 5579-5592.

van Opijnen, T., M. C. Boerlijst and B. Berkhout (2006). "Effects of random mutations in the human immunodeficiency virus type 1 transcriptional promoter on viral fitness in different host cell environments." Journal of virology **80**(13): 6678-6685.

Wang, S. and L. Rong (2014). "Stochastic population switch may explain the latent reservoir stability and intermittent viral blips in HIV patients on suppressive therapy." Journal of theoretical biology **360**: 137-148.

Wang, Z., E. E. Gurule, T. P. Brennan, J. M. Gerold, K. J. Kwon, N. N. Hosmane, M. R. Kumar, S. A. Beg, A. A. Capoferri and S. C. Ray (2018). "Expanded cellular clones carrying replication-competent HIV-1 persist, wax, and wane." Proceedings of the National Academy of Sciences **115**(11): E2575-E2584.

Weinberger, A. D. and L. S. Weinberger (2013). "Stochastic fate selection in HIV-infected patients." Cell **155**(3): 497-499.

Weinberger, L. S., J. C. Burnett, J. E. Toettcher, A. P. Arkin and D. V. Schaffer (2005). "Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity." Cell **122**(2): 169-182.

Wong, J. K., M. Hezareh, H. F. Gunthard, D. V. Havlir, C. C. Ignacio, C. A. Spina and D. D. Richman (1997). "Recovery of replication-competent HIV despite prolonged suppression of plasma viremia." Science **278**(5341): 1291-1295.

Yeh, Y.-H. J., K. Yang, A. Razmi and Y.-C. Ho (2021). "The Clonal Expansion Dynamics of the HIV-1 Reservoir: Mechanisms of Integration Site-Dependent Proliferation and HIV-1 Persistence." Viruses **13**(9): 1858.

Zaikos, T. D., M. M. Painter, N. T. Sebastian Kettinger, V. H. Terry and K. L. Collins (2018). "Class 1-selective histone deacetylase (HDAC) inhibitors enhance HIV latency reversal while preserving the activity of HDAC isoforms necessary for maximal HIV gene expression." Journal of Virology **92**(6): e02110-02117.