2022

Integration Sites in the Persistence of Latent HIV-1

Amy S. Huang

INTEGRATION SITES IN THE PERSISTENCE OF LATENT HIV-1

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Amy S. Huang

June 2022

INTEGRATION SITES IN THE PERSISTENCE OF LATENT HIV-1

Amy S. Huang, Ph.D.

The Rockefeller University 2022

Human immunodeficiency virus (HIV-1), the pathogen that causes acquired immune deficiency syndrome (AIDS), remains one of the world's most pressing health issues. Since the beginning of the HIV/AIDS epidemic, over 32M people have succumbed to AIDS-related illnesses. Despite remarkable advances in HIV-1 biology, neither a vaccine or cure have been achieved. While antiretroviral therapy (ART) has significantly improved disease outcomes for people with HIV-1 and reduced transmission, treatment is often accompanied by long-term side effects or stigma, or impeded by limited access to health care. Furthermore, because viral load quickly rebounds upon treatment interruption, ART is required to be a life-long medication. The major barrier to HIV cure is the persistence of long-lived latently infected CD4+ T cells. Collectively known as the latent reservoir, these cells carry integrated HIV-1 proviruses that are transcriptionally quiescent and are thus able to evade host immunity and virus-induced cell death. Cellular and molecular characterization of the latent reservoir is challenging because latently infected cells are exceedingly rare and express no known surface marker. Moreover, the infected cell pool is dominated by cells containing defective proviruses which cannot contribute to rebound viremia.

The work herein elucidates the contribution of proviral integration site to HIV-1 latency and the maintenance of the replication-competent, or intact, reservoir. Using an innovative single-cell sequencing technique that provides paired proviral sequence and integration site information, as well as enables selective analysis of replication-competent HIV-1, I interrogate the integration landscape of HIV infected individuals whose reservoirs are dominated by a small number of large expanded clones. By performing viral outgrowth assays, I also show that replication-competent proviruses harbored in the expanded clones readily produce infectious virion upon stimulation. I report an integrative analysis of the clonal dynamics, inducibility, and genomic position of intact proviruses in ART-suppressed individuals, and demonstrate that proviruses in expanded clones across patients are significant more likely to be mapped to Krüppel-associated box (KRAB) domain-containing zinc finger (ZNF) genes on chromosome 19. Transcriptional and epigenetic meta-analysis of primary CD4+ T cells reveal that these specific chromosomal locations harboring integrated provirus are associated with genes downregulated upon cellular activation. Taken together, the data indicate that selected sites in the genome, including ZNF genes, can be especially permissive for maintaining HIV-1 latency during memory CD4+ T cell clonal expansion. These findings demonstrate that gene activity at the integration site impacts the survival and persistence of intact, expanded HIV proviruses in infected cells, and provide evidence that the quality, not only the quantity, of the latent reservoir must a key consideration in HIV-1 cure strategies.

ACKNOWLEDGEMENTS

First, I would like to thank my adviser, Michel Nussenzweig, for his mentorship and support. Michel, I am continually inspired by your fierce pursuit of excellent science and the sharpness of your mind. It has been a privilege to learn from you.

I owe my deepest gratitude to my thesis committee: Daniel Mucida and Paul Bieniasz. Your fresh perspectives and thoughtful feedback has made all the difference for this thesis and my growth as a scientist. Brad Jones, thank you for serving as my external examiner.

It is not an exaggeration when I say that this work would not have been possible without two brilliant women, Mila Jankovic and Lillian Cohn. In addition to the optimism she brings to the lab, Mila's knowledge and experience is unparalleled. Mila, without you, there would be a lot less data and a lot more panicking. Lillie, I am so incredibly lucky to have worked with you from day one. Thank you for showing me the ropes, sharing your ideas, and listening to mine.

Throughout my time in the Nussenzweig Lab, I have been fortunate to work alongside and learn from numerous talented scientists whose work has paved the way for my own: Ching-lan Lu, Yotam Bar-On, Julio Lorenzi, Pilar Mendoza, Christian Gaebler, Alice Cho, Johannes Weymar and everyone who has been a part of the "HIV-1 Reservoir Support Group"; Harry Hartweger the CRISPR guru. Thomas Hagglöf and Spencer Chen, thank you for the lively journal club discussions. Thiago Olivera and Victor Ramos, the bioinformatics dream team, thank you for helping me make sense of the data and for your friendship.

Graduate school can be a formidable beast, and I am immensely indebted to Lauren Rosenblum and Dr. Nisha Mehta-Naik for protecting my mental health these past years. I am also grateful to The Rockefeller University Dean's Office: thank you for always advocating for students and keeping your doors open.

I wouldn't have dared to embark on this journey without the training and encouragement of my undergraduate research mentor, Dr. Peter Bradley. Peter, thank you for nurturing my curiosity and helping build my foundation in molecular biology.

I am also extremely fortunate to have had the opportunity to be a part of the Celdara Medical NYC Team. Brandy Houser, Jean-Loup Romet-Lemonne, Jing Gong, Ramsey Bekdash, thank you all for a fantastic learning environment and the meaningful discussions.

To my wolfpack: thank you for everything. I am so lucky to know you all. Masa Jankovic, who has made me more adventurous and Julian Kraft, who reminds me to see the glass as half-full. Marianna Agudelo, I am so grateful we both ended up in Michel's lab. Thank you for your magic, your balance, the countless cups of tea and wonderful conversations. Jennifer McQuillan, thank you for the stories, your wisdom and strength, and the book

recommendations. Lilian Nogueira, you have become a sister to me. Thank you for always being up for an adventure and welcoming me into your home. Darwin Nogueira-Oliveira, thank you for your unconditional love and joy. You are, without a doubt, the best boy!

Lastly, I dedicate this thesis to my parents Chen Donghang and Huang Weijian, and to the memory of Chen Dongyan and Ou Qining. Your love has sustained me, and I am here today because of your sacrifices. Thank you.

SIGNIFICANT CONTRIBUTIONS

The work presented here was originally published in *Nature Medicine (11)* and *Journal of Experimental Medicine (12)*. Experiments described in Chapter 2 were performed as a part of Dr. Lillian B. Cohn's thesis project. In Chapter 3, high throughput sequencing was conducted by the Rockefeller University Genomics Resource Center and bioinformatic analysis of RNA-seq and RNAPII ChIP-seq data was carried out by Dr. Israel Tojal da Silva and Renan Valieris. Thank you to the Nussenzweig Lab's Clinical Team and Clinical Processing Team, and most of all, our study participants for making this work possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ART	antiretroviral therapy
ARV	antiretroviral drugs
ATI	analytical treatment interruption
bnAb	broadly neutralizing antibody
DEG	differentially expressed gene
EC	elite controller
FDR	false discovery rate
HIV-1	immunodeficiency virus-1
IN	HIV-1 integrase protein
KRAB-ZNF	Krüppel-associated box zinc finger
LTR	long terminal repeat
MIP-seq	matched integration site and proviral sequencing
NFL	near full length
NHP	non-human primate
NRTI	nucleoside reverse transcriptase inhibitor
PBMC	peripheral blood mononuclear cell
PHA	phytohemagglutinin
PCA	principal component analysis
PIC	HIV-1 pre-integration complex
QVOA	quantitative viral outgrowth assay
Q2VOA	quantitative and qualitative viral outgrowth assay
Q4PCR	quadruplex qPCR
RNAPII	RNA polymerase II
RT	reverse transcriptase
SIV	simian immunodeficiency virus
TCR	T cell receptor
TSS	transcription start site
WGA	whole genome amplification
ZNF	zinc finger

CHAPTER 1: INTRODUCTION

40 years of the HIV-1/ AIDS epidemic

In 1981, young and previously healthy gay men began dying from opportunistic infections and a rare and unusually aggressive cancer. The disease would become known as Acquired Immune Deficiency Syndrome (AIDs) and soon, be determined to be caused by a retrovirus – Human Immunodeficiency Virus type 1 (HIV-1) (*13, 14*). By the end of the decade, the World Health Organization estimated there were up to 400,000 cases of AIDS worldwide and between 5 to 10 million individuals living with HIV-1 (*15, 16*). Today, the global prevalence of HIV-1 infection is nearly 38 million.

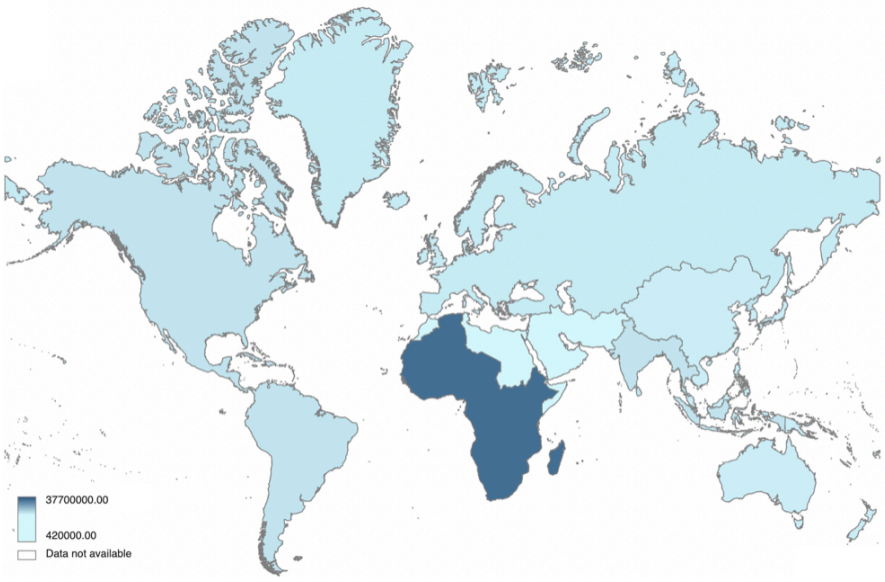


Figure 1.1. Global prevalence and distribution of HIV-1

Despite remarkable advances in our understanding of HIV biology, there is still no cure or vaccine for HIV-1 infection. While prophylactic or post-exposure treatment with antiretroviral drug cocktails has reduced transmission and antiretroviral therapy (ART) has significantly improved disease outcomes for infected individuals, treatment is impeded by long-term side effects or stigma, or limited by poor access to health care. Furthermore, because viral load quickly rebounds upon treatment interruption, ART is a life-long medication. Unfortunately, regions where populations are most affected by HIV-1 are also most resource-poor, making ART penetrance, adherence, and limiting disease spread a major ongoing challenge. As such, HIV-1 continues to present a serious threat to public health.

The HIV-1 life cycle

HIV-1 is a human-tropic, enveloped, single-strand RNA retrovirus that gains entry into host cells through interactions between viral surface glycoproteins and cellular receptor, CD4, and chemokine co-receptors CCR5 or CXCR4 (*17, 18*). HIV-1 preferentially infects CD4+ T cells, but is also known to target cells from the monocyte/ macrophage lineage (*19, 20*). Upon fusion of viral and cellular membranes, the HIV-1 virion sheds its capsid

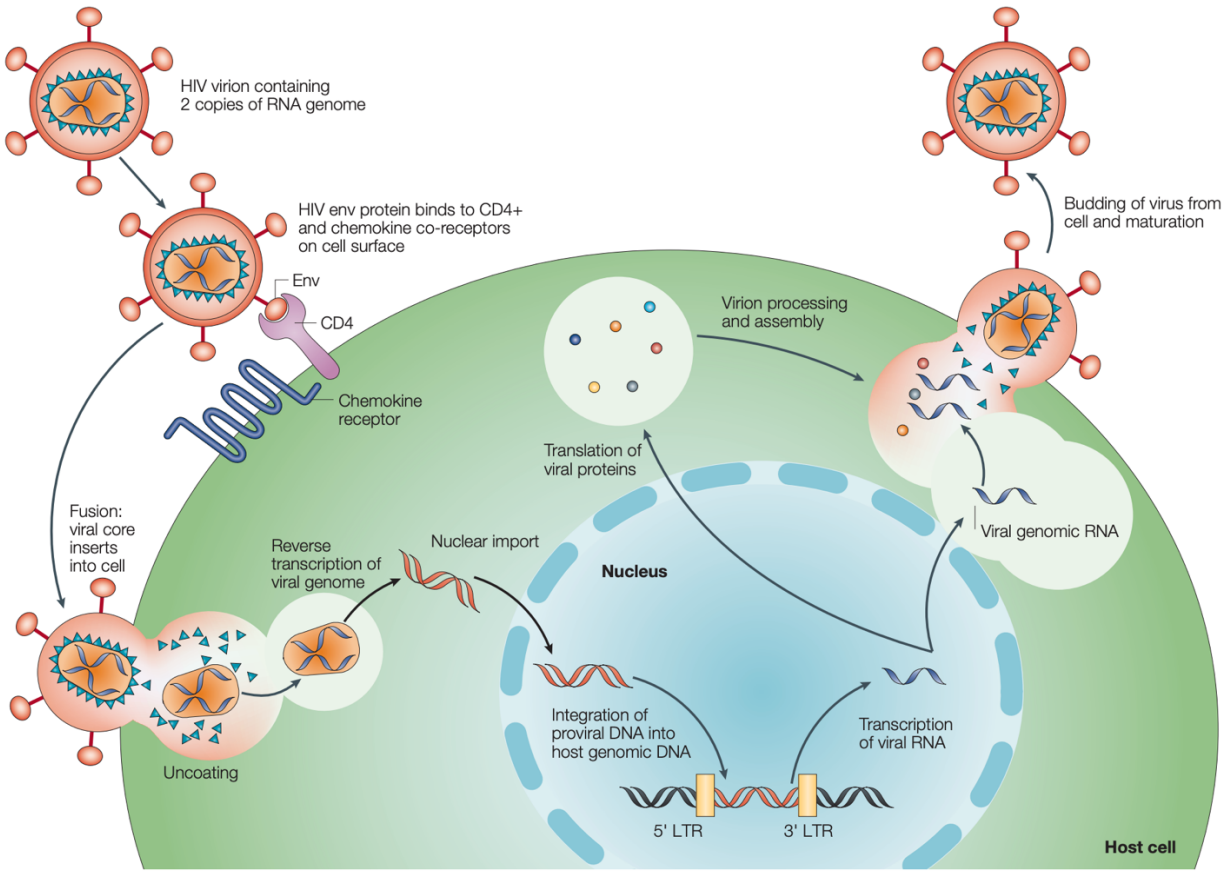


Figure 1.2. Key events in the HIV life cycle. The HIV-1 virion attaches to the target cell surface through binding of the host CD4 receptor and CCR5 or CXCR4 co-receptor. Viral entry is facilitated by fusion of the HIV-1 envelope and cellular membrane. Upon entry, the virion uncoats and uses viral enzyme Reverse Transcriptase to convert its RNA genome into DNA. The newly formed DNA genome is transported into the nucleus where it is stably inserted into the host chromosome by the viral enzyme Integrase. After integration, cellular transcription machinery can be recruited to transcribe viral genes and replicate the viral genome. In the cytoplasm, viral proteins and RNA are assembled into immature virions. The HIV-1 Protease cleaves viral polyproteins to convert the virion into is mature, infectious form before release. Adapted from Rambaut et al (*10*).

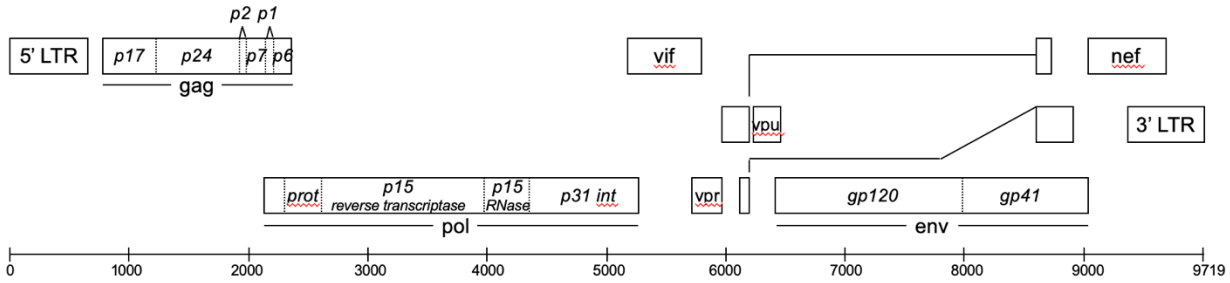


Figure 1.3. HIV-1 genome organization, HXB2 reference strain

and the viral genome is reverse transcribed by virally encoded Reverse Transcriptase (RT) into double stranded DNA. Because RT lacks proofreading activity, reverse transcription is highly error-prone and drives rapid viral genetic evolution through introduction of mutations (*21, 22*). Such mutations contribute to the exceptional diversity in HIV-1 genomes that enables immune escape or drug resistance (*10*). Next, the newly formed double stranded DNA is transported into the nucleus where viral protein Integrase (IN) catalyzes its irreversible integration into the host chromosome. The integrated provirus can then undergo one of two process. Cellular transcription machinery may engage with the integrated provirus to produce HIV-1 RNAs that are translated into viral proteins or packaged as full-length RNA genomes to form new infectious virions that bud from the host membrane to perpetuate the replication cycle (*10*). Alternatively, the integrated provirus may remain transcriptionally quiescent. In this latent state, the infected cell generates little to no viral RNA and expresses no known surface marker and as such, is undetectable by immune surveillance (*23*). Collectively, these latently infected cells make up the latent reservoir, which is unable to be eliminated by ART (*24*), is long-lived, with an estimated half-life of 4-18 years (*25-28*), and likely gives rise to viremia upon ART interruption. This latent reservoir represents the major barrier to HIV-1 cure.

HIV-1 transmission and host immune responses to HIV-1 infection

HIV-1 is transmitted through three major routes: sexual contact at mucosal surfaces, maternal-infant exposure, and percutaneous inoculation (*29, 30*). Historically, local epidemics have been associated with sex work, intravenous drug use, and infection clusters among men who have sex with men (MSM), although today, approximately 70% of HIV-1 infections are attributed to heterosexual transmission (*31*). Efficiency of transmission primarily depends on viral load in the transmitter, and is directly impacted by genital inflammation or ulcers from other sexually transmitted disease, pregnancy, and circumcision. Unsurprisingly, socioeconomic status is also an indirect, but major factor in HIV-1 transmission (*32*).

Following infection, clinical progression is classified based on viral load and antibody seroconversion using the Fiebig staging system (*33*). The Fiebig system marks the sequential appearance of viral RNA, viral p24 antigen, and HIV-specific antibodies in plasma. During the initial eclipse phase, which spans 7 to 21 days after exposure, virus

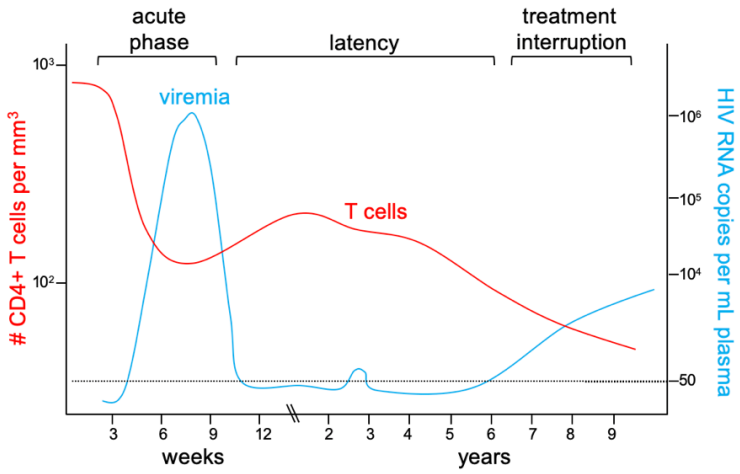


Figure 1.4. Clinical progression of HIV-1 infection

replicates in infected cells of the mucosa, submucosa and draining lymph nodes, but is not yet detectable in the blood. After eclipse, exponential propagation of virus and acute infection begin, and is associated with symptoms including fever, malaise, and generalized lymphadenopathy. After peak viremia occurs, plasma viremia falls to a stable level. This set point represents the balance between immune engagement against the virus's capacity for immune evasion, and is correlated with disease progression to AIDS in the absence of ART (*34*).

Humoral response to HIV-1

B-cell responses to HIV-1 are initially detected as immunoglobulin (IgM, IgG):viral complexes within one week of infection. Circulating antibodies against the gp41 and gp120 subunits of HIV-1 Envelope (Env) appear subsequently, approximately 2 weeks after plasma viremia. The Env glycoprotein is the sole viral surface protein and its subunits form trimeric spikes on the virion surface to interact with host co-receptors and facilitate viral entry (*35*). While these early antibodies do not neutralize the transmitted-founder virus strain and do not effectively control acute-phase viremia (*36*), non-neutralizing antibodies (nnAbs) were the only reported correlate of protection in the RV144 vaccine efficacy trial (*37-40*). Studies in a humanized mouse model of HIV-1 have suggested that protection by nnAbs is conferred, at least in part, by Fc-receptor mediated clearance of infected cells (*41*). Autologous neutralizing antibodies, which hold the ability to disrupt receptor/co-receptor binding or viral fusion with permissible target cells, develop within several months after infection and drive rapid viral evolution by exerting selective pressure on the transmitted-founder virus, which in turn produces escape mutations. As a consequence of the emergence of genetically diverse heterologous strains (*42, 43*), cross-neutralizing antibodies targeting the CD4 binding site and various epitopes of trimeric Env can develop on average, 2.5 years after infection. In rare cases (less than 1% of the HIV-1 infected population), individuals develop broadly neutralizing antibodies (bnAbs) with exceptional breadth and potency that represent clonal B cell responses against multiple epitopes on gp120 (*44, 45*). To date, numerous monoclonal bnAbs have been successfully cloned using single-cell techniques, characterized, and evaluated as potential prophylactic treatments and long-acting alternatives to antiretroviral therapy (*46-48*). Understanding the evolution of bnAb lineages in nature infection has also guided the immunogen design and sequential immunization strategies which are currently being pursued in HIV-1 vaccine development (*49-51*).

T cells in HIV-1 infection

T cells derive from two main lineages in the thymus, and each have distinct but complementary functions. Cytotoxic CD8+ T eliminate cells expressing foreign antigen while helper CD4+ T cells regulate the antigen-specific immune response (*52*). In response to pathogens, T cells undergo activation, differentiation and expansion, and memory formation.

HIV-1 specific cytotoxic CD8+ T cell responses arise early in infection, preceding the first detectable plasma antibodies, and occur in the majority of infected individuals, even those who cannot control infection (*53-55*). Initially, narrowly directed T cell responses against regulatory and accessory viral proteins (i.e. Tat, Vpr, Nef) emerge as viremia peaks. This T cell response peaks 1-2 weeks later, coinciding reductions in viral loads to 'set points', although it remains uncertain whether cytotoxic T lymphocyte (CTL) responses are predictive of viral set point or disease progression. The immune pressure exerted by HIV-1-specific CD8+ T cells gives rise to and selects for escape mutants that carry changes in dominant CTL epitopes (*56*). As infection progresses, persistent immune activation induces a number of phenotypic changes in the CD8+ T cell population, including sustained heightened expression of exhaustion markers (i.e. PD-1, CTLA-4) (*57*), diminished cytolytic activity, loss of polyfunctionality, and loss of proliferative potential (*57-59*).

Because CD4+ T cells are the principal target of HIV-1 and circulating memory CD4+ T cells are massively depleted during acute infection, the CD4+ T cell response is weaker than the CD8+ response (*60*). The magnitude of CD4 responses has been correlated with the rate of disease progression (*61*). HIV-specific CD4 responses are mainly directed against Gag and Nef epitopes (*62*). While depleted effector memory CD4+ T cell populations can be partially replenished by CD4+ central memory precursors, immune function is not completely restored, as chronic immune activation results in not only the upregulation of inhibitory receptors, but the production of short-lived cells that are not stable over time (*63, 64*). *Ex vivo* genome-wide transcriptional studies indicate that in chronic infection, HIV-specific CD4+ T cell are characterized by a T follicular helper cell (Tfh)-like gene signature, whereas viral control in elite controllers is characterized by Th22 and Th17 responses (*65*).

The latent HIV-1 reservoir

The latent reservoir comprises of long-lived HIV-1 infected cells harboring replication competent but transcriptionally quiescent provirus. The reservoir is seeded very early in infection; treatment initiation within 10 days of infection (Fiebig I) is insufficient to prevent reservoir establishment (*66, 67*). In the closely related simian immunodeficiency virus (SIV) infection, viral reservoirs were found to be seeded during the eclipse phase before detectable viremia. Rhesus macaques infected SIV and treated within three days of infection inevitably rebounded, albeit later than macaques that had delayed treatment initiation (*68*).

In humans, latently infected cells are rare in blood, accounting for approximately only 1 in 10^6 resting CD4+T cells (*69*). The majority of HIV-1 proviruses harbor lethal mutations such as single-nucleotide insertions and deletions, large internal deletions, hypermutations, or major splice donor deletions, and are considered defective due to their inability to replicate or directly contribute to viral rebound (*70*). Because there is no surface marker known to distinguish latently infected cells from uninfected cells, investigation of the latent reservoir largely focuses on the provirus or depends on indirect cellular surrogates, such as exhaustion markers or gene profiles (*11, 71*) .

Historically, the gold standard for reservoir measurement has been the quantitative viral outgrowth assay (QVOA) (*72*). In this assay, *ex vivo* CD4+ T cells are plated in limiting dilution and stimulated with mitogens or other potent latency reversal agents in culture. Target cells are added to allow for the amplification of the newly produced virus. Wells containing intact HIV-1 are identified by detection of viral p24 by ELISA. Viral outgrowth assays provide key information about the replication competent, inducible reservoir but consistently underestimate reservoir size because proviruses may require multiple rounds of in vitro stimulation to produce new virions, or may not be reactivatable at all (*73*). Akin to QVOA is the Tat/rev Induced Limiting Dilution Assay (TILDA), which detects the multiply spliced HIV-1 RNA only found in reactivated latent cells (*74*). DNA PCR-based methods like near full length genome sequencing and multiplex qPCR, provide global snapshots of reservoir sequence composition but not assessment of the proviruses' productivity and often overestimate the size of the latent compartment (*4, 70, 75, 76*). Together, these assays form a toolkit for exploring reservoir formation and dynamics in a clinical context.

Clonal expansion of latently infected cells

Several different explanations have been considered for the long half-life of the latent reservoir, including infected cell proliferation. However, this idea is counter-intuitive because stimuli that induce extensive CD4+ T cell division also activate HIV-1 gene expression, which in turn inhibits cell division and, presumably, should lead to cell death through viral cytopathy or immune-mediated elimination (*10, 77*). Nevertheless, several independent studies have presented evidence that support the estimate that 50-60% of all latent cells in chronically infected individuals are found in large, expanded clones and that the proportional contribution of latent cells in expanded clones increases over time (*2, 73, 78-81*). The CD4+ T cells that comprise a given clone share the same HIV-1 proviral integration site and express a unique T cell receptor (*11, 82*). Notably, a significant fraction of these clonal T cells express antigen receptors that recognize viruses which cause chronic or recurrent infections suggesting that expansion of latent clones is driven, at least in part, by antigen exposure (*82, 83*).

The factors that determine whether an integrated provirus adopts a latent phenotype and its subsequent potential to be re-activated are not well understood. When considering all integrated proviruses, including the majority that are defective, integrations into *BACH2*,

MKL2, and *STAT5B* are enriched (*1, 78, 84-86*). However, whether these genes contribute to integration, selection, or latency remains unknown.

In contrast to defective proviruses, the integration landscape of intact proviruses has been more difficult to document in part because the cells that carry latent viruses represent a very minor fraction of all CD4+ T cells (*70, 87*). However, recent work has demonstrated that in chronically infected individuals on long-term ART, intact proviruses with the potential to be re-activated are enriched in non-genic regions, in an opposite transcriptional orientation to host genes (*1*). In contrast, integrated proviruses in individuals that control infection in the absence of ART, elite controllers, were enriched in regions of centromeric satellite DNA and in KRAB-ZNF genes (*3*). The elite controllers also showed high level enrichment of large clonal integrations (*3, 88-90*). Whether the distinct integration profile found in elite controllers represents a correlate of control or simply a feature that is especially permissive for clonal expansion has not been determined.

Animal and cell models of HIV-1 latency

In-depth characterization of the molecular mechanisms of HIV-1 latency has been constrained by the limitations of animal models. Humanized mice are commonly used in basic and preclinical research, fail to mimic the HIV-1 disease progression observed in humans. On the other hand, non-human primates, which do mimic HIV-1 disease progression, come with significant bioethical and cost barriers (*91*). *In vitro* cell models have been indispensable for elucidating details of HIV-1 entry and replication, but nevertheless, are insufficient tools to accurately study the intricacies of te cellular landscape in HIV-1 infection.

Efficient expression of the essential cofactors in HIV-1 infection and replication has not been possible in traditional small animal models. As an alternative, many studies employ genetically immunodeficient mice that have been engrafted with human immune tissues (i.e. fetal thymus) or immune cells (i.e. peripheral blood cells, “PBLs”; hematopoietic stem cells). The various models recapitulate human lymphocyte development and distribution to varying degrees, as well as differ in their permissiveness to HIV-1 infection. For example, *scid*-hu-PBL mice, which can produce human antibody responses after immunization and have been employed to test passive immunization with Env-specific monoclonal antibodies (*92, 93*), but must be infected via intraperitoneal injection with cell-free virus or with HIV-1 infected human cells and are not valid studying mucosal transmission. While humanized mice are instrumental for evaluating the effectiveness of novel therapeutic agents and approaches in viral suppression, the incompleteness of their immune systems limit their use in studying HIV-1 reservoir formation and dynamics. Asian macaques are the most widely accepted non-human primate (NHP) model of HIV/AIDs, as infection of macaques with certain SIV strains results in high viral loads, progressive CD4+ T cell depletion, and susceptibility to opportunistic infections. While NHP models are immunocompetent, immunogenetic differences in the sequence diversity and copy number of major histocompatibility genes, must be considered when modeling the human

immune response to HIV-1. Additionally, there are major differences between the SIV and HIV-1 nucleotide identity and genome organization resulting in phenotypic differences (i.e. accessory proteins, co-receptor usage) and thus precluding the use of macaques in testing certain inhibitors and vaccines based on HIV-1 Env (*94*). To this end, there is ongoing efforts to engineer recombinant simian-human immunodeficiency viruses (SHIVs) that express key HIV-1 components and better replicate HIV-1 pathogenesis in NHPs.

Studies using patient derived HIV-1 infected cells obtained through leukapheresis have expanded our understanding of reservoir dynamics and reactivation but such samples are limited in availability and cannot be used in large-scale assays. As an alternative, chronically infected T cell and monocytic cell lines are widely used to define mechanisms regulating proviral transcription (*95, 96*), HIV-1 integration landscapes (*97*), as well as screen latency reversal agents (*98-100*). *Ex vivo* models derived from uninfected primary CD4+ T cells have also gained traction as tools to dissect signaling pathways involved in latency (*101-103*). Nevertheless, a number of caveats must be noted. First, the transformed nature of cell lines is fundamentally divergent, metabolically and transcriptionally, from the signature of quiescence of latent cells. Salient features of HIV-1 latency, including proviral expression levels, T cell phenotype, and mechanism of host transcriptional interference, have been reported to vary across latent cell lines, and between the cell lines and latently infected cells *in vivo*.

The experiments discussed in this thesis rely on the J1.1 cell line, which was established by infecting the Jurkat immortalized human T cell line with the LAV strain of HIV-1 (*104*). Despite deficiency IL-2 production, J1.1 cells contain replication-competent HIV-1 provirus and are consistently activated via mitogen stimulation (*104*).

HIV-1 integration

As a retrovirus, HIV-1 stably integrates a DNA copy of its RNA genome into the infected cell's chromosome. Following reverse transcription, viral cDNA associates with viral integrase (IN) at its 5' and 3' DNA ends, as well as capsid proteins, and cellular co-factors to form a large nucleoprotein complex called the pre-integration complex (PIC). Viral capsid components of the PIC interact with cellular nucleoporins to facilitate nuclear import and IN, a specialized DNA recombinase, directly interacts with host transcriptional mediator LEDGF/p75 to tether the PIC to the chromosome (*105, 106*). IN then catalyzes 3' processing of the viral DNA to produce reactive terminal 3'-OH groups and strand transfer (*107*). In some cases, however, HIV-1 DNA is not integrated into the host genome and instead, forms circular molecules that contain one or both LTRs. These viral copies, called LTR circles, do not support HIV-1 transcription or ongoing replication.

While there is no definitive consensus sequence for retroviral integration targeting, the process is not entirely random. Large scale studies of retroviral integration sites reveal target site preferences to vary across retrovirus genera and correlate with integrase sequence and structure, suggesting that in addition to chromatin accessibility and cellular cofactors, integrase is a major factor in integration site selection,. Murine leukemia virus

(MLV), a gammaretrovirus, favors transcriptional start sites and CpG islands, whereas avian sarcoma-leukosis virus (ASLV), an alpharetrovirus, shows a weak preference for active genes but little preference for transcriptional start sites (*108, 109*). For human T-cell virus type I (HTLV-1), a deltaretrovirus genus, integrations are "nearly random", showing no clear preference for genomic elements like TSS, CpG island, or genic regions (*110*). Unsurprisingly, SIV exhibits similar integration pattern to that of HIV-1 (*111*). It is well accepted that HIV-1 favors integration within transcriptionally active parts of the genome and highly expressed genes, in particular intronic regions (*84, 85, 112-116*). Thus, genomic features like high G/C content, high gene density, permissive histone modifications, and Alu repeats are all associated with high occurrences of HIV-1 integration (*117*). There is also evidence that spatially, HIV-1 preferentially integrates into chromatin at the nuclear periphery, proximal to the nuclear pore (*118, 119*).

The extent to which integration location influences latent state remains unclear. During latency, repression of the integrated provirus is maintained through a combination of mechanisms. Cytosine methylation of the CpG islands flanking the proviral transcription start site has been shown to be a epigenetic regulator in latently infected Jurkat cells as well as primary CD4+ T cells (*120*). Silencing may also occur through transcriptional interference, in which the upstream RNA polymerase II complex to fails to terminate and prevents initiation from the downstream HIV-1 promoter by hindering binding of the preinitiation complex and necessary transcription factors. (*121*). Other mechanisms for modulation of HIV-1 expression include presence of transcription repressors or limited availability of host factors, such as NFAT or NF-κB, that activate the proviral long terminal repeats (LTRs) (*77*).

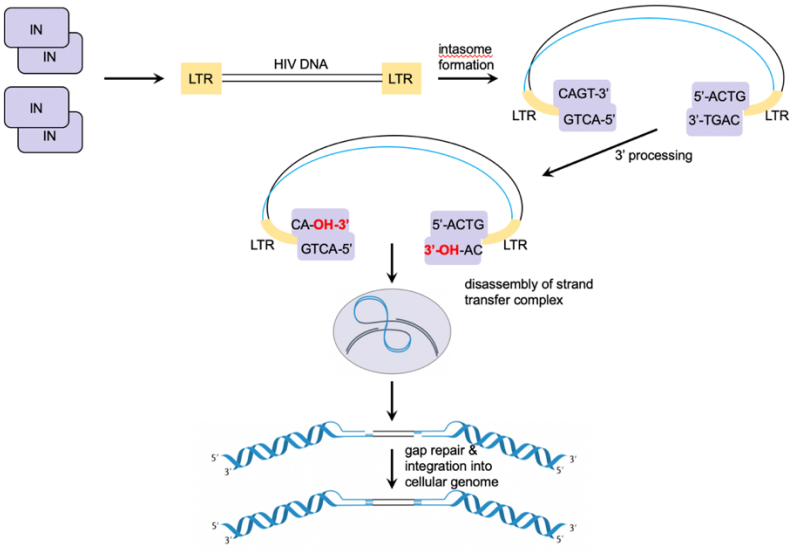


Figure 1.5. The HIV-1 integration process. IN multimerizes on the ends of the viral LTR and then cleaves two nucleotides from each 3′ end of the viral DNA during 3′ processing, which results in a recessed and chemically reactive hydroxyl group. After nuclear import of the pre-integration complex, integrase binds to target DNA and strand transfer occurs. After disassembly of the strand transfer complex, the DNA recombination intermediate is repaired. The provirus, flanked by host DNA, is now an integrated component of the cellular chromosome. Adapted from Lusic et al (*8*).

In this thesis, I describe evidence for clonal expansion of the latent reservoir and approach HIV-1 persistence from two distinct, but complementary perspectives. First, I employed high-throughput TCR receptor sequencing to demonstrate that reactivated latent cells largely express an identical TCR. Second, I generated a latent T cell model to interrogate cellular characteristics in latency and proviral reactivation. Finally, I performed integrative analysis of integration features associated with clonally expanded infected cells. Collectively, this work expands our understanding of reservoir dynamics and maintenance.

CHAPTER 2: CLONAL EXPANSION OF LATENTLY INFECTED CELLS

Reactivated latently infected CD4+ T cells can be isolated based on their surface expression of HIV-1 Env in a process called latent cell capture (LURE), which consists of a series of antibody staining, magnetic enrichment, and flow cytometry. Single-cell RNA sequencing (scRNA-seq) of LURE cells, alongside uninfected and actively infected controls, revealed a distinct gene signature in reactivated latent cells. In our study of three ART-suppressed HIV-infected individuals, reactivated latent cells, compared to control cells, differed in their expression of transcriptional regulators and genes involved in antiviral immunity, in particular, the response to type I interferon and the regulation of type I interferon production. In addition to transcriptome analysis, scRNA-seq was used to recover full-length, intact HIV-1 sequences from single reactivated latent cells. In two of the three individuals, all reconstructed intact viruses were attributed to a single expanded viral clone; in the third individual, four unique viruses were identified, including one dominant clone.

Persistent HIV-infected cells containing identical intact HIV-1 may stem from T cell clonal expansion (*2, 73, 78-80, 84, 85, 122, 123*) or viral replicative burst, in which identical viruses infect a diverse set of T cells. To distinguish how our detected clones arose, we analyzed the T cell receptor (TCR) sequences of the reactivated latent cells. T cells express unique antigen receptors through somatic recombination of their variable, diversity, and joining (VDJ) gene segments during early stages of thymocyte development (*124*). The potential diversity of the resultant repertoire is approximately $2x10^9$ TCRs (*125*), and it is accepted that T cells with identical TCRs are produced by clonal expansion. Across all three study participants, all TCR sequences retrieved from scRNA-seq reads from reactivated latent cells with identical proviruses were clonal. As a control, I sequenced TCRs from ~600 single CD4+ T cells from three non-infected and three ART-treated HIV-1 infected individuals. In contrast, 99.9% of control TCR sequences were unique. Only a single two-member clone was identified in one of the six donors. To exclude the possibility that clonality was due to T cell division during *in vitro* activation, we labelled cells with carboxyfluorescein succinimidyl ester, a fluorescent intracellular dye, at the beginning of culture. No measurable division was detected. Together, our data demonstrates that latent cells harboring identical replication-competent HIV-1 are products of in vivo CD4+ T cell clonal expansions, and provides further evidence that the latent reservoir is maintained through clonal expansion.

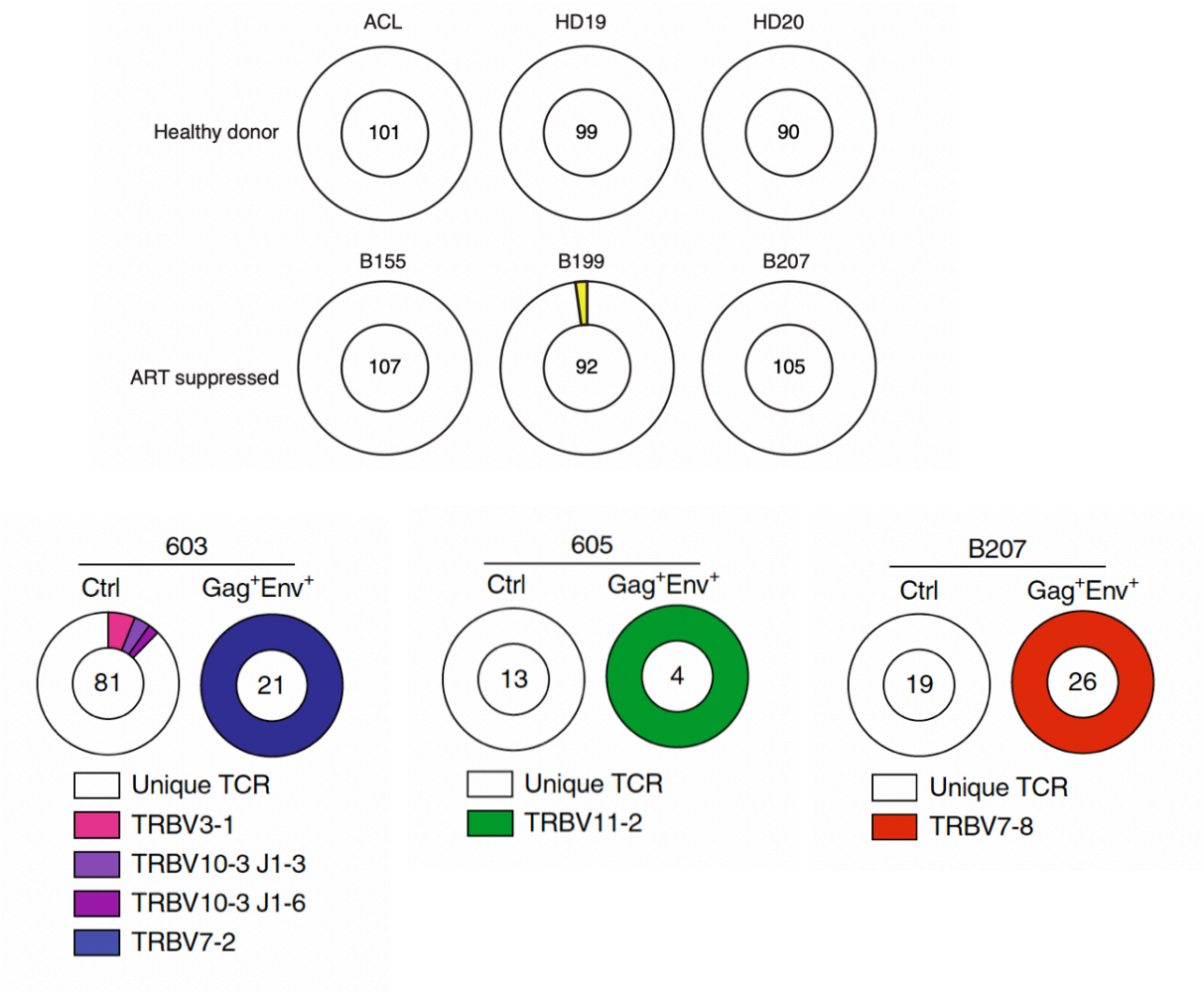


Figure 2.1 Cells containing identical proviruses represent expanded clones of CD4+ T cells. (a) PCR-amplified TCR sequences in single sorted CD4+ T cells from uninfected and ART-suppressed HIV-infected donors. Number of cells sequences noted in center of pie. Yellow section is a unique two-member clone. **(b)** TCR sequences recovered from reactivated latent cells. Control ("Ctrl") cells are unfractionated CD4+ T cells, not enriched by Env and Gag expression. White slices indicate unique TCRs, colored slices indicate clonal TCRs. Slices are proportional to clone sizes. Clones were determined by shared TCRα and β sequences.

CHAPTER 3: THE CELLULAR RESPONSE TO HIV-1 TRANSCRIPTION

This original aim of this project was to contribute to the field's understanding of the cellular mechanisms governing latency and viral persistence. HIV-1 persistence is thought to be maintained through survival and proliferation of infected cells rather than stochastic low level viral replication, which is thought to be absent during effective therapy. However, shared metabolic programs drive cell division and latent viral reactivation. Thus, how infected cells can divide without succumbing to the cytopathic effects of the virus is unknown. I sought to illuminate the particular aspects of T cell biology that the enable HIV-1 infected cells to maintain viral quiescence and divide without being cleared by host immunity or HIV-1 induced cell death. I hypothesized that latently infected CD4+ T cells may express a gene program to maintain viral suppression. Alternatively, these infected cells may possess a unique ability to detect and respond to low-level viral transcription. To address these hypotheses, I generated a latent HIV-1 cell line and interrogated the transcriptional and epigenetic landscapes of these cells at rest and at various time points after activation.

Characterization of a latent HIV-1 cell line

The multifactorial nature of HIV-1 latency has made it difficult to recapitulate *in vitro*. In the cases of primary cell models as well as ones derived from immortalized T lymphocyte lines, the proliferative nature of the transformed cells diverges from the general quiescence of in vivo latent cells. To study the cellular response to HIV-1 transcription, I required a cell harboring a single copy of replication competent HIV-1 that is relatively transcriptionally silent in culture, but consistently activated with mitogens or latency reversal agents. Due to their accessibility, we procured a number of latent cell lines through the NIH AIDS Reagent Program. In preliminary experiments, we found that J1.1 cells exhibit lowest levels of baseline viral activity as assayed by surface staining with biotinylated broadly neutralizing antibodies against the HIV-1 Envelope (Env) glycoprotein, compared to other readily available latent cell line models: ACH-2, OM-10.1 (*126-129*). We were able to isolate the most suppressed cells of the J1.1 population by bulk sorting for $Env^{-/low}$ cells or magnetic depletion of Env+ cells. We found that an antiretroviral drug (ARV) combination of nucleoside/nucleotide analog reverse transcriptase inhibitors (NRTIs) Emtricitabine and Tenofovir could further suppress viral replication and prevent superinfection *in vitro* without affecting cell viability. Proviruses in these cells are inducible with an array of cellular stimuli: phorbol 12-myristate 13-acetate (PMA) and ionomycin; phytohemagglutinin (PHA); and αnti-CD3/CD28 tetramers (Fig 3.1). Because J1.1s exhibit relatively low levels of proviral transcription at rest and their proviruses are readily reactivated upon cellular activation, we conclude J1.1s to be an acceptable model for our investigations of HIV-1 latency.

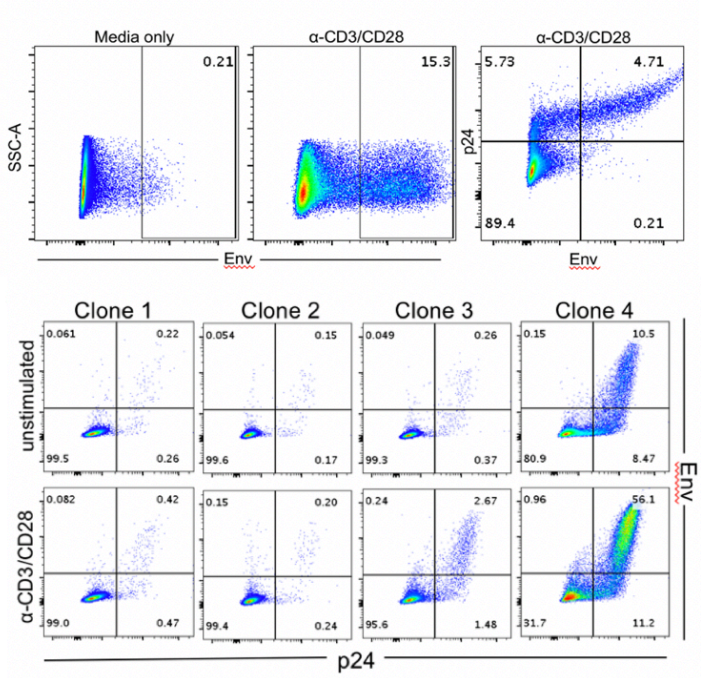


Figure 3.1. Basal viral transcription in J1.1 model of HIV-1 latency. (a) Flow cytometry of surface Env in J1.1s at rest and in presence of a T cell activation agent. **(b)** Env+ cells are also positive for intracellular Gag. Gated on live cells. All cells maintained on NRTIs.

To obtain homogeneous sub-clones for downstream studies, we single-cell sorted resting, ARV-treated J1.1s. It is possible that prior to ARV initiation, low-level virion production by the cells could have resulted in reinfection events and thus, loss of cellular and proviral homogeneity. Flow cytometry analysis of these cells revealed variability in levels of proviral expression, as shown by p24 and Env production, at rest. Furthermore, proviruses in different clones exhibited ranges of inducibility that did not appear to correlate with resting levels of provirus activity (Figure 3.2). While the original J1.1 cell line is derived from a single Jurkat subclone, ongoing replication during numerous passages in the absence of antiretroviral agents potentially allowed accumulation of mutations in the viral genome that affects viral fitness or quiescence. To confirm the intactness of the integrated provirus, we extracted genomic DNA from each subclone and PCR amplified a region spanning from the U5 region of 5' LTR to the U5 region of the 3' LTR. Amplicons were sequenced and confirmed to be intact and identical across subclones. Limiting dilution *gag* PCRs were performed to verify that only a single copy of HIV-1 was present in each subclone.

Integration into certain genomic loci may also influence viral gene expression. To assess heterogeneity in proviral positions, we employed a molecular cloning based technique to identify the integration sites within each subclone (*130*). In each subclone assayed, we detected the provirus to be integrated at same position within *DEAF1* in all subclones. These findings suggest that HIV-1 activity was not provirus-dependent, but perhaps determined by cellular environments.

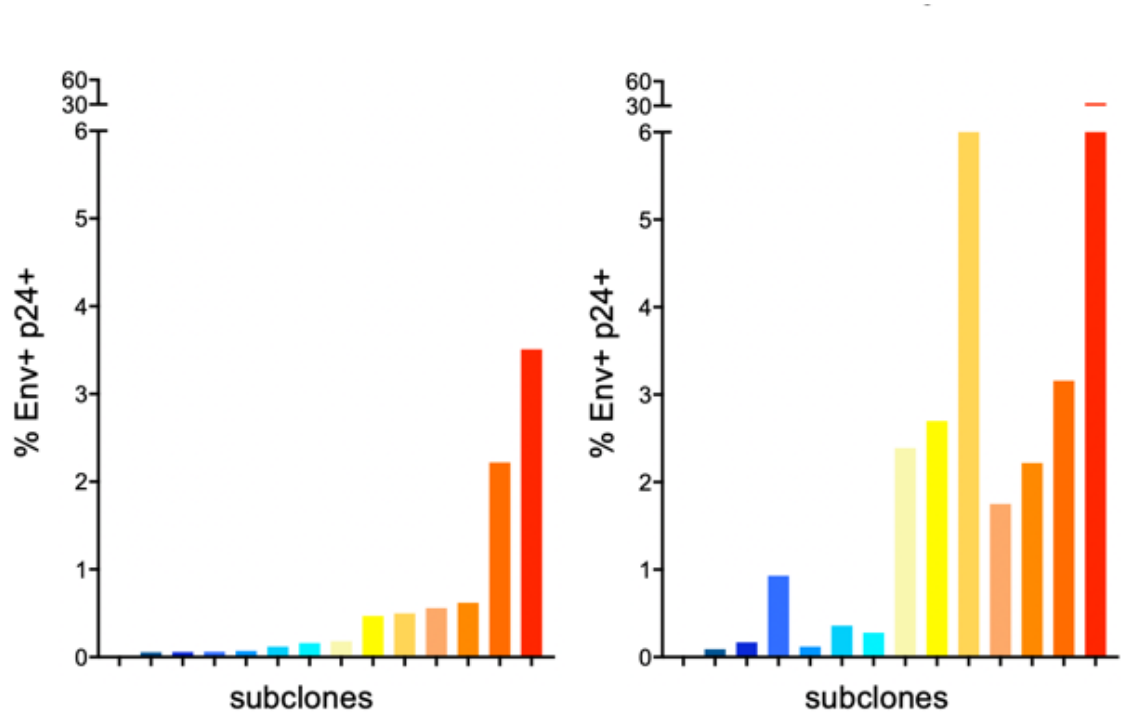


Figure 3.2 Variability in J1.1 subclones. (a) Flow cytometry of representative subclones of J1.1s exhibit different levels of proviral transcription at rest and have diverse propensities for proviral reactivation. **(b)** Distribution of levels of proviral transcription at rest and after stimulation.

During new viral infection events, germline-encoded pattern recognition receptors detect foreign nucleic acids, trigger downstream innate immune responses, and activate adaptive immunity. In particular, cytoplasmic cDNA from HIV-1 and other retroviruses is sensed by cyclic GMP-AMP synthase , which produces cGAMP for signal transduction and secretion of Type I interferons (*131*). However whether latently infected cells sense reactivated virus using similar mechanisms and whether the cell responds to by suppressing viral transcription to avoid apoptosis remains to be demonstrated.

Cellular transcriptional response to proviral activity

To understand the host cell response to nascent HIV-1 transcription, we performed bulk RNA sequencing on J1.1 subclone populations representing different conditions in latency. From our pool of J1.1 subclones, we selected two highly inducible subclones ("C5", "C21") and two lowly inducible subclones ("C6" and "C30"). Sorted $Env^{negative}$ fractions of resting lowly inducible subclones representing physiologically resting latent cells. We hypothesized that the cells within these subclones which remained $Env^{negative}$ after stimulation express a distinct gene program functioning to detect and suppress proviral transcription in order to maintain a latent state. In parallel, we sorted the $Env^{negative}$

cells from lowly inducible subclones that had been stimulated for 4 and 24 hours with anti-CD3/CD28 tetramers. For comparison, we sorted Env^{positive} cells from the highly inducible J1.1 subclones at rest and after 4 and 24 hours of stimulation. As a control for virus-independent responses to cellular stimulation, we also activated J1.1s in which we excised the integrated provirus using a Cas9-RNP approach (termed “J1.1 knockout”; Appendix 1), and sorted live cells for analysis.

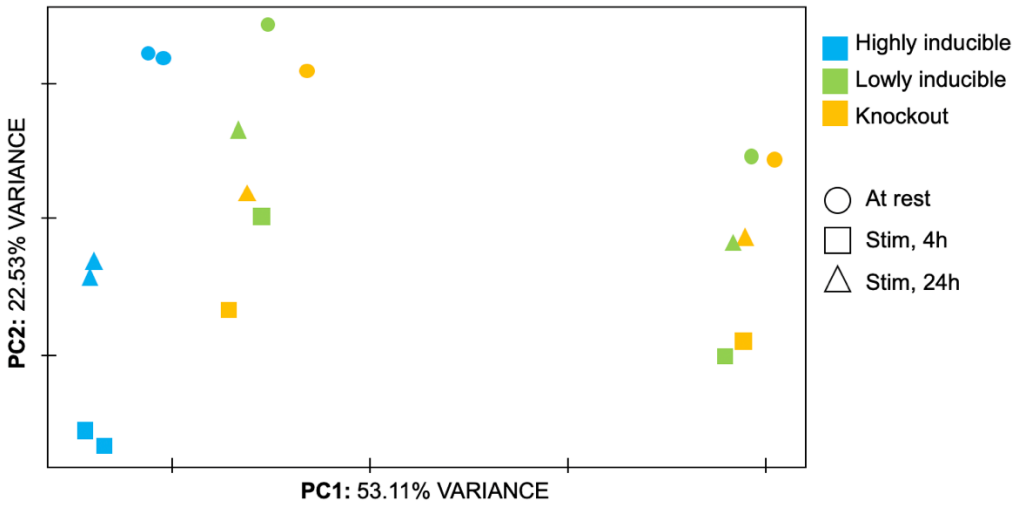


Figure 3.3. Principal component analysis segregates J1.1 populations based on gene expression. Blue, green and yellow denote subclone characteristics. Circles, squares and triangles indicate culture conditions of cells. Stim, stimulated.

RNA-seq gene expression data was visualized using principal component analysis (PCA). Differentially expressed genes (DEGs) were identified using unsupervised hierarchical clustering. Within the total transcriptome as well when restricted to the top 1000 differentially expressed genes in each cell population, lowly inducible J1.1s and J1.1 knockouts cluster closely at rest and after stimulation, indicating the two populations share appreciable similarities in transcriptome despite the fundamental difference in proviral status. Furthermore, each lowly inducible subclone clustered with a knockout subclone rather than its biological replicate (i.e. the other lowly inducible clone). As expected, highly inducible latent cells stimulated to produce virus displayed a divergent gene profile.

Env^{negative} and Env^{positive} cells exhibited a number of differentially expressed genes relatively to Env^{positive} cells. To understand the functional differences between these, we performed gene set enrichment analysis using the Gene Ontology knowledgebase on the 240 top differentially expressed genes (*132-134*). Biological processes found to be

enriched in Env^{positive} cells were all associated with immune function and pathogen response, including: cytokine activity, chemokine signaling, TNFα signaling via NF-kB, IFNγ response, and the NFAT pathway. Inexplicably, Env^{negative} cells were found to be enriched for pathways implicated in tissue morphogenesis and pattern specification. We conclude that Env^{positive} cells from highly inducible subclones differ from Env^{negative} cells by exhibiting heightened antiviral immune responses. However, the observed responses appear to be broadly antiviral and not specific to HIV-1. It is likely that latency is not strictly mediated by DNA sensing, but other alternative cellular mechanisms as well.

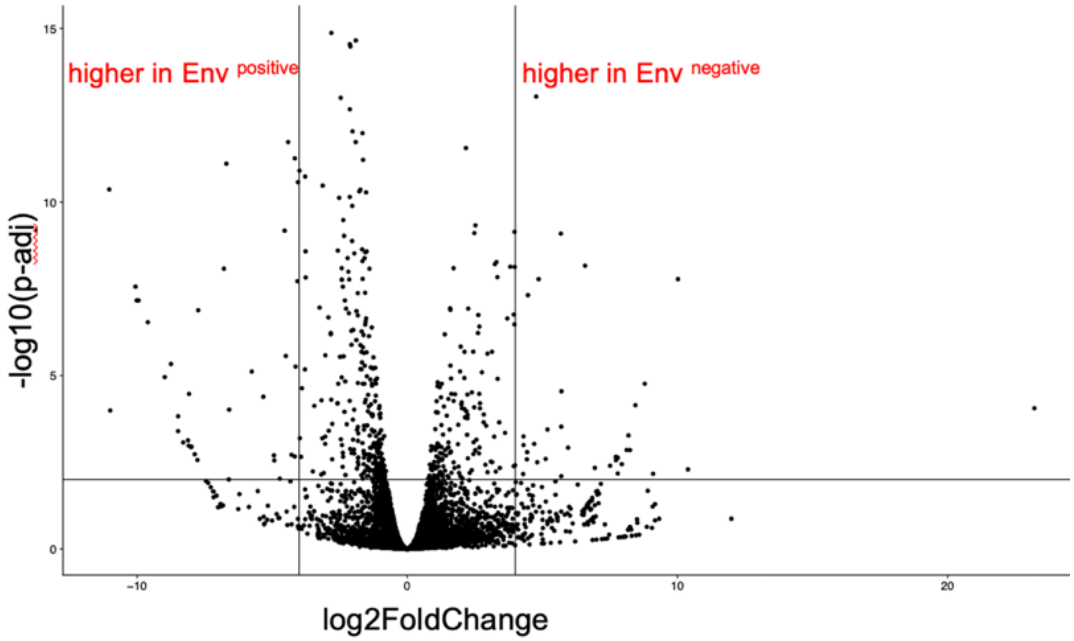


Figure 3.4. Volcano plot of transcriptome-wide differences between Env^{negative} lowly inducible cells and Env^{positive} highly inducible cells after 4h stimulation. Horizontal line represents statistical significance threshold (FDR adjusted p-value <0.01). Vertical lines indicate threshold of magnitude (log2 fold change > 4).

Epigenetic regulation of proviral transcription

Epigenetic mechanisms have been demonstrated to mediate pre-integration latency as well as establishment of post-integration latency in HIV-1 infected cells (*135*). For example, binding of repressive transcription factors to the HIV-1 5' LTR, the proviral promoter recruits histone deacetylases and facilitates the repression of HIV-1 expression (*136-138*). Components of the NURD/ Mi-2/ CHD chromatin remodeling complex, which functions in nucleosome remodeling and histone deacetylation, have been linked to HIV-1 latency (*138, 139*). We hypothesized that epigenetic regulation may also present a potential explanation for differences in inducibility in our J1.1 subclones.

To investigate how native chromatin structure and transcription factor binding may modulate proviral inducibility –or more broadly, latency reversal— we profiled RNA

polymerase II occupancy across the entire genome. We postulated that lack of HIV-1 transcription may be associated with RNA polymerase II (RNAPII) stalling along the proviral sequence. We performed RNA polymerase II chromatin immunoprecipitation (RNAPII ChIP) on each of the previously characterized lowly inducible and highly inducible subclones at rest and after stimulation with anti-CD3/CD28 tetramers for 1 hour. The most inducible subclone, C5, was also stimulated for 24 hours as a positive control. For each condition, one million live, unsorted cells were collected and crosslinked. Genomic DNA was precipitated with an antibody against the RNAPII C-terminal domain repeat YSPTSPS, a key domain that coordinates transcription initiation, elongation, and transcript processing (*140*). The bound DNA was purified and deep sequenced. In parallel, for each condition, fractions of cells were subject to analysis by flow cytometry to assess HIV-1 expression.

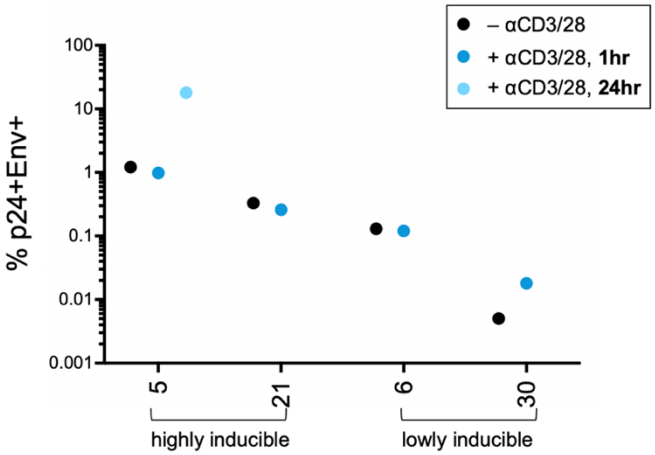


Figure 3.5. Comparison of HIV-1 activity, as indicated by % of cells positive for intracellular p24 and surface Env staining. Lowly and highly inducible subclones are indicated on x-axis. Colored dots represent cellular activation conditions.

By flow cytometric analysis, there were no difference in HIV-1 expression after 1 hour of stimulation across all subclones (Figure 3.5). As expected, the highly inducible subclone C5 showed a log-fold increase in proportion of cells expressing HIV-1 after 24 hours. Despite this congruency in phenotype, we rationalized that some difference must still exist that primes the cells for HIV-1 expression. Indeed, when we examined RNAPII occupancy of HIV-1, highly inducible and lowly inducible subclone after activation, at 1hr, RNAPII is more abundant at the LTR and along the remainder of the proviral genome in highly inducible cells. (Figure 3.6). This finding provided evidence of divergences in the early cellular response to activation that underlie proviral reactivation from latency.

We then evaluated genome-wide RNAPII binding to identify genes whose altered expression or regulation may in turn be modulating proviral transcription. ChIP-seq analysis revealed a number of regions with differential RNAPII occupancy. As expected, differences between lowly inducible and highly inducible J1.1 subclones were more pronounced following cellular activation (Figures 3.7 and 3.8). At rest, the most

significantly enriched region in highly inducible cells occurred on chromosome 12, within the second intron of chromobox protein homolog 5, *CBX5*. Following activation, numerous

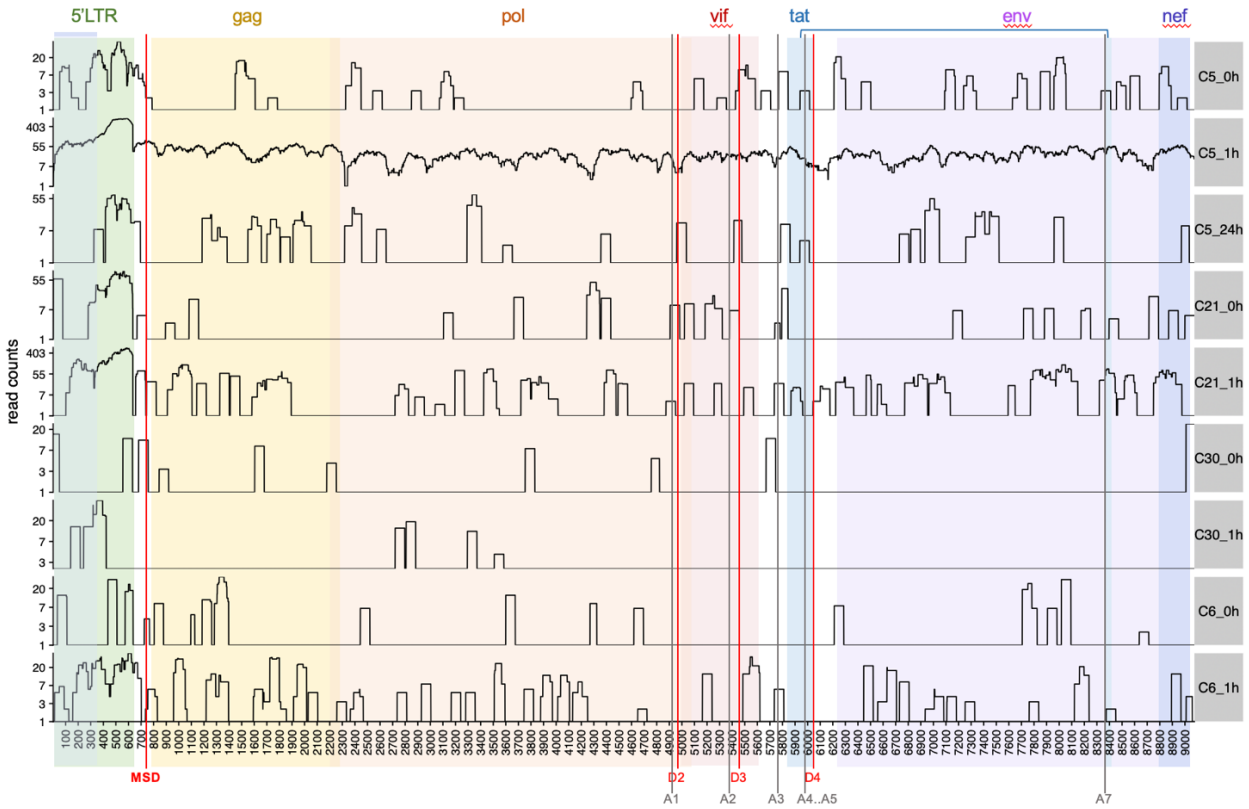


Figure 3.6. RNAPII occupancy in the proviral genome in J1.1 subclones at rest and after stimulation, as indicated by read count (left). J1.1 subclone noted on right axis. Genome position and positions of HIV-1 major splice donor and downstream splice acceptor sits are denoted on bottom axis. Colored blocks and text label indicate HIV-1 coding genes.

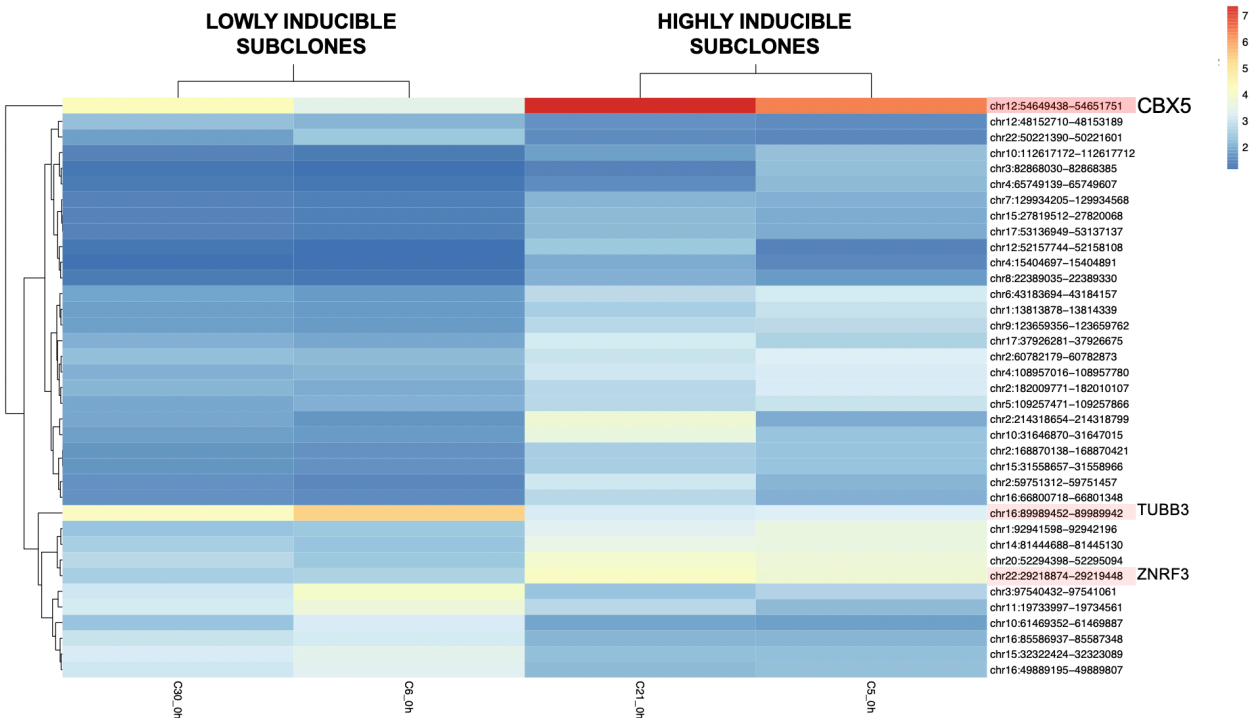


Figure 3.7. Heat map displaying genomic positions (GRCh37/hg38) with highest levels of differential binding of RNA polymerase II (FDR adjusted p-value < 0.01). Comparison of lowly inducible J1.1 subclones (C6, C30) and highly inducible J1.1 subclones (C5, C21) at rest. Chromosomal positions mapped to GRCh37/hg19 Human Assembly.

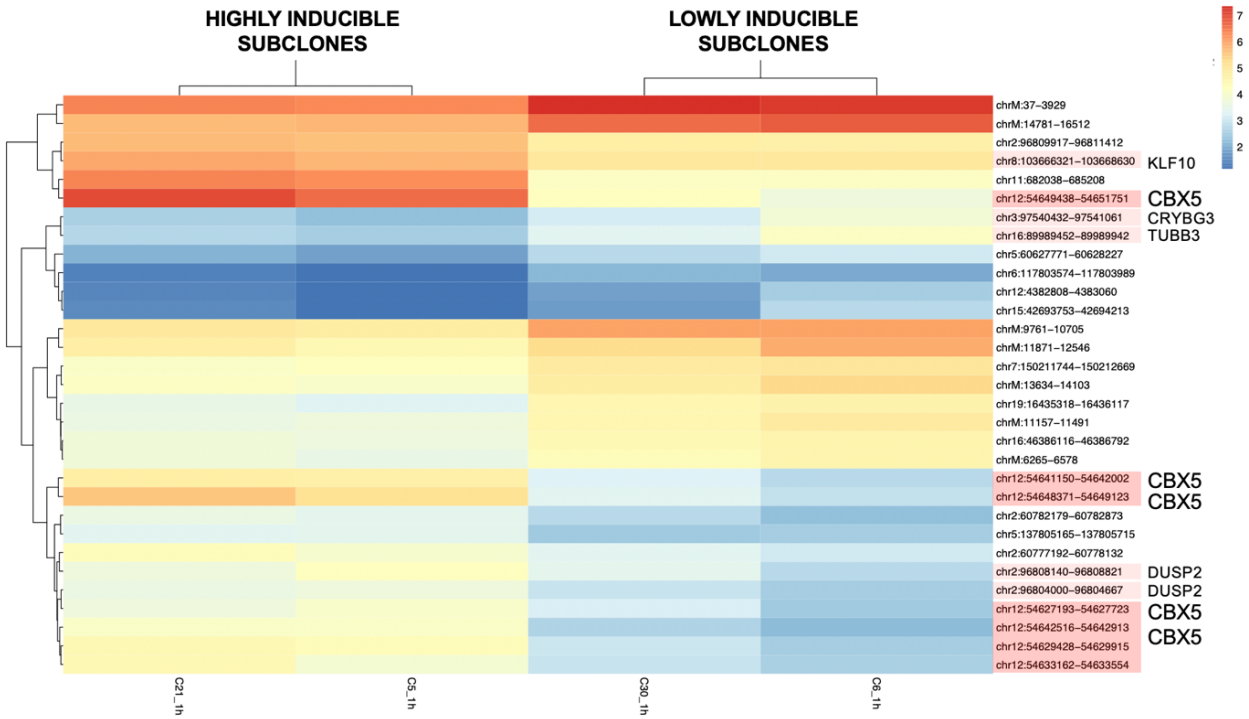


Figure 3.8. Heat map displaying genomic positions (GRCh37/hg38) with highest levels of differential binding of RNA polymerase II (FDR adjusted p-value < 0.01). Comparison of lowly inducible J1.1 subclones (C6, C30) and highly inducible J1.1 subclones (C5, C21) after 1h stimulation with anti-CD3/CD28 tetramers. Chromosomal positions mapped to GRCh37/hg19 Human Assembly.

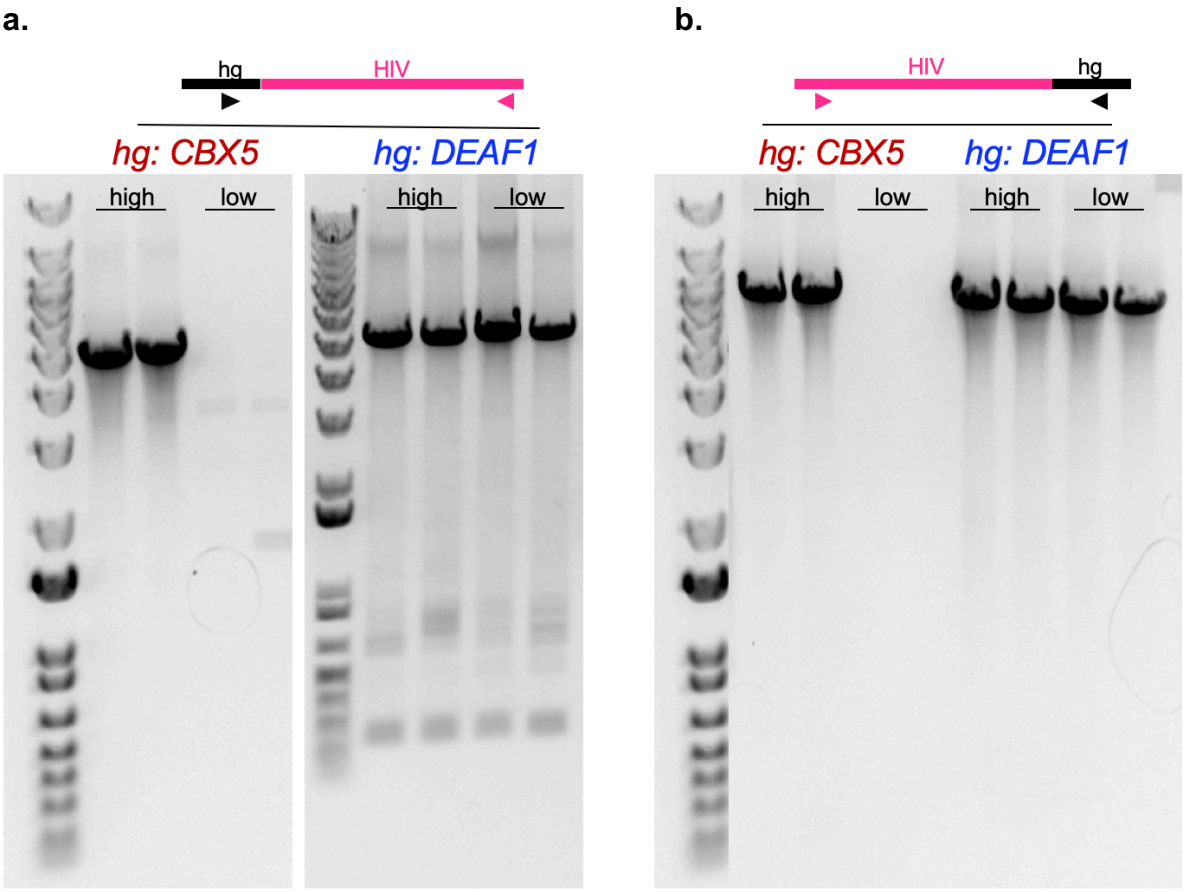


Figure 3.9. PCR confirmation of integration sites in highly inducible ("high") and lowly inducible ("low") J1.1 subclone. **(a)** Amplification of 5' integration site using forward primer in *CBX5* or *DEAF1* and reverse primer in HIV-1. **(b)** Amplification of 3' integration site using forward primer in HIV-1 and reverse primer in *CBX5* or *DEAF1*. Each lane represents PCR reaction using genomic DNA extracted from one subclone. "High" denotes highly inducible subclones and "low" indicates lowly inducible subclones.

additional regions mapping to *CBX5* appeared as regions of differential RNAPII binding between highly and lowly inducible cells (Figure 3.8). Because transcription from the HIV-1 LTR is known to disrupt or drive expression of host genes downstream of the integration site (*141*), it was surprising that regions in *DEAF1* proximal to the integrated provirus were not identified as differentially bound by RNAPII. Closer genome browser analysis of mapped human reads revealed that in highly inducible subclones, there were similar RNAP II occupancy patterns, albeit of different magnitudes, at the integration site in DEAF1 and within the second intron of *CBX5*. We postulated that there exists a previously undetected proviral integration within *CBX5* in the highly inducible subclones.

To verify an integrated provirus in *CBX5*, we extracted genomic DNA from each of the four subclones and performed site specific PCRs. For 5' amplification of the integration site, we designed forward primers corresponding to *DEAF1* and *CBX5* sequences genome and used reverse primers based on published primer sequences corresponding to HIV-1 pol. For 3' amplification of the integration site, we used forward primers in HIV-1 and designed reverse primers corresponding to *DEAF1* and *CBX5*. Amplicons were visualized by gel electrophoresis. In all subclones, both 5' and 3' fragments of the correct size were observed for *DEAF1*-specific integration, further confirming integration in *DEAF1*. In highly inducible, but not lowly inducible subclones, 5' and 3' fragments were observed for CBX5. This definitively identified a proviral integration in *CBX5* present in highly inducible but absent in lowly inducible clones.

The discovery of an additional proviral integration within *CBX5* in highly inducible subclones, along with the observation that DEAF1 activity is enhanced in highly inducible subclones compared the lowly inducible subclones led us to further posit that proviral transcription at the more permissive *CBX5* locus may transactivate proviral transcription from silenced *DEAF1* locus. While the incongruence in integrations amongst these latent subclones preclude their use in studies of the cellular environment during latency, our work using these cells brought us back to the topic of integration sites, and the contribution of integration site to proviral quiescence, clonal expansion of latently infected cells, and ultimately HIV-1 persistence.

CHAPTER 4: INTEGRATION FEATURES OF INTACT LATENT HIV-1 PROVIRUSES IN EXPANDED CD4+ T CELLS CONTRIBUTE TO VIRAL PERESISTENCE

Integrative analysis of latent reservoirs

To characterize the integration features of latently infected CD4+ T cells, we used a modified MIP-seq protocol (*1, 142, 143*). The analysis focused on 6 individuals whose intact and replication competent proviruses had been extensively characterized through direct sequencing (Q4PCR) (*4, 144*) and viral outgrowth (Q2VOA) (*2, 11, 145*) (Figure 3.1). The selected individuals were chronically infected and had been suppressed on ART for 4-21 years (median 13 years, Table 1). As expected, the latent reservoirs in these volunteers were dominated by large-expanded clones of CD4+ T cells (*4, 11, 145*).

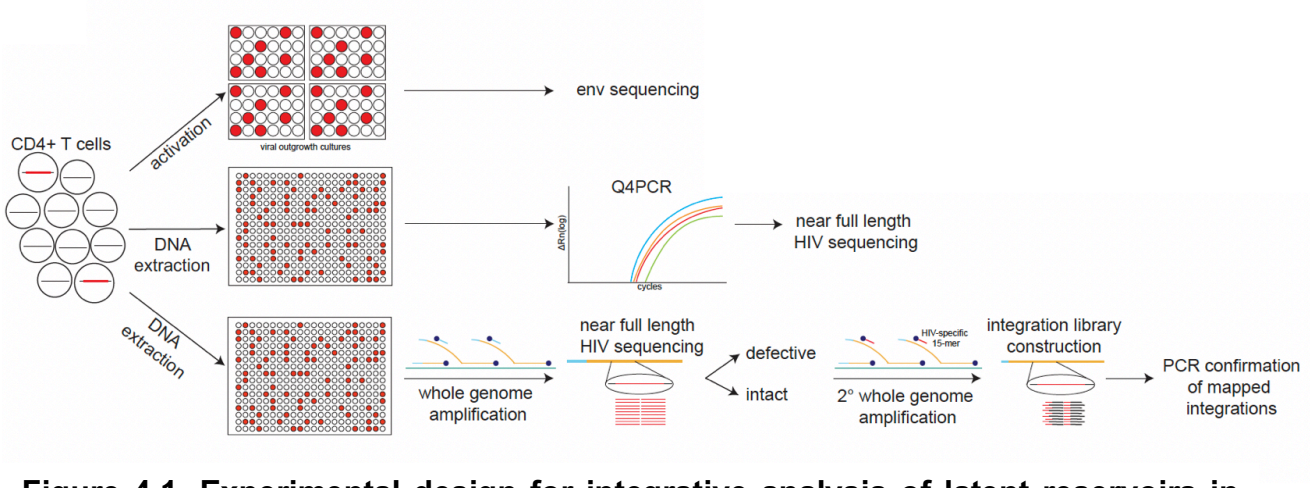


Figure 4.1. Experimental design for integrative analysis of latent reservoirs in ART-treated participants. Env sequences from inducible proviruses are obtained by limiting dilution viral outgrowth cultures (*2*) and near full length (NFL) sequences are obtained by Q4PCR (*4*). Paired NFL sequencing and integration site mapping is performed by modified MIP-seq (*1*).

Table 1. Demographics and clinical features of study participants. *Abbreviations: EVG-elvitegravir, cobi - cobicistat, TDF-tenofovir disoproxil fumarate, FTC -emtricitabine, RPV-rilpivirine, TAF-tenofivir alafenamide fumarate, BIC-bictegravir. ART, antiretroviral treatment; Uninterr., uninterrupted; Dx, diagnosis.

ID	Age	Gender	Race	Year HIV-1 Dx	Uninterr. ART (yrs)	CD4 T cell count	Reported nadir	ART regimen*
9254	48	M	White	1996	21	860	590	EVG/ cobi/ TAF/ FTC
9255	30	M	White	2013	4	1360	779	EVG/ cobi/ TAF/FTC
B207	48	M	Not reported/ Hisp	2006	14	724	50	EVG/ TDF/ FTC
603	38	M	White/ Hisp	2001	15	524	372	RPV/TDF/ FTC
5104	35	M	Black	2011	7	606	400	BIC/ TAF/ FTC
5203	59	M	White	1995	21	483	500	DTG/ RPV
9242	43	M	White/ Hisp	2013	2	654	450	EVG/ cobi/ TDF/ FTC
9252	51	F	Black	-	11	598	270	EFV/ TDF/ FTC
9241	40	M	White/ Hisp	-	5	515	500	EVG/ cobi/ TDF/ FTC

To analyze HIV-1 sequences and determine their corresponding site of chromosomal integration, genomic DNA from CD4+ T cells was diluted to single proviral genome level and subjected to whole genome amplification (WGA) (Figure 4.1). *Env* and NFL HIV-1 sequencing was performed on the WGA products. Viral sequences were compared to previously identified intact viruses (*4, 83*), and proviruses were selected for integration site analysis by ligation-mediated PCR based on whether they were intact or defective. Through this approach, we focused the analysis on intact proviruses and proviral clones. Mapped integration

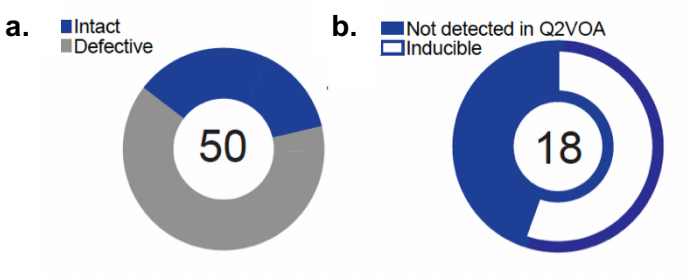


Figure 4.2. Integrative analysis of latent reservoirs in ART-treated participants. (a) Pie chart summarizes the total number of intact and defective proviral sequences mapped, center. Intact proviruses in blue, and defective in gray. **(b)** Pie chart shows the proportion of intact proviruses that are inducible *in vitro* in white. Intact proviruses not detected by Q2VOA represented in blue. The number in center indicates the total number of intact proviral integrations mapped.

sites were subsequently verified by integration site-specific PCR of WGA products. We mapped 50 unique integrations: 32 from defective proviruses and 18 from intact proviruses, 55.5% of which were inducible in outgrowth cultures (Figure 4.2a, b) (*145*). Integration sites for identical sequences were counted as a single integration to avoid bias due to clonal expansion. We define unique or non-clonal proviruses as sequences detected only once in *ex vivo* assays while acknowledging the caveat that they may be members of small clones.

Integrations of highly expanded and inducible proviruses

Phylogenetic analysis was performed to compare the paired intact NFL proviral sequences obtained through MIP-seq to NFL viral genomes obtained by Q4PCR (*4, 83*). Intact proviral sequences matching Q4PCR sequences were found in each of the 6 participants, confirming the absence of selection bias during WGA. Importantly, we captured members of the dominant clones from the reservoir of all 6 participants (Fig. 4.3). Across all individuals, the dominant clone corresponds to *env* sequences identified through outgrowth assays (*11, 145, 146*). Since *env* identity is a robust predictor of proviral clonality (*146*), our analysis indicates that these large clones can produce infectious HIV-1.

Genomic characteristics of mapped integrations

In agreement with studies of global retroviral DNA integration, HIV-1 integration was enriched on chromosome 19 (*116, 147*) (Figure 4.4a and 4.5). Integrations were more frequent in genic regions but there was no significant difference in thihs regard (p=0.38, two-tailed Fisher's exact test). In addition, we found no difference between the intact and defective proviruses with respect to their position in introns and exons (p=0.495, two-tailed Fisher's exact test) (Figure 4.4b). Furthermore, between intact and defective proviruses, there were not significant differences in the distributions of orientation relative to the proximal transcriptional start site (TSS) (Figure 3.4c). Intact and defective integrations occurred at similar distances from the TSS (Figure 4.4d). Both defective and intact proviruses favored integration in the proximity of repetitive short interspersed nuclear elements (SINEs) (Figure 4.4e), with no difference in relative distance to the proximal repetitive element (Figure 4.4f). Whereas we found only one intact and no defective proviruses in centromeric satellite region DNA, this site of integration is characteristic of intact proviruses in elite controllers (Figure 4.4e) (*3*).

Expression and epigenetics of genes associated with mapped integrations

We used publicly available data on memory CD4+ T cells to assess gene expression and epigenetic features in the vicinity of mapped integrations because the latent reservoir resides primarily in this T cell subset (*148, 149*). Across all categories, integrations were predominantly mapped within genes that are moderately or highly expressed in memory CD4+ T cells (Figure 4.9). While there was no difference in gene expression levels between integrations of intact vs. defective proviruses, three-dimensional chromosomal interaction data from Hi-C seq shows intact proviruses are less frequently found in the active chromatin compartments, compared to their defective counterparts (Figure 4.4g)

(*7*). Despite the transcriptional activity, the sites of both intact and defective integrations were associated with high levels of methylation that typically indicate epigenetic repression (Figure 4.4h) (*9*).

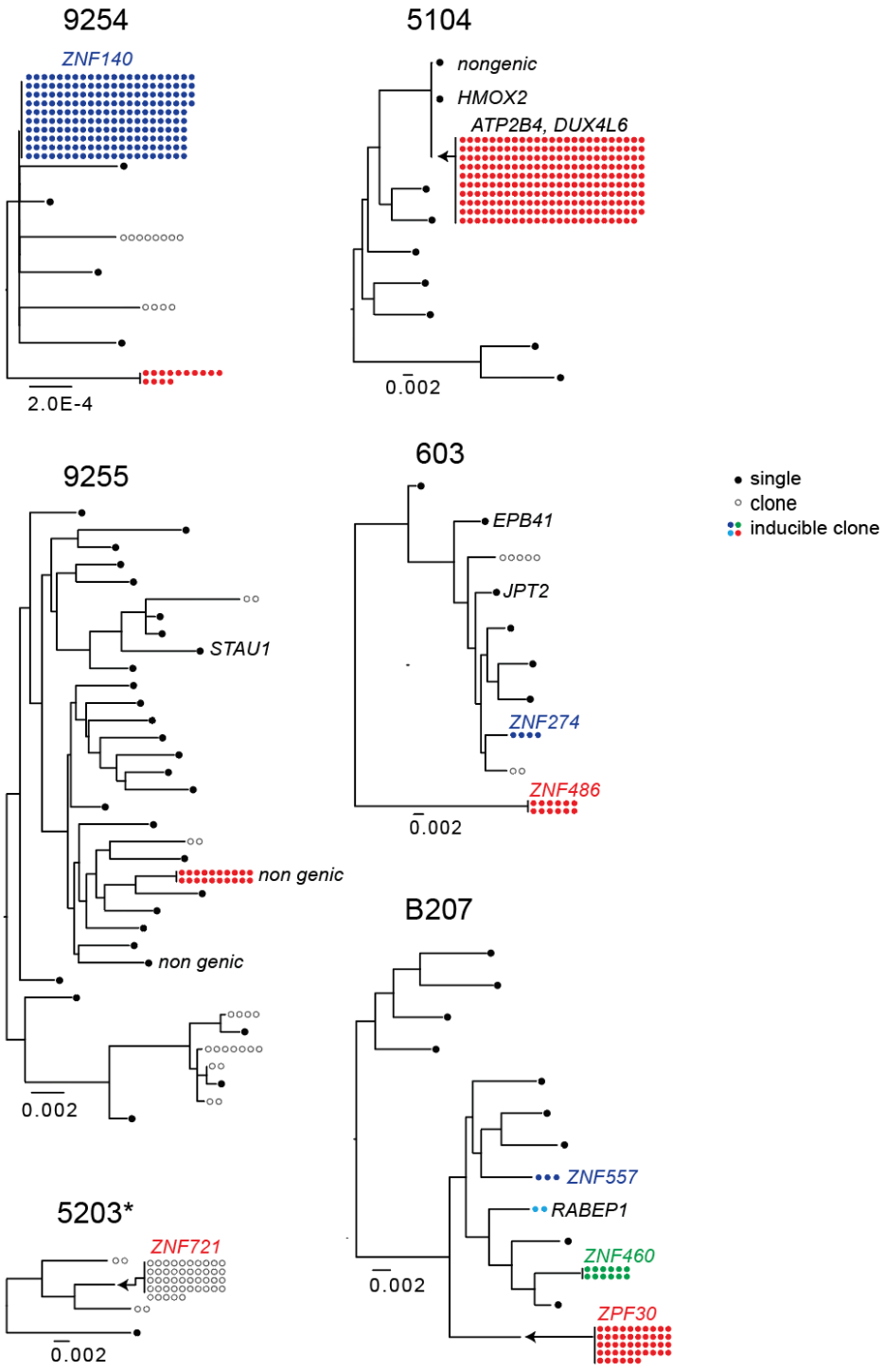


Figure 4.3. Maximum likelihood phylogenetic trees of intact NFL sequences obtained by Q4PCR and MIP-seq (*1, 2*). Filled circles represent inducible clones and the number of circles is proportional to the size of the clone. Outlined circles represent NFL HIV-1 clones detected by PCR, but not by outgrowth assays. Black circles represent unique proviruses detected only once. Integration sites indicated by adjacent text. *Viral outgrowth assay not performed.

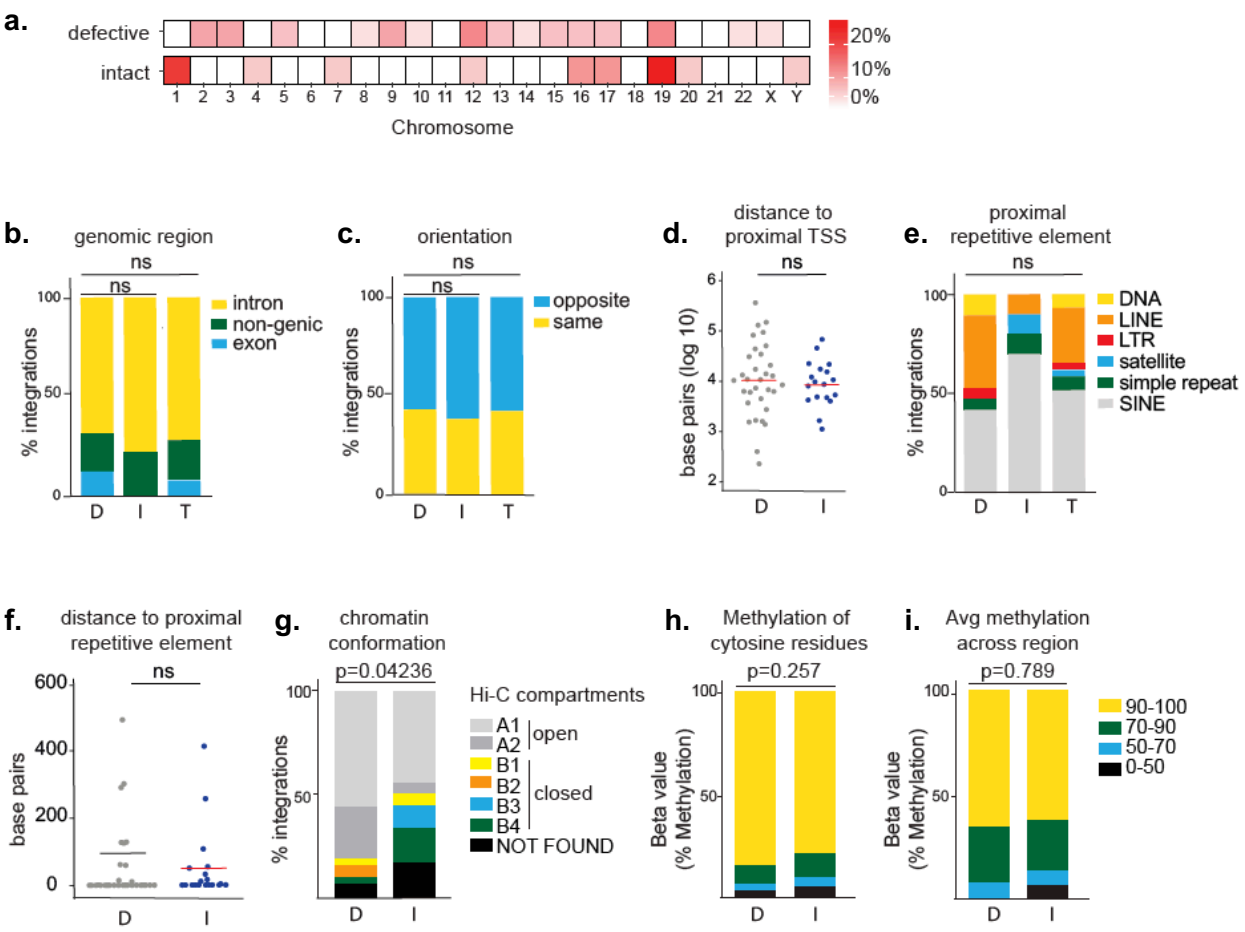


Figure 4.4. Genomic features of defective (D), intact (I), and total (T) integration sites in ART-treated individuals. (a) Upper heat map depicts proportion of HIV-1 defective integration sites in each chromosome; bottom heat map depicts proportion of HIV-1 intact integration sites in each chromosome (see Appendix 2). **(b)** Proportion of integrations in introns, exons, or non-genic regions. **(c)** Proportion of proviruses integrated in opposite or the same direction as host gene transcription. **(d)** Distance, in base pairs, of integration from nearest transcriptional start site (TSS). **(e)** Proportion of integrations in different repetitive elements classified by UCSC RepeatMasker (*6*). **(f)** Distance, in base pairs, of integrations from nearest repetitive element. **(g)** Proportion of intact and defective proviral sequences mapped within chromatin structural compartments A and B and their respective sub-compartments as determined by Hi-C seq data (see Appendix 2) (*7*). P-value refers to proportion of integrations in compartment A as determined by two-proportion Z-test. **(h)** Methylation 1000bp upstream of HIV-1 proviral promoter integration site in CD4+ T cells (*9*). Proportion of intact and defective proviral integrations with average number of cytosine residues with indicated levels of methylation. **(i)** Proportion of intact and defective proviral integrations with indicated methylation levels for all residues. ns = not significant. P-values determined by two tailed Fisher's exact test.

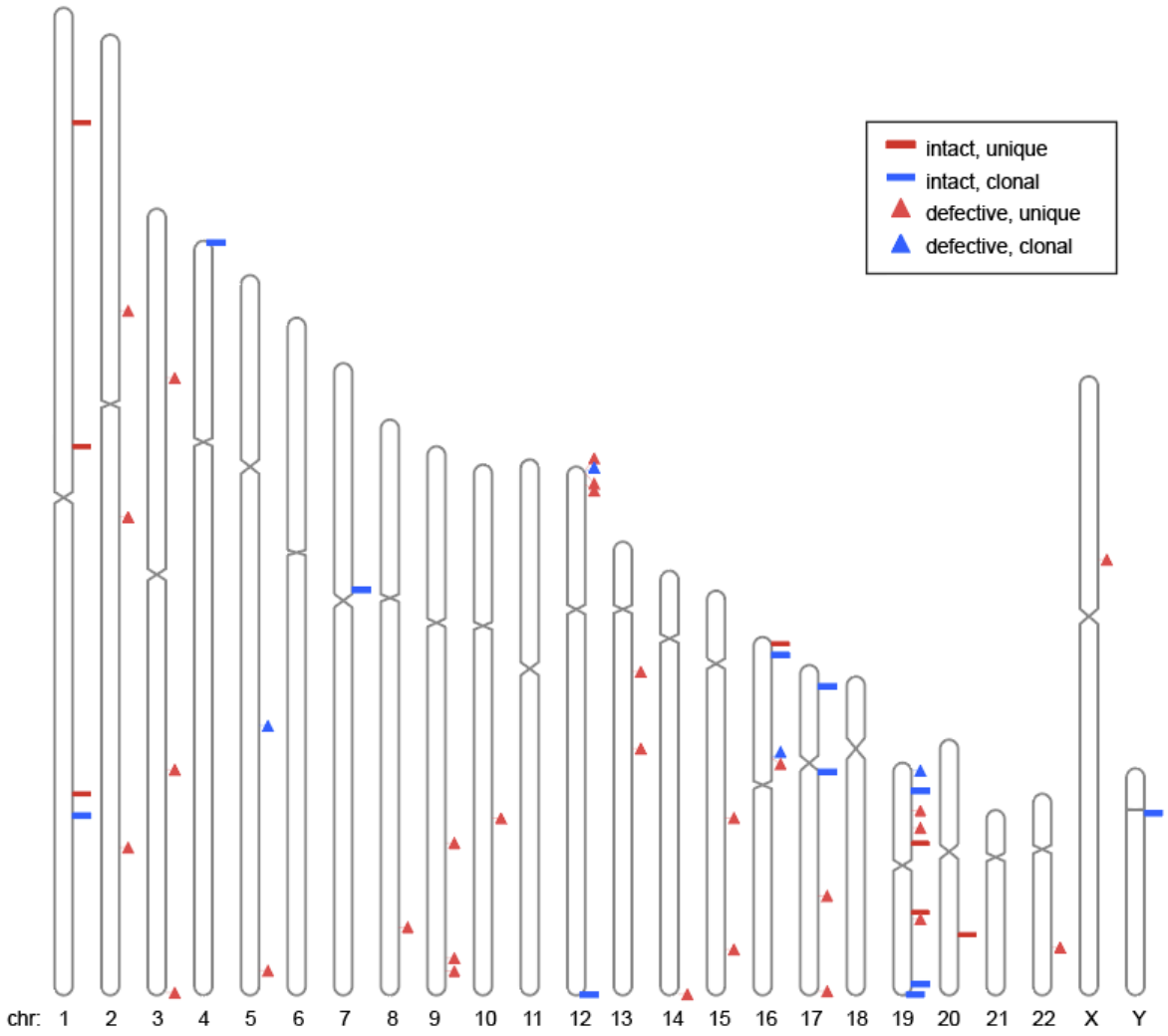


Figure 4.5. Ideogram of positions of mapped integrations. Diagrammatic representation of chromosomal locations of intact and defective; clonal and unique proviruses. Chr; chromosome.

Preponderance of intact, expanded integrations within ZNF genes

When examined individually, there was a notable preponderance of intact provirus integrations into ZNF family genes, particularly amongst expanded clones (Figure 4.3). 61% of intact and 13% of defective proviral integrations were found in expanded clones (Figure 4.3 and 4.6c). Among these clonal proviral integrations 39% and 6% of the intact and defective proviruses respectively were integrated into ZNF genes (Figure 4.3 and Fig. 4.7). This finding is best illustrated in participant B207, whose latent reservoir is largely populated by clones of various sizes. All four intact proviral clones in this individual were inducible *in vitro*, and three of the four integrated proviruses mapped to genes in the ZNF family (Figure 4.3). Moreover, we note that in study participants whose reservoirs contain proviral clones mapped within ZNF genes, the specific clones with integrations in ZNF genes appear to be the most expanded (Figure 4.3).

To better understand the apparent association between intact clonally expanded proviruses and ZNF genes, we combined our data set with other published data obtained from long-term ART-suppressed HIV-infected individuals (*1*). Our intact proviral integrations are significantly enriched in ZNF genes (p=0.0069, Figure 4.6a). However, when all the available data is combined this enrichment is less pronounced (p=0.068, Figure 4.6b). To understand this potential discrepancy, we analyzed clonality in the two datasets. In contrast to our dataset in which there is a preponderance of clonal integrations (61% of intact and 13% of defective integrations), only a minority of the 199 unique integrations in the combined dataset (18.8% of intact and 9.4% of defective integrations) were clonal (Figure 4.6c and d).

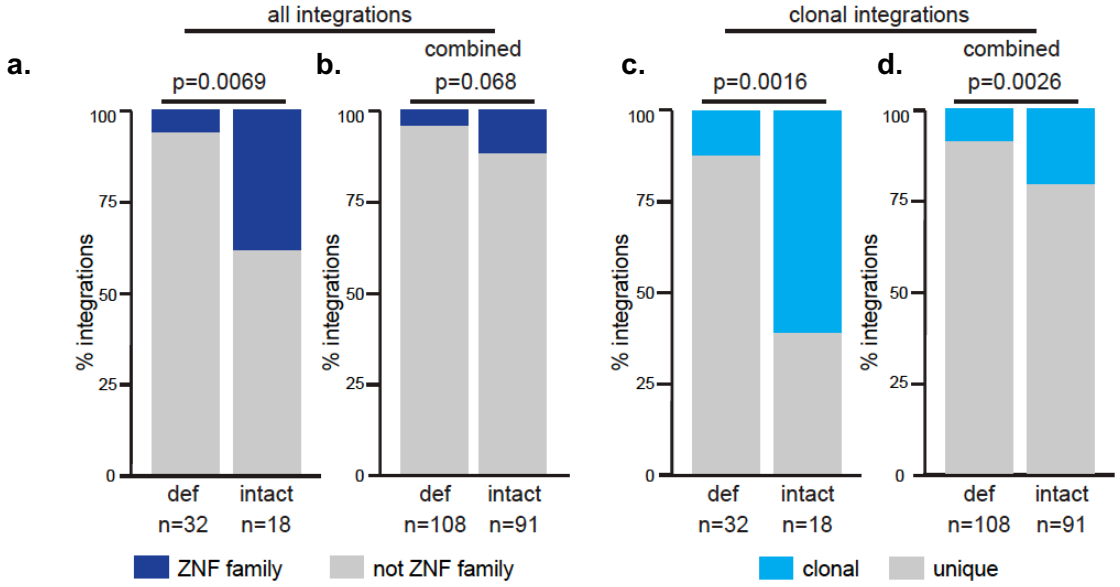


Figure 4.6. Frequency of integrations into zinc finger (ZNF) genes. Proportion of intact and defective proviral integrations mapped to ZNF genes in long-term ART-treated individuals mapped **(a)** in our data and **(b)** in combination with long-term ART-treated individuals reported by others (*1*). Proportion of clonal vs. non-clonal defective and intact integrations in long-term ART-treated individuals mapped **(c)** in our data and **(d)** in combination with long-term ART-treated individuals reported by others (*1*). All p-values determined by two-tailed Fisher's exact test.

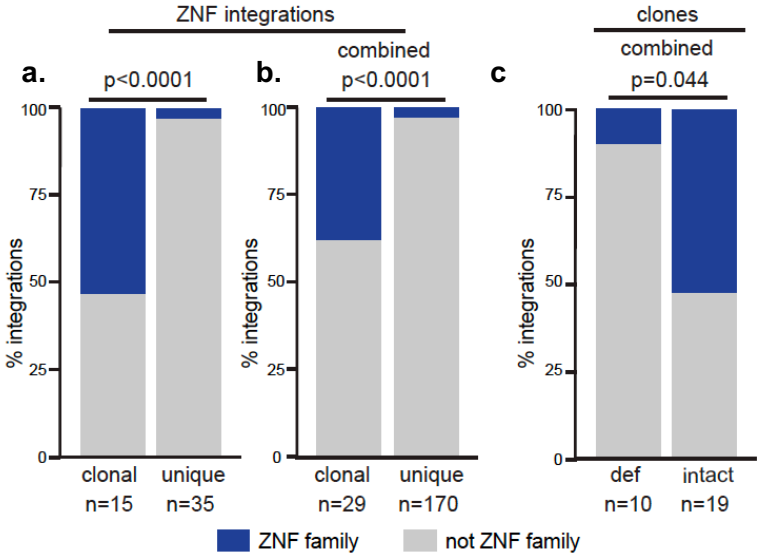


Figure 4.7. Clonal integrations into zinc finger (ZNF) genes. Proportion of clonal and non-clonal integrations mapped in ZNF genes found in long-term ART-treated individuals **(a)** in this study and **(b)** in combination with long-term ART-treated individuals reported by others (*1*). **(c)** Proportion of defective and intact clones in ZNF genes from combined analysis with long-term ART-treated individuals reported by others (*1*). All p-values determined by two-tailed Fisher's exact test.

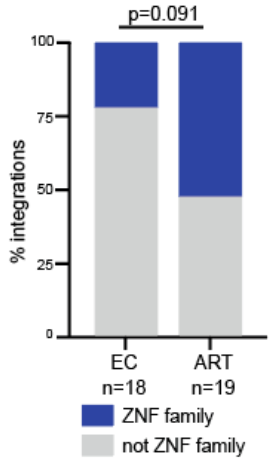


Figure 4.8. Comparison of proviral integration sites in intact proviral clones from elite controllers (EC) and ART-treated individuals (*1, 3*). Fraction of clones obtained from chronically infected ART-treated individuals and EC integrated in ZNF genes represented in blue. ns = not significant, two tailed Fisher's exact test.

We found the proportion of ZNF integration within clones is significantly enriched ($p<0.001$) compared to unique integrations (Figures 4.7a and b). When all intact clonal integrations are combined 11 of 29 were in ZNF genes which is a significant enrichment when compared to defective clonal integrations ($p=0.044$, Figure 4.7c). Thus, intact proviruses found in expanded clones of CD4+ T cells are preferentially integrated in distinct sites in the host genome.

Preferential integration into ZNF family genes on chromosome 19 is associated with viral reservoirs in HIV-1 elite controllers that durably control viral replication without therapy (*3*). The reservoir in these individuals is also enriched in large expanded clones. To determine whether this is specific to the integration landscape of elite controllers or can be generalized to all integrations in expanded clones of CD4+ T cells, we compared the relative proportion of intact clonal proviruses found in ZNF genes in ART-treated individuals and elite controllers. There was no significant difference in the proportion of ZNF integrations between the two groups, suggesting that this is a feature of expanded clones of proviruses and not limited to elite controllers (Figure 4.8).

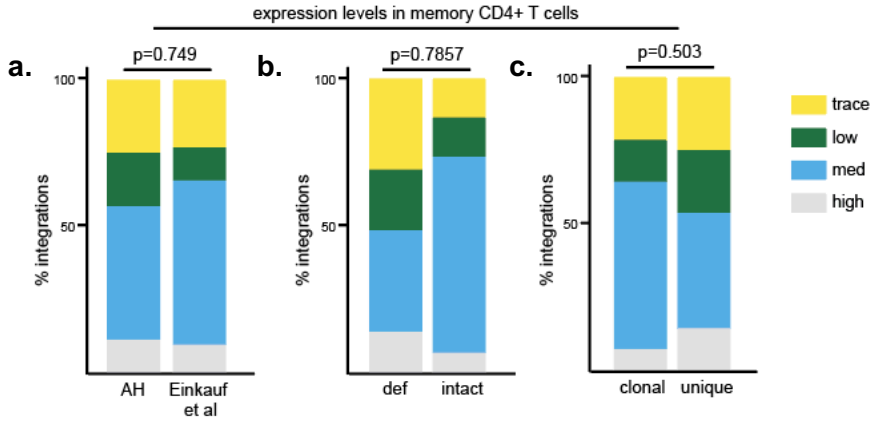


Figure 4.9. Gene expression in resting in memory CD4+ T cells in the vicinity of integration. (a) Proportion of all proviral integrations in genes with low, med, or high expression. Cut-offs determined by upper quartile, median, lower quartile, and minimum expression values. Comparison of integrations in this study (AH) and previous publications (*1*). **(b)** Proportion of defective and intact proviral integrations mapped to genes with trace, low, medium, or high expression (see Appendix 2). **(c)** Proportion of clonal and nonclonal proviral integrations mapped to genes with trace, low, medium, or high expression (see Appendix 2). All p-values refer to proportion of integrations in highly expressed genes, determined by two-proportion Z-test.

ZNF genes are associated with repressive chromatin marks in CD4+ memory T cells. Nevertheless, ZNF genes are expressed in these cells (Figure 4.9). To determine whether ZNF genes associated with clonal integrations might be remodeled to repress transcription upon cellular activation we examined RNA-seq data comparing resting and activated memory CD4+ T cells (*5*). ZNF genes that carry intact integrations are repressed upon memory T cell activation (p= 0.0007, Figures 4.10a and 4.11). In contrast, the transcriptional activity of non-ZNF genes that carry proviral integrations is unaltered by memory CD4+ T cell activation (Figure 4.10b and 4.11). To assess whether this transcriptional pattern is characteristic of both intact and defective integrations, we analyzed expression of ZNF genes associated with defective proviruses (Figure 4.12a) and separately, those associated with replication competent HIV-1 (Figure 4.12b). In this combined data set, activity of ZNF genes harboring defective proviruses remain stable upon cellular activation (ns, p=0.0625), whereas ZNF genes containing intact proviruses become downregulated (p=0.0005). This observation supports the idea that defective and intact proviruses differ in their biology (*27*) and integration site may contribute to differential selective pressures in these proviruses.

In summary, the data suggest that integration into ZNF genes provides a survival advantage that favors clonal expansion of intact HIV-1.

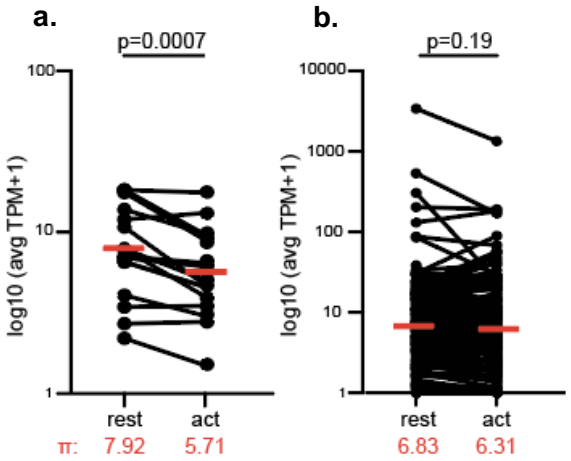


Figure 4.10 (a) RNA expression levels of genes with mapped integrations from this study as well as previously published integration data from elite controllers and individuals on long-term ART (*1, 3*). Normalized expression at rest and upon activation ("act") in primary memory CD4+ T cells for ZNF genes with mapped integrations. Geometric mean (π) of expression levels noted in red (*5*). **(b)** Normalized expression at rest and upon activation ("act") in primary memory CD4+ T cells for non-ZNF genes with mapped integrations (*5*). Geometric mean (π) of expression levels noted in red. P-values

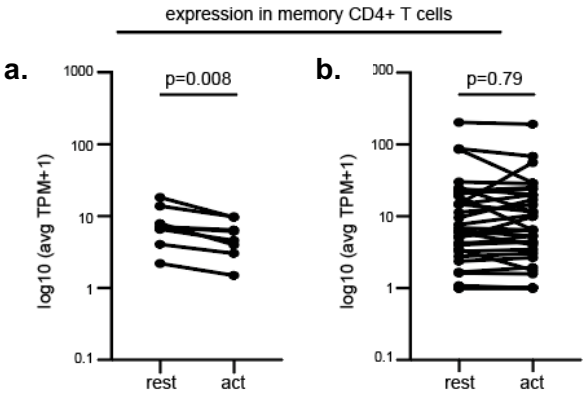


Figure 4.11. Comparison of expression levels in ZNF and non ZNF genes in memory CD4+ T cells with integrations in this study alone. (a) Expression of ZNF genes with mapped integrations, at rest and upon activation with α-CD3/ α-CD28 human T-Activator Dynabeads ("act"). **(b)** Expression of non-ZNF genes with mapped integrations, at rest and upon activation ("act"). P-values determined by Wilcoxon signed rank test.

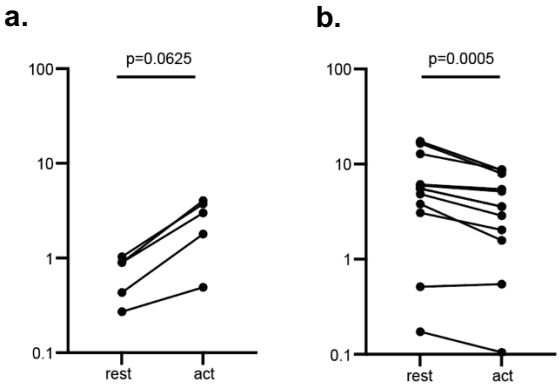


Figure 4.12 (a) RNA expression of ZNF genes harboring defective HIV-1 at rest and upon activation with α-CD3/ α-CD28 human T-Activator Dynabeads ("act"). **(b)** Expression of ZNF genes harboring intact HIV-1, at rest and upon activation ("act"). P-values determined by Wilcoxon signed rank test.

CHAPTER 5: DISCUSSION

Understanding the mechanisms that mediate HIV-1 latency is critical to developing curative strategies. Recent advances in sequencing technologies have demonstrated that the persistent reservoir is dynamic, and it is maintained by proliferation of latent HIV-1 infected cells (*2, 73, 78, 82, 150-152*). Indeed, the majority of intact replication competent latent HIV-1 proviruses are found in expanded clones of CD4+ T cells in chronically infected individuals that are suppressed on ART (*2, 73*). Changes in the quality of the reservoir over time reflect fluctuations in the size of CD4+ T cell clones infected with genetically identical HIV-1 proviruses (*11, 82*). This study provides evidence that integration of latent HIV-1 proviruses is significantly enriched in ZNF genes in clonally expanded CD4+ T cells.

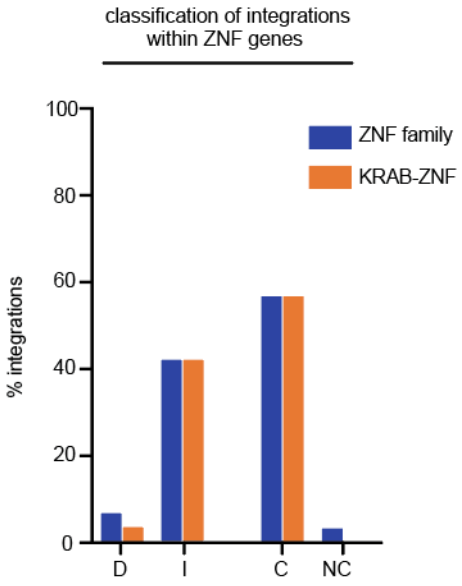


Figure 5.1. Sub-classification of ZNF integrations. Proportion of defective (D) or intact (I); clonal (C) or non-clonal (NC) ZNF integrations found in KRAB-ZNF genes.

Expression of latent HIV-1 proviruses can be induced in CD4+ T cells by stimuli that induce cell division (*72, 153*). However, the latent reservoir evolves over time in individuals on suppressive ART such that an initially diverse collection of latent HIV-1 proviruses tends to become more and more clonal with time (*2, 73, 78, 81, 154, 155*). Reservoir size remains relatively stable, in part, because proliferation and clonal expansion of infected memory CD4+ T cells in response to chronic antigen exposure is offset by cell death (*82, 83, 151*). Given the dynamic nature of the reservoir, the surviving latent proviruses must be selected for resistance to reactivation by signals that induce CD4+ T cell division. Consistent with this idea, the probability of latent HIV-1 reactivation is inversely related to the size of the expanded CD4+ T cell population (*2, 156*).

Given the cytopathic effects of HIV-1 gene expression, expansion of infected cells must be accompanied by persistent silencing of viral transcription. Whether silencing is mediated by repressors of HIV-1 transcription, or the genomic context of HIV-1 integration or both remains to be determined (*1, 11, 114, 157*). To understand how integration into the genome might contribute to survival of clonal latent cells we studied chronically infected participants with large intact reservoirs dominated by expanded clones of latent HIV-1 proviruses. Overall, the genomic characteristics of our intact integrations are consistent with previous studies (*158*). Integrations are enriched in genic regions, and in chromosomes 1 and 19, with no differences between intact and defective proviruses in terms of introns and exons, relative orientation or distance to TSSs (*116, 158*).

In contrast to other HIV-1 infected individuals, intact proviral integrations were shown to favor KRAB-ZNF genes in elite controllers suggesting that this type of integration might contribute to maintenance of HIV-1 infection by favoring survival in the presence of the immune pressure responsible for virologic control (*3*). Our data indicate that this is a more general feature of integrated intact proviruses in expanded clones of CD4+ T cells. In contrast, intact non-clonal and defective proviral integrations are not enriched in KRAB-ZNF family of transcriptional repressors (Figure 4.1). Defective provirus integrations are enriched in proximity to some cancer-associated genes such as *BACH2*, *MKL2*, and *STAT5B* but whether integration in the proximity of such genes favors clonal expansion remains to be determined (*84, 85, 141*).

HIV-1 integration favors actively transcribed regions of the genome (*116, 117*). Consistent with this idea, ZNF genes are actively transcribed in resting memory CD4+ T cells (*1, 3, 5*). Despite a distribution of overall expression levels, ZNF genes in resting memory CD4+ T cells collectively appear to decrease in activity when the cells are activated (Appendix 3) (*5*). Whether this pattern is present in patient-derived latently infected cells is still unknown. Although additional mechanistic studies are necessary to show how ZNF gene expression might impact proviral transcription, we hypothesize that silencing KRAB-ZNF family locus transcription during cellular activation favors latency thereby preventing HIV-1 mediated cytopathic effects and enabling clonal expansion. Conversely, proviral integrations in genes that are transcriptionally activated during CD4+ T cell stimulation might be eliminated over time (*159*).

The extent to which expanded clones with proviral integrations contributes to viral rebound following treatment interruption remains unclear. Rebound viruses are typically distinct from clonally expanded and outgrowth viruses possibly due to suppression by autologous neutralizing antibodies (*160*) and CD8+ T cell responses (*145, 161-163*). In conclusion, the data indicate that the site of proviral integration provides a selective advantage for clonal expansion and survival of latently infected CD4+ T cells .

Looking forward, functional studies are required to understand the basis for enrichment of integrated proviruses in ZNF genes within clonally expanded CD4+ T cells of ART-suppressed individuals. A limitation of our analysis is its basis on the datasets generated from uninfected memory CD4+ T cells. However, studying host transcriptional changes in bona fide latently infected cells is unfeasible: these cells are exceedingly rare and we currently lack the methodology to purify them in their physiological state. As an alternative, a latent cell model may be used to assess how local gene environment impacts proviral transcription. Targeted insertion of a reporter construct (i.e. a fluorescent marker protein driven by HIV-1 LTR and Tat) into previously identified ZNF loci may enable measurement of local transcriptional changes at cellular rest and activation.

Another major area of ongoing research is the determinants of HIV-1 expression, particularly as it relates to latency reversal (*164*). *In vitro*, maximal activation of patient-derived cells is often insufficient to reactivate the latent provirus (*73*). Based on this

observation, it is postulated that reactivation latency reversal is stochastic. Bayesian inference of viruses detected by outgrowth show that larger proviral clones are less likely to be reactivated *ex vivo (2)*. Understanding the relative contributions of integration site and clone size to latent provirus reactivation will be critical in evaluating “shock and kill” as well as “block and lock” modalities of HIV-1 treatment (*165, 166*).

Moreover, it has been demonstrated that latency reversal of infected cells alone does not necessarily induce cell death through viral-induced cytopathic effects (*167*). As such, shock and kill strategies will require robust engagement of HIV-specific antibodies or HIV-specific CTLs (*168, 169*). In individuals where ART is initiated during chronic infection, it has been observed that >98% of proviruses are variants that are resistant to common CTL responses against HIV-1 Gag and that these escape mutations are present in replication competent proviruses (*170*). How integration site contributes to the selection and persistence of these CTL-resistant (or CTL-responsive) proviruses requires further study. It is possible that proviral integration site in particular genes, including ZNF genes, shapes an infected cell’s susceptibility to CTL-mediated elimination by directly disrupting antigen presentation, or indirectly by affecting expression of HIV-1 Nef, which can in turn affect MHC-1 presentation of HIV-1 antigens (*171*). Integration locations that facilitate proviral transcriptional silencing early in infection may also harbor intact proviruses with nonmutated epitopes that can be recognized by CTLs in the event of latency reversal.

This thesis is a first step toward understanding how integration location dampens proviral expression and alters cell intrinsic susceptibility to killing by CD8+ T cells. Our work with J1.1 cells is consistent with the idea that integration site is a key factor in a latent cell’s ability reactivate, and the results of our integrative analyses evidence a role for integration sites in clonal expansion of latently infected cells. Altogether, this thesis highlights the heterogeneity and dynamic nature of latent reservoirs in HIV-1 infected individuals, and illustrates the importance of addressing the interplay between genomic context, latency and immune evasion as we develop eradicative or therapeutic strategies for HIV-1.

CHAPTER 6: MATERIALS AND METHODS

CD4+ T cell isolation

Frozen aliquots of PBMC were thawed in 37°C water bath and washed once in cold RPMI medium. CD4+ T cells were isolated from bulk PBMCs through negative selection magnetic separation (Miltenyi Biotec, 130-096-533).

Latent cell capture

Purified CD4+ T were cultured at 2×10^6/mL in RPMI supplemented with 10% heat inactivated FCS, 10mM HEPES, 100U/mL penicillin/ streptomycin), and 25% volume conditioned media. Conditioned media was made by culturing healthy PBMCs in R10 with PHA and IL-2 for 2 days, followed by a wash and 5 days in culture with IL-2 alone. The conditioned media was then collected and frozen at –80C until use. 100U/mL IL-2 (Peprotech), 1ug/mL PHA (Sigma), 10uM Z-VAD-FKM (R&D), 10uM Ritonavir, 10uM Dolutegravir, 10uM Emtricitabine, 5uM Tenofovir, and 10uM Maraviroc (Selleckchem) were added to the media. 36h later, cells were labeled with 5ug/mL each of biotinylated 3BNC117, 10-1074, PG16, followed by Streptavidin PE (1:500, BD) and anti-PE magnetic beads (Miltenyi Biotech). Cells were then passed over a magnetic column and bound cells were eluted for downstream analysis. For FACS sorting, cells were labeled with the following antibodies, all Biolegend: CD1c (cat. no. 331510), CD3 (cat. no. 300430), CD4 (cat. no. 317444), CD8 (cat. no. 344726), CD14 (cat. no. 301812), CD20 (cat. no. 302318), CD32a (cat. no. 303204), and CD56 (cat. no. 318314).

TCR sequencing

Briefly, single CD4+ T cells were sorted into 5µL TCL buffer (Qiagen) with 1% 2-mercaptoethanol. RNA were purified using RNA SPRI beads (Agencourt A63987) and cDNA was generated using primer Bio-T22VN (5'Bio- TTTTTTTTTTTTTTTTTTTTTTVN). TCRVα and TCRVβ were amplified as previously described (*172*).

Cell culture

All primary CD4+ T cells and T cell derived cell lines were cultured in R10: RPMI 1640 (RPMI, Gibco) supplemented with 10% (vol/vol) fetal bovine serum, 1% penicillin/ streptomycin, 1% L-glutamine, 1% antibiotic-antimycotic and 1% HEPES. J1.1 cells and all J1.1-derived subclones were maintained in 10µM emtricitabine (Selleckchem) and 5µM tenofovir (Selleckchem) to prevent superinfection.

Generation of cell lines

J1.1 and associated cell lines

The HIV-1 Lymphadenopathy-Associated Virus (LAV)-infected Jurkat E6 (J1.1) cell line was obtained through the NIH HIV Reagent Program (*104*). To remove non-latent, HIV-expressing cells J1.1 were labeled with with 5µg/mL each of biotinylated 3BNC117, 10-1074, and PG16 antibodies against Env, followed by Streptavidin PE (1:500, BD) and Fixable Viability Stain (BD). Live, non-labelled ($Env^{negative}$) cells were sorted and cultured

for subsequent experiments. To account for cellular heterogeneity within the J1.1 population, we single cell sorted Env- J1.1s into 96-well plates and propagated the cells as subclones for subsequent experiments.

As a control for cellular conditions in studies involving J1.1 cells, we generated J1.1-derived cell lines in which HIV-1 was knocked out by nucleofection of Cas9 ribonucleoproteins with guide RNAs targeting the HIV-1 LTR and *gag*. Complete deletion of the full-length provirus was confirmed by PCR as well as p24 intracellular staining following stimulation with αCD3/CD28 tetramers.

Single cell sorting
Sorts were performed on the BD FACS Aria into 96-well plates containing 5µL TCL buffer (Qiagen) with 1% 2-mercaptoethanol. Sorted plates were immediately centrifuged, immediately frozen on dry ice, and stored at -80°C.

Single cell RNA sequencing
RNA-seq libraries were generated based on the protocol of Trombetta et al (*173*), using primers from Islam et al (*174*).

RNA was converted to full-length cDNA using oligo(dT) priming (5'-Bio-AATGATACGGCGACCACCGATCGTTTTTTTTTTTTTTTTTTTTTTTTTTTT) and SMART template switching technology (5'-Bio-AAUGAUACGGCGACCACCGAUNNNNNGGG) followed by 24 cycles of PCR preamplification of cDNA (primer 5'-Bio-GAATGATACGGCGACCACCGAT). Amplified cDNA was used to construct standard Illumina sequencing libraries. Samples were sequenced by Illumina NextSeq.

Bulk RNA sequencing
Two HIV-1 knockout J1.1 clones, two highly inducible J1.1 clones (C5, C21), and two lowly inducible J1.1 clones (C6, C30) were stimulated with αCD3/CD28 tetramers. After 4hr and 24hr, cells were labelled with a biotinylated antibody cocktail against HIV-1 Env, as previously described. Live cells were sorted from the knockout clones; Env- cells were sorted from the lowly inducible clones; Env+ cells were sorted from the highly inducible clones. For each condition,150,000 cells were sorted into cold R10 buffer. RNA was extracted using RNeasy Plus Mini Kit (Qiagen). Standard Illumina sequencing libraries were constructed by The Rockefeller University Genomics Resource Center and sequenced by Illumina NextSeq.

RNA Polymerase II Chromatin Immunoprecipitation
RNAPII-ChIP was performed based on a protocol developed by Yamane et al (*175*).

Two lowly inducible J1.1 subclones (C6, C30) and two highly inducible J1.1 subclones (C5, C21) were assayed at rest and after 1hr stimulation with αCD3/CD28 tetramers. As a positive control, the highly inducible subclone C5 was also stimulated for 24h. For each

sample, $20x10^6$ cells were fixed in 1% (vol/vol) formaldehyde and quenched with 1/20 volume of 2.5M glycine in 1xPBS. Cell pellets were lysed in RIPA buffer (10mM pH7.6 Tris-HCl, 1mM EDTA, 0.1% SDS, 1% Triton X-100) supplemented with EDTA-free protease inhibitor, and sonicated using the Covaris S220 (peak power 105; duty factor 5; cycles per burst 200; duration 10 min).

Immunoprecipitation and sequencing
Lysates were incubated with 50ul Dynabeads Protein G (ThermoFisher) only for 30mins at 4°C to pre-clear chromatin. Antibody-bound beads were generated by combining washed Dynabeads Protein G with anti-RNAPII CTD repeat YSPTSPS (phospho S5) antibody (4H8, Abcam). To enrich for DNA fragments bound by RNAPII, pre-cleared chromatin was incubated with antibody-bound Dynabeads overnight at 4°C, with gentle rotation. Beads were washed as follows: 2x with 1mL RIPA buffer; 2x with 1mL RIPA buffer + 0.3M NaCl; 2x with 1mL LiCl buffer (0.25M LiCl, 0.5% NP-40, 0.5% sodium deoxycholate); 1x with 1mL TE + 0.2% Triton X-100; 1x with 1mL TE, and resuspended in TE. To reverse crosslink, 3µl 10% SDS and 5µl 20mg/mL proteinase K were added and the mixture was incubated for 4hr at 65°C. Supernatants enriched for RNAPII-bound DNA fragments were extracted using phenol-chloroform and ethanol precipitation. Fragments were blunted (Lucigen, ER81050), A-tailed, and NextFlex DNA barcodes (diluted 1:1000) were ligated. Libraries analyzed on 2% agarose gel, purified, pooled, and sequenced on the Illumina HiSeq.

Study participants and sample collection
All study participants were recruited by the Rockefeller University Hospital, New York, NY, and the University Hospital Cologne, Cologne, Germany. Informed consent was obtained from all subjects. All relevant ethical regulations were followed. Peripheral blood mononuclear cells (PBMCs) were isolated from leukapheresis performed according to protocols approved at The Rockefeller University by the Rockefeller Internal Review Board. Individuals 9254 and 9255 were participants in analytical treatment interruption (ATI) trial NCT02825797. Individuals 5203 and 5104 are participants in ATI trial NCT03526848. Individuals 603 and B207 did not undergo ATI.

Participant eligibility criteria included adults aged 18-65 with HIV-1 infection and undetectable plasma HIV-1 RNA (<20 copies/ ml) during ART. Participants were confirmed to be aviremic at time of sample collection. PBMCs were purified by Ficoll separation, frozen in 90% fetal bovine serum/ 10% DMSO, and stored in liquid nitrogen.

DNA extraction
Purified CD4+ T cells were briefly frozen at -80°C, thawed on ice, and resuspended in 100µl 1x PBS. For lysis, 1mL Proteinase K buffer (100mM pH 8 Tris, 0.2% SDS, 200mM NaCl, 5mM EDTA) and 10ul Proteinase K (20mg/mL) were added per 2 million cells. Genomic DNA (gDNA) was extracted using phenol-chloroform, followed by ethanol precipitation, and incubated with 20mg/ml RNAse A (1µl per 1 million cells) at room temperature for one hour.

Quadruplex PCR (Q4PCR)
Q4PCR was performed as previously described (*4*). Briefly, to quantify total HIV-1 proviral frequency, gDNA was assayed for gag by qPCR. gDNA was diluted to single viral genome levels based on gag qPCR results such that <30% of wells were *gag*+. Individual reactions containing a single proviral copy as template were PCR amplified to produce near full length HIV-1 amplicons. Individual amplicons were assayed for reactivity with each of four qPCR probes covering HIV-1 packaging signal, *gag*, *pol*, and *env*. Samples reactive with ≥2 probes were sequenced and analyzed (*70, 163*).

Whole genome amplification
gDNA was diluted to single viral genome levels and subjected to multiple displacement amplification (MDA) with ϕ29 polymerase at 30°C for 4 hours (Qiagen, 150345). Subsequently, MDA products in each well were diluted 1:10 with nuclease-free water and subject to near full length HIV-1 sequencing (*70, 163*). Different plates were processed independently and no difference was observed between plates within each individual sampled.

HIV-1 near-full-length sequencing
Near full length HIV-1 sequences were amplified from 1.5ul of diluted MDA products using a nested two-amplicon (*70*) or a nested five-amplicon method (*1*). PCR amplicons were subject to paired-end sequencing using Illumina MiSeq Nano 300 V2 cycle kits at a concentration of 12pM. HIV-1 sequences were reconstructed and classified as intact or defective using our in-house pipeline (Defective and Intact HIV Genome Assembler), as previously described (*83*). Clones were defined by aligning sequences classified as Intact, MSD mutation, Non-functional, or Missing Internal Genes to HXB2 calculating the Hamming distance. Sequences containing three or fewer differences between the first nucleotide of *gag* and the last nucleotide of *nef* were classified as members of the same clone. Maximum likelihood phylogenetic trees were generated using FigTree (https://github.com/rambaut/figtree/releases/).

Integration library construction
To obtain sufficient material for integration library construction, a second-round whole genome amplification was performed from the first round of MDA products. To enrich for amplicons containing the 3' HIV-1 and human sequence, 1µl of 50µM 15UTRi-F (5'-CTAGGGAACCCACTG) was spiked into the MDA reaction. Second-round MDA products were purified using AMPure XP beads (Beckman Coulter) and 10µg of DNA was sonicated to 300-1500bp. Fragments were ligated to 200pM of annealed linker, digested with BglII (NEB) to remove HIV-1 sequences, and bead purified For each sample, DNA was divided into 300ng aliquots and amplified by linear PCR with biotinylated LTR1 (5'-CTTAAGCCTCAATAAAGCTTGCCTTGAG) [1 × (98°C-1 min) 17 × (98°C-15 s, 62°C-30 s, 72°C-30 s) 1 × (72°C-5 min)]. Reactions were spiked with p115 primer (5'-GCAGCGGATAACAATTTCACACAGGAC)' and further amplified as follows: [1 × (98°C-1 min) 25 × (98°C-15 s, 62°C-30 s, 72°C-30 s) 1 × (72°C-5 min)]. Products larger than

300bp were purified using AMPure XP beads and enriched by magnetic streptavidin T1 beads. DNA was amplified off the magnetic beads by semi-nested PCR using LTR2 (5'-AGACCCTTTTAGTCAGTGTGGAAAATC; same cycling conditions as above) before a spike-in of p115 and additional cycles. PCR products were clean-up with AMPureXP beads (Beckman Coulter), blunted (Lucigen, ER81050), and ligated with NextFlex paired-end adapters (PerkinElmer) diluted 1:10. Adapter-ligated fragments were PCR-enriched with NextFlex primers: [1 × (98°C-1 min) 27× (98°C-15 s, 66°C-30 s, 72°C-30 s) 1 × (72°C-5 min)]. Amplicons from 300-1500 base pairs were gel purified and sequenced by 150bp paired-end sequencing on Illumina MiSeq or Illumina NextSeq 500. All PCR reactions were performed with High Fidelity Phusion polymerase (ThermoFisher).

Computational pipeline for mapping HIV-1 integration sites
Reads starting with the last 34 nucleotides of the HIV-1 LTR (bait reads) and corresponding pairs starting with the first 43 nucleotides of the p115 linker (target reads) were extracted from raw sequencing file using cutadapt v2.10 (https://cutadapt.readthedocs.io/en/stable/) and seqtk (https://github.com/lh3/seqtk).

Bait reads were mapped to the HIV-1 LTR using Smalt (https://www.sanger.ac.uk/tool/smalt-0/), and only read pairs containing identical 34 nucleotides were selected for downstream processing. The paired-end reads were mapped to the human genome assembly GRCh38 after trimming both LTR and the p115 linker sequences by BBduk from BBTools (https://jgi.doe.gov/data-and-tools/bbtools/). Once mapped, reads were subjected to a two-step filtering process using in-house Perl scripts. The first step, a distance-based filter, selects the most reliable PCR fragments generated by the 3' LTR and p115 primers. It consists of selecting reads overlapping a maximum of 15 nucleotides or mapped at a maximum distance of 2 kilobases from each other. The second step, an alignment-based filter, determines reads resembling bona-fide HIV-1 integrations in the human genome. Paired-end reads are selected if the bait read met the following criteria: (1) at least 30 perfect matches with the human genome, (2) soft-clipping between 1 and 3 nucleotides, at least 27 identical nucleotides in the remaining sequence, and finally (3) one match in the first nucleotide followed by 1 or 2 mismatches and more then 28 matches in the remaining sequence. Reads mapping in the same position and orientation were combined into a single putative PCR fragment amplified by the LTR2 and p115 primers. The number of reads collapsed into one fragment reflects the score of the respective fragment. Fragments supported by a single alignment were not considered in the analysis. At the end of the pipeline execution, integration hotspots were calculated based on a definition of hotspots established by previous work on Translocation-Capture Sequencing (*176*).

Parallel execution and computing resource management were carried out by the Snakemake framework (*177*).

PCR verification
To exclude sequencing artifacts, mapped integration sites were confirmed by PCR.

LTR1 and LTR2 primers were used in combination with genome-specific primers for in nested PCRs using high fidelity Phusion polymerase. Products were visualized by gel electrophoresis, purified, and subject to Sanger sequencing.

Combined analysis with published integration datasets

The published dataset for integration sites during long-term ART was generated from PBMCs of three virally suppressed HIV-infected individuals recruited at Massachusetts General Hospital (Boston, MA), Brigham and Women's Hospital (Boston, MA), and the National Institutes of Health Clinical Center (Bethesda, MD), and who have been on continuous ART for 8–13 years at the time of PBMC sampling (*1*). Integration sites were mapped using integration site loop amplification, nonrestrictive linear amplification-mediated PCR, or ligation-mediated PCR (Lenti-X Integration Site Analysis Kit; Clontech; cat. 631263) (*84, 178*). The published dataset for integration sites in ECs was generated from PBMCs of 11 HIV-infected ECs recruited at Massachusetts General Hospital, Brigham and Women's Hospital, and at the University of California, San Francisco, at the Zuckerberg San Francisco General Hospital (*3*). In both studies, DNA from enriched CD4+ T cells was amplified and quantified using digital droplet PCR (BioRad) to detect a 127-bp 5' LTR-gag amplicon (*79*).

Computational analysis of HIV-1 integration sites

HIV-1 integration sites were annotated by in-house Perl scripts using GENCODE v32 and RepeatMasker tracks downloaded from the University of California, Santa Cruz Genome Browser (https://genome.ucsc.edu/cgi-bin/hgTables). Because more than one transcript may be equally distant from the integration site, a priority level was defined to select one transcript representing a possible affected gene. We selected transcripts if the HIV-1 integration was (1) inside an exon, (2) inside an intron, or (3) in a non-genic region. For overlapping transcripts, we selected the one containing the closest TSS.

We obtained information on gene expression in different T cell subtypes from the Database of Immune Cell Expression, Expression of quantitative trait loci and Epigenomics (*179*) and from bulk RNA sequencing data generated by Cano-Gamez et al (*5*). Genes were divided into silent (0 transcripts per million) and expressed genes. We observed a bimodal distribution on the expressed genes. Expressed genes were subdivided into trace-expressed genes and more highly expressed genes using a normal mixture model implemented by the mclust R package (*180*). The higher group was further divided into tertiles, representing low, medium, and highly expressed genes.

Epigenetics analyses were performed by integrating data from three different sources: (1) methylation from the iMethyl database (*9*), (2) chromatin immunoprecipitation sequencing data for H3K4me1 and H3K9me3 histone marks from the Roadmap Epigenomics Consortium repository (*181*), and Hi-C data generated by Rao et al. (*7*).

Qualitative and quantitative viral outgrowth assay and sequence analysis

CD4+ T cells purified from cryopreserved PBMCs were cultured at 300,000 cells per well in 24-well plates in R10 with 1µg/mL phytohemagglutinin (PHA; LifeTechnologies) and 100 U/mL of IL-2 (Peprotech). $2.5x10^6$ of irradiated heterologous PBMCs in 1mL R10 were also added to each well as feeder cells. After 24h, 1.5ml of media was removed from each well, and 1.5mL of Molt4-CCR5 cells in R10 and 100 U/mL IL-2 were added to each well as target cells. After 9 days, 1 mL of cell suspension was removed from each well and replaced with 1mL R10 with IL-2. At day 14, the supernatant of each well was tested for HIV-1 production using an in-house ELISA to detect p24 antigen. Under these conditions, using $0.3x10^6$ donor cells, less than 30% of all cultures were positive for all individuals.

Supernatants from p24-positive wells were collected and RNA extraction was performed using the Qiagen MinElute Virus Spin kit QIACube. cDNA was generated using 1:10 diluted RNA with SuperScript III reverse transcriptase (Invitrogen Life Technologies) and antisense primer envB3out (5'-TTGCTACTTGTGATTGCTCCATGT), followed by RNase H digestion at 37°C. cDNA was diluted 1:40 for amplification of full gp160 env using envB5out (5'-TAGAGCCCTGGAAGCATCCAGGAAG) and envB3out (5'-TTGCTACTTGTGATTGCTCCATGT) in the first round and nested primers envB5in (5′-TTAGGCATCTCCTATGGCAGGAAGAAG) and envB3in (5′-GTCTCGAGATACTGCTCCCACCC) in the second round. PCR assays were performed using High Fidelity Platinum Taq (Invitrogen) at 94°C for 2 min; 94°C for 15 s, 55°C for 30 s, and 68°C for 4 min × 35; and 68°C for 15 min. Second-round nested PCR was performed using 1 µL of PCR1 product as a template and High Fidelity Platinum Taq at 94°C for 2 min; 94°C for 15 s, 57°C for 30 s, and 68°C for 4 min × 45; and 68°C for 15 min. PCR2 products were checked using 1% 96-well E-Gels (Invitrogen). PCR bands with the expected HIV-1 envelope size were quantified and subjected to library preparation using the Illumina Nextera DNA Sample Preparation Kit (Illumina). Sequence adapters were removed using Cutadapt v1.8.3. Read assembly for each virus was performed in three steps. First, de novo assembly was performed using Spades v3.6.1 to yield long contig files. Contigs longer than 255 base pairs were subsequently aligned to an HIV-1 envelope reference sequence, and a consensus sequence was generated using Geneious 8. Finally, reads were realigned to the consensus sequence to close gaps, and a final read consensus was generated for each sequence. Sequences with double peaks (cutoff consensus identity for any residue <75%), with stop codons, or shorter than the expected envelope sizes were omitted from downstream analyses.

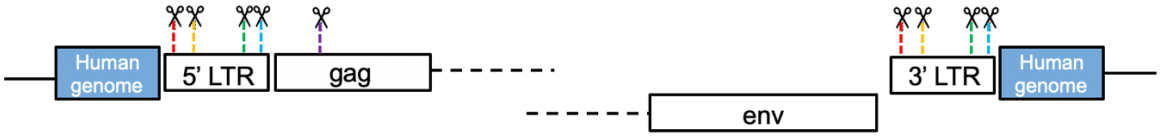
Data availability

Source code for integration analysis pipeline can be accessed at the following link: https://github.com/victor-ramos/HIVIntegrationPipeline . Due to privacy concerns regarding patient genomic information, sequence data will only be made available to investigators upon reasonable request and signing of a coded tissue agreement.

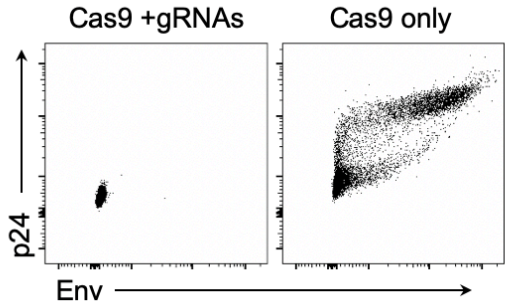
APPENDIX 1

Generation of J1.1 knockout cell line. **(a)** Experimental design for excision of integrated HIV-1 **(b)** Flow cytometry validation of knockout.

a.



b.



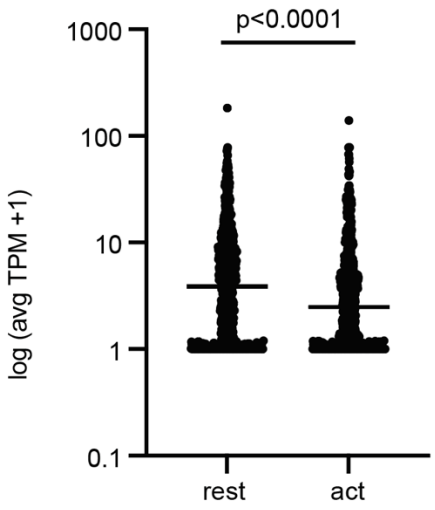
APPENDIX 2

Characteristics of integration sites mapped from CD4+ T cells from individuals on long-term ART.

sample	status	chr	integration_start	gene_symbo	orientation	Intron_or_Exon	distance_tss	rep_element_class	expression level
1	defective	chrX	46198904	LINC01186	same	Non-genic region	128745	LINE	trace
2	defective	chr12	4488235	C12orf4	same	Exon	50234	DNA	low
3	defective	chr10	89118459	AL513533.1	opposite	Non-genic region	20584	LTR	not found
9	defective	chr9	132298647	SETX	opposite	Intron	13209	LINE	med
10	defective	chr12	874762	WNK1	opposite	Intron	6195	SINE	med
11	defective	chr22	38712522	GTPBP1	same	Intron	6003	LINE	low
12	defective	chr13	32733756	PDS5B	same	Intron	147209	LINE	low
14	defective	chr19	1899111	SCAMP4	same	Non-genic region	6260	SINE	med
17	defective	chr16	30464869	Y_RNA	opposite	Non-genic region	6717	SINE	trace
19	defective	chr12	1703793	ADIPOR2	opposite	Intron	12631	SINE	high
20	defective	chr19	16389851	EPS15L1	opposite	Intron	13894	LINE	low
21	defective	chr5	113557188	YTHDC2	opposite	Intron	43495	LINE	med
22	defective	chr8	127983564	PVT1	opposite	Intron	92850	LINE	trace
23	defective	chr12	1221135	ERC1	same	Intron	17170	LINE	trace
24	defective	chr19	12039504	ZNF433-AS1	opposite	Intron	3682	SINE	trace
25	defective	chr5	175301400	ARL2BPP6	same	Non-genic region	10670	LINE	trace
26	defective	chr17	58171055	OR4D2	opposite	Exon	1655	SINE	trace
28	defective	chr3	141401063	ZBTB38	same	Intron	8345	DNA	med
29	defective	chr9	100017375	ERP44	same	Intron	81625	SINE	med
30	defective	chr9	128978248	NUP188	same	Intron	30550	Simple_repeat	not found
31	defective	chr3	197614126	AC132008.2	opposite	Intron	1534	LTR	med
32	defective	chr17	82208352	CCDC57	same	Intron	4461	SINE	med
33	defective	chr2	69598933	B3GALNT1P1	same	Exon	1581	SINE	med
34	defective	chr13	52150248	NEK3	opposite	Intron	7237	LINE	low
36	defective	chr2	121571845	Not Found	opposite	Intron	34215	SINE	low
37	defective	chr15	90398610	IQGAP1	same	Intron	10334	DNA	high
38	defective	chr19	39353346	SAMD4B	opposite	Intron	10879	DNA	med
41	defective	chr16	30510737	ITGAL	opposite	Intron	1379	SINE	high
42	defective	chr2	204905856	PARD3B	opposite	Intron	359857	SINE	trace
43	defective	chr14	106728842	IGHV3-69-1	opposite	Non-genic region	227	SINE	not found
44	defective	chr3	42629897	NKTR	same	Exon	2712	Simple_repeat	high
45	defective	chr15	57248438	TCF12	opposite	Intron	397	LTR	trace
4	intact	chr17	26936842	AC069061.1	opposite	Non-genic region	44851	Satellite	high
5	intact	chr1	203635309	ATP2B4	opposite	Intron	8523	SINE	med
6	intact	chrY	11302961	DUX4L16	same	Non-genic region	3956	Simple_repeat	med
7	intact	chr16	4491128	NMRAL1	same	Intron	4635	DNA	med
8	intact	chr4	456927	ZNF721	opposite	Intron	17124	SINE	med
13	intact	chr1	28988709	EPB41	same	Intron	1099	SINE	med
15	intact	chr19	58196520	ZNF274	opposite	Intron	9574	SINE	med
16	intact	chr16	1685129	JPT2	opposite	Intron	4793	SINE	not found
18	intact	chr19	20187849	ZNF486	same	Intron	20622	LINE	low
27	intact	chr12	133102882	ZNF140	opposite	Intron	21298	SINE	low
35	intact	chr1	110669697	KCNA3	opposite	Non-genic region	5243	SINE	high
39	intact	chr7	57155190	AC099654.11	same	Non-genic region	15079	Simple_repeat	not found
40	intact	chr20	49143597	STAU1	same	Intron	10471	SINE	med
46	intact	chr19	37646812	ZFP30	opposite	Intron	7917	SINE	trace
47	intact	chr19	57283507	ZNF460	same	Intron	1621	SINE	med
48	intact	chr1	198161290	NEK7	opposite	Intron	4170	DNA	low
49	intact	chr19	7081618	ZNF557	opposite	Intron	11914	Simple_repeat	med
50	intact	chr17	5349173	RABEP1	opposite	Intron	66687	LINE	med

APPENDIX 3

Comparison of resting state (“rest”) and activated state (“act”) gene expression of over 700 C2H2-type zinc finger family (ZNF) genes in resting memory CD4+ T cells. P-value calculated using Wilcoxon signed rank test in Prism 9.



REFERENCES

1. K. B. Einkauf *et al.*, Intact HIV-1 proviruses accumulate at distinct chromosomal positions during prolonged antiretroviral therapy. *J Clin Invest* **129**, 988-998 (2019).
2. J. C. Lorenzi *et al.*, Paired quantitative and qualitative assessment of the replication-competent HIV-1 reservoir and comparison with integrated proviral DNA. *Proc Natl Acad Sci U S A* **113**, E7908-E7916 (2016).
3. C. Jiang *et al.*, Distinct viral reservoirs in individuals with spontaneous control of HIV-1. *Nature* **585**, 261-267 (2020).
4. C. Gaebler *et al.*, Combination of quadruplex qPCR and next-generation sequencing for qualitative and quantitative analysis of the HIV-1 latent reservoir. *J Exp Med* **216**, 2253-2264 (2019).
5. E. Cano-Gamez *et al.*, Single-cell transcriptomics identifies an effectorness gradient shaping the response of CD4(+) T cells to cytokines. *Nat Commun* **11**, 1801 (2020).
6. J. Jurka, Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet* **16**, 418-420 (2000).
7. S. S. Rao *et al.*, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680 (2014).
8. M. Lusic, R. F. Siliciano, Nuclear landscape of HIV-1 infection and integration. *Nat Rev Microbiol* **15**, 69-82 (2017).
9. S. Komaki *et al.*, iMETHYL: an integrative database of human DNA methylation, gene expression, and genomic variation. *Hum Genome Var* **5**, 18008 (2018).
10. A. Rambaut, D. Posada, K. A. Crandall, E. C. Holmes, The causes and consequences of HIV evolution. *Nat Rev Genet* **5**, 52-61 (2004).
11. L. B. Cohn *et al.*, Clonal CD4(+) T cells in the HIV-1 latent reservoir display a distinct gene profile upon reactivation. *Nat Med* **24**, 604-609 (2018).
12. A. S. Huang *et al.*, Integration features of intact latent HIV-1 in CD4+ T cell clones contribute to viral persistence. *J Exp Med* **218**, (2021).
13. R. C. Gallo *et al.*, Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science* **220**, 865-867 (1983).
14. F. Barre-Sinoussi *et al.*, Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **220**, 868-871 (1983).
15. W. H. O. (WHO), HIV seropositivity and AIDS prevention and control. (1989).
16. W. H. O. (WHO), 'Global Statistics' in Weekly epidemiological record **62(49):372**, (1987).
17. P. J. Maddon *et al.*, The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell* **47**, 333-348 (1986).
18. J. S. McDougal *et al.*, Binding of HTLV-III/LAV to T4+ T cells by a complex of the 110K viral protein and the T4 molecule. *Science* **231**, 382-385 (1986).
19. C. B. Wilen, J. C. Tilton, R. W. Doms, HIV: cell binding and entry. *Cold Spring Harb Perspect Med* **2**, (2012).
20. Z. Kruize, N. A. Kootstra, The Role of Macrophages in HIV-1 Persistence and Pathogenesis. *Front Microbiol* **10**, 2828 (2019).
21. J. D. Roberts, K. Bebenek, T. A. Kunkel, The accuracy of reverse transcriptase from HIV-1. *Science* **242**, 1171-1173 (1988).

22. B. D. Preston, B. J. Poiesz, L. A. Loeb, Fidelity of HIV-1 reverse transcriptase. *Science* **242**, 1168-1171 (1988).
23. D. Finzi *et al.*, Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science* **278**, 1295-1300 (1997).
24. T. W. Chun *et al.*, Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proc Natl Acad Sci U S A* **94**, 13193-13197 (1997).
25. A. M. Crooks *et al.*, Precise Quantitation of the Latent HIV-1 Reservoir: Implications for Eradication Strategies. *J Infect Dis* **212**, 1361-1365 (2015).
26. N. Bachmann *et al.*, Determinants of HIV-1 reservoir size and long-term dynamics during suppressive ART. *Nat Commun* **10**, 3193 (2019).
27. M. J. Peluso *et al.*, Differential decay of intact and defective proviral DNA in HIV-1-infected individuals on suppressive antiretroviral therapy. *JCI Insight* **5**, (2020).
28. J. D. Siliciano *et al.*, Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nat Med* **9**, 727-728 (2003).
29. D. M. Phillips, V. R. Zacharopoulos, X. Tan, R. Pearce-Pratt, Mechanisms of sexual transmission of HIV: does HIV infect intact epithelia? *Trends Microbiol* **2**, 454-458 (1994).
30. C. f. D. C. a. Prevention. (U.S. Department of Health & Human Services).
31. G. M. Shaw, E. Hunter, HIV transmission. *Cold Spring Harb Perspect Med* **2**, (2012).
32. W. M. El-Sadr, K. H. Mayer, S. L. Hodder, AIDS in America--forgotten but not gone. *N Engl J Med* **362**, 967-970 (2010).
33. E. W. Fiebig *et al.*, Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. *AIDS* **17**, 1871-1879 (2003).
34. J. W. Mellors *et al.*, Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science* **272**, 1167-1170 (1996).
35. K. T. Arrildt, S. B. Joseph, R. Swanstrom, The HIV-1 env protein: a coat of many colors. *Curr HIV/AIDS Rep* **9**, 52-63 (2012).
36. G. D. Tomaras *et al.*, Initial B-cell responses to transmitted human immunodeficiency virus type 1: virion-binding immunoglobulin M (IgM) and IgG antibodies followed by plasma anti-gp41 antibodies with ineffective control of initial viremia. *J Virol* **82**, 12449-12463 (2008).
37. M. Bonsignori *et al.*, Antibody-dependent cellular cytotoxicity-mediating antibodies from an HIV-1 vaccine efficacy trial target multiple epitopes and preferentially use the VH1 gene family. *J Virol* **86**, 11521-11532 (2012).
38. B. F. Haynes *et al.*, Immune-correlates analysis of an HIV-1 vaccine efficacy trial. *N Engl J Med* **366**, 1275-1286 (2012).
39. S. Zolla-Pazner *et al.*, Vaccine-induced IgG antibodies to V1V2 regions of multiple HIV-1 subtypes correlate with decreased risk of HIV-1 infection. *PLoS One* **9**, e87572 (2014).
40. G. D. Tomaras *et al.*, Vaccine-induced plasma IgA specific for the C1 region of the HIV-1 envelope blocks binding and effector function of IgG. *Proc Natl Acad Sci U S A* **110**, 9019-9024 (2013).
41. J. A. Horwitz *et al.*, Non-neutralizing Antibodies Alter the Course of HIV-1 Infection In Vivo. *Cell* **170**, 637-648 e610 (2017).
42. X. Wei *et al.*, Antibody neutralization and escape by HIV-1. *Nature* **422**, 307-312 (2003).

43. D. D. Richman, T. Wrin, S. J. Little, C. J. Petropoulos, Rapid evolution of the neutralizing antibody response to HIV type 1 infection. *Proc Natl Acad Sci U S A* **100**, 4144-4149 (2003).
44. J. F. Scheid *et al.*, Broad diversity of neutralizing antibodies isolated from memory B cells in HIV-infected individuals. *Nature* **458**, 636-640 (2009).
45. S. G. Deeks, B. D. Walker, Human immunodeficiency virus controllers: mechanisms of durable virus control in the absence of antiretroviral therapy. *Immunity* **27**, 406-416 (2007).
46. M. Caskey, F. Klein, M. C. Nussenzweig, Broadly neutralizing anti-HIV-1 monoclonal antibodies in the clinic. *Nat Med* **25**, 547-553 (2019).
47. B. F. Haynes, D. R. Burton, J. R. Mascola, Multiple roles for HIV broadly neutralizing antibodies. *Sci Transl Med* **11**, (2019).
48. Y. Liu, W. Cao, M. Sun, T. Li, Broadly neutralizing antibodies for HIV-1: efficacies, challenges and opportunities. *Emerg Microbes Infect* **9**, 194-206 (2020).
49. A. Escolano *et al.*, Sequential Immunization Elicits Broadly Neutralizing Anti-HIV-1 Antibodies in Ig Knockin Mice. *Cell* **166**, 1445-1458 e1412 (2016).
50. P. Dosenovic *et al.*, Immunization for HIV-1 Broadly Neutralizing Antibodies in Human Ig Knockin Mice. *Cell* **161**, 1505-1515 (2015).
51. D. C. Malherbe *et al.*, Sequential immunization with a subtype B HIV-1 envelope quasispecies partially mimics the in vivo development of neutralizing antibodies. *J Virol* **85**, 5262-5274 (2011).
52. K. Murphy, Weaver, C., Janeway, C. , *Janeway's Immunobiology* (Garland Science, ed. 9th, 2017).
53. J. C. Gea-Banacloche *et al.*, Maintenance of large numbers of virus-specific CD8+ T cells in HIV-infected progressors and long-term nonprogressors. *J Immunol* **165**, 1082-1092 (2000).
54. M. R. Betts *et al.*, Analysis of total human immunodeficiency virus (HIV)-specific CD4(+) and CD8(+) T-cell responses: relationship to viral load in untreated HIV infection. *J Virol* **75**, 11983-11991 (2001).
55. D. R. Collins, G. D. Gaiha, B. D. Walker, CD8(+) T cells in HIV control, cure and prevention. *Nat Rev Immunol* **20**, 471-482 (2020).
56. P. Borrow *et al.*, Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. *Nat Med* **3**, 205-211 (1997).
57. C. Fenwick *et al.*, T-cell exhaustion in HIV infection. *Immunol Rev* **292**, 149-163 (2019).
58. V. Appay *et al.*, HIV-specific CD8(+) T cells produce antiviral cytokines but are impaired in cytolytic function. *J Exp Med* **192**, 63-75 (2000).
59. M. Lichterfeld *et al.*, Loss of HIV-1-specific CD8+ T cell proliferation after acute HIV-1 infection and restoration by vaccine-induced HIV-1-specific CD4+ T cells. *J Exp Med* **200**, 701-712 (2004).
60. B. Walker, A. McMichael, The T-cell response to HIV. *Cold Spring Harb Perspect Med* **2**, (2012).
61. E. S. Rosenberg *et al.*, Vigorous HIV-1-specific CD4+ T cell responses associated with control of viremia. *Science* **278**, 1447-1450 (1997).

62. D. E. Kaufmann *et al.*, Comprehensive analysis of human immunodeficiency virus type 1-specific CD4 responses reveals marked immunodominance of gag and nef and the presence of broadly recognized peptides. *J Virol* **78**, 4463-4477 (2004).
63. A. A. Okoye, L. J. Picker, CD4(+) T-cell depletion in HIV infection: mechanisms of immunological failure. *Immunol Rev* **254**, 54-64 (2013).
64. D. E. Kaufmann *et al.*, Upregulation of CTLA-4 by HIV-specific CD4+ T cells correlates with disease progression and defines a reversible immune dysfunction. *Nat Immunol* **8**, 1246-1254 (2007).
65. A. Morou *et al.*, Altered differentiation is central to HIV-specific CD4(+) T cell dysfunction in progressive disease. *Nat Immunol* **20**, 1059-1070 (2019).
66. T. J. Henrich *et al.*, HIV-1 persistence following extremely early initiation of antiretroviral therapy (ART) during acute HIV-1 infection: An observational study. *PLoS Med* **14**, e1002417 (2017).
67. D. J. Colby *et al.*, Rapid HIV RNA rebound after antiretroviral treatment interruption in persons durably suppressed in Fiebig I acute HIV infection. *Nat Med* **24**, 923-926 (2018).
68. J. B. Whitney *et al.*, Rapid seeding of the viral reservoir prior to SIV viraemia in rhesus monkeys. *Nature* **512**, 74-77 (2014).
69. S. Eriksson *et al.*, Comparative analysis of measures of viral reservoirs in HIV-1 eradication studies. *PLoS Pathog* **9**, e1003174 (2013).
70. Y. C. Ho *et al.*, Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* **155**, 540-551 (2013).
71. R. Fromentin *et al.*, CD4+ T Cells Expressing PD-1, TIGIT and LAG-3 Contribute to HIV Persistence during ART. *PLoS Pathog* **12**, e1005761 (2016).
72. J. D. Siliciano, R. F. Siliciano, Enhanced culture assay for detection and quantitation of latently infected, resting CD4+ T-cells carrying replication-competent virus in HIV-1-infected individuals. *Methods Mol Biol* **304**, 3-15 (2005).
73. N. N. Hosmane *et al.*, Proliferation of latently infected CD4(+) T cells carrying replication-competent HIV-1: Potential role in latent reservoir dynamics. *J Exp Med* **214**, 959-972 (2017).
74. F. A. Procopio *et al.*, A Novel Assay to Measure the Magnitude of the Inducible Viral Reservoir in HIV-infected Individuals. *EBioMedicine* **2**, 874-883 (2015).
75. S. D. Falcinelli, C. Ceriani, D. M. Margolis, N. M. Archin, New Frontiers in Measuring and Characterizing the HIV Reservoir. *Front Microbiol* **10**, 2878 (2019).
76. K. M. Bruner *et al.*, A quantitative approach for measuring the reservoir of latent HIV-1 proviruses. *Nature* **566**, 120-125 (2019).
77. R. F. Siliciano, W. C. Greene, HIV latency. *Cold Spring Harb Perspect Med* **1**, a007096 (2011).
78. L. B. Cohn *et al.*, HIV-1 integration landscape during latent and active infection. *Cell* **160**, 420-432 (2015).
79. G. Q. Lee *et al.*, Clonal expansion of genome-intact HIV-1 in functionally polarized Th1 CD4+ T cells. *J Clin Invest* **127**, 2689-2696 (2017).
80. J. K. Bui *et al.*, Proviruses with identical sequences comprise a large fraction of the replication-competent HIV reservoir. *PLoS Pathog* **13**, e1006283 (2017).
81. A. A. Antar *et al.*, Longitudinal study reveals HIV-1-infected CD4+ T cell dynamics during long-term antiretroviral therapy. *J Clin Invest* **130**, 3543-3559 (2020).

82. F. R. Simonetti *et al.*, Antigen-driven clonal selection shapes the persistence of HIV-1-infected CD4+ T cells in vivo. *J Clin Invest* **131**, (2021).
83. P. Mendoza *et al.*, Antigen-responsive CD4+ T cell clones contribute to the HIV-1 latent reservoir. *J Exp Med* **217**, (2020).
84. T. A. Wagner *et al.*, HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science* **345**, 570-573 (2014).
85. F. Maldarelli *et al.*, HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science* **345**, 179-183 (2014).
86. G. J. Bedwell, S. Jang, W. Li, P. K. Singh, A. N. Engelman, rigrag: high-resolution mapping of genic targeting preferences during HIV-1 integration in vitro and in vivo. *Nucleic Acids Res*, (2021).
87. K. M. Bruner *et al.*, Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nat Med* **22**, 1043-1049 (2016).
88. E. A. Boritz *et al.*, Multiple Origins of Virus Persistence during Natural Control of HIV Infection. *Cell* **166**, 1004-1015 (2016).
89. R. T. Veenhuis *et al.*, Long-term remission despite clonal expansion of replication-competent HIV-1 isolates. *JCI Insight* **3**, (2018).
90. B. A. Woldemeskel, A. K. Kwaa, J. N. Blankson, Viral reservoirs in elite controllers of HIV-1 infection: Implications for HIV cure strategies. *EBioMedicine* **62**, 103118 (2020).
91. K. K. A. Van Rompay, Tackling HIV and AIDS: contributions by non-human primate models. *Lab Anim (NY)* **46**, 259-270 (2017).
92. M. C. Gauduin *et al.*, Passive immunization with a human monoclonal antibody protects hu-PBL-SCID mice against challenge by primary isolates of HIV-1. *Nat Med* **3**, 1389-1393 (1997).
93. J. T. Safrit *et al.*, hu-PBL-SCID mice can be protected from HIV-1 infection by passive transfer of monoclonal antibody to the principal neutralizing determinant of envelope gp120. *AIDS* **7**, 15-21 (1993).
94. T. Hatziioannou, D. T. Evans, Animal models for HIV/AIDS research. *Nat Rev Microbiol* **10**, 852-867 (2012).
95. E. J. Duh, W. J. Maury, T. M. Folks, A. S. Fauci, A. B. Rabson, Tumor necrosis factor alpha activates human immunodeficiency virus type 1 through induction of nuclear factor binding to the NF-kappa B sites in the long terminal repeat. *Proc Natl Acad Sci U S A* **86**, 5974-5978 (1989).
96. G. Nabel, D. Baltimore, An inducible transcription factor activates expression of human immunodeficiency virus in T cells. *Nature* **326**, 711-713 (1987).
97. F. D. Bushman, T. Fujiwara, R. Craigie, Retroviral DNA integration directed by HIV integration protein in vitro. *Science* **249**, 1555-1558 (1990).
98. H. Liu *et al.*, Identification of celastrol as a novel HIV-1 latency reversal agent by an image-based screen. *PLoS One* **16**, e0244771 (2021).
99. J. Kulkosky *et al.*, Prostratin: activation of latent HIV-1 expression suggests a potential inductive adjuvant therapy for HAART. *Blood* **98**, 3006-3015 (2001).
100. K. Fujinaga, D. C. Cary, Experimental Systems for Measuring HIV Latency and Reactivation. *Viruses* **12**, (2020).
101. M. Kim *et al.*, A primary CD4(+) T cell model of HIV-1 latency established after activation through the T cell receptor and subsequent return to quiescence. *Nat Protoc* **9**, 2755-2770 (2014).

102. A. Bosque, V. Planelles, Induction of HIV-1 latency and reactivation in primary memory CD4+ T cells. *Blood* **113**, 58-65 (2009).
103. S. Saleh *et al.*, 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* **110**, 4161-4164 (2007).
104. V. L. Perez *et al.*, An HIV-1-infected T cell clone defective in IL-2 production and Ca2+ mobilization after CD3 stimulation. *J Immunol* **147**, 3145-3148 (1991).
105. C. Marchand, A. A. Johnson, E. Semenova, Y. Pommier, Mechanisms and inhibition of HIV integration. *Drug Discov Today Dis Mech* **3**, 253-260 (2006).
106. E. M. Poeschla, Integrase, LEDGF/p75 and HIV replication. *Cell Mol Life Sci* **65**, 1403-1424 (2008).
107. A. N. Engelman, P. K. Singh, Cellular and molecular mechanisms of HIV-1 integration targeting. *Cell Mol Life Sci* **75**, 2491-2507 (2018).
108. X. Wu, Y. Li, B. Crise, S. M. Burgess, Transcription start regions in the human genome are favored targets for MLV integration. *Science* **300**, 1749-1751 (2003).
109. R. S. Mitchell *et al.*, Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol* **2**, E234 (2004).
110. D. Derse *et al.*, Human T-cell leukemia virus type 1 integration target sites in the human genome: comparison with those of other retroviruses. *J Virol* **81**, 6731-6741 (2007).
111. B. Crise *et al.*, Simian immunodeficiency virus integration preference is similar to that of human immunodeficiency virus type 1. *J Virol* **79**, 12199-12204 (2005).
112. Y. Han *et al.*, Resting CD4+ T cells from human immunodeficiency virus type 1 (HIV-1)-infected individuals carry integrated HIV-1 genomes within actively transcribed host genes. *J Virol* **78**, 6122-6133 (2004).
113. S. D. Barr *et al.*, HIV integration site selection: targeting in macrophages and the effects of different routes of viral entry. *Mol Ther* **14**, 218-225 (2006).
114. M. K. Lewinski *et al.*, Retroviral DNA integration: viral and cellular determinants of target-site selection. *PLoS Pathog* **2**, e60 (2006).
115. S. Sherrill-Mix *et al.*, HIV latency and integration site placement in five cell-based models. *Retrovirology* **10**, 90 (2013).
116. A. R. Schroder *et al.*, HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* **110**, 521-529 (2002).
117. R. Craigie, F. D. Bushman, HIV DNA integration. *Cold Spring Harb Perspect Med* **2**, a006890 (2012).
118. A. Albanese, D. Arosio, M. Terreni, A. Cereseto, HIV-1 pre-integration complexes selectively target decondensed chromatin in the nuclear periphery. *PLoS One* **3**, e2413 (2008).
119. B. Marini *et al.*, Nuclear architecture dictates HIV-1 integration site selection. *Nature* **521**, 227-231 (2015).
120. S. E. Kauder, A. Bosque, A. Lindqvist, V. Planelles, E. Verdin, Epigenetic regulation of HIV-1 latency by cytosine methylation. *PLoS Pathog* **5**, e1000495 (2009).
121. T. Lenasi, X. Contreras, B. M. Peterlin, Transcriptional interference antagonizes proviral gene expression to promote HIV latency. *Cell Host Microbe* **4**, 123-133 (2008).
122. J. I. Mullins, L. M. Frenkel, Clonal Expansion of Human Immunodeficiency Virus-Infected Cells and Human Immunodeficiency Virus Persistence During Antiretroviral Therapy. *J Infect Dis* **215**, S119-S127 (2017).

123. F. R. Simonetti *et al.*, Clonally expanded CD4+ T cells can produce infectious HIV-1 in vivo. *Proc Natl Acad Sci U S A* **113**, 1883-1888 (2016).
124. D. Jung, F. W. Alt, Unraveling V(D)J recombination; insights into gene regulation. *Cell* **116**, 299-311 (2004).
125. P. Barennes *et al.*, Benchmarking of T cell receptor repertoire profiling methods reveals large systematic biases. *Nat Biotechnol* **39**, 236-245 (2021).
126. T. M. Folks *et al.*, Tumor necrosis factor alpha induces expression of human immunodeficiency virus in a chronically infected T-cell clone. *Proc Natl Acad Sci U S A* **86**, 2365-2368 (1989).
127. K. A. Clouse *et al.*, Monokine regulation of human immunodeficiency virus-1 expression in a chronically infected human T cell clone. *J Immunol* **142**, 431-438 (1989).
128. S. T. Butera, V. L. Perez, B. Y. Wu, G. J. Nabel, T. M. Folks, Oscillation of the human immunodeficiency virus surface receptor is regulated by the state of viral activation in a CD4+ cell model of chronic infection. *J Virol* **65**, 4645-4653 (1991).
129. S. T. Butera *et al.*, Extrachromosomal human immunodeficiency virus type-1 DNA can initiate a spreading infection of HL-60 cells. *J Cell Biochem* **45**, 366-373 (1991).
130. A. Ciuffi, S. D. Barr, Identification of HIV integration sites in infected host genomic DNA. *Methods* **53**, 39-46 (2011).
131. X. Cai, Y. H. Chiu, Z. J. Chen, The cGAS-cGAMP-STING pathway of cytosolic DNA sensing and signaling. *Mol Cell* **54**, 289-296 (2014).
132. A. Subramanian *et al.*, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).
133. C. Gene Ontology, The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res* **49**, D325-D334 (2021).
134. M. Ashburner *et al.*, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29 (2000).
135. S. Hakre, L. Chavez, K. Shirakawa, E. Verdin, Epigenetic regulation of HIV latency. *Curr Opin HIV AIDS* **6**, 19-24 (2011).
136. K. S. Keedy *et al.*, A limited group of class I histone deacetylases acts to repress human immunodeficiency virus type 1 expression. *J Virol* **83**, 4749-4756 (2009).
137. S. A. Williams *et al.*, NF-kappaB p50 promotes HIV latency through HDAC recruitment and repression of transcriptional initiation. *EMBO J* **25**, 139-149 (2006).
138. R. Verdikt, O. Hernalsteens, C. Van Lint, Epigenetic Mechanisms of HIV-1 Persistence. *Vaccines (Basel)* **9**, (2021).
139. S. A. Denslow, P. A. Wade, The human Mi-2/NuRD complex and gene regulation. *Oncogene* **26**, 5433-5438 (2007).
140. J. P. Hsin, J. L. Manley, The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev* **26**, 2119-2137 (2012).
141. R. Liu *et al.*, Single-cell transcriptional landscapes reveal HIV-1-driven aberrant host gene transcription as a potential therapeutic target. *Sci Transl Med* **12**, (2020).
142. S. C. Patro *et al.*, Combined HIV-1 sequence and integration site analysis informs viral dynamics and allows reconstruction of replicating viral ancestors. *Proc Natl Acad Sci U S A* **116**, 25891-25899 (2019).
143. E. K. Halvas *et al.*, HIV-1 viremia not suppressible by antiretroviral therapy can originate from large T cell clones producing infectious virus. *J Clin Invest* **130**, 5847-5857 (2020).

144. C. Gaebler *et al.*, Sequence Evaluation and Comparative Analysis of Novel Assays for Intact Proviral HIV-1 DNA. *J Virol* **95**, (2021).
145. P. Mendoza *et al.*, Combination therapy with anti-HIV-1 antibodies maintains viral suppression. *Nature* **561**, 479-484 (2018).
146. S. B. Laskey, C. W. Pohlmeyer, K. M. Bruner, R. F. Siliciano, Evaluating Clonal Expansion of HIV-Infected Cells: Optimization of PCR Strategies to Predict Clonality. *PLoS Pathog* **12**, e1005689 (2016).
147. G. Vansant *et al.*, The chromatin landscape at the HIV-1 provirus integration site determines viral expression. *Nucleic Acids Res* **48**, 7801-7817 (2020).
148. A. J. Murray, K. J. Kwon, D. L. Farber, R. F. Siliciano, The Latent Reservoir for HIV-1: How Immunologic Memory and Clonal Expansion Contribute to HIV-1 Persistence. *J Immunol* **197**, 407-417 (2016).
149. R. Fromentin, N. Chomont, HIV persistence in subsets of CD4+ T cells: 50 shades of reservoirs. *Semin Immunol*, 101438 (2020).
150. P. Gantner *et al.*, Single-cell TCR sequencing reveals phenotypically diverse clonally expanded cells harboring inducible HIV proviruses during ART. *Nat Commun* **11**, 4089 (2020).
151. R. Liu, F. R. Simonetti, Y. C. Ho, The forces driving clonal expansion of the HIV-1 latent reservoir. *Virol J* **17**, 4 (2020).
152. L. B. Cohn, N. Chomont, S. G. Deeks, The Biology of the HIV-1 Latent Reservoir and Implications for Cure Strategies. *Cell Host Microbe* **27**, 519-530 (2020).
153. G. M. Laird *et al.*, Rapid quantification of the latent reservoir for HIV-1 using a viral outgrowth assay. *PLoS Pathog* **9**, e1003398 (2013).
154. M. R. Pinzone *et al.*, Longitudinal HIV sequencing reveals reservoir expression leading to decay which is obscured by clonal expansion. *Nat Commun* **10**, 728 (2019).
155. D. B. Reeves *et al.*, A majority of HIV persistence during antiretroviral therapy is due to infected cell proliferation. *Nat Commun* **9**, 4811 (2018).
156. Z. Wang *et al.*, Expanded cellular clones carrying replication-competent HIV-1 persist, wax, and wane. *Proc Natl Acad Sci U S A* **115**, E2575-E2584 (2018).
157. M. K. Lewinski *et al.*, Genome-wide analysis of chromosomal features repressing human immunodeficiency virus transcription. *J Virol* **79**, 6610-6619 (2005).
158. E. M. Anderson, F. Maldarelli, The role of integration and clonal expansion in HIV infection: live long and prosper. *Retrovirology* **15**, 71 (2018).
159. H. C. Chen, J. P. Martinez, E. Zorita, A. Meyerhans, G. J. Filion, Position effects influence HIV latency reversal. *Nat Struct Mol Biol* **24**, 47-54 (2017).
160. L. N. Bertagnolli *et al.*, Autologous IgG antibodies block outgrowth of a substantial but variable fraction of viruses in the latent reservoir for HIV-1. *Proc Natl Acad Sci U S A* **117**, 32066-32077 (2020).
161. K. J. Bar *et al.*, Effect of HIV Antibody VRC01 on Viral Rebound after Treatment Interruption. *N Engl J Med* **375**, 2037-2050 (2016).
162. Y. Z. Cohen *et al.*, Relationship between latent and rebound viruses in a clinical trial of anti-HIV-1 antibody 3BNC117. *J Exp Med* **215**, 2311-2324 (2018).
163. C. L. Lu *et al.*, Relationship between intact HIV-1 proviruses in circulating CD4(+) T cells and rebound viruses emerging during treatment interruption. *Proc Natl Acad Sci U S A* **115**, E11341-E11348 (2018).

164. Y. J. Yeh, Y. C. Ho, Shock-and-kill versus block-and-lock: Targeting the fluctuating and heterogeneous HIV-1 gene expression. *Proc Natl Acad Sci U S A* **118**, (2021).
165. S. G. Deeks, HIV: Shock and kill. *Nature* **487**, 439-440 (2012).
166. G. Vansant, A. Bruggemans, J. Janssens, Z. Debyser, Block-And-Lock Strategies to Cure HIV Infection. *Viruses* **12**, (2020).
167. L. Shan *et al.*, Stimulation of HIV-1-specific cytolytic T lymphocytes facilitates elimination of latent viral reservoir after virus reactivation. *Immunity* **36**, 491-501 (2012).
168. Y. Kim, J. L. Anderson, S. R. Lewin, Getting the "Kill" into "Shock and Kill": Strategies to Eliminate Latent HIV. *Cell Host Microbe* **23**, 14-26 (2018).
169. R. B. Jones, B. D. Walker, HIV-specific CD8(+) T cells and HIV eradication. *J Clin Invest* **126**, 455-463 (2016).
170. K. Deng *et al.*, Broad CTL response is required to clear latent HIV-1 due to dominance of escape mutations. *Nature* **517**, 381-385 (2015).
171. S. H. Huang *et al.*, Latent HIV reservoirs exhibit inherent resistance to elimination by CD8+ T cells. *J Clin Invest* **128**, 876-889 (2018).
172. A. Han, J. Glanville, L. Hansmann, M. M. Davis, Linking T-cell receptor sequence to functional phenotype at the single-cell level. *Nat Biotechnol* **32**, 684-692 (2014).
173. J. J. Trombetta *et al.*, Preparation of Single-Cell RNA-Seq Libraries for Next Generation Sequencing. *Curr Protoc Mol Biol* **107**, 4 22 21-17 (2014).
174. S. Islam *et al.*, Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat Methods* **11**, 163-166 (2014).
175. A. Yamane *et al.*, RPA accumulation during class switch recombination represents 5'-3' DNA-end resection during the S-G2/M phase of the cell cycle. *Cell Rep* **3**, 138-147 (2013).
176. I. A. Klein *et al.*, Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. *Cell* **147**, 95-106 (2011).
177. J. Koster, S. Rahmann, Snakemake-a scalable bioinformatics workflow engine. *Bioinformatics* **34**, 3600 (2018).
178. A. Paruzynski *et al.*, Genome-wide high-throughput integrome analyses by nrLAM-PCR and next-generation sequencing. *Nat Protoc* **5**, 1379-1395 (2010).
179. B. J. Schmiedel *et al.*, Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell* **175**, 1701-1715 e1716 (2018).
180. L. Scrucca, M. Fop, T. B. Murphy, A. E. Raftery, mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *R J* **8**, 289-317 (2016).
181. C. Roadmap Epigenomics *et al.*, Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330 (2015).