Proliferation-based Enrichment of HIV-infected Cells Reveals Distinct Clonally Expanded Cell Populations

Lizette Carrasco

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2021

Committee:

Lisa Frenkel

Florian Hladik

Martin Prlic

Program Authorized to Offer Degree:

Global Health

University of Washington

Abstract

Proliferation-based Enrichment of HIV-infected Cells Reveals Distinct Clonally Expanded Cell Populations

Lizette Carrasco

Chair of the Supervisory Committee:

Lisa Frenkel

Department of Global Health

Antiretroviral therapy (ART) effectively controls HIV-1 disease and prevents transmission through suppression of viral replication to clinically undetectable levels. Yet, ART is not curative due to integration of the virus into the host genome which is maintained as a long-lived reservoir of HIV-infected cells. HIV-1 infected cells can undergo clonal expansion over time on ART and this clonal expansion is thought to be driven by three non-mutually exclusive mechanisms: antigen-driven proliferation, homeostatic proliferation, and integration site (IS)-driven proliferation. By developing and using methods that support *ex vivo* homeostatic proliferation of cells from HIV-infected donors, I was able to enrich for infected cell populations with a proliferative advantage. I compared approximately 720 HIV IS from CD4+ T cells from individuals on ART to approximately 106 IS from *ex vivo* expanded CD4+ T cells from the same participants to determine if HIV IS might contribute to survival advantage by altering gene expression of the integrated gene. I demonstrate that distinct populations of clonal cells emerge as a

consequence of proliferation-based enrichment indicating a preferred fitness of these cells to expand in homeostatic proliferation conditions. I demonstrate that HIV-host chimeric transcripts are formed between the HIV LTR and the major clone, *CPNE1.* I discovered that HIV integration into the *CPNE1* gene disrupts normal cellular splicing pathways and results in a reading frame shift and intron retention of the *CPNE1* gene. Intron retention may lead to altered gene expression, thus representing a novel mechanism of HIV-driven dysregulation of host gene expression.

Table of Contents

List of Figures

List of Tables

Acknowledgements

This project was only made possible by the constant support and guidance of Dr. Lisa Frenkel, my dissertation advisor. Thank you for always providing invaluable advice, guidance, and career support. I will always value your unique insights as an HIV/AIDS physician-scientist as they provided real-world context and understanding beyond the lab bench. A special thank you also goes to Dr. Ana Gervassi and Samantha Hardy. It is so powerful and inspirational to have strong female mentors and collaborators and I was incredibly lucky to have that support during my graduate tenure.

It truly takes a village as they say – additional essential support was provided by the following individuals and groups:

- *The study participants.* Thank you for your contribution to HIV research. I always worked with your best interest in mind and hope one day soon we will have a cure.
- My graduate committee: Lisa Frenkel (CGIDR), Florian Hladik (Fred Hutch), Martin Prlic (Fred Hutch). Thank you all for your ideas, advice, and encouragement.
- My extended Frenkel Lab group, past and present. Thank you for providing a fun, collaborative and comfortable work environment. Thank you for always being there for coffee breaks, brainstorming chats, and venting sessions.
- The extended Mullins Lab group (UW). A special thanks to Jim Mullins, Michael Dapp and Lennie Chen for entertaining all of my questions and for sharing time, reagents and ideas.
- The Edlefsen Group (Fred Hutch). Paul Edlefsen, Ewelina Kośmider, and Danica Shao - thank you for assisting with the statistical analyses and gene set enrichment analyses for this work.
- The extended Pathobiology student group and faculty members. The camaraderie and quality of education provided by this program are truly unmatched.

Finally, I want to express my endless love and gratitude for my entire family. To my parents - throughout our lives you've worked hard manual-labor jobs so Dany and I could have the most "successful" life by whatever metric we used to define it. You are both the truest embodiment of the American Dream and I am endlessly grateful for your support and love.

Chapter 1. Introduction

1.1 Global epidemiology of HIV-1

Human immunodeficiency virus type 1 (HIV-1), the causative agent of acquired immunodeficiency syndrome (AIDS), continues to be one of the world's most critical public health challenges. Since the start of the HIV epidemic in the 1980s, over 33 million people have died from HIV/AIDS-related illnesses globally (WHO, 2009). Currently, there are an estimated 38 million people living with HIV worldwide (UNAIDS, 2020b) and this number is estimated to increase to over 42 million individuals by 2030 (Dybul et al. 2020). Decades of aggressive public health campaigns to reduce HIV acquisition, research to develop effective antiretroviral and effective treatment regimens, and global treatment- and implementation-programs have accelerated progress towards ending the HIV epidemic (UNAIDS, 2020a). However, there is still not an intervention that will reliably cure an individual from HIV infection, signaling the need for progress towards novel and innovative therapeutic strategies. Adding to this challenge, the ongoing COVID-19 pandemic has veritably reversed gains made in the global HIV response through disruptions to HIV care, testing and medical supply chains (The Lancet, 2020; UNAIDS, 2020a)

The greatest burden of HIV disease is disproportionately shouldered by lower- and middle-income countries (LMICs) in terms of HIV prevalence, highest degrees of viral diversity, and economic and social inequalities that exacerbate the HIV/AIDS epidemic (**Figure 1.1**) (Shao and Williamson, 2012; UNAIDS, 2019). This is evident in the case of sub-Saharan Africa which accounts for 57% of new HIV infections and 68% of the people living with HIV despite accommodating only 15% of the world's population (Dybul et al.

2020). Additionally, the global spread and rapid evolution of HIV has resulted in a divergent epidemic wherein multiple distinct subtypes of HIV circulate, with the ability to recombine, and form sub-epidemics (Taylor et al. 2008). Adolescent girls, sex workers, transgender people, people who inject drugs, and men who have sex with men are disproportionately affected by HIV, highlighting the deep social, cultural, and political barriers to HIV/AIDS care and prevention (UNAIDS, 2019, 2020a).

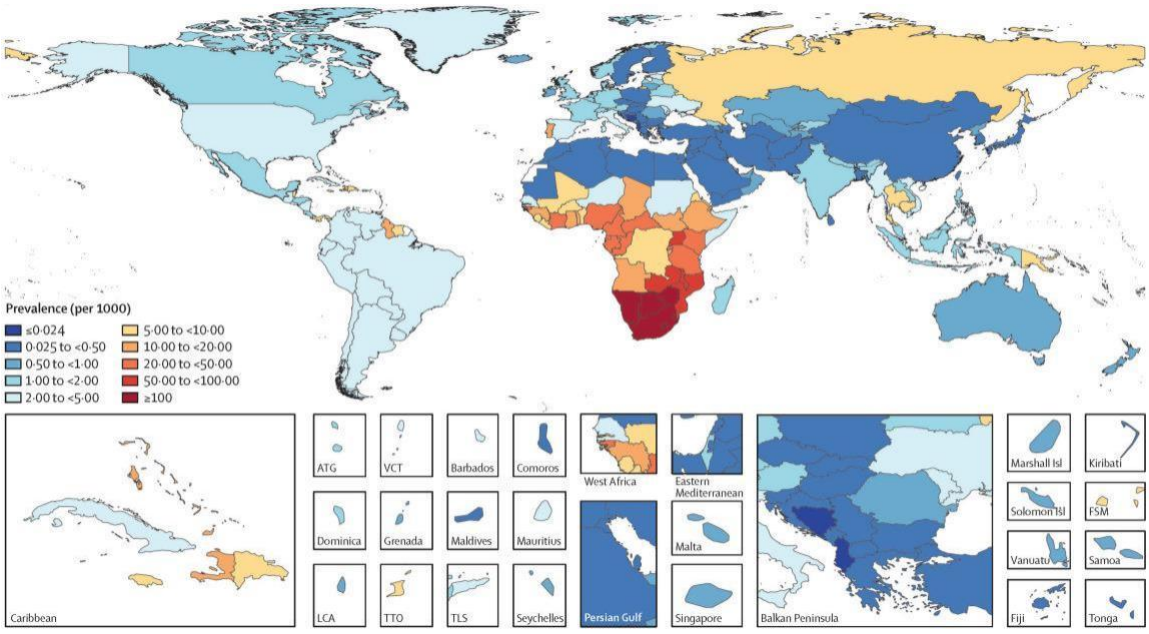


Figure 1.1 Global prevalence of HIV-1. Data from the 2017 Global Burden of Diseases, Injuries, and Risk Factors Study (Frank et al. 2019). Global map of age-standardized HIV prevalence for both sexes. Data range is from 1980-2017 for male and female sexes.

1.2 HIV-1 life cycle

HIV-1 is an enveloped, single-stranded, positive sense RNA virus in the *Retroviridae* family. Humans are the natural hosts for HIV-1 and virus can be detected in a variety of human tissues and cell types with the primary cellular target being CD4+ T-lymphocytes and to a lesser degree, cells from the monocyte/macrophage lineage (Ho, Rota, and Hirsch 1986; Klatzmann et al. 1984).

The HIV-1 life cycle begins with binding of the surface-expressed HIV-1 envelope glycoproteins (gp120 and gp41) to the primary cellular receptor CD4 and a cellular coreceptor, CCR5 or CXCR4 (Chan and Kim 1998; Wilen, Tilton, and Doms 2012). Upon fusion of the viral and host cell membranes, the viral core is delivered into the cytoplasm of the target cell where the viral RNA genome is transcribed into DNA by the viral reverse transcriptase protein (Lusic and Siliciano 2017). The reverse transcription process is extremely error-prone and results in a diverse pool of genetically intact and defective viruses (Pulsinelli and Temin 1994). Co-localization of the viral DNA genome with viral integrase and capsid proteins and a number of cellular proteins form what is known as the pre-integration complex (PIC) (Lusic and Siliciano 2017).

Retroviruses like HIV-1 critically depend on integration of the viral DNA into the host genome to perpetuate the retroviral replication cycle. Thus, to enter the nucleus, viral proteins associated with the PIC interact with host nucleoporin proteins to dock and import the virion through the nuclear pore complex and into the nucleus (Lusic and Siliciano 2017). Integration of the viral genome into a host chromosome depends on the enzymatic activity of HIV-1 integrase and can take place at many locations in the host genome (Craigie and Bushman 2012). Factors influencing sites of viral integration in the host

genome are complex and well-studied; these factors are explored in more detail in the section below, “Integration site selection”. Once integrated, the integrated form of HIV-1, known as the provirus, is replicated alongside cellular DNA for the lifetime of the cell during cycles of cellular division (Craigie and Bushman 2012).

In productively infected cells, the viral genome is actively transcribed followed by splicing, nuclear export of viral RNA, and production and assembly of new virus particles, generally with cell lysis. In a small number of cells, the proviral genome undergoes transcriptional silencing due to a multitude of factors including, but not limited to, integration into a transcriptionally repressive chromatin setting. The transcriptional silencing of proviral genomes leads to the generation of the latent viral reservoir which represents the principal barrier to HIV-1 cure.

1.3 Treatment and vaccines against HIV

Currently there is neither a cure nor an effective vaccine available against HIV/AIDS. The development of an HIV vaccine presents unique challenges, owing in part to the lack of natural immunity to HIV and the variability of HIV subtypes. Induction of neutralizing antibodies against HIV using recombinant Env protein-based vaccines has been explored, but so far, attempts have been unsuccessful at preventing HIV-1 infection or delaying disease progression (Flynn et al. 2005; Pitisuttithum et al. 2006).

While induction of sterilizing humoral immunity against pathogens is historically highly correlated with protection by vaccination (Plotkin 2001), a vaccine that stimulates HIV-specific CD8+ cytotoxic T lymphocyte (CTL) responses is a possible alternative to vaccines that target B cell responses. During natural infection with HIV-1, strong CTL responses are associated with viremic control (Borrow et al. 1994; Koup et al. 1994). However, in most individuals, this immune response is incapable of eliminating the viral reservoir (Collins, Gaiha, and Walker 2020) due at least partially to: (1) reduced or absent viral antigen expression in quiescent latent cells, (2) viral escape from CD8+ T cell recognition (Deng et al. 2015), (3) CD8+ T cell dysfunction (Day et al. 2006; Kaufmann et al. 2007) and (4) compartmentalization of the viral reservoir in immune-privileged tissues (Chaillon et al. 2020; Ganor et al. 2019; Wong and Yukl 2016). In a small subset of infected individuals known as elite controllers, who can suppress viral replication without ART, there is a strong association between viremic control and amino acid polymorphisms within the major histocompatibility complex (HLA) class I peptide binding groove, which may favor optimal presentation of critical viral peptides to HIV-specific CTLs (The International HIV Controllers Study 2010). Therefore, it was thought that an

epitope-based vaccine would stimulate CD8+ T cell responses against immunodominant epitopes and induce broad, novel and protective T cell responses. However, human clinical trials testing this approach have yielded disappointing results; in the STEP trial, a vaccine based on use of recombinant adenovirus type 5 vector expressing HIV-1 Gag, Pol, and Nef proteins, vaccine treatment did not prevent HIV infection or reduce viral load (Buchbinder et al. 2008; McElrath et al. 2008).

A weak signal of an effective T-cell targeting vaccine emerged after the RV144 clinical trial showed, for the first time, efficacy against HIV-1 infection (31.2% efficacy in the modified intention-to-treat analysis) using a prime-boost (Gag, Pol, Env prime; gp120 boost) vaccination strategy (Rerks-Ngarm et al. 2009). Although these results showed only a modest benefit in preventing HIV-1, they offered valuable insights into the immune correlates of protection against HIV (Haynes et al. 2012).

HIV infection can be effectively treated with combinations of antiretroviral therapy (ART) drugs. ART consists of four broad classes of drugs that work by blocking various stages of the HIV-1 life cycle including viral entry (entry inhibitors), reverse transcription (non-nucleoside and nucleoside reverse transcriptase inhibitors), viral integration (integrase inhibitors), and viral protease activity required for making processed HIV proteins (protease inhibitors) (Gandhi and Gandhi 2014). With proper adherence, ART can effectively reduce plasma HIV RNA loads to clinically undetectable levels resulting in improved survival rates and complete prevention of transmission between serodiscordant couples (Cohen et al. 2011). However, interruption of suppressive therapy almost inevitably results in viral rebound which originates from the viral reservoir that persists during treatment. Thus, lifelong ART adherence is required.

1.4 Establishment and maintenance of the persistent HIV reservoir

CD4+ T lymphocytes are the primary cellular target of HIV infection and are recognized as the predominant cellular reservoir in HIV-infected individuals (Chun, Stuyver, et al. 1997; Chun et al. 1998). Macrophages and monocytes express low levels of CD4 and CCR5 and thus, are susceptible to infection by R5-tropic HIV viruses, but these cell types are not thought to be the predominant contributors to the persistent reservoir (Koenig et al. 1986; Kruize and Kootstra 2019; Sengupta and Siliciano 2018).

The HIV reservoir is heterogeneous, consisting of various CD4+ T cell subsets including central memory (T_{cm}), effector memory (T_{em}), naïve (T_n), and stem cell-like memory (T_{scm}) T cells (Brenchley et al. 2004; Kulpa and Chomont 2015), with resting memory CD4+ T cells harboring most of the persistent HIV reservoir (Chomont et al. 2009; Chun, Carruth, et al. 1997). Activated CD4+ T cells, most likely those specific for HIV antigens (Douek et al. 2002), are the preferential targets for new HIV infection. This is due in part to increased expression of the HIV entry coreceptor, CCR5, in activated T cells (Joag et al. 2016; Sengupta and Siliciano 2018). However, most newly infected activated T cells are rapidly eliminated by direct viral cytopathic effects or elimination of infected cells by CTLs (McCune 2001), leaving the precise mechanisms driving the establishment of the persistent reservoir undefined.

An important clarification to make is that the pool of HIV-1 proviruses detected in an HIV-infected individual on ART can be defective or intact (more likely to be replication-competent), with defective proviruses comprising greater than 95% of proviruses in the peripheral blood (Ho et al. 2013; Imamichi et al. 2020). Most HIV-1 proviruses are defective owing to the error-prone reverse transcription process and APOBEC-mediated

G-to-A hypermutations that render these proviruses replication-incompetent (Abram et al. 2010). Defective proviruses are thought to not be recognized and cleared by the host due to inadequate antigen presentation by the MHC-I, as defective proviruses are not thought to make HIV protein (Imamichi et al. 2020). However, emerging evidence indicates that defective proviruses can produce viral proteins, have pathobiological significance and are important for HIV pathogenesis (Imamichi et al. 2020; Pollack et al. 2017). *In the present study, I will refer to the persistent reservoir as an all-encompassing term that includes all forms of integrated proviral DNA (intact and defective proviruses) that persist despite effective ART.* Persistent proviruses can be subdivided based on three distinct states of viral expression. First, persistent proviruses that can produce LTR-driven RNA transcripts in spite of suppressive ART are referred to as the *active reservoir.* Second, persistent proviruses can exist in a functionally null state, that is, proviruses cannot initiate LTR-driven transcription either due to a major genomic defect or irreversible epigenetic silencing. Finally, persistent proviruses can exist in a *reversibly quiescent,* or latent form due to a reversible transcriptional block; the *latent reservoir*, as its referred to, is defined as cells that harbor intact, replication-competent proviruses that persist despite effective ART and that can re-initiate infection upon treatment cessation.

Latent infection is hypothesized to occur by two non-mutually exclusive mechanisms: (1) a rare event in which an HIV-infected cell survives long enough to revert from an active to a resting phenotype, consistent with the normal physiology of CD4+ T cell memory (Chun et al. 1995; Shan et al. 2017), (2) direct infection of quiescent or resting T cells (Cameron et al. 2010; Zack et al. 1990). The latter is likely not the major mechanism by which latent infection of resting T cells is established due to blocks at the

level of viral entry (Pierson et al. 2000), reverse transcription (Baldauf et al. 2012; Zack et al. 1990), and viral gene expression (Shan et al. 2017).

The latent reservoir is established extremely early in HIV-infected humans and simian immunodeficiency virus-infected nonhuman primates (Chun et al. 1998; Whitney et al. 2014), but can be reduced in size with immediate ART initiation (Colby et al. 2018; Whitney et al. 2014). The latent reservoir is extremely durable and long-lived with a half-life of 43.9 months and thus, it is estimated it would take approximately 73 years of consistent ART to eradicate this reservoir (Siliciano et al. 2003).

Latency is maintained through multiple cellular and molecular mechanisms and is reinforced at both the transcriptional and post-transcriptional levels. These mechanisms include chromatin-mediated HIV transcriptional silencing (Coull et al. 2000; G. Jiang et al. 2007; Williams et al. 2006) and sequestration of critical cellular transcription initiation factors (such as NF-κB and NFAT) and elongation factors (such as P-TEFb) that normally associate with the HIV LTR promoter or HIV Tat protein to stimulate viral gene expression (Budhiraja et al. 2013; Kinoshita et al. 1997; Nabel and Baltimore 1987; Nguyen et al. 2001; Yang et al. 2001). At the post-transcriptional level, miRNA regulation of viral protein expression can contribute to the maintenance of latency (Huang et al. 2007) as well as depletion of Rev cofactors such as MATR3, PSF, and PTB that are required for Rev-mediated export of HIV RNA (Kula et al. 2011; Kula, Gharu, and Marcello 2013).

The orientation of the provirus with respect to the host genome may also reinforce latency. Integration of the provirus in the same transcriptional orientation as the host gene may result in transcriptional interference, which is defined by the displacement of transcription factors from the HIV LTR promoter due to read-through transcription from a

stronger upstream host gene promoter (Greger et al. 1998; Han et al. 2008). HIV integration in the convergent orientation can reduce HIV gene expression through promoter occlusion, which occurs when two RNA Pol II complexes, one initiating transcription from the LTR promoter, one initiating transcription from the host promoter, collide during elongation leading to premature termination of transcription (Han et al. 2008).

Latency may also be maintained as a direct consequence of the distinct chromosomal location of persisting HIV proviruses (Einkauf et al. 2019; C. Jiang et al. 2020). In one study using a group of elite controllers, HIV control was associated with a) a smaller reservoir and b) chromosomal integration into transcriptionally inaccessible regions of the genome (C. Jiang et al. 2020). In effect, in elite controllers, immune selection forces preferentially eliminate proviruses that are more permissive to viral transcription and what is left are latent proviruses in transcriptionally inactive regions of the genome (“deep latency”), effectively precluding HIV gene expression, which some hypothesize could be emulated to achieve a functional HIV cure.

1.5 Integration site selection

Integration of viral DNA is an obligate step during productive HIV-1 infections and is facilitated by the interaction between viral integrase protein, host DNA, and host cell factors such as LEDGF/p75 (Craigie and Bushman 2012). Analysis of HIV-1 integration locations following acute infection of *in vitro* infected T cells, find HIV IS most commonly in the introns of actively expressed genes (Brady et al. 2009; Schröder et al. 2002; Wu et al. 2003) and includes preference for genomic regions with transcription-associated chromatin modifications (G. P. Wang et al. 2007).

Examination of HIV integration sites from individuals on long-term suppressive ART, where selective forces that enable cell survival or evasion of immune clearance have shaped the population of persisting proviruses, revealed similar global patterns for integration (Cohn et al. 2015; Maldarelli et al. 2014; Wagner et al. 2014) with a few notable differences. A modest bias for integration in the opposite orientation of gene transcription exists *in vivo* (Cohn et al. 2015; Edlefsen et al. 2020). Particular integration sites are rarely selected for *in vivo,* but a few unique genes appear to be overrepresented in *in vivo* IS datasets (Ikeda et al. 2007; Maldarelli et al. 2014; Wagner et al. 2014). Integration into notable genes, such as BACH2, STAT5B and MKL2, are biased for integration in the forward orientation relative to gene transcription and are located upstream of the first coding exon (Edlefsen et al. 2020; Maldarelli et al. 2014; Wagner et al. 2014). This unique pattern, which is only observed *in vivo* after prolonged ART, suggests that proviruses persisting in specific cellular loci may alter T cells in a manner that promotes cell survival by proviral induced misexpression of the host target gene product; this hypothesis is

explored in-depth in the section below, “HIV-1 mediated insertional activation of host genes”.

1.6 Clonal expansion of the persistent HIV reservoir

The persistence of HIV-1 proviruses in individuals on suppressive ART reflects the longevity of memory CD4+ T cell subsets (Chomont et al. 2009; Imamichi et al. 2014) and clonal proliferation of infected cells (Cohn et al. 2015; Maldarelli et al. 2014; Wagner et al. 2014). The first direct evidence for clonal expansion of infected cells came from the study of residual viremia, defined as virions produced in spite of suppressive ART that are detected at low frequencies in the plasma (Dornadula et al. 1999). In these studies, a homogenous population of cells with identical sequences were hypothesized to have arisen as clones of infected cells, that were repeatedly identified within and across timepoints (Bailey et al. 2006; Tobin et al. 2005). Additionally, the proportion of identical sequences without evident viral evolution increased over time in these individuals. Collectively, these results provided initial evidence for clonal proliferation of HIV-infected cells.

Longitudinal examination of HIV integration sites and proviral sequences from individuals on long-term suppressive ART confirmed the clonal expansion of HIV-infected CD4+ lymphocytes based on repeat recovery of identical integration sites or proviruses with identical sequences (Cohn et al. 2015; Maldarelli et al. 2014; Pinzone et al. 2019; Wagner et al. 2014). The proportion of clonally expanded HIV-1 infected cells increased over time, with estimates that over 40% of the persistent reservoir consisted of expanded clones (Wagner et al. 2014). Because these studies analyzed integration sites of proviral sequences without distinguishing intact from defective proviruses, it was unknown if clonally expanded cells primarily harbored defective or replication-competent proviruses. The latter is problematic because it would imply that the proportion of HIV-infected cells

capable of reinitiating infection upon ART cessation is increasing over time. Three independent follow-up studies demonstrated, using a combination of quantitative viral outgrowth and qualitative sequence analysis of clonal outgrowth viruses, that cells harboring replication-competent viral genomes clonally expand *in vivo* (Bui et al. 2017; Cohn et al. 2016; Hosmane et al. 2017). For perspective, although replication-competent proviruses only make up 5-10% of the persistent reservoir (Ho et al. 2013; Imamichi et al. 2020), up to 57% of HIV-infected cells harboring replication-competent proviruses are thought to be clonally expanded (Cohn et al. 2016; Hosmane et al. 2017).

Clonal expansion also represents a major obstacle for the clinical management of HIV infection. While ART can suppress viremia to below the limit of clinical detection, viremia can remain detectable in individuals who are fully adherent and who have no known drug resistance mutations (Tobin et al. 2005). Work from our lab and others demonstrated that *expanded clones* can be responsible for the production and release of potentially infectious virions during ART, this can result in detection of viremia either below or in the detectable range of clinical assays **(Figure 1.2)** (Aamer et al. 2020; Halvas et al. 2020; Siliciano and Siliciano 2020)**.** When viremia is at detectable levels, it presents the clinical dilemma of whether “virologic failure” from non-adherence to ART or selection of drug-resistant virus or residual viremia released from previously infected cells that have created an expanded clone over time. It remains unknown why certain clonally expanded cells are selected to persist, and in some cases, produce residual viremia.

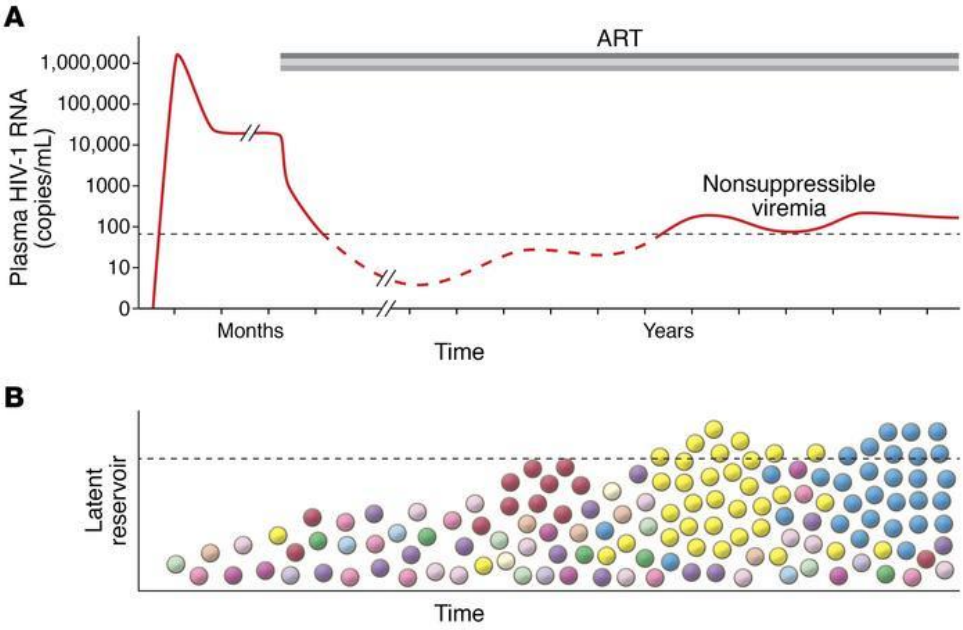


Figure 1.2 Residual viremia can result from large clones of infected cells. Figure taken from (Siliciano and Siliciano 2020). (A) Plasma viremia is represented as the solid red line. ART initiation, represented by the grey bars, reduces plasma viremia to below the limit of clinical assays (dotted black line). However, low-level viremia persists (dashed red line), representing virus release from cells in the persistent reservoir that have become activated. (B) The size and composition of the reservoir is such that the level of virus remains below the limit of detection (dashed black line). However, in some individuals, clones can massively expand (yellow and blue cells) and release a sufficient amount of virus to produce detectable viremia which cannot be suppressed by ART.

In ART-suppressed individuals, HIV-infected cells can expand and contract over time due to viral cytopathic effects (contract), immune clearance (contract) and clonal expansion (expand) (Liu, Simonetti, and Ho 2020; Z. Wang et al. 2018). There are thought to be 3 non mutually exclusive mechanisms driving clonal expansion dynamics: (1) antigen-driven proliferation, (2) homeostatic proliferation and (3) integration site driven proliferation **(Figure 1.3)**.

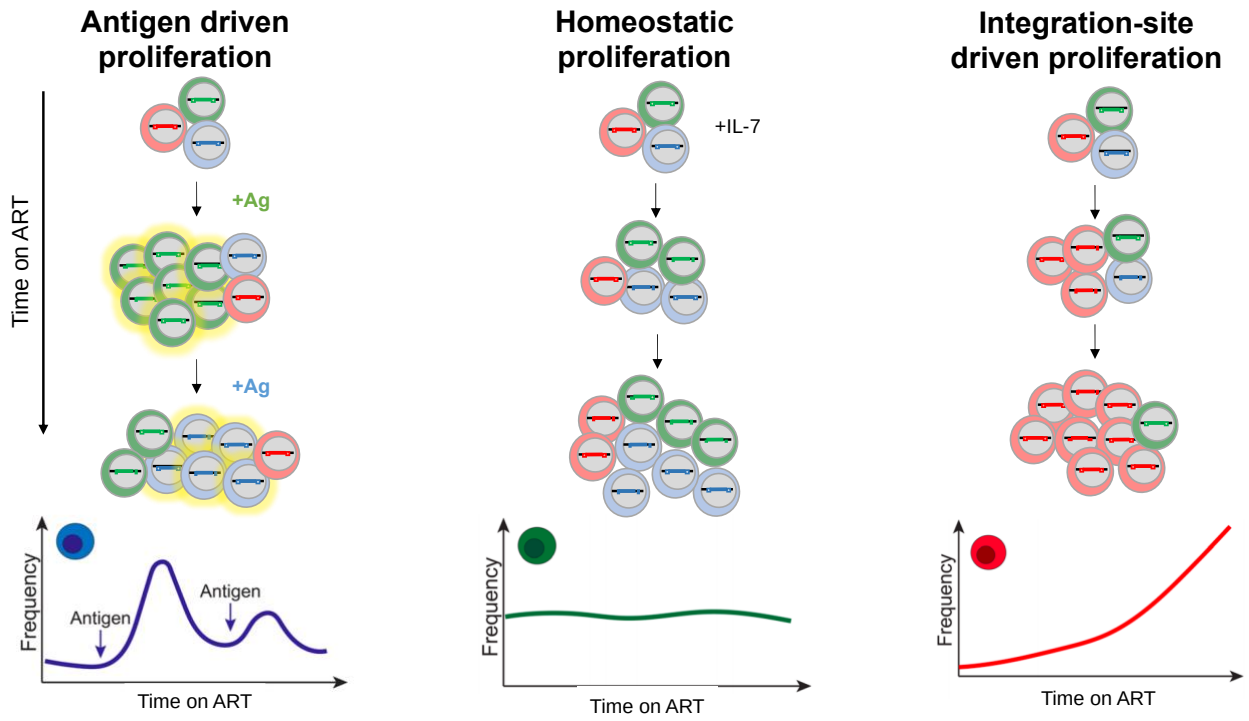


Figure 1.3. Three non-mutually exclusive mechanisms may drive *in vivo* HIV clonal expansion. The persistent reservoir is shaped by viral cytopathic effects, immune clearance and clonal expansion. Three major mechanisms are thought to drive *in vivo* clonal expansion of HIV-infected cells: (1) antigen driven proliferation, (2) homeostatic proliferation and (3) integration-site driven proliferation. Figure adapted from (Cohn 2020; Liu, Simonetti, and Ho 2020).

Antigen-driven proliferation of HIV-1 infected cells is defined by the periodic expansion of antigen-specific HIV-infected cells in response to antigen stimulation. The first evidence to suggest that clonal expansion of HIV-infected cells may be driven by repeated antigen exposure came from the study of a HIV-infected clone carrying the *AMBI-1* provirus (Simonetti et al. 2016). The *AMBI-1* clone, which was identified from an HIV-infected individual with metastatic squamous cell carcinoma, seemed to expand in parallel with tumor growth and contract as the tumor regressed in size after chemotherapy suggesting that *AMBI-1* clones may expand in response to tumor antigen. Secondly, it was noted that the pattern of emergence and disappearance of clonal populations

seemed to follow the typical expansion and contraction phases of a canonical adaptive immune response (Z. Wang et al. 2018). However, the most definitive evidence supporting antigen-driven clonal expansion of infected cells came from complementary studies that used linkage of a combination of antigen responsiveness, T cell receptor β-chain (TCRβ) sequence, full proviral sequence, and/or HIV integration site to infer that clonal HIV DNA sequences can be maintained in antigen-specific CD4+ T cell subsets in patients on suppressive ART (Gantner et al. 2020; Mendoza et al. 2020; Simonetti et al. 2020). The magnitude of infected cell expansion by antigen driven proliferation is thought to be proportional to the magnitude of the antigen response (Liu, Simonetti, and Ho 2020).

Clonal expansion of HIV-infected cells can also result from T cell homeostatic proliferation, a tightly regulated dynamic physiological process induced by lymphopenia that is critical for restoring and maintaining T cell homeostasis (Jameson 2002). Homeostatic proliferation is regulated by diverse stimuli, including signals to the T cell receptor (TCR) and/or common γ chain (γ_C) cytokines such as interleukin (IL)-2, IL-7, and IL-15 (Boyman et al. 2007; Jameson 2002; Min 2018). In simplified terms, IL-2 is thought to control T cell maintenance, while IL-7 and IL-15 are thought to control cell survival and proliferation, respectively (Ma, Koka, and Burkett 2006; Surh and Sprent 2008). In the context of HIV infection, production of IL-7 is clinically correlated with restoration of normal CD4+ T cell counts after HIV-induced lymphopenia, suggesting a role for homeostatic cytokines such as IL-7 in maintaining the latent reservoir (Camargo et al. 2009; Napolitano et al. 2001; Vandergeeten et al. 2013). Proliferation of infected cells in response to homeostatic cytokines was later confirmed *in vitro* (Bosque et al. 2011) and *in vivo* (Chomont et al. 2009). Interestingly it was found that homeostatic proliferation of

HIV-infected cells can occur without viral antigen expression and thus, does not induce immune recognition (Bosque et al. 2011; Katlama et al. 2016).

Finally, integration-site mediated proliferation may also promote clonal expansion of HIV-infected cells. As was discussed in the previous sections, cells containing proviruses integrated into certain genes such as *STAT5B, BACH2* and MKL2 are often clonally expanded. It is thought that integration into these genes (and likely others yet to be defined) may impart infected cells with a survival advantage due to altered expression of the integrated gene driven as a consequence of proviral insertional mutagenesis. The process of proviral insertional mutagenesis is complex, but relatively well-defined, owing to years of research on the subject from the study of other retroviral pathogens (see the section below, HIV-1 mediated insertional activation of host genes). In contrast to the expansion dynamics observed in cases of homeostatic proliferation and antigen-driven proliferation which fluctuate over time **(Figure 1.3)**, integration-site driven proliferation drives a slow but steady increase in HIV-infected cells over time on ART. However, only a handful of specific HIV integration sites have been linked to this “positive selection” phenotype and it is unknown if these proviruses are intact and capable of reinitiating infection.

In summary, many, sometimes overlapping mechanisms can contribute to HIV persistence. However, the relative contributions of these mechanisms (antigen-driven, homeostatic proliferation-driven, integration site-driven) remains unclear.

1.7 HIV-1 mediated insertional activation of host genes

Retroviruses critically depend on proviral integration in order to propagate infection. As a consequence of proviral integration into host genomes, retroviral DNA insertion can induce oncogenic activation of cellular proto-oncogenes in the host genome (Maeda, Fan, and Yoshikai 2008). Insertion of proviral DNA within the vicinity of a proto-oncogene can result in overexpression of the proto-oncogene as a consequence of long terminal repeat (LTR) promoter activation (Coffin, Hughes, and Varmus 1997). Tumor induction by retroviruses is relatively long and multiple cooperative changes in infected cell phenotype need to occur in order to initiate the oncogenic process (Coffin, Hughes, and Varmus 1997).

HIV-1 has not been shown to cause cancer in infected cells (Liu, Simonetti, and Ho 2020). However, some patterns of HIV integration bear resemblance to the integration patterns of non-acute retroviruses such as preference for integration within gene bodies and in some more interesting cases, preference for integration at or near transcription start sites (Fan and Johnson 2011). As previously mentioned, specific integration sites are rarely selected for *in vivo,* but a handful of genes (such as *BACH2, STAT5B,* and MKL2) appear to positively influence the survival of infected cells. Based on the critical function of these genes in important immune signaling pathways and based on the location of the provirus directly upstream of the transcription start site, it was hypothesized that viral integration into specific cellular genes can promote cell proliferation through HIV-1 mediated insertional activation. Indeed, the concept of HIV-induced insertional activation of host genes was supported by seminal work from (Cesana et al. 2017) who detected HIV-human chimeric spliced RNA products from the blood of ART-treated, HIV-

infected individuals. Chimeric HIV-human RNA products can be formed by (1) cellular gene transcription reading into the integrated provirus (host-virus), (2) HIV transcription through the viral polyadenylation site within the 5' and 3' LTR into the cellular gene (virus-host) and (3) splicing between human and viral splice sites (Sherrill-Mix, Ocwieja, and Bushman 2015). The latter mechanism has been explored extensively in recent work and is thought to be a major factor driving aberrant HIV-1 promoter driven transcription in *in vivo* clonally expanded cells (Cesana et al. 2017; Liu et al. 2020; Pinzone et al. 2019).

Multiple species of chimeric HIV-human RNA products can be produced as a consequence of HIV integration either within or upstream of a protein coding sequence **(Figure 1.4 A)** and as a consequence of integration in the opposite orientation as gene transcription **(Figure 1.4 B).** The functional consequences of these chimeric RNA products remain unknown. It is important to note that ART treatment does not inhibit LTR function. Therefore, intact LTR promoter activity may be a driver of cell-associated HIV RNA expression, seen in cases of residual viremia.

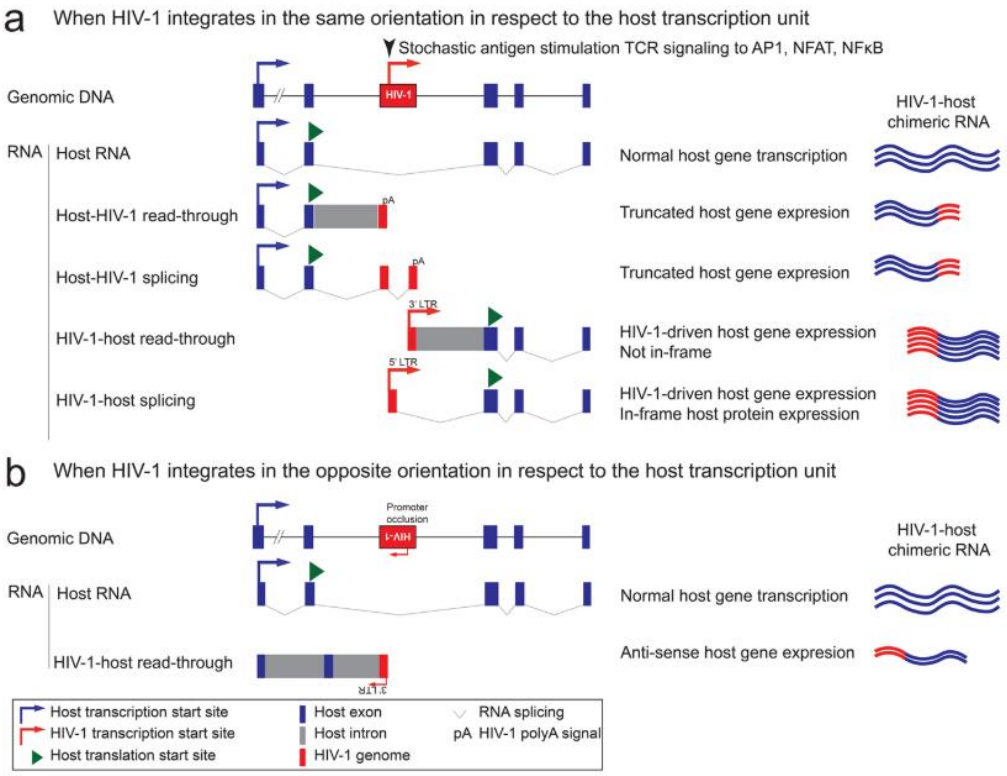


Figure 1.4. A wide array of chimeric HIV-host RNA products can be produced as a consequence of HIV integration. Potential chimeric RNA products that can be produced as a consequence of HIV integration in the same orientation as gene transcription (A) and as a consequence of HIV integration in the opposite orientation as gene transcription (B). Image taken from (Liu, Simonetti, and Ho 2020).

1.8 Thesis summary

The goal of this thesis is to enrich for and characterize HIV-infected cells with a proliferative advantage in order to characterize how HIV IS and drivers of homeostatic proliferation contribute to the survival advantage of cultured CD4+ T cells enriched from HIV-infected donors. The overall goal of this work is to better understand the mechanisms by which HIV-infected cells persist by selecting for phenotypes associated with *in vivo* persistence. Chapter 2 describes the method I developed and used to enrich for HIV-infected cell populations with proliferative survival advantage. I go in depth describing the method, the reproducibility of the method and the HIV IS I recovered from cultured vs. uncultured cells. I present data describing a gene set enrichment analysis that was performed to determine if cultured cells are enriched for particular biological pathways that may be conducive for HIV persistence and clonal expansion. Finally, I characterize HIV-host chimeric RNA products I detected from the cultured cells of an HIV-infected individual. Chapter 3 describes the greater implications of this work in the context of developing and evaluating strategies for an HIV cure.

Chapter 2. Characterization of HIV-infected cell populations with proliferative survival advantage

2.1 Introduction

There are thought to be three non-mutually exclusive mechanisms driving clonal expansion dynamics of HIV-infected cells *in vivo*: (1) antigen driven proliferation, (2) integration site driven proliferation and (3) homeostatic proliferation. The main focus for these experiments was to select for cells that survive and proliferate under homeostatic cytokine conditions and then examine if HIV integration site contributes to this survival advantage by altered gene expression of the gene with integrated provirus.

Homeostatic proliferation is a well-documented mechanism by which the mature, naïve and memory T cells are maintained *in vivo* (see the previous chapter, “Clonal expansion of the persistent reservoir”). Homeostatic proliferation drives relatively non-specific expansion of T cells as a result of signals from homeostatic cytokines. We hypothesized that we could capitalize on this phenomenon to induce non-specific proliferation of cells from an HIV-infected individual with minimal manipulation as to more accurately reflect the *in vivo* gene expression profile of infected cells. We are very interested in characterizing HIV-induced changes to host gene expression in populations of minimally manipulated cells, so we devised a method that could (1) avoid the use of pan T cell activation agents such as phytohemagglutinin (PHA), and (2) induce the outgrowth of clonal populations of HIV-infected cells, thus enabling future single cell transcriptome studies. In doing so, we’d be able to in tandem study the relative contribution of homeostatic proliferation and potential contributing effects of integration site-driven proliferation to *in vivo* HIV persistence. To add an additional layer of selection

we also included a serum starvation period in our experiment to promote the outgrowth of HIV-infected cells resistant to apoptosis and autophagy, based on recent evidence that some HIV-infected cells exhibit an intrinsic resistance to apoptosis/autophagy (Mahlknecht et al. 2000; Ren et al. 2020).

2.2 Methods

Participant selection

Participants for this study were selected from the Seattle Primary Infection Cohort (PIC). The PIC cohort has enrolled individuals with primary and early HIV-1 infection (refers to the six-month time period after infection) into an observational study based at the University of Washington since 1992 (Schacker et al. 1996). Inclusion criteria for the current study include participants that started ART during acute or chronic HIV infection and achieved suppression of viremia to <50 copies of HIV-1 RNA/mL for at least 7 years of ART. Additional inclusion criteria include availability of leukapheresis specimens post-ART suppression.

Selection of CD4+ T cells with a proliferative advantage

We developed a method to support *ex vivo* homeostatic proliferation of cells from HIV-infected donors to determine if we can detect and enrich for HIV-infected cells that survive and proliferate under homeostatic cytokines **(Figure 2.1)**. Directly downstream from our cellular proliferation selection method is examining if HIV integration sites (**Figure 2.1,** HIV integration site analysis) might contribute to this survival advantage by altering gene expression of the integrated genes (**Figure 2.1,** HIV-human chimeric transcript analysis).

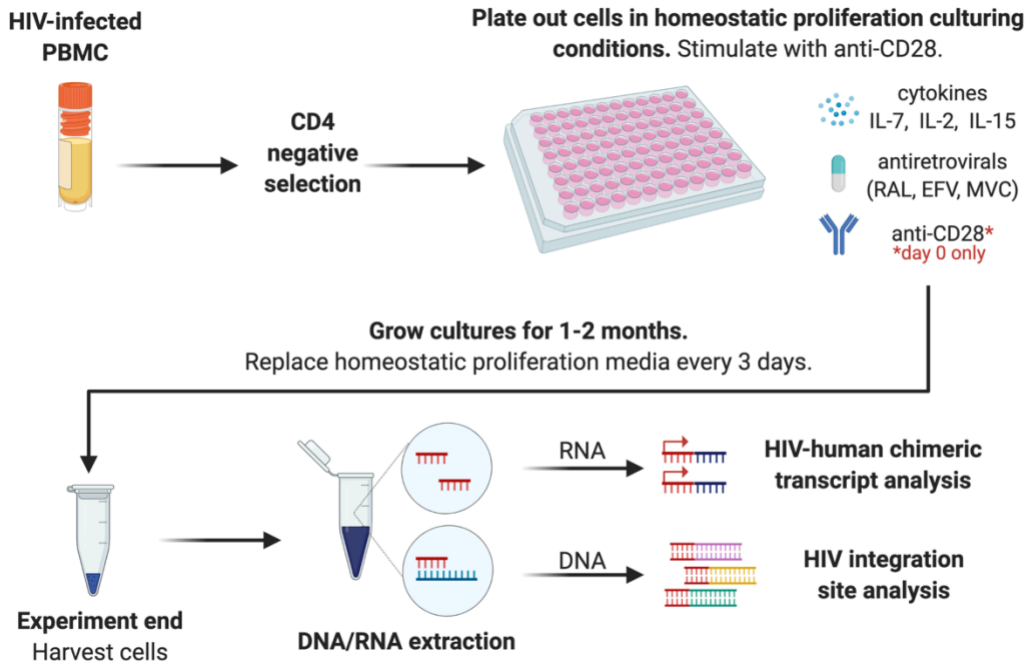


Figure 2.1. Experimental overview of the *ex vivo* culturing method used to detect and enrich for HIV-infected cells that survive and proliferate using homeostatic cytokines. CD4+ T cells from an HIV-infected donor were cultured for 27-74 days (or approximately 1-2 months) in homeostatic proliferation culturing conditions which included addition of homeostatic proliferation cytokines (interleukin (IL)-7, IL-2, IL-15) and antiretroviral therapy drugs (raltegravir (RAL), efavirenz (EFV), maraviroc (MVC)). CD4+ T cell cultures were stimulated with antibodies to CD28 (anti-CD28) on day zero only. At the end of the culturing experiment, CD4+ T cell cultures were harvested and prepared for DNA and RNA extraction.

Cryopreserved HIV-infected peripheral blood mononuclear cells (PBMC) obtained from HIV-infected individuals who had been virologically suppressed by ART for minimum of 7 years had CD4+ T lymphocytes isolated using negative selection by the EasySep immunomagnetic selection kit (STEMCELL Technologies, Vancouver, Canada). Purified CD4+ T cells were cultured at a concentration of $2x10^6$ cells per well in 24-well plates with "homeostatic proliferation media" which consisted of 2mL RPMI (Fisher Scientific, Waltham, MA) supplemented with 10% heat-inactivated human serum AB (Gemini Bio-Products, West Sacramento, CA), 1% penicillin G/streptomycin sulfate/L-glutamine

solution (Fisher Scientific, Waltham, MA), 10U/mL IL-2 (PeproTech, Rocky Hill, NJ), 10ng/mL IL-7 (PeproTech, Rocky Hill, NJ) and 10ng/mL IL-15 (PeproTech, Rocky Hill, NJ). 1µg/mL of soluble anti-human CD28.2 antibody (BioLegend, San Diego, CA) was added to all wells at culture day 0 to provide an initial activation signal to CD4+ T cell cultures. Every 3 days homeostatic proliferation media was replenished.

An aliquot of cells was periodically collected by pooling all wells and harvesting $2x10^6$ to $10x10^6$ cells from CD4+ T cell cultures to measure HIV-1 DNA load over the course of the culturing period. At the termination of the culturing period (range: 27 – 74 days), CD4 T cell cultures were submitted to dual DNA/RNA extraction using the AllPrep DNA/RNA Mini Kit (Qiagen, Valencia, CA). Purified DNA was used for HIV integration site analysis whereas total RNA was used for HIV-human chimeric transcript analysis. A RNeasy spin column (Qiagen, Valencia, VA) was used during the RNA extraction step to bind RNA and remove the majority of DNA and all other contaminants.

HIV DNA load quantification by LTR qPCR

HIV DNA copy number was quantified by LTR real-time qPCR using iTaq Universal Probes Supermix (Bio-Rad, Hercules, CA). We used previously validated (Arvold et al. 2007) cross-subtype primers and probe targeting the LTR into gag genes: 20µM LTR 5' (5'GCC TCA ATA AAG CTT GCC TTG A 3'), 20µM LTR 3' (5' GGC GCC ACT GCT AGA GAT TTT 3'), 100µM LTR probe (5' 6FAM-CTG GTA ACT AGA GAT CCC T –MGB; Applied Biosystems, Foster City, CA).

Integration site analysis by TMO MDA ISLA

TherMOstable (TMO) multiple displacement amplification (MDA) integration site loop amplification (ISLA) is an extension of our previously described method, ISLA (Wagner et al. 2014). ISLA amplifies single-stranded templates originating in the HIV LTR and extending across the integration site. For TMO-MDA-ISLA, we used a diversified primer set spanning across the HIV genome **(Figure 2.2)** to (1) amplify the full HIV genome, and (2) increase the number of templates for ISLA. A thermostable polymerase allows multiple cycles of DNA denaturation, which should reduce amplification bias that might occur when a proviral genome is integrated into an inaccessible region of the genome. This method is optimized to capture HIV integration sites from both the 3' and 5' ends of HIV. The input of HIV DNA is diluted so that no more than 30% of the reactions are positive for an HIV integration site based on the Poisson distribution to ensure reactions are likely to be at the single proviral copy level.

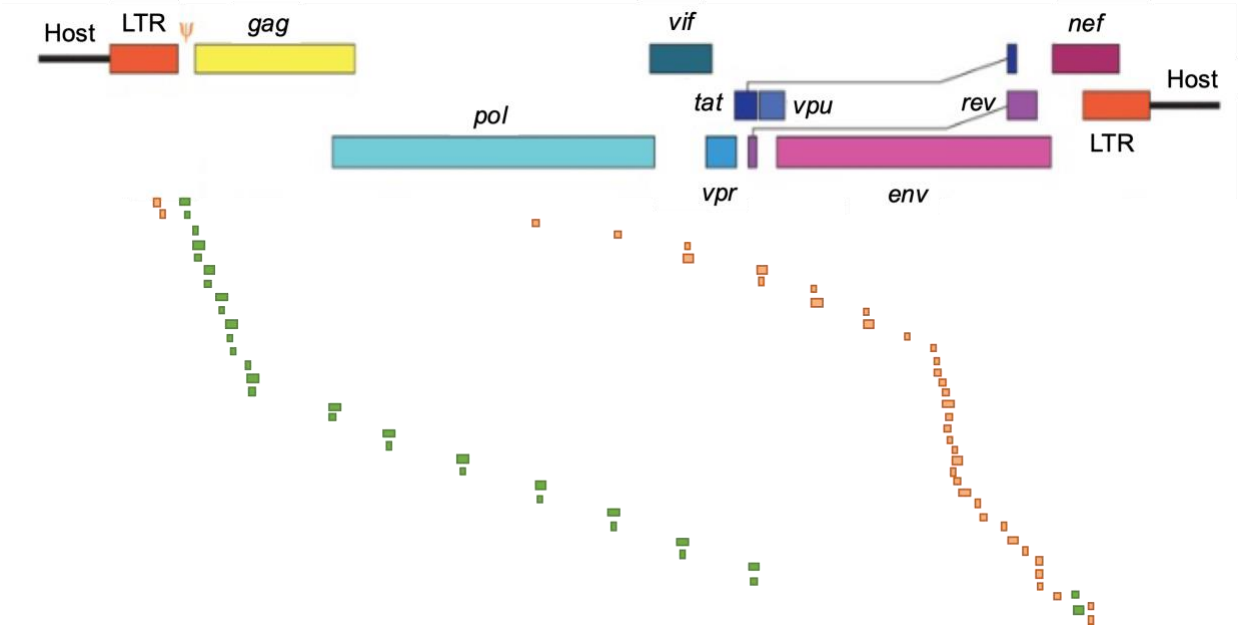


Figure 2.2. Approximate positions of MDA primers used in the TMO-MDA-ISLA assay. Shown in green are the forward primers used for 3' TMO-MDA-ISLA to capture HIV integration sites from the 3' LTR. Shown in orange are the reverse primers used for 5' TMO-MDA-ISLA to capture HIV integration sites from the 5' LTR. To create a reverse primer site for loop formation of the proviral integration site and adjacent human sequence, a U5-specific primer tailed with random decamers (deca1.U5) was used.

In brief for TMO MDA, genomic DNA is combined with 10mM dNTPs, 10X Advantage buffer (Takara Bio, Kusatsu, Japan), 20uM deca 1.U5 primer (5' TCA AGT AGT GTG TGC CCG TCT GTN NNN NNN NNN 3'), 50U/mL SD polymerase (Boca Scientific, Westwood, MA) and 20µM MDA primers in 50µL total reaction volumes. **(Table 2.1)**. For amplification from the 3' LTR, MDA primers include 31 primers: (1) 2 primers amplifying from HIV Nef, (2) 16 short primers and (3) 13 long primers spanning from 5'UTR to early Env. For amplification from the 5' LTR, MDA primers include 38 primers: (1) 2 primers amplifying from the 5' UTR of HIV, (2) 26 short primers and (3) 10 long primers spanning from Pol to Nef. The reactions conditions are as followed: 2 minutes (min) at 92ºC, 20 seconds (sec) at 92ºC, 30 sec at 30ºC for 10 cycles. Then the reactions

were progressively warmed to 50ºC by 1ºC per second until reaching 68ºC. The reactions were then incubated at 68ºC for 2 min, followed by 20 sec at 92ºC, 30 sec at 55ºC, 2 min at 68ºC, then a final hold at 12ºC.

The TMO-MDA products went through three rounds of PCR with nested LTR primers for ISLA as previously described (Wagner et al. 2014). Third round-PCR products were identified by gel electrophoresis and Sanger sequenced using 2.U5 (5'-GTTGTGTGACTCTGGTAACTAGAGAT-3'), a U5-specific primer located downstream of the looping site.

Table 2.1. TMO-MDA-ISLA primers used to derive HIV integration sites from the 3' LTR or 5' LTR. Shaded in white are the short primers, in yellow are the long primers, and in grey are the special primers amplifying from HIV Nef (for 3' TMO) or from the 5' UTR of HIV (for 5' TMO).

3' TMO-MDA-ISLA		
Primer	**Sequence (5' to 3')**	**HIV region cover**
MDA.F10s	CACCAAATG*A*A	Gag
MDA.F11s	CTGTACCA*G*T	Pol
MDA.F12s	TCCATCCT*G*A	Pol
MDA.F13s	AACAAATCAG*A*A	Pol
MDA.F14s	ATTCCCTAC*A*A	Pol
MDA.F15s	CATAGAATG*G*A	Vif
MDA.F16s	GCAGGAAG*A*A	Tat / Rev
MDA.F1s	GACTCGGC*T*T	5' UTR
MDA.F2s	GAGGCTAG*A*A	5' UTR
MDA.F3s	GGAGAGAG*A*T	5' UTR / Gag
MDA.F4s	GTATGGGC*A*A	Gag
MDA.F5s	CAGAAGAACT*T*A	Gag
MDA.F6s	GAAGCTTTA*G*A	Gag
MDA.F7s	AACAAAAGTAA*G*A	Gag
MDA.F8s	TGGGTAAAA*G*T	Gag
MDA.F9s	TACCCATG*T*T	Gag
MDA.687.F	CGACGCAGGACTCGGCTTG*C*T	5' UTR
MDA.779.F	GAAGGAGAGAGATGGGTGCGA*G*A	5' UTR / Gag
MDA.881.F	TAAAACATATAGTATGGGCAAG*C*A	Gag
MDA.991.F	CTTCAGACAGGATCAGAAGAACT*T*A	Gag
MDA.1070.F	TAAAAGACACCAAGGAAGCTTTA*G*A	Gag
MDA.1283.F	GCCCAGAAGTAATACCCATG*T*T	Gag
MDA.2036.F	GTGGAAAGGAAGGACACCAAATG*A*A	Gag
MDA.2550.F	CCCATTAGTCCTATTGAAACTGT*A*C	Pol
MDA.3235.F	GGATGGGTTATGAACTCCATCCT*G*A	Pol
MDA.3952.F	CCCTAACTGACACAACAAATCAG*A*A	Pol
MDA.4637.F	GCAGGAATTTGGCATTCCCTAC*A*A	Pol
MDA.5283.F	GGGTCAGGGAGTCTCCATAGAATG*G*A	Vif
MDA.5954.F	GGCTTAGGCATCTCCTATGGCAGGAAG*A*A	Tat / Rev
up3.1	CCAGTCAGACCTCAGGTACCTTTAAGACCAATG	Nef
up3.2	CCAATGCTGATTGTGCCTGGCTAGAAGCA	Nef

5' TMO-MDA-ISLA		
Primer	**Sequence (5' to 3')**	**HIV region covered**
MDA.R10s	TTACTGCTT*T*G	Rev
MDA.R11s	TTCTGAAAAA*C*A	Vif
MDA.R12s	GTACTGCT*G*T	Pol
MDA.R13s	GTCTGTTACT*A*T	Pol
MDA.R1s	TGACTGGA*A*A	Nef
MDA.R2s	AAGTCTCTC*A*A	Env / Rev
MDA.R3s	AACCCAAG*G*A	Env
MDA.R4s	CCACTCTT*C*T	Env
MDA.R5s	CTAATGGTTC*A*A	Env
MDA.R6s	CAGGTCTG*A*A	Env
MDA.R7s	CTCCACAATT*A*A	Env
MDA.R8s	AGTTGAGTT*G*A	Env
MDA.R9s	TGTTCTACC*A*T	Env
MDA.R26	CACCATCTC*T*T	Env
MDA.R25	CTACTTTATATTTA*T*A	Env
MDA.R24	TCTTTTTTCT*C*T	Env
MDA.R23	AAGAACCC*A*A	Env
MDA.R22	CATAGTGC*T*T	Env
MDA.R21	TACCAGACAA*T*A	Env
MDA.R20	CAACCCCA*A*A	Env
MDA.R19	TTACTCCAAC*T*A	Env
MDA.R18	AACCAATTC*C*A	Env
MDA.R17	CTATCATTATGAA*T*A	Env
MDA.R16	CTTCGATTC*C*T	Env / Tat / Rev
MDA.R15	TCAAGAGTAA*G*T	Env / Rev
MDA.R14	CCCTATCT*G*T	Env
MDA.5409.R	GGCHTTTCTTATAGCAGAKTCTGA*A*A	Vif
MDA.6064.R	CATTACATGTACTACTTACTGCTTTG*R*T	Rev
MDA.6550.R	ATTATATCYTCATGCATCTGTTCTACC*A*T	Env
MDA.7015.R	CTGCCATTTAACAGCAGTTGAGT*T*G	Env
MDA.DR8new	CCCTCATATCTCCTCCTCCAGGTCTG*A*A	Env
MDA.BH2	CCTTGGTGGGTGCTACTCCTAATGGTT*C*A	Env
MDA.7751.R	TTCTCTYTGCACCACTCTT*C*T	Env
MDA.ED12	AGTGCTTCCTGCTGCTCCCAAGAACCCA*A*G	Env
MDA.8556.R	CAATCAAGAGTAAGTCTCTC*A*A	Env / Rev
MDA.9020.R	GGTACCTGAGGTCTGACTGGA*A*A	Nef
UTR.629.R	CCCTGTTCGGGCGCCACTGCTA	5' UTR
UTR.682.R	GCCGAGTCCTGCGTCGAGAGA	5' UTR

Preparation of integration site data

Raw reads were edited using Geneious R11.1.5 to remove poor quality sequence, manually call ambiguous bases, and extract any mixed sequences prior to genome location mapping. Cleaned sequences were mapped to the human reference genome GRCh38.p2 with the Integration Sites pipeline developed in the Mullins Lab at the University of Washington (indra.mullins.microbiol.washington.edu/integrationsites/). A conserved sequence from the 3' LTR (TCT CTA GCA) or the 5' LTR (GCC CTT CCA) was used to pinpoint the site of proviral integration into the human genome. Sequences that mapped to an ambiguous location, due to HIV integration into a repetitive region, were excluded. Gene and exon information was derived from Ensembl version 101 corresponding to GENCODE release 35 through annotations extracted from Ensembl's BioMart data service. Genes and exons/introns were associated with IS by computing the overlaps with IS locations. *Unique IS* were determined by de-duplicating on the tuple of (participant, chromosome/landmark, location, provirus orientation). The *multiplicity* of an IS was defined as the number of times that exact (landmark, location, orientation) tuple is observed independently within a participant. IS with a multiplicity greater than one are said to be found within *demonstrably proliferating cells.*

Gene set enrichment of integration sites

Gene set enrichment analysis was performed using a set of compiled pathways from the MSigDB database to discern relationships between the genes harboring integration sites and the biological processes regulated by those genes. Three MSigDB collections were included (hallmark gene sets (h), canonical pathways (c2.cp), chemical and genetic

perturbations (c2.cgp)), amounting to 6,279 total gene sets. Pathway enrichment of integration sites was assessed by Fisher's exact test and assessed as "Tier 1" significant when Holm-adjusted p-values (over all pathways evaluated) satisfied Holm $p \leq 0.05$, and as "Tier 2" significant when (not Tier 1 significant but) Benjamini-Hochberg False Discovery Rate q-values ≤ 0.20 in addition to unadjusted p-values ≤ 0.05. We excluded any dataset for which the contingency table contains any marginal totals less than 4. In other words, we excluded any gene set containing fewer than 4 genes associated with the observed unique IS (pooling across comparison groups). However, when this exclusion resulted in a small number of gene sets considered in the analysis, a post-hoc analysis was performed in which gene sets were excluded only if they contained exactly zero of the observed IS (pooling across comparison groups)

Detection of HIV-human chimeric spliced RNA products

Extracted RNA was reverse transcribed into complementary DNA using the SuperScript III system (Invitrogen, Carlsbad, CA). cDNA synthesis was followed by nested PCR using integration-site specific and LTR-specific primer sets based on the predominant IS clone in each individuals' IS dataset. LTR-specific primers have the ability to prime off of the 5' or 3' LTR. Here, the *predominant clone* is said to be the IS with the highest multiplicity in each individual. For Participant 2 whose predominant clone was *CPNE1*, we designed nested primer sets complementary to the R region of the LTR and *CPNE1* exon 3/4 to detect HIV-CPNE1 chimeric spliced RNA products. Amplified bands were gel extracted and Sanger sequenced.

2.3 Results

Proliferation-based selection of HIV-infected cells results in enrichment of HIV DNA

Clonal expansion of HIV-infected T cells is an important mechanism of HIV persistence that could be driven by a multitude of factors including homeostatic proliferation, antigen-driven proliferation, and integration-site (IS) mediated proliferation. To study the relative contribution of homeostatic proliferation and IS-driven proliferation in contributing to clonal expansion and *in vivo* persistence of HIV-infected cells, we designed an assay to select for HIV-infected cells that survive and proliferate under conditions mimicking *in vivo* homeostatic proliferation with or without additional selection for resistance to apoptosis/autophagy. We then examined if HIV IS might contribute to this survival advantage of infected cells by altering gene expression of the integrated gene.

PBMC from six HIV-1 infected individuals on suppressive ART for a median of 11.125 years (IQR: 8.2 – 16.3 years) were obtained and used in an *ex vivo* homeostatic proliferation culturing experiment. For experiment 1, cells were cultured for 74 days **(Figure 2.3)**. The duration of the experiment was based on previous work from our collaborators (unpublished) that demonstrated that targeted insertion of HIV LTR at the *BACH2* locus in primary CD4+ T cells drives cellular proliferation of the targeted cell population over a culturing period of 56 days. We maintained CD4+ T cells in tissue culture for an extended period of time without T cell restimulation to promote outgrowth of HIV-infected cells with a survival advantage.

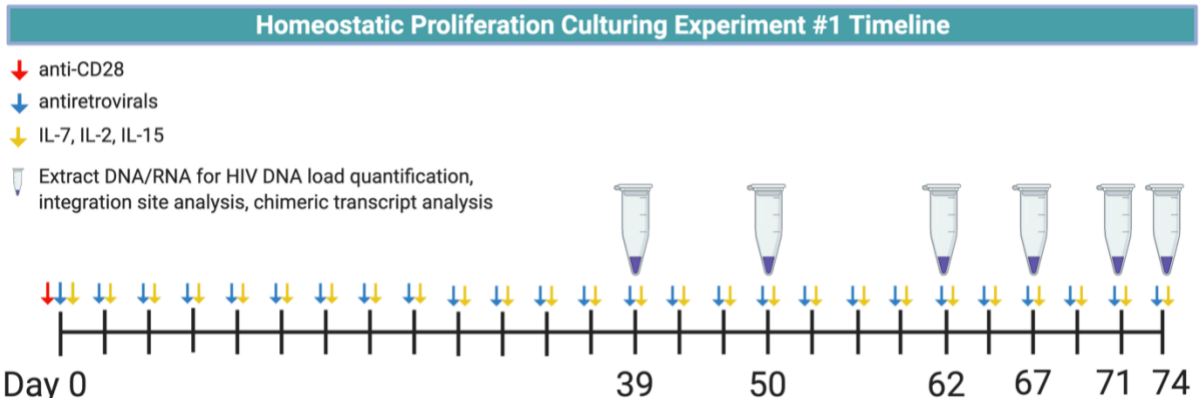


Figure 2.3. Experiment #1 timeline for culture of primary CD4 T cells under conditions promoting homeostatic proliferation. Negatively-selected CD4+ T cell cultures were stimulated with anti-human CD28 to promote cell expansion on day zero only and then were replenished with antiretrovirals and homeostatic cytokines (IL-7, IL-2, IL-15) every three days. Periodically a small number of cells were taken from each culture for DNA and RNA extraction in preparation for downstream HIV DNA load quantification, integration site analysis and chimeric transcript analysis.

Total cell-associated HIV DNA loads, reported as LTR copies per µg DNA, were determined from negatively selected CD4+ T cells from the six participants' leukapheresis specimens. These "uncultured cells" are referred to as "**total CD4+ T cells" (Figure 2.4 A).** Of six total participants tested, only two participants had enrichment of total HIV DNA after 39 days of selection by our cell proliferation method compared to HIV DNA load from total CD4 T cells **(Figure 2.4)**. Consequently, only Participants 1 and 2 were included in subsequent analyses.

A

Participant ID	LTR cp/ug DNA Total CD4+ T cells	LTR cp/ug DNA Culturing Day 39
1	34.6	82.07 ± 13.75
2	117.7	514.77 ± 170.77
3	18.1	8.4
4	2333.7	409.28
5	441.9	19
6	48.3	5.23

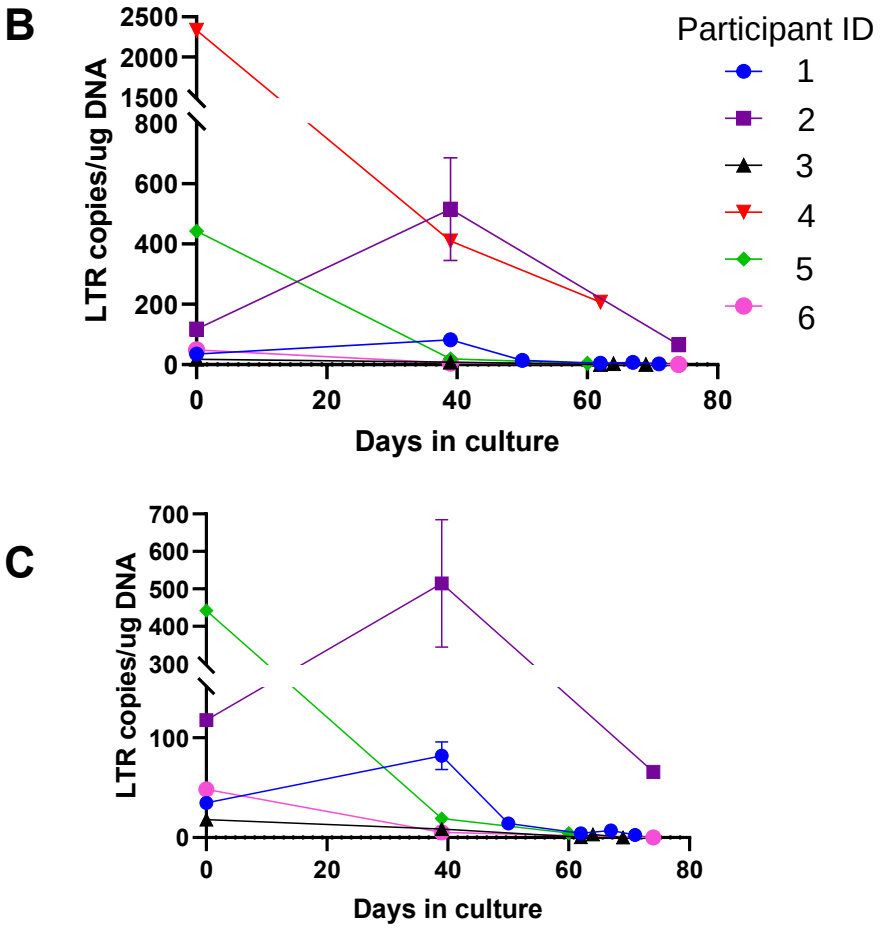


Figure 2.4. HIV DNA load over the course of a homeostatic proliferation selection experiment. (A) HIV DNA load was previously determined from uncultured, or total CD4+ T cells and compared to HIV DNA load from cultured CD4+ T cells taken after 39 days of culture. HIV DNA load is represented as copies of LTR per 1µg of DNA. Participants with enrichment of HIV DNA load after 39 days are indicated in green, whereas participants with no enrichment of HIV DNA load are indicated in red. (B) HIV DNA load over the 74 day culturing period for all participants. Combined experimental values are represented as mean ± SEM. Error bar for Participant 1 day 39 replicate values was shorter than the size of data symbol (●) and could not be plotted without adjusting the scale of the y-axis. (C) The lower portion of plot B is magnified to better resolve viral load dynamics in cultured CD4+ T cells over the course of the experiment. Note that Participant 4 is excluded and that the error bar for Participant 1 is now distinguishable.

The demographic and clinical characteristics of individuals selected for further analyses are shown in **Table 2.2**. Participants 1 and 2 were adults who started ART during acute infection, Fiebig stage II (range: 16 and 19 days, or 0.53 and 0.63 months post HIV infection, respectively) and who at the time of analysis, had maintained ART suppression for 12.75 and 16.3 years, respectively.

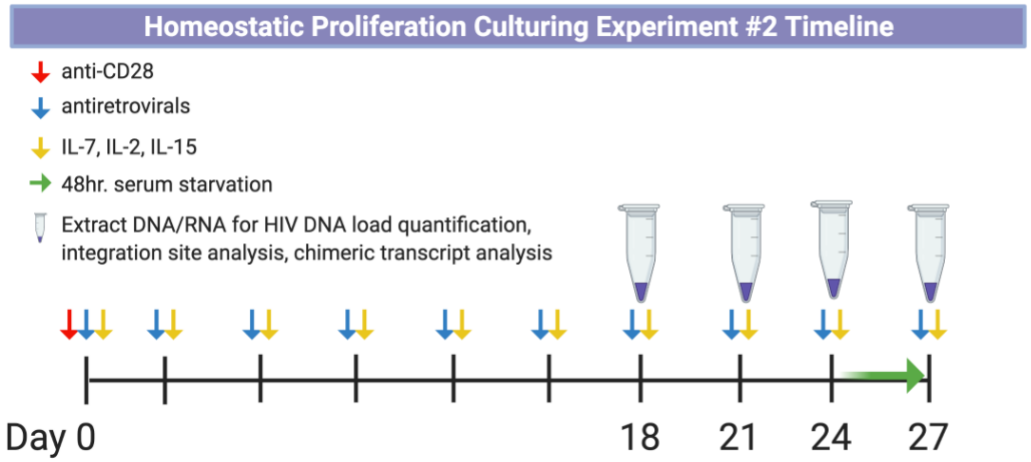
Table 2.2. Demographic and clinical characteristics of study population. Study population profile including demographic and clinical HIV and antiretroviral treatment (ART) details.

				At Time of Analysis			
Participant ID	**Gender**	**Interval Between HIV Infection and ART Initiation (months)**	**Time of ART Initiation Relative to Time of HIV Infection**	**ART Regimen**	**Date of Sample Collection**	**Years of Continuous ART-suppression**	**CD4 T cells/uL**
1	Male	0.53	Acute	TDF, FTC, EVG	12/2014	16.3	935
2	Male	0.63	Acute	3TC, ABC, EFV	01/2015	12.75	655

TDF: tenofovir, FTC: emtricitabine, EVG: evitelgravir, 3TC: lamivudine, ABC: abacavir, EFV: efavirenz

To compare the reproducibility across independent culturing experiments, we repeated the homeostatic proliferation culturing experiment with Participants 1 and 2 only. The greatest enrichment of HIV DNA load was observed after 39 days of culturing in experiment 1; therefore, we predicted that the HIV DNA load would reach its peak within ≤ 39 days of culturing for experiment 2. The experimental set-up for experiment 2 was identical to experiment 1 except for a reduced culturing period of only 27 days and the addition of culturing with medium with low levels of human serum to mimic “starvation” (serum starvation) during the final 48 hours of the experiment end **(Figure 2.5 A)**. Serum starvation was employed to induce cell apoptosis and autophagy in CD4+ T cell cultures in an effort to select for and characterize HIV-infected cells resistant to apoptosis/autophagy.

A



B

		LTR copies/ug DNA	
		Participant 1	Participant 2
Total CD4+ T cells		34.6	117.7
Experiment #2 Culturing Day	18	115.66	4.63
	21	224.74	0.13
	24	0	47.47 ± 47.28
	27	20.53	0

Figure 2.5. Experiment #2 timeline for culture of primary CD4 T cells under conditions promoting homeostatic proliferation and apoptosis/autophagy. (A) The homeostatic proliferation assay was repeated (culturing experiment #2) to evaluate the reproducibility of the assay with Participants 1 and 2 only. CD4+ T cell cultures were maintained for a culturing period of 27 days. 48 hours before the end of the culturing period, cells were starved of human serum to promote selection of cells resistant to apoptosis/autophagy. (B) HIV DNA load (copies of LTR per 1µg of DNA) from total CD4+ T cells (grey) was compared to HIV DNA load from cultured CD4+ T cells taken after 18, 21, 24, and 27 days of culture (blue) in Participant 1 and 2. Samples with enrichment of HIV DNA load are indicated by red text.

HIV DNA load was found to be enriched after 18 - 21 days of selection by our cell proliferation method for Participant 1 only, providing some validation of the approach to enrich for HIV-infected cell populations with a proliferative advantage **(Figure 2.5 B)**. Due to the serum starvation period, we expected DNA load to have decreased by the end of

the assay (day 27) compared to HIV DNA load from total CD4+ T cells due to death of HIV-infected and -uninfected cells by starvation-induced apoptosis and autophagy. As expected, HIV DNA load decreased by day 27 compared to HIV DNA load from total CD4+ T cells (20.53 vs 34.6 LTR copies/µg DNA, respectively) in Participant 1, likely due to selection of HIV-infected cells most resistant to apoptosis and autophagy, but increased susceptibility to apoptosis/autophagy compared to uninfected cells. However, we cannot explain why HIV DNA was not detected on day 24, which was 1 day before the 48-hour starvation **(Figure 2.5 B)**. It is likely that too little of the specimen was measured because DNA clearly was present in these cells as evidenced by detection of HIV DNA on day 27. In experiment 1, HIV DNA load for Participant 2 reached a peak of 514.77 ± 170.77 LTR copies/µg DNA after 39 days of culture **(Figure 2.4 A)**; however, in experiment 2, no HIV DNA enrichment was detected after 27 days of culture. This indicates inter-individual variability in culturing length; Participant 2 required longer culturing time to promote outgrowth of HIV-infected cells.

Proliferation-based enrichment detected in cultured HIV-infected T cells

We hypothesized that the site of HIV integration into the human genome could impact viral persistence during ART-suppression. Therefore, our experiments aimed to identify whether HIV integration site (IS) could potentially impact the survival and outgrowth of particular HIV-infected clones after homeostatic proliferation and/or apoptotic selection in a reproducible manner. We determined the distribution of HIV IS in cultured CD4+ T cells from Participants 1 and 2 and compared this to the distribution of IS in their uncultured total CD4+ T cells obtained from the same timepoint. We predicted that there would be a high degree of overlap between cultured and uncultured IS datasets and that there will be an expansion of particular HIV-infected clones after homeostatic proliferation selection.

A total of N = 36 and N = 70 IS were determined from Participants 1 and 2, respectively, across both culturing experiments and were compared to a total of N = 582 and N = 138 IS from the same individuals' uncultured CD4+ T cells **(Figure 2.6)**. Intra-individual comparisons of HIV IS between cultured CD4+ T cells and uncultured CD4+ T cells revealed that overlap of IS between the datasets was rare. After proliferation-based enrichment, novel clones (i.e., unique HIV IS not observed in the uncultured CD4+ T cell IS data set) were detected. The only clone detected in cultured and uncultured cells was the *ARIH2* IS in Participant 1, which represented the majority clone in both data sets.

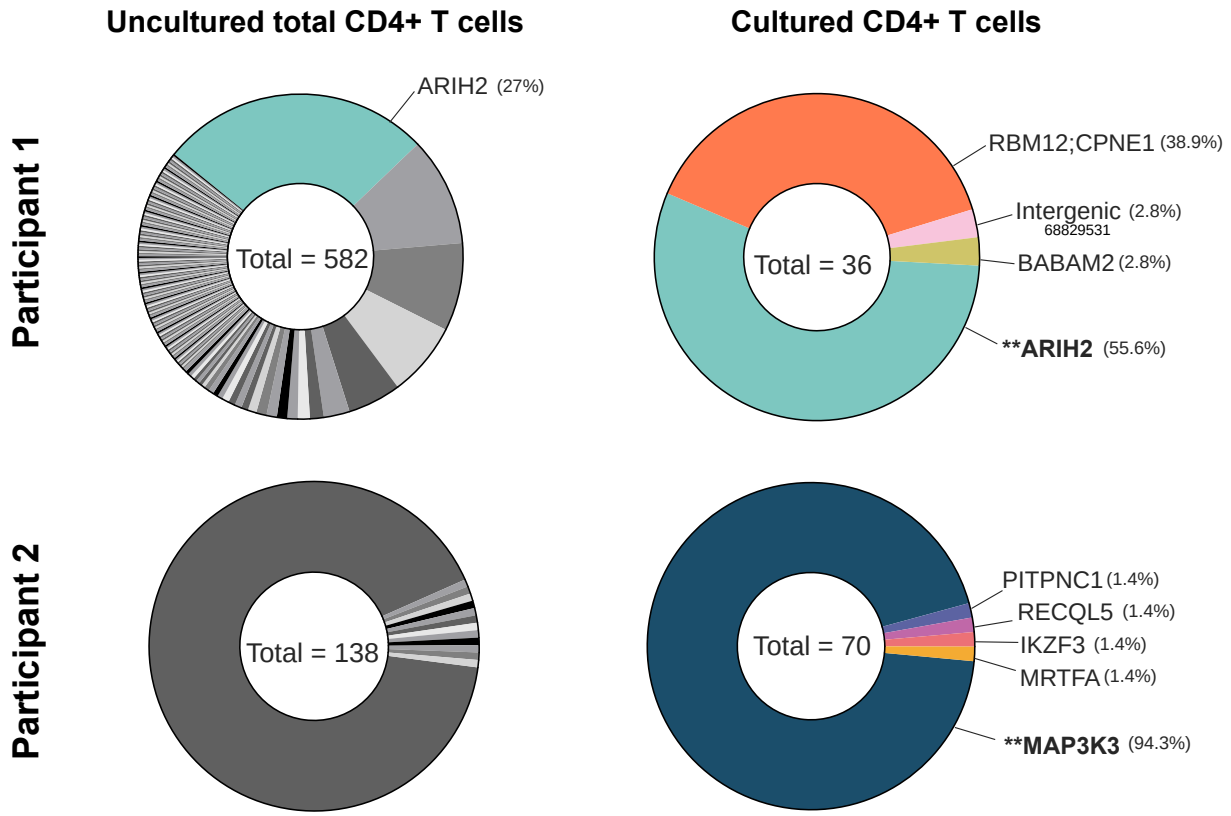


Figure 2.6. Integration sites detected in CD4+ T cells before and after culture with homeostatic cytokines and selection by apoptosis/autophagy. TMO MDA ISLA-derived integration sites (IS) are shown for Participants 1 (top) and 2 (bottom) from uncultured CD4+ T cells (left) and from CD4+ T cells after culture with homeostatic cytokines (right). Total IS obtained from each participant is indicated by the number in the center of the doughnut plot. Each unique IS found in cultured CD4+ T cells is represented by a distinct color. Greyscale coloring indicates unique integration sites found only in uncultured CD4+ T cells. The frequency of a particular IS among total IS is represented by the percentage shown adjacent to the gene name. IS in overlapping genes are separated by semicolon. Intergenic IS are marked by their unique chromosomal integration site location. *IS that were reproducibly detected in experiments 1 and 2 are noted in bold and by the asterisk* **(**)**.

High-frequency clones, or major clones, were defined by identical IS comprising >35% of total IS. For Participant 1, *ARIH2* (55.6% of cultured cells) and the overlapping genes *RBM12/CPNE1* (38.9% of cultured cells) persisted in cultured CD4+ T cells. For Participant 2, clones with provirus integrated in *MAP3K3* persisted and proliferated such

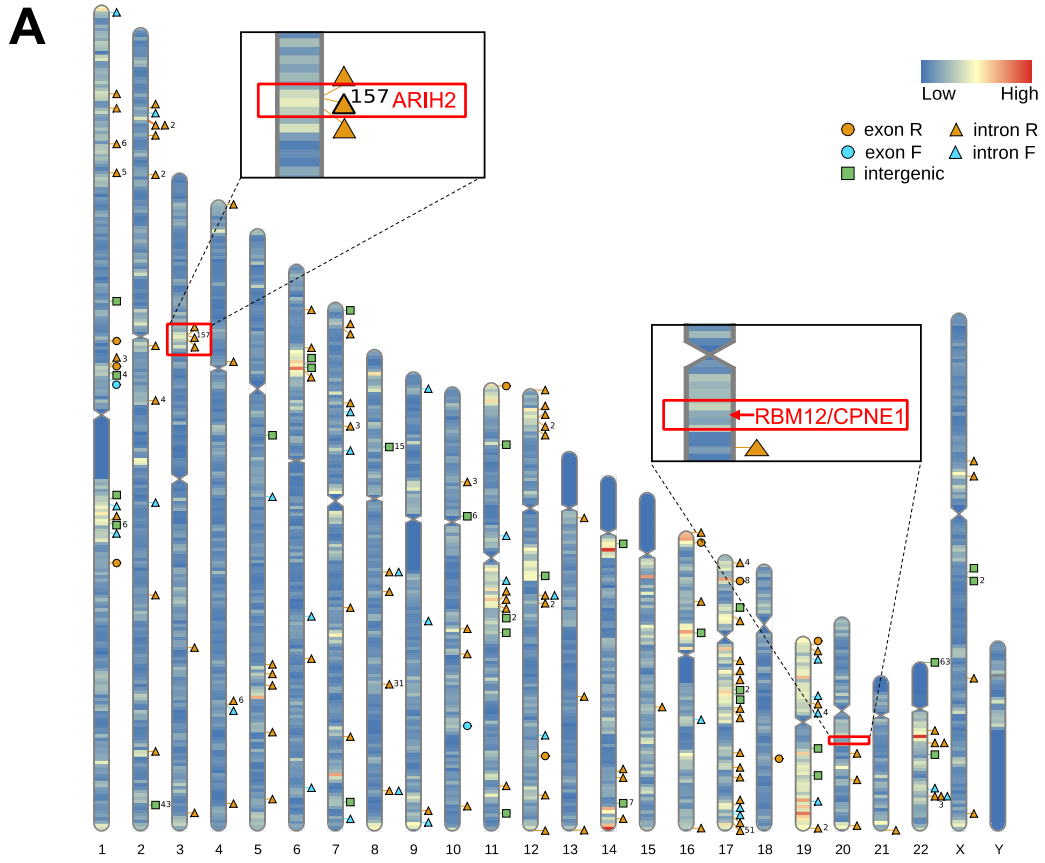
that this IS represented roughly 94% of cells from which provirus was detected from cultured cells in both experiments 1 and 2. To determine whether integration into these genes may have potentiated cell proliferation, we analyzed the orientation and chromosomal features of these integrations within the human genome **(Figure 2.7)**.

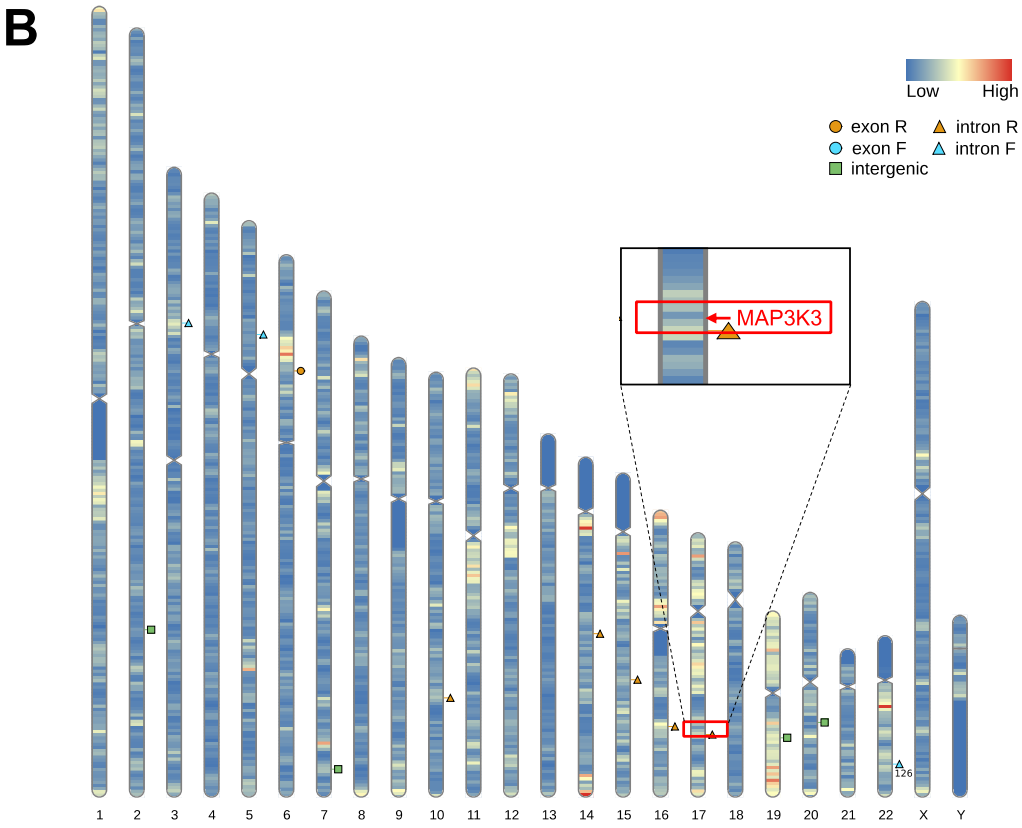
For each major clone found within the cultured CD4+ cells IS datasets, the IS were found in the introns of genes and in the same orientation as host gene transcription. Proviruses were detected in the *ariadne homolog 2 (ARIH2)* gene in both cultured and uncultured CD4+ T cells at the same chromosome position in Participant 1. The *ARIH2* IS was located within the protein coding sequence (CDS) of the gene. *ARIH2* is a E3 ubiquitin-protein ligase that is linked to NLRP3 inflammasome activation (see Discussion).

HIV integrations were mapped to one position within intron 1 of the overlapping genes *copine 1 (CPNE1)* and *RNA binding motif protein (RBM12)* in Participant 1 **(Figure 2.8 B)**. This IS was oriented in the forward orientation relative to the gene(s) and was the only major clone detected in the cultured CD4 T cells found to be integrated upstream from the start codon. We hypothesized this orientation could lead to HIV promoter-driven insertional activation and thus, dysregulation of the *CPNE1/RBM12* gene(s). *CPNE1/RBM12* are partially overlapping genes that share the same promoter and most 5’ exons, but that encode two proteins with different functions in different reading frames. *CPNE1* is a calcium-dependent phospholipid-binding protein associated with cell proliferation, inhibition of apoptosis, and metastasis (see Discussion). The exact function of *RBM12,* a protein containing RNA binding motifs and potentially transmembrane domains, has not been clearly defined.

A majority of Participant 2's IS in cultured CD4+ T cells were in one position within the *mitogen-activated protein kinase kinase kinase 3 (MAP3K3)* gene indicating that integration into *MAP3K3* may promote cell proliferation. HIV integration into *MAP3K3* occurs downstream of the start codon, within the CDS. *MAP3K3* is a critical regulatory kinase that regulates many cellular functions including cell proliferation, apoptosis, and differentiation through its ability to activate a diverse array of MAPKs. *MAP3K3* overexpression has also been shown to promote cell proliferation and inhibit of apoptosis. In the context of HIV persistence, the potential for HIV-induced dysregulation of *MAP3K3* gene presents a rational mechanism for clonal expansion of HIV-infected cells due to the well-defined role of *MAP3K3* in cell proliferation and apoptosis.

Figure 2.7. Chromosome distribution of HIV integration sites from uncultured CD4+ T cells. Location of HIV integration sites (IS) derived from uncultured CD4+ T cells from Participant 1 (A) and Participant (2). Overall gene distribution across the human genome is represented by chromosome heatmaps. Forward- vs reverse-oriented IS and genic (intron/exon) vs intergenic IS is noted. Relative size of clones can be inferred by the number of IS noted from a total of 582 and 138 IS derived from participants. The location of major clones observed in the cultured IS dataset are noted in red and the enlarged boxes.


A
157 ARIH2
RBM12/CPNE1
Low
High
exon R
intron R
exon F
intron F
intergenic
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y


B
MAP3K3
Low
High
exon R
intron R
exon F
intron F
intergenic
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y

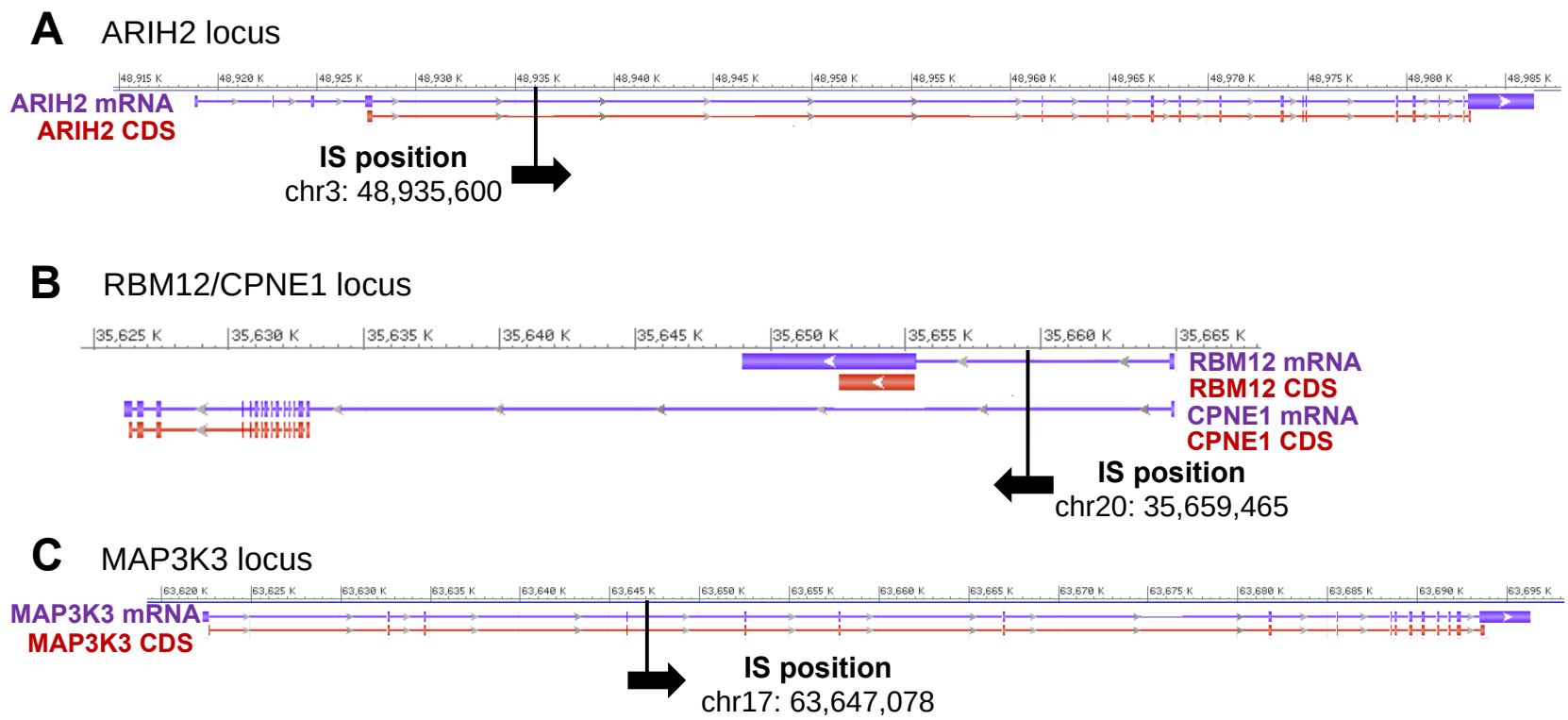


Figure 2.8. Genomic locus of clonal proviruses identified from cultured CD4+ T cells. The gene locus of integration sites (IS) defined as major clones are mapped against the corresponding human chromosome. The exons corresponding to the primary mRNA transcript are shown as purple boxes. Intron sequence is represented by the thin purple line. The protein coding sequence (CDS) of the gene is shown as red boxes. The direction of the arrow on the red and purple boxes indicates the orientation of host gene transcription whereas the black arrow represented the orientation of IS with respect to the genome

Biological pathways associated with unique IS enriched in cultured compared to uncultured CD4+ T cells

Biological pathway analysis was performed to determine if IS identified in cultured CD4+ T cells are enriched in genes associated with proliferation compared to uncultured (i.e., “total”) CD4+ T cells. After post-hoc analysis, several pathways were enriched. In Participant 1 these included pathways associated with regulation of cell proliferation and survival were enriched ($p < 0.0475$ or $p > 1.32$ of -log10 in Participant 1) in IS from cultured CD4+ T cells versus uncultured CD4+ T cells **(Figure 2.9 A)**. In Participant 2, only one gene set was enriched in IS from cultured CD4+ T cells versus uncultured CD4+ T cells (NIKOLSKY_BREAST_CANCER_17Q21_Q25_AMPLICON, $p < 0.012$ or $p > 1.91$ of -log10; data not shown). In neither participant did enriched gene sets have tier 1 or tier 2 significance (see Methods; more stringent requirements for statistical significance) Due to the low number of unique integration sites observed in the cultured CD4+ IS data sets, our analysis included any gene set with any IS at all across comparison groups (cultured versus uncultured IS) in these analyses (see Methods for more details). These results were therefore based on very small numbers of unique IS (N = 4 unique IS for Participant 1; N = 5 unique IS for Participant 2) and should be interpreted with caution.

In the cultured CD4+ T cell IS data sets, 1-2 clones make up the majority of the IS data set. Therefore, we performed an analysis to assess the collective power of the clones in contributing to gene set enrichment. Many gene sets appear to be statistically significant at tier 1 including those involved in cell proliferation and survival for Participant 1 **(Figure 2.9 B)**; and those involved in IL-1 signaling pathway and MAPK signaling cascade for Participant 2 **(Figure 2.9 C)**.

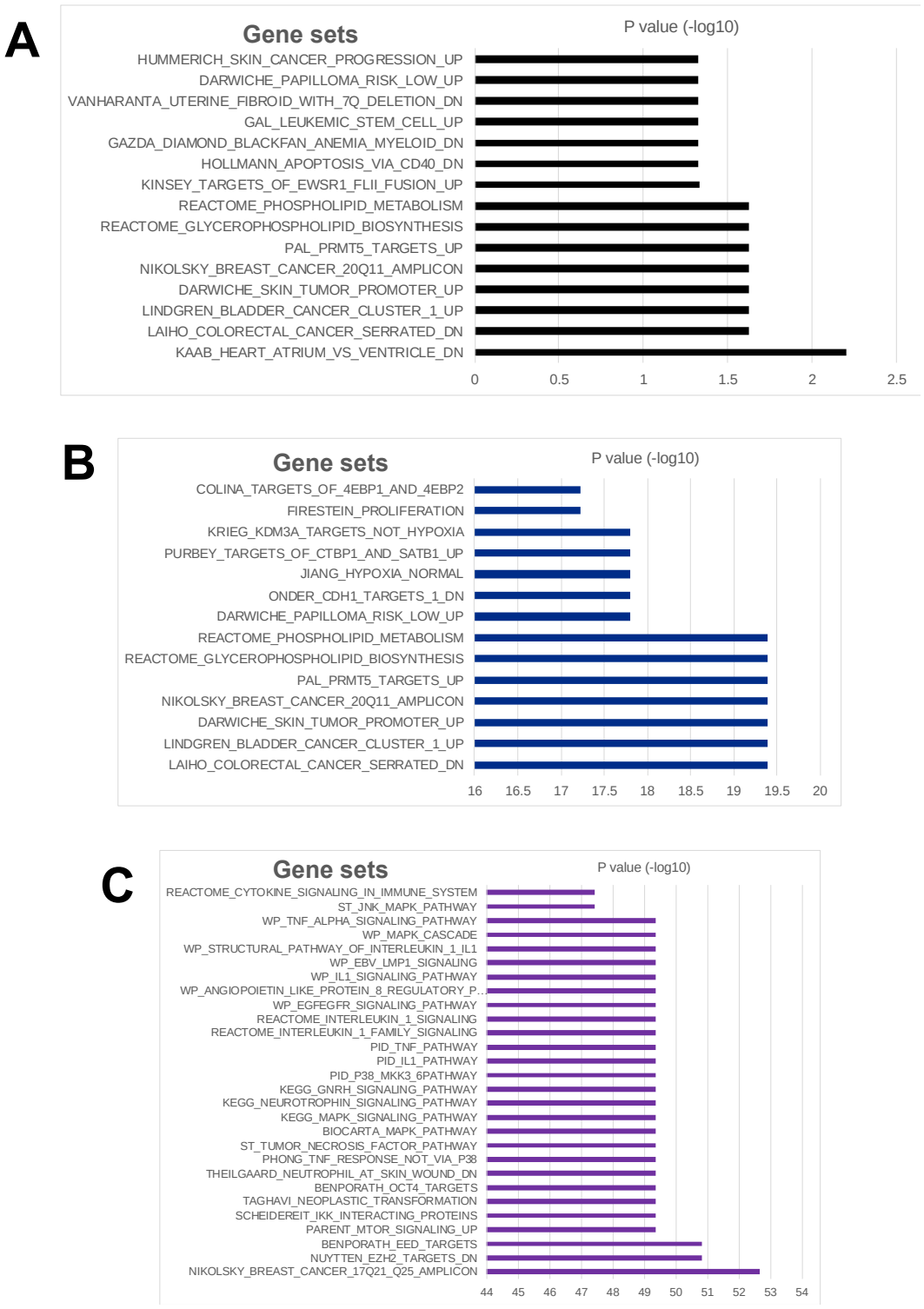


Figure 2.9. Biological pathways associated with cell proliferation, survival and immune signaling cascades were enriched in cultured CD4+ T cells. Gene set enrichment of unique IS from Participant 1 from the cultured vs uncultured CD4+ T cell IS data set (A). Gene set enrichment of all IS from Participant 1 (B) and Participant 2 (C) from the cultured vs uncultured CD4+ T cell IS data set.

Detection of HIV-human hybrid transcripts

We hypothesize that the high frequency clones (*ARIH2, RBM12/CPNE1, MAP3K3)* observed in the cultured CD4+ T cells preferentially persist during culture due to HIV-1-mediated insertional activation of the host gene resulting in either enhanced transcription of the integrated gene or altered protein structure of the integrated gene product. We focused our analysis on HIV-1 insertions targeting *CPNE1*, which occurred in the same orientation of gene transcription and upstream of the first protein-coding exon, which could provide a selective advantage to infected cell clones. We quantified chimeric HIV-*CPNE1* transcripts by nested RT-PCR using primers that targeted the R region of the viral long terminal repeat (LTR) promoter and the 3^{rd}/4^{th} exons of *CPNE1.* Alternative splicing of *CPNE1* results in nine mRNA transcript variants encoding different *CPNE1* proteins. Therefore, we designed primers to detect HIV-*CPNE1* aberrant transcripts based on the primary *CPNE1* transcript, transcript 1, which encodes 16 exons, 15 of which are coding exons, with the protein coding sequence beginning at exon 2 **(Figure 2.10 A)**.

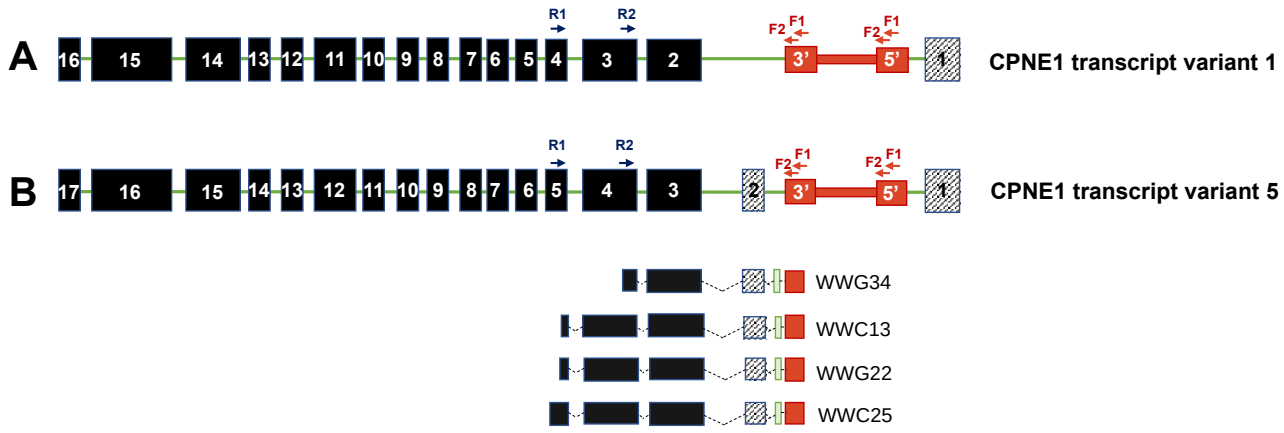


Figure 2.10. Chimeric HIV/*CPNE1* transcripts in negatively selected CD4+ T cells after culture with homeostatic cytokines. (A) Strategy to amplify putative HIV/*CPNE1* chimeric transcripts based on the primary *CPNE1* transcript variant 1. The *CPNE1* gene is in the reverse orientation with respect to the genome, whereas HIV is integrated in the forward orientation with respect to the gene. The exon numbers are noted; coding exons are shown by the black boxes whereas non-coding exons are noted by the striped boxes. Intronic regions are indicated by the green line. Nested primer sets are shown as the navy and red arrows. HIV LTR primers can prime from either the 5' or 3' LTR. Approximate location of the HIV provirus within *CPNE1* is noted in red in reverse orientation. A full-length, intact HIV provirus is depicted with both LTRs for illustration purposes only; we did not evaluate proviral sequences for replication competency. (B) Structure of the alternative *CPNE1* splice variant 5 observed in all chimeric HIV-*CPNE1* splice products. Below that are the approximate structures of the four HIV-*CPNE1* transcripts detected.

We detected human-HIV chimeric transcripts with aberrant transcription between the viral promoter and *CPNE1* within the expected size of 400 - 600bp. Sequence analysis of the RT-PCR products revealed the expression of an alternative *CPNE1* splice variant, *CPNE1* transcript variant 5 as opposed to the primary *CPNE1* transcript variant 1. *CPNE1* transcript variant 5 has 17 total exons which includes one additional non-coding exon prior to the coding sequence (**Figure 2.10 B,** exon 2). HIV-host chimeric RNAs incorporated roughly ~90 nucleotides of the LTR sequence (corresponds to the location of forward primers; F1, F2) but lacked the major splice donor (MSD) sequence. The MSD sequence is located downstream of the 5' LTR in the untranslated 5' leader region of the

viral RNA and is thought to play a role in HIV-human splicing due to presence of the HIV major splice donor, D1.

Interestingly, chimeric transcripts contained viral sequence (LTR), intronic sequence, the non-coding exon 2 of *CPNE1,* and coding exons 3-5 of *CPNE1* **(Figure 2.10 B)**. The retention of intronic sequence in chimeric mRNA transcripts was highly unusual and is discussed in more detail in the Discussion below. Retention of intronic sequence resulted in a shifted reading frame that started at an out-of-frame start codon from within *CPNE1* intron sequence as opposed to the proper translation start site in exon 3. This led us to believe that the reading frame had shifted, potentially as a result of HIV integration.

2.4 Discussion

Studies into the mechanisms driving *in vivo* HIV persistence are hindered by the extremely low frequency of HIV-infected cells in the blood of ART-treated individuals (~10-100 per $1x10^6$ CD4 T cells) (Eriksson et al. 2013; Ho et al. 2013) and the lack of a surface marker to uniquely identify latently infected cells (Osuna et al. 2018). To study how distinct clones contribute to HIV persistence and clonal expansion dynamics, we needed to find a way to reliably and repeatedly isolate clonal cell populations with minimal manipulation as to minimally impact the *in vivo* gene expression profile of infected cells. Therefore, we designed an assay to enrich for HIV-infected cells selected *in vitro* for phenotypes likely to persist during ART-suppression, such as proliferation and resistance to apoptosis and autophagy. By using leukapheresis samples and by developing a strategy that might enrich for clones of HIV-infected cells, we address some of these technical challenges presented by the low frequency of HIV-infected cells in the blood.

When analyzing the integration site distribution in cultured CD4+ T cells in HIV-infected individuals on long-term ART, we chose to compare against the IS observed to each individuals' uncultured CD4+ T cells characterized from an HIV-infected individual on long-term ART. Examining HIV integration sites from uncultured CD4+ T cells from an on-ART time point, particularly a 12+ years on-ART timepoint (**Table 2.2**) will provide a snapshot of the persisting proviruses that have survived through *in vivo* selection forces. We predicted that through our homeostatic proliferation selection method we'd be able to enrich for *clonal cell populations* that have persisted during ART-suppression and/or enrich for cells by potentially IS induced properties.

Assay reproducibility was shown by repeated detection of the same IS from experiments 1 and 2 in Participants 1 and 2 (**Figure 2.6,** bolded IS). Although enrichment of HIV DNA load in cultured CD4+ T cells across experiments was variable (**Figure 2.4 and 2.5**), we were encouraged to see the same integration sites emerge multiple times across unique experiments in each individual (**Figure 2.6**). For example, the major clone in uncultured cells, *ARIH2* was also observed in Participant 1 across pooled experiments (Experiments 1 and 2). Given that *ARIH2* was the major clone in the uncultured cells, it could have remained simply because cells died slowly in culture. However, cells with this HIV IS increased in proportion to the total DNA. Reproducible enrichment of specific HIV IS is more convincing for Participant 2. The *MAP3K3* IS was observed across both experiments 1 and 2, but not amongst IS from uncultured cells. This indicates there was no scenario in which different major clones emerge in each culturing experiment. The fact that we observe the same clones emerge across experiments 1 and 2 indicates a preferred fitness of these HIV-infected clones to expand in homeostatic proliferation conditions. We conclude that our enrichment method, based on outgrowth of cells in homeostatic proliferation conditions, can reliably expand distinct HIV-infected clones with a proliferative survival advantage.

We observed a poor overlap between IS that emerged from cultured versus uncultured cells, on a per individual basis. In fact, 2 major clones, *RBM12/CPNE* and *MAP3K3* were not observed in the uncultured CD4 IS data set at all. The limited overlap between integration sites in cultured and uncultured cells could be due to low sampling depth in the uncultured CD4 IS data set. The true diversity of the persistent reservoir is always underestimated due to incomplete sampling. IS analysis of hundreds of thousands

of infected cells from a single time point would be required to capture the true population richness of the persistent reservoir, an effort that can only perhaps be achieved by analyzing IS from postmortem tissues or from a massive volume of blood (Reeves et al. 2018). Although we identified 582 and 138 total IS in the uncultured CD4 IS data set for Participants 1 and 2, respectively, this sampling depth provides only a limited picture the reservoir. Additionally, we do not believe that the IS identified only in cultured CD4s were a product of *de novo* infection of cells in culture. A combination of antiretroviral drugs directed at multiple distinct HIV steps in viral replication (efavirenz to target HIV reverse transcription, raltegravir to target HIV integration and maraviroc to target HIV entry) were renewed in the cultured T cells every three days at an effective concentration, which should prevent full cycles of HIV replication during the culturing period, and thus should prevent new integration events.

Given the length of our experiment time of 24 - 74 days, we believe that the proliferation of cultured cells is primarily driven by homeostatic proliferation as a result from signaling from homeostatic cytokines such as IL-2, IL-7 and IL-15 (Min 2018; Surh and Sprent 2008). It is recognized that homeostatic proliferation can occur in varying degrees: *fast homeostatic proliferation* which is characterized by IL-7-independent rapid cellular division mediated by low avidity TCR-contact (antigen-driven proliferation), and *slow proliferation* which is characterized by IL-7-driven slow cellular division of T cells that does not result in a change of phenotype (Min et al. 2005; Rosado-Sánchez et al. 2018). It is difficult to know if expanded HIV-infected T cells in our assay were driven to proliferate by a fast or slow form of homeostatic proliferation because the rates of homeostatic proliferation can massively differ between CD4+ T cell subsets that receive the same

homeostatic stimuli (IL-7, IL-15, IL-2) and we did not determine the phenotype of CD4+ T cells before or after culturing. However, I would argue that in our relatively controlled *in vitro* experiment, there should be little to no antigen exposure, arguing against fast proliferation as a mechanism as it requires low avidity TCR contact with either self Ag or foreign Ag (Min et al. 2005; Rosado-Sánchez et al. 2018). Additionally, it was previously described that homeostatic proliferation of HIV-infected cells can occur without viral antigen expression (Bosque et al. 2011; Katlama et al. 2016) thus leading me to think our culturing method should not induce viral antigen production that could trigger fast homeostatic proliferation

Another alternative for the cell proliferation observed in CD4+ T cell cultures could be due to HIV integration site driven clonal expansion, rather than homeostatic proliferation driven clonal expansion. Proviruses integrated into genes thought to confer a selective advantage for the survival of infected cells (i.e. *BACH2, STAT5B*) can be observed as early as 12 months into ART treatment (Cesana et al. 2017). Forced expression of these genes can confer a proliferative advantage to CD4+ T cells as early as 10 days after expansion (Cesana et al. 2017). Therefore, it is possible that misexpression of a gene (identified as a major clone in cultured CD4+ T cells such as *CPNE1*) as a consequence of insertional mutagenesis could have driven the clonal outgrowth we observed during the relatively short culturing period.

We found that multiple gene pathways associated with cell proliferation and cell survival were in enriched in cultured CD4 T cells as compared to uncultured CD4 T cells. This was largely driven by enrichment of molecular pathways regulating cell proliferation, survival and cell metabolism in cancer cells, suggesting that proviruses identified in

cultured cells potentially modulated cell function to promote excessive cell proliferation. We believe that upregulation of these pathways may promote the survival of infected clones under conditions of homeostatic proliferation and cell stress such as apoptosis and autophagy. When we factored the multiplicity of a particular clone into the analysis, we detected enrichment of pathways related to cell proliferation and immune signaling cascades (such as IL-1 and MAPK signaling). We think that large clones of HIV-infected cells up or down regulate the expression of genes controlling immune responses. This is potentially supported by this finding as these specific pathways were not enriched when the analysis was based on unique IS only as opposed to total IS.

The *ARIH2* IS may promote HIV persistence *in vivo* and in cultured CD4 T cells*. ARIH2* was the only IS detected both in cultured and uncultured CD4+ T cell IS datasets (**Ie**). *ARIH2* is a E3 ubiquitin-protein ligase that is linked to the immune response and NLRP3 inflammasome activation (Kawashima et al. 2017; Lin et al. 2013). The *ARIH2* IS was oriented in the forward orientation relative to the gene and interrupts the protein coding region which could lead to the expression of truncated protein. Protein truncation has occurred in other examples where HIV integration occurs downstream of the translation start codon (Liu et al. 2020). While we did not perform sequencing to evaluate full length proviruses, in complementary experiments, our collaborators detected and sequenced this clone and detected a small deletion in LTR-gag. They also determined that cells harboring the *ARIH2* IS (1) quickly outgrow uninfected cells in culture and (2) upregulate expression of genes associated with T cell exhaustion and immune checkpoint inhibition such as *LAG3, TIGIT* and *CTLA-4* using single-cell transcriptome sequencing (10x Genomics, Pleasanton, CA). These markers are closely associated with HIV

persistence during ART; upregulation of these markers likely prevents targeting for elimination by cytotoxic T lymphocytes (Fromentin et al. 2016).

The *CPNE1* IS identified in cultured CD4s may drive persistence of cultured cells through insertional mutagenesis by HIV. We detected aberrant HIV-*CPNE1* mRNA transcripts in the cultured CD4+ T cells of Participant 1. The finding of retained intronic sequence from the untranslated region of *CPNE1* in the chimeric transcripts was unexpected and perplexing. In addition, we observed that HIV integration into the *CPNE1* gene disrupts normal cellular splicing pathways, enabling detection of a minority *CPNE1* transcript variant (*CPNE1* transcript variant 5) as opposed to the primary *CPNE1* transcript, transcript variant 1. Not much is known about the specific function of the translated transcript variant 5, but it is thought that each *CPNE1* transcript variant encodes for a different protein (O'Leary et al. 2016). Intron retention (IR), defined as the retention of unspliced introns in mature polyadenylated transcripts, was previously thought to be a less common form of alternative splicing (Braunschweig et al. 2014). However, IR is becoming increasingly recognized as an important regulatory mechanism of mammalian gene expression (Braunschweig et al. 2014) and has been previously observed in aberrant HIV-1-host chimeric transcripts (Liu et al. 2020; Sherrill-Mix, Ocwieja, and Bushman 2015). Intron retention is most often associated with down-regulation of gene expression via nonsense mediated decay (Ge and Porse 2014; Jacob and Smith 2017). Interestingly, intron retention of *CPNE1* has been found in studies of aging to contribute to immunosenescence-associated phenotypes (Yao et al. 2020). In effect, properly spliced *CPNE1* transcripts compete with intron-retained *CPNE1* transcripts, which are rapidly degraded, resulting in decreased normal translation

template and reduced *CPNE1* protein production, ultimately leading to senescence-associated phenotypes (Yao et al. 2020). While it remains to be further investigated, HIV IS-mediated intron retention could present a novel mechanism of HIV-driven dysregulation of host gene expression.

Overall, the patterns observed in chimeric HIV-*CPNE1* transcripts such as retention of intron sequence and out-of-frame transcription are consistent with LTR-driven host gene transcription (Liu, Simonetti, and Ho 2020). We did not detect inclusion of the MSD sequence in the HIV-*CPNE1* chimeric transcripts, arguing against HIV-human splicing as the MSD sequence is thought to be important for HIV-host splicing. The LTR sequence we detected in the HIV-*CPNE1* chimeric transcripts does not cover the HIV-1 polyadenylation (polyA) signal site so we cannot infer if transcription was initiated from the 5' or 3' LTR. The 5' and 3' LTR sequences flanking the proviral genome are duplicated copies of each other and contain identical polyA signal sites, which primarily function to control viral transcription. However, to allow proper viral expression of the entire HIV-1 genome, the use of the 3' LTR poly(A) site must be restricted (Ashe et al. 1995). If the 3' LTR polyA site is absent or nonfunctional in the cells harboring the *CPNE1* IS, this restriction is released and transcription of HIV-1-host chimeric RNAs can proceed from the 3' LTR. Conversely, if the 3' LTR polyA site is intact, transcription of HIV-1-host chimeric RNAs can proceed from the 5' LTR (Ashe et al. 1995; Liu, Simonetti, and Ho 2020). The size of the chimeric transcripts (400 – 600bp) would argue against transcription from the 5'LTR into the *CPNE1* gene as a transcript promoted from the 5' LTR should be >10kb (the approximate size of the HIV-1 genome). This led us to believe that the transcription of HIV-human chimeric transcripts is promoted by bypass of the viral

polyadenylation site within the 5’ LTR, enabling transcription from the 3’ LTR into the *CPNE1* gene.

Due to experimental challenges related to simultaneous analysis of antigen responsiveness, HIV IS, and proviral sequence, we cannot differentiate between factors leading to enrichment (i.e. antigen-driven, IS-driven, homeostatic proliferation-driven) in our *ex vivo* selection method. We cannot conclusively define the relative contribution of IS-driven vs homeostatic proliferation-driven mechanisms of clonal expansion. Additionally, we cannot eliminate antigen-driven clonal selection as a major factor for these results because we do not know the antigen specificity of our clonally expanded cells.

Chapter 3. Conclusions and Study Significance

3.1 Summary of findings

The importance of homeostatic proliferation and integration-site proliferation in driving clonal selection is supported by this work through 1) identification of a subpopulation of HIV-infected cells with a preferred fitness to expand in homeostatic proliferation conditions and 2) the identification of intron retention as a potentially novel mechanism of HIV-driven dysregulation of host gene expression.

We have developed an experimental system to enrich for HIV-infected cells by selecting for phenotypes associated with *in vivo* persistence such as proliferation and resistance to apoptosis and autophagy. This strategy reproducibly enriched clones of HIV-infected cell populations. In addition, we were able to promote expansion of rare cell populations not observed in uncultured cells. The HIV-infected clones that only emerged after homeostatic proliferation culturing likely reflect a preferred fitness of these HIV-infected cells to expand in homeostatic proliferation conditions.

We originally hypothesized that HIV IS might contribute to survival advantage of *ex vivo* expanded CD4+ T cells by altering gene expression of the integrated gene. To that end, we found that HIV-1 drives aberrant host gene transcription at the integration site of an expanded cellular clone (*CPNE1)*, identified in cultured CD4+ T cells. Sequencing of the provirus in these clones could provide additional support as to whether hybrid transcripts were promoted from the 3'LTR.

Gene pathways associated with regulation of cell proliferation and survival were enriched in cultured compared to each participant's uncultured CD4+ T cell populations, which suggested that the provirus modulated cell function. Therefore, further study of

these unique HIV-infected clones, including gene expression may allow us to identify the gene pathways that distinguish these cells from other uninfected CD4 to better understand the mechanisms that favored their survival under conditions of homeostatic proliferation.

3.2 Future directions

With this experimental system in place, we plan to perform single cell analysis to understand the effects of HIV-1 mediated insertion on the cell transcriptome. We also plan to sequence the proviral genome in future work. Additionally, enrichment of HIV DNA to concentrations suitable for single cell transcriptome analysis will enable, to our knowledge, the first linkage between *in vivo* HIV IS to a specific transcriptional profile that could alter CD4 T cell function and/or lead to maintenance of the persistent reservoir. By virtue of the diverse primer sets spanning the entire HIV genome for TMO MDA, we can use the MDA products to perform full length proviral sequencing of cultured CD4 T cells. Full length proviral sequencing will enable us to (1) determine if expanded clones harbor intact proviruses, (2) progressively sequence the entire LTR sequence and thus provide better support as to whether hybrid transcripts were promoted from the 5' or 3' LTR. In future work, we also plan to investigate the effect of HIV-induced intron retention in promoting cell survival and HIV persistence using the *HIV/CPNE1* chimeric transcripts as one potential example. We are interested in determining the primary fate of intron retained *CPNE1* hybrid transcripts – are these aberrant transcripts translated into aberrant protein, or are they degraded by a mRNA surveillance pathway such as nonsense mediated decay? Ultimately the goal is to determine the effect of chimeric HIV-host transcripts on infected cell phenotype.

3.3 Conclusion

In conclusion, we provide additional support for HIV-integration site-driven as a potent mechanism for clonal expansion due to the identification of intron retention as a potentially novel mechanism of HIV-driven dysregulation of host gene expression. The intention is to develop and evaluate novel strategies for an HIV cure based on the knowledge that a diverse repertoire of mechanisms can drive HIV-driven dysregulation of integrated gene expression. We anticipate use of our experimental system to continue to study HIV-1 persistence mechanisms in a greater number of HIV-infected individuals to further understand phenotypes of HIV-infected cells likely to persist during ART-suppression.

References

Aamer, Hadega A. et al. 2020. “Cells Producing Residual Viremia during Antiretroviral Treatment Appear to Contribute to Rebound Viremia Following Interruption of Treatment.” *PLoS Pathogens* 16(8): 1–24. http://dx.doi.org/10.1371/journal.ppat.1008791.

Abram, Michael E. et al. 2010. “Nature, Position, and Frequency of Mutations Made in a Single Cycle of HIV-1 Replication.” *Journal of Virology* 84(19): 9864–78.

Arvold, Nils D. et al. 2007. “Maternal HIV-1 DNA Load and Mother-to-Child Transmission.” *AIDS Patient Care and STDs* 21(9): 638–43.

Ashe, M. P., P. Griffin, W. James, and N. J. Proudfoot. 1995. “Poly(A) Site Selection in the HIV-1 Provirus: Inhibition of Promoter- Proximal Polyadenylation by the Downstream Major Splice Donor Site.” *Genes and Development* 9(23): 3008–25.

Bailey, Justin R. et al. 2006. “Residual Human Immunodeficiency Virus Type 1 Viremia in Some Patients on Antiretroviral Therapy Is Dominated by a Small Number of Invariant Clones Rarely Found in Circulating CD4+ T Cells.” *Journal of Virology* 80(13): 6441–57.

Baldauf, Hanna Mari et al. 2012. “SAMHD1 Restricts HIV-1 Infection in Resting CD4 + T Cells.” *Nature Medicine* 18(11): 1682–87.

Borrow, Persephone et al. 1994. "Virus-Specific CD8+ Cytotoxic T-Lymphocyte Activity Associated with Control of Viremia in Primary Human Immunodeficiency Virus Type 1 Infection." *Journal of Virology* 68(9): 6103-10.

Bosque, Alberto et al. 2011. “Homeostatic Proliferation Fails to Efficiently Reactivate HIV-1 Latently Infected Central Memory CD4+ T Cells.” *PLoS Pathogens* 7(10).

Boyman, Onur, Jared F. Purton, Charles D. Surh, and Jonathan Sprent. 2007. “Cytokines and T-Cell Homeostasis.” *Current Opinion in Immunology* 19(3): 320–26.

Brady, Troy et al. 2009. “HIV Integration Site Distributions in Resting and Activated CD4+ T Cells Infected in Culture.” *AIDS* 23(12): 1461–71.

Braunschweig, Ulrich et al. 2014. “Widespread Intron Retention in Mammals Functionally Tunes Transcriptomes.” *Genome Research* 24(11): 1774–86.

Brenchley, Jason M. et al. 2004. “T-Cell Subsets That Harbor Human Immunodeficiency Virus (HIV) In Vivo: Implications for HIV Pathogenesis.” *Journal of Virology* 78(3): 1160–68.

Buchbinder, Susan P. et al. 2008. “Efficacy Assessment of a Cell-Mediated Immunity HIV-1 Vaccine (the Step Study): A Double-Blind, Randomised, Placebo-Controlled,

Test-of-Concept Trial." *The Lancet* 372(9653): 1881–93.

Budhiraja, S. et al. 2013. "Cyclin T1 and CDK9 T-Loop Phosphorylation Are Downregulated during Establishment of HIV-1 Latency in Primary Resting Memory CD4+ T Cells." *Journal of Virology* 87(2): 1211–20.

Bui, John K. et al. 2017. "Proviruses with Identical Sequences Comprise a Large Fraction of the Replication-Competent HIV Reservoir." *PLoS Pathogens* 13(3): 1–18.

Camargo, Jose F. et al. 2009. "Responsiveness of T Cells to Interleukin-7 Is Associated with Higher CD4 + T Cell Counts in HIV-1-Positive Individuals with Highly Active Antiretroviral Therapy-Induced Viral Load Suppression." *Journal of Infectious Diseases* 199(12): 1872–82.

Cameron, Paul U. et al. 2010. "Establishment of HIV-1 Latency in Resting CD4+ T Cells Depends on Chemokine-Induced Changes in the Actin Cytoskeleton." *Proceedings of the National Academy of Sciences of the United States of America* 107(39): 16934–39.

Cesana, Daniela et al. 2017. "HIV-1-Mediated Insertional Activation of STAT5B and BACH2 Trigger Viral Reservoir in T Regulatory Cells." *Nature Communications* 8(1).

Chaillon, Antoine et al. 2020. "HIV Persists Throughout Deep Tissues with Repopulation from Multiple Anatomical Sources." *Journal of Clinical Investigation* 130(4): 1699-712.

Chan, David C, and Peter K Kim. 1998. "HIV Entry and Its Inhibition." *Cell* 93: 681–84.

Chomont, Nicolas et al. 2009. "HIV Reservoir Size and Persistence Are Driven by T Cell Survival and Homeostatic Proliferation." *Nature Medicine* 15(8): 893–900.

Chun, Tae-Wook et al. 1995. "In Vivo Fate of HIV-1-Infected T Cells: Quantitative Analysis of the Transition to Stable Latency." *Nature Medicine* 1(12): 1284–90.

Chun, Tae-Wook, Lieven Stuyver, et al. 1997. "Presence of an Inducible HIV-1 Latent Reservoir during Highly Active Antiretroviral Therapy." *Proceedings of the National Academy of Sciences of the United States of America* 94(24): 13193–97.

Chun, Tae-Wook, Lucy Carruth, et al. 1997. "Quantification of Latent Tissue Reservoirs and Total Body Viral Load in HIV-1 Infection." *Nature* 387: 183–88.

Chun, Tae-Wook et al. 1998. "Early Establishment of a Pool of Latently Infected, Resting CD4+ T Cells during Primary HIV-1 Infection." *Proceedings of the National Academy of Sciences of the United States of America* 95(15): 8869–73.

Coffin, John, et al. 1997. *Retroviruses.* Cold Spring Harbor, NY: Cold Spring Harbor

Laboratory Press.

Cohen, Myron S. et al. 2011. “Prevention of HIV-1 Infection with Early Antiretroviral Therapy.” *New England Journal of Medicine* 365(6): 493–505.

Cohn, Lillian B. et al. 2015. “HIV-1 Integration Landscape during Latent and Active Infection.” *Cell* 160(3): 420–32. http://dx.doi.org/10.1016/j.cell.2015.01.020.

———. 2020. “Antigen Driven Clonal Expansion of HIV-Infected Cells.” In *Strategies for an HIV Cure Conference*,.

Cohn, Lillian B et al. 2016. “Paired Quantitative and Qualitative Assessment of the Replication-Competent HIV-1 Reservoir and Comparison with Integrated Proviral DNA.” *Proceedings of the National Academy of Sciences of the United States of America* 113(49): E7908–16.

Colby, Donn J. et al. 2018. “Rapid HIV RNA Rebound after Antiretroviral Treatment Interruption in Persons Durably Suppressed in Fiebig i Acute HIV Infection Brief-Communication.” *Nature Medicine* 24(7): 923–26.

Collins, David R. et al. 2020. "CD8+ T Cells in HIV Control, Cure and Prevention." *Nature Reviews Immunology* 20(8): 471-82.

Coull, Jason J. et al. 2000. “The Human Factors YY1 and LSF Repress the Human Immunodeficiency Virus Type 1 Long Terminal Repeat via Recruitment of Histone Deacetylase 1.” *Journal of Virology* 74(15): 6790–99.

Craigie, Robert, and Frederic D. Bushman. 2012. “HIV DNA Integration.” *Cold Spring Harbor Perspectives in Medicine* 2(7).

Day, Cheryl L. et al. 2006. "PD-1 Expression on HIV-Specific T Cells is Associated with T-Cell Exhaustion and Disease Progression." *Nature* 443(7109): 350-54.

Deng, Kai et al. 2015. "Broad CTL Response is Required to Clear Latent HIV-1 Due to Dominance of Escape Mutations." *Nature* 517(7534): 381-85.

Dornadula, Geethanjali et al. 1999. “Residual HIV-1 RNA in Blood Plasma of Patients Taking Suppressive Highly Active Antiretroviral Therapy.” *Journal of the American Medical Association* 282(17): 1627–32.

Douek, Daniel C. et al. 2002. “HIV Preferentially Infects HIV-Specific CD4+ T Cells.” *Nature* 417(6884): 95–98.

Dybul, Mark et al. 2020. “The Case for an HIV Cure and How to Get There.” *The Lancet HIV* 3018(20).

Edlefsen, Paul T et al. 2020. *Unpublished*.

Einkauf, Kevin B. et al. 2019. “Intact HIV-1 Proviruses Accumulate at Distinct Chromosomal Positions during Prolonged Antiretroviral Therapy.” *The Journal of clinical investigation* 129(3): 988–98.

Eriksson, Susanne et al. 2013. “Comparative Analysis of Measures of Viral Reservoirs in HIV-1 Eradication Studies.” *PLoS Pathogens* 9(2).

Fan, Hung, and Chassidy Johnson. 2011. “Insertional Oncogenesis by Non-Acute Retroviruses: Implications for Gene Therapy.” *Viruses* 3(4): 398–422.

Flynn, N. M. et al. 2005. “Placebo-Controlled Phase 3 Trial of a Recombinant Glycoprotein 120 Vaccine to Prevent HIV-1 Infection.” *Journal of Infectious Diseases* 191(5): 654–65.

Frank, Tahvi D. et al. 2019. “Global, Regional, and National Incidence, Prevalence, and Mortality of HIV, 1980-2017, and Forecasts to 2030, for 195 Countries and Territories: A Systematic Analysis for the Global Burden of Diseases, Injuries, and Risk Factors Study 2017.” *The Lancet HIV* 6(12): e831–59.

Fromentin, Rémi et al. 2016. “CD4+ T Cells Expressing PD-1, TIGIT and LAG-3 Contribute to HIV Persistence during ART.” *PLoS Pathogens* 12(7): 1–19.

Gandhi, Monica, and Rajesh T. Gandhi. 2014. “Single-Pill Combination Regimens for Treatment of HIV-1 Infection.” *New England Journal of Medicine* 371(3): 248–59.

Ganor, Yonatan et al. 2019. "HIV-1 Reservoirs in Urethral Macrophages of Patients Under Suppressive Antiretroviral Therapy." *Nature Microbiology* 4(4): 633-44.

Gantner, Pierre et al. 2020. “Single-Cell TCR Sequencing Reveals Phenotypically Diverse Clonally Expanded Cells Harboring Inducible HIV Proviruses during ART.” *Nature Communications* 11(1): 1–9. http://dx.doi.org/10.1038/s41467-020-17898-8.

Ge, Ying, and Bo T. Porse. 2014. “The Functional Consequences of Intron Retention: Alternative Splicing Coupled to NMD as a Regulator of Gene Expression.” *BioEssays* 36(3): 236–43.

Greger, Ingo H., Francesca Demarchi, Mauro Giacca, and Nick J. Proudfoot. 1998. “Transcriptional Interference Perturbs the Binding of Sp1 to the HIV-1 Promoter.” *Nucleic Acids Research* 26(5): 1294–1300.

Halvas, Elias K. et al. 2020. “HIV-1 Viremia Not Suppressible by Antiretroviral Therapy Can Originate from Large T-Cell Clones Producing Infectious Virus.” *Journal of Clinical Investigation* 130(11): 5847–57.

Han, Yefei et al. 2008. “Orientation-Dependent Regulation of Integrated HIV-1 Expression by Host Gene Transcriptional Readthrough.” *Cell Host and Microbe* 4(2): 134–46.

Haynes, Barton F. et al. 2012. "Immune-Correlates Analysis of an HIV-1 Vaccine Efficacy Trial." *New England Journal of Medicine* 366(14): 1275-86.

Ho, David D. et al. 1986. "Infection of Monocyte/Macrophages by Human T Lymphotrophic Virus Type III." *Journal of Clinical Investigation* 77(5): 1712-15.

Ho, Ya Chi et al. 2013. “Replication-Competent Non-Induced Proviruses in the Latent Reservoir Increase Barrier to HIV-1 Cure.” *Cell* 155(3): 540–51.

Hosmane, Nina N. et al. 2017. “Proliferation of Latently Infected CD4+ T Cells Carrying Replication-Competent HIV-1: Potential Role in Latent Reservoir Dynamics.” *Journal of Experimental Medicine* 214(4): 959–72.

Huang, Jialing et al. 2007. “Cellular MicroRNAs Contribute to HIV-1 Latency in Resting Primary CD4 + T Lymphocytes.” *Nature Medicine* 13(10): 1241–47.

Ikeda, Terumasa et al. 2007. “Recurrent HIV-1 Integration at the BACH2 Locus in Resting CD4+ T Cell Populations during Effective Highly Active Antiretroviral Therapy.” *Journal of Infectious Diseases* 195(5): 716–25.

Imamichi, Hiromi et al. 2014. “Lifespan of Effector Memory CD4+ T Cells Determined by Replication-Incompetent Integrated HIV-1 Provirus.” *Aids* 28(8): 1091–99.

———. 2020. “Defective HIV-1 Proviruses Produce Viral Proteins.” *Proceedings of the National Academy of Sciences of the United States of America* 117(7): 3704–10.

Jacob, Aishwarya G., and Christopher W.J. Smith. 2017. “Intron Retention as a Component of Regulated Gene Expression Programs.” *Human Genetics* 136(9): 1043–57.

Jameson, Stephen C. 2002. “Maintaining the Norm: T-Cell Homeostasis.” *Nature Reviews Immunology* 2(8): 547–56.

Jiang, Chenyang et al. 2020. “Distinct Viral Reservoirs in Individuals with Spontaneous Control of HIV-1.” *Nature* 585(7824): 261–67.

Jiang, Guochun, Amy Espeseth, Daria J. Hazuda, and David M. Margolis. 2007. “C-Myc and Sp1 Contribute to Proviral Latency by Recruiting Histone Deacetylase 1 to the Human Immunodeficiency Virus Type 1 Promoter.” *Journal of Virology* 81(20): 10914–23.

Joag, V. R. et al. 2016. “Identification of Preferential CD4+ T-Cell Targets for HIV Infection in the Cervix.” *Mucosal Immunology* 9(1): 1–12.

Katlama, Christine et al. 2016. “Treatment Intensification Followed by Interleukin-7 Reactivates HIV without Reducing Total HIV DNA: A Randomized Trial.” *Aids* 30(2): 221–30.

Kaufmann, Daniel E. et al. 2007. "Upregulation of CTLA-4 by HIV-Specific CD4+ T Cells Correlates with Disease Progression and Defines a Reversible Immune Dysfunction." *Nature Immunology* 8(11): 1246-54.

Kawashima, Akira et al. 2017. “ARIH2 Ubiquitinates NLRP3 and Negatively Regulates NLRP3 Inflammasome Activation in Macrophages.” *The Journal of Immunology* 199(10): 3614–22.

Kinoshita, Shigemi et al. 1997. “The T Cell Activation Factor NF-ATc Positively Regulates HIV-1 Replication and Gene Expression in T Cells.” *Immunity* 6(3): 235–44.

Klatzmann, David et al. 1984. “Selective Tropism of Lymphadenopathy Associated Virus (LAV) for Helper-Inducer T Lymphocytes.” *Science* 225(4657): 59–63.

Koenig, Scott et al. 1986. “Detection of AIDS Virus in Macrophages in Brain Tissue from AIDS Patients with Encephalopathy.” *Science* 233(4768): 1089–93.

Koup, R A et al. 1994. “Temporal Association of Cellular Immune Responses with the Initial Control of Viremia in Primary Human Immunodeficiency Virus Type 1 Syndrome.” *Journal of virology* 68(7): 4650–55.

Kruize, Zita, and Neeltje A. Kootstra. 2019. “The Role of Macrophages in HIV-1 Persistence and Pathogenesis.” *Frontiers in Microbiology* 10(December): 1–17.

Kula, Anna et al. 2011. “Characterization of the HIV-1 RNA Associated Proteome Identifies Matrin 3 as a Nuclear Cofactor of Rev Function.” *Retrovirology* 8: 1–15.

Kula, Anna, Lavina Gharu, and Alessandro Marcello. 2013. “HIV-1 Pre-MRNA Commitment to Rev Mediated Export through PSF and Matrin 3.” *Virology* 435(2): 329–40. http://dx.doi.org/10.1016/j.virol.2012.10.032.

Kulpa, Deanna A., and Nicolas Chomont. 2015. “HIV Persistence in the Setting of Antiretroviral Therapy: When, Where and How Does HIV Hide?” *Journal of Virus Eradication* 1(2): 59–66.

Lin, Amy E. et al. 2013. “ARIH2 Is Essential for Embryogenesis, and Its Hematopoietic Deficiency Causes Lethal Activation of the Immune System.” *Nature Immunology* 14(1): 27–33.

Liu, Runxia et al. 2020. “Single-Cell Transcriptional Landscapes Reveal HIV-1-Driven Aberrant Host Gene Transcription as a Potential Therapeutic Target.” *Science Translational Medicine* 12(543): 1–14.

Liu, Runxia, Francesco R. Simonetti, and Ya Chi Ho. 2020. “The Forces Driving Clonal Expansion of the HIV-1 Latent Reservoir.” *Virology Journal* 17(1): 1–13.

Lusic, Marina, and Robert F. Siliciano. 2017. “Nuclear Landscape of HIV-1 Infection and

Integration.” *Nature Reviews Microbiology* 15(2): 69–82.

Ma, Averil, Rima Koka, and Patrick Burkett. 2006. “Diverse Functions of IL-2, IL-15, and IL-7 in Lymphoid Homeostasis.” *Annual Review of Immunology* 24: 657–79.

Maeda, Naoyoshi, Hung Fan, and Yasunobu Yoshikai. 2008. “Oncogenesis by Retroviruses: Old and New Paradigms.” *Reviews in Medical Virology* 18(6): 387–405.

Mahlknecht, Ulrich et al. 2000. “ Resistance to Apoptosis in HIV-Infected CD4 + T Lymphocytes Is Mediated by Macrophages: Role for Nef and Immune Activation in Viral Persistence .” *The Journal of Immunology* 165(11): 6437–46.

Maldarelli, F. et al. 2014. “Specific HIV Integration Sites Are Linked to Clonal Expansion and Persistence of Infected Cells.” *Science* 345(6193): 179–83.

McCune, J. M. 2001. “The Dynamics of CD4+ T-Cell Depletion in HIV Disease.” *Nature* 410(6831): 974–79.

McElrath, M. Juliana et al. 2008. “HIV-1 Vaccine-Induced Immunity in the Test-of-Concept Step Study: A Case-Cohort Analysis.” *The Lancet* 372(9653): 1894–1905. http://dx.doi.org/10.1016/S0140-6736(08)61592-5.

Mendoza, Pilar et al. 2020. “Antigen Responsive CD4+ T Cell Clones Contribute to the HIV-1 Latent Reservoir.” *bioRxiv* 217(7).

Min, Booki. 2018. “Spontaneous T Cell Proliferation: A Physiologic Process to Create and Maintain Homeostatic Balance and Diversity of the Immune System.” *Frontiers in Immunology* 9(547).

Min, Booki, Hidehiro Yamane, Jane Hu-Li, and William E. Paul. 2005. “Spontaneous and Homeostatic Proliferation of CD4 T Cells Are Regulated by Different Mechanisms.” *The Journal of Immunology* 174(10): 6039–44.

Nabel, Gary, and David Baltimore. 1987. “An Inducible Transcription Factor Activates Expression of Human Immunodeficiency Virus in T Cells.” *Nature* 326: 711–13.

Napolitano, Laura A. et al. 2001. “Increased Production of IL-7 Accompanies HIV-1-Mediated T-Cell Depletion: Implications for T-Cell Homeostasis.” *Nature Medicine* 7(1): 73–79.

Nguyen, Van Trung, Tamàs Kiss, Annemieke A. Michels, and Olivier Bensaude. 2001. “7SK Small Nuclear RNA Binds to and Inhibits the Activity of CDK9/Cyclin T Complexes.” *Nature* 414: 322–25.

O’Leary, Nuala A. et al. 2016. “Reference Sequence (RefSeq) Database at NCBI: Current Status, Taxonomic Expansion, and Functional Annotation.” *Nucleic Acids Research* 44(D1): D733–45.

Osuna, Christa E. et al. 2018. “Evidence That CD32a Does Not Mark the HIV-1 Latent Reservoir.” *Nature* 561(7723): E9–16.

Pierson, Theodore et al. 2000. “Characterization of Chemokine Receptor Utilization of Viruses in the Latent Reservoir for Human Immunodeficiency Virus Type 1.” *Journal of Virology* 74(17): 7824–33.

Pinzone, Marilia Rita et al. 2019. “Longitudinal HIV Sequencing Reveals Reservoir Expression Leading to Decay Which Is Obscured by Clonal Expansion.” *Nature Communications* 10(1): 1–12.

Pitisuttithum, Punnee et al. 2006. “Randomized, Double-Blind, Placebo-Controlled Efficacy Trial of a Bivalent Recombinant Glycoprotein 120 HIV-1 Vaccine among Injection Drug Users in Bangkok, Thailand.” *Journal of Infectious Diseases* 194(12): 1661–71.

Plotkin, Stanley A. 2001. “Immunologic Correlates of Protection Induced by Vaccination.” *Pediatric Infectious Disease Journal* 20(1): 63–75.

Pollack, Ross A. et al. 2017. “Defective HIV-1 Proviruses Are Expressed and Can Be Recognized by Cytotoxic T Lymphocytes, Which Shape the Proviral Landscape.” *Cell Host and Microbe* 21(4): 494-506.e4.

Pulsinelli, Gary A., and Howard M. Temin. 1994. “High Rate of Mismatch Extension during Reverse Transcription in a Single Round of Retrovirus Replication.” *Proceedings of the National Academy of Sciences of the United States of America* 91(20): 9490–94.

Reeves, Daniel B. et al. 2018. “A Majority of HIV Persistence during Antiretroviral Therapy Is Due to Infected Cell Proliferation.” *Nature Communications* 9(1).

Ren, Yanqin et al. 2020. “BCL-2 Antagonism Sensitizes Cytotoxic T Cell-Resistant HIV Reservoirs to Elimination Ex Vivo.” *Journal of Clinical Investigation* 130(5): 2542–59.

Rerks-Ngarm, Supachai et al. 2009. "Vaccination with ALVAC and AIDSVAX to Prevent HIV-1 Infection in Thailand." *New England Journal of Medicine* 361(23): 2209-20.

Rosado-Sánchez, Isaac et al. 2018. “An in Vitro System of Autologous Lymphocytes Culture That Allows the Study of Homeostatic Proliferation Mechanisms in Human Naive CD4 T-Cells.” *Laboratory Investigation* 98(4): 500–511.

Schacker, Timothy et al. 1996. “Clinical and Epidemiologic Features of Primary HIV Infection.” *Annals of Internal Medicine* 125(4): 257–64.

Schröder, Astrid R W et al. 2002. “HIV-1 Integration in the Human Genome Favors Active Genes and Local Hotspots.” *Cell* 110(4): 521–29. http://www.ncbi.nlm.nih.gov/pubmed/12202041.

Sengupta, Srona, and Robert F. Siliciano. 2018. “Targeting the Latent Reservoir for HIV-1.” *Immunity* 48(5): 872–95. https://doi.org/10.1016/j.immuni.2018.04.030.

Shan, Liang et al. 2017. “Transcriptional Reprogramming during Effector-to-Memory Transition Renders CD4+ T Cells Permissive for Latent HIV-1 Infection.” *Immunity* 47(4): 766-775.e3. https://doi.org/10.1016/j.immuni.2017.09.014.

Shao, Yiming, and Carolyn Williamson. 2012. "The HIV-1 Epidemic: Low- to Middle-Income Countries." *Cold Spring Harbor Perspectives in Medicine* 2(a007187): 1-17.

Sherrill-Mix, Scott, Karen E. Ocwieja, and Frederic D. Bushman. 2015. “Gene Activity in Primary T Cells Infected with HIV89.6: Intron Retention and Induction of Genomic Repeats.” *Retrovirology* 12(1): 1–19.

Siliciano, Janet D. et al. 2003. “Long-Term Follow-up Studies Confirm the Stability of the Latent Reservoir for HIV-1 in Resting CD4+ T Cells.” *Nature Medicine* 9(6): 727–28.

Siliciano, Janet D., and Robert F. Siliciano. 2020. “Nonsuppressible HIV-1 Viremia: A Reflection of How the Reservoir Persists.” *Journal of Clinical Investigation* 130(11): 5665–67.

Simonetti, Francesco R. et al. 2016. “Clonally Expanded CD4+ T Cells Can Produce Infectious HIV-1 in Vivo.” *Proceedings of the National Academy of Sciences of the United States of America* 113(7): 1883–88.

———. 2020. “Antigen-Driven Clonal Selection Shapes the Persistence of HIV-1 Infected CD4+ T Cells in Vivo.” *Journal of Clinical Investigation* 131(3).

Surh, Charles D., and Jonathan Sprent. 2008. “Homeostasis of Naive and Memory T Cells.” *Immunity* 29(6): 848–62.

Taylor, Barbara S., Magdalena E. Sobieszczyk, Francine E. Cutchan, and Scott M. Hammer. 2008. “The Challenge of HIV-1 Subtype Diversity.” *New England Journal of Medicine* 358(15): 1590–1602.

The International HIV Controllers Study. 2010. “The Major Genetic Determinants of HIV-1 Control Affect HLA Class I Peptide Presentation.” *Science* 330: 1551–58.

The Lancet. 2020. “Maintaining the HIV Response in a World Shaped by COVID-19.” *The Lancet* 396(10264): 1703.

The World Health Organization. 2009. “HIV/AIDS Fact Sheet.” (November 30).

Tobin, Nicole H. et al. 2005. “Evidence That Low-Level Viremias during Effective Highly Active Antiretroviral Therapy Result from Two Processes: Expression of Archival Virus and Replication of Virus.” *Journal of Virology* 79(15): 9625–34.

United Nations Joint Programme on HIV/AIDS. 2019. “AIDS Info.”

http://aidsinfo.unaids.org/.

———. 2020a. “Global AIDS Update 2020 - Seizing the Moment.” July 6: 1–380.

———. 2020b. “Global HIV Statistics.” *Fact Sheet - World AIDS Day*.

Vandergeeten, Claire et al. 2013. “Interleukin-7 Promotes HIV Persistence during Antiretroviral Therapy.” *Blood* 121(21): 4321–29.

Wagner, Thor A. et al. 2014. “Proliferation of Cells with HIV Integrated into Cancer Genes Contributes to Persistent Infection.” *Science* 345(6196): 570–73.

Wang, Gary P. et al. 2007. “HIV Integration Site Selection: Analysis by Massively Parallel Pyrosequencing Reveals Association with Epigenetic Modifications.” *Genome Research* 17(8): 1186–94.

Wang, Zheng et al. 2018. “Expanded Cellular Clones Carrying Replication-Competent HIV-1 Persist, Wax, and Wane.” *Proceedings of the National Academy of Sciences* 115(11): E2575–84.

Whitney, James B. et al. 2014. “Rapid Seeding of the Viral Reservoir Prior to SIV Viraemia in Rhesus Monkeys.” *Nature* 512(1): 74–77.

Wilen, Craig B, John C Tilton, and Robert W Doms. 2012. “HIV: Cell Binding and Entry.” *Cold Spring Harbor Perspective Med* 2(a006866): 1–13.

Williams, Samuel A. et al. 2006. “NF-KB P50 Promotes HIV Latency through HDAC Recruitment and Repression of Transcriptional Initiation.” *EMBO Journal* 25(1): 139–49.

Wong, Joseph K., and Steven A. Yukl. 2016. "Tissue Reservoirs of HIV." *Current Opinion in HIV and AIDS* 11(4): 362-70.

Wu, Xiaolin, Yuan Li, Bruce Crise, and Shawn M. Burgess. 2003. “Transcription Start Regions in the Human Genome Are Favored Targets for MLV Integration.” *Science* 300(5626): 1749–51.

Yang, Zhiyuan, Qingwei Zhu, Kunxin Luo, and Qiang Zhou. 2001. “The 7SK Small Nuclear RNA Inhibits the CDK9/Cyclin T1 Kinase to Control Transcription.” *Nature* 414: 317–22.

Yao, Jun et al. 2020. “Prevalent Intron Retention Fine-Tunes Gene Expression and Contributes to Cellular Senescence.” *Aging Cell* 19(12): 1–14.

Zack, Jerome A. et al. 1990. “HIV-1 Entry into Quiescent Primary Lymphocytes: Molecular Analysis Reveals a Labile, Latent Viral Structure.” *Cell* 61(2): 213–22.