2023

Enrichment and Transcriptome Analysis of CD4+ T Cell Clones Harboring Intact HIV-1 Proviruses

Georg Hartmut Johannes Weymar

Enrichment and transcriptome analysis of $CD4^+$ T cell clones harboring intact HIV-1 proviruses

A Thesis Presented to the Faculty of
The Rockefeller University
in Partial Fulfillment of the Requirements for
the degree of Doctor of Philosophy

by
Georg Hartmut Johannes Weymar
June 2023

Enrichment and transcriptome analysis of CD4$^+$ T cell clones harboring intact HIV-1 proviruses

Georg Hartmut Johannes Weymar, Ph.D.

The Rockefeller University 2023

More than 40 years after its onset, the Human Immunodeficiency Virus Type 1 (HIV-1) epidemic remains a major global health burden. HIV-1 is a retrovirus that infects and destroys predominantly CD4$^+$ T cells. Since CD4$^+$ T cells are crucial for the innate and adaptive immune system, their progressive loss in the course of HIV-1 infection leads to the Acquired Immune Deficiency Syndrome (AIDS). Antiretroviral therapy can suppress the virus in people living with HIV and prevent progression to AIDS, but does not provide a cure. Therefore, lifelong antiretroviral therapy is necessary.

The barrier to HIV-1 cure is the HIV-1 reservoir, a scarce pool of HIV-1 infected CD4$^+$ T cells that has a very long lifetime, can multiply, and is not affected by antiretroviral therapy. These persisting infected cells have been the focus of HIV research, with the aim to identify potential biomarkers and characterize their gene expression profile in order to tailor a therapeutic approach to HIV-1. Since these cells are exceedingly rare and cannot be distinguished from uninfected CD4$^+$ T cells, *ex vivo* analysis on a single cell level has not been feasible.

In the first part of this thesis, a method for the enrichment of clonally expanded cells harboring an intact HIV-1 provirus that does not require latency reversal is presented. The enrichment is based on CD45RA, a marker that distinguishes naïve from memory CD4$^+$ T cells, as well as on the variable and constant domain of the T cell receptor (TCR) β chain, that are shared by all members of a given infected clone.

The resulting enrichment enables the identification of the unique TCRαβ sequence of infected clones, which is presented in the second part. The TCRαβ sequence is a unique molecular identifier of a given infected clone and can be read out by commercially available high throughput single cell gene expression sequencing platforms.

In the third part, the gene expression profile of six enriched infected clones harboring an intact provirus from six individuals living with HIV-1 is analyzed by single cell gene expression and TCR sequencing on the 10x Genomics platform. Cells pertaining to expanded infected clones harboring an intact provirus are predominantly found in the $CD4^+$ T effector memory compartment, but also belong to other $CD4^+$ memory cell populations. Thus, expanded infected clones display a diverse gene expression profile and are not a homogenous cell population. Furthermore, the infected clones analyzed in this study did not display a unique gene expression profile that would set them apart from uninfected $CD4^+$ T cells. Thus, proviral integration does not appear to shape the host cell neither in a uniform, nor in a unique way.

For Ann Tennenbaum and Dickie Brodersen

ACKNOWLEDGMENTS

First and foremost, I would like to thank Michel Nussenzweig for having me as a student in his lab, and for being such a great advisor and teacher.

I would also like to thank my Faculty Committee, Daniel Mucida and Paul Bieniasz, for their thoughtful discussions and valuable input, and R. Brad Jones as the external examiner.

I thank all the study participants who enable the great research at the Nussenzweig lab and at the Rockefeller University by dedicating their time.

Many thanks also go to the all members of the Nussenzweig lab, especially:

To Mila Jankovic, for her incredible support, many constructive discussions, and her never ending willingness to help.

To Yotam Bar-On, Thiago Y. Oliveira, Christian Gaebler, Victor Ramos, Harald Hartweger, Gaëlle Breton, Marina Caskey, and Lillian B. Cohn, who all contributed greatly to this work.

To Kristie Gordon and Kalsang Chhosphel for operating the cell sorters.

Thanks to Connie Zhao, Hong Duan, Christine Lai, and Sophie Huang from the Genomics Resource Center at the Rockefeller University for the preparation and the sequencing of the 10x Genomics libraries, and to the Rockefeller University Hospital Research support office and nursing staff.

My gratitude goes out to the members of the Dean's Office, Sidney Strickland, Kristen Cullen, Stephanie Fernandez, Marta Delgado, Emily Harms, Tasnia Islam, Andrea Morris, and Cristian

Rosario; to Marnel Herbert from the Student Housing Department; and to Ashton Murray and Kelly Ann Turner from the Diversity, Equity and Inclusion Department. Thank you all for your continuous support and all your efforts that make the David Rockefeller Graduate Program so special.

Lastly, I would like to thank my parents, my sister and her wonderful family who bring me so much joy, Adam, and my friends for their unconditional love and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AIDS Acquired Immune Deficiency Syndrome

ART Antiretroviral Therapy

CTL Cytotoxic lymphocyte

env Envelope gene

gp Glycoprotein

HIV Human Immunodeficiency Virus

LRA Latency-reversing agent

TCM T central memory

TCR T cell receptor

TEM T effector memory

TRAC TCRα constant gene

TRAJ TCRα joining gene

TRAV TCRα variable gene

TRBC TCRβ constant gene

TRBD TCRβ diversity gene

TRBJ TCRβ joining gene

TRBV TCRβ variable gene

CHAPTER 1. Introduction

1.1 Epidemiology of HIV-1

On June 5th 1981, the Center for Disease Control and Prevention reported cases of Pneumocystis Pneumonia in young homosexual men (CDC, 1981). On July 3rd 1981, an article was published in The New York Times about "41 cases of a rare and often rapidly fatal form of cancer" that affected homosexual men in New York and California. At that time, medical investigators saw indicators against a contagion, and specifically "no apparent danger to nonhomosexuals from contagion" (Altman, 1981). The Human Immunodeficiency Virus Type 1 (HIV-1) was later discovered to be the causative agent of those conditions that are entities of the Acquired Immune Deficiency Syndrome (AIDS).

Today, the HIV-1 epidemic is still a global health crisis. In 2020, there were an estimated 37.7 million people living with HIV, and approximately 680,000 AIDS-related deaths (UNAIDS, 2021). Sub-Saharan Africa accounted for 60 % of approximately 1.5 million new infections in 2017, with adolescent girls and young women at increased risk (Karim and Baxter, 2019). Other disproportionally affected populations are sex workers, men who have sex with men, transgender individuals, and intravenous drug users (UNAIDS, 2021).

HIV-1 can be transmitted sexually, parenterally, or perinatally and by breast-feeding (Shaw and Hunter, 2012). Combination of antiretroviral drugs (ART) can suppress the replication of the virus and detection of the virus in blood by clinical assays. Thus, prevention of progression to AIDS and prevention of HIV transmission (undetectable = untransmittable, U = U) can be achieved (Bavinton et al., 2018; Cohen et al., 2016; Rodger et al., 2019). More recently, combination of broadly neutralizing antibodies has been shown to be effective in suppressing viraemia in the

absence of ART (Gaebler et al., 2022). Nonetheless, a cure for HIV remains elusive (Arts and Hazuda, 2012; Saag et al., 2020).

HIV prevention includes a multitude of approaches, such as sexual education, condom use, pre-exposure prophylaxis, post-exposure prophylaxis, provision with clean needles for intravenous drug users, or voluntary male circumcision (Deeks et al., 2015; Krishnaratne et al., 2016; Spinner et al., 2016). The development of a vaccine against HIV as a preventative measure has not been successful yet (Kim et al., 2021).

In summary, HIV-1 remains a global health crisis. Access to adequate treatment can prevent disease progression and HIV transmission, but a cure from the virus has not been achieved yet.

1.2 The HIV-1 replication cycle

1.2.1 Attachment, fusion, and uncoating

HIV is a human-tropic virus of the *Retroviridiae* family and subdivided into two main strains, HIV-1, which accounts for the majority of infections worldwide and is more pathogenic, and HIV-2, which is endemic in western Africa (Deeks et al., 2015).

The genetic information of HIV-1 is encoded in two positive-sense, single stranded, genomic RNA molecules of approximately 9 kbp length that are encased by a conical capsid core and a viral membrane (Engelman and Cherepanov, 2012; Moore and Hu, 2009). The membrane carries the viral envelope (env) protein, a non-covalently linked heterotrimer that consists of the surface glycoprotein (gp) gp120 and the transmembrane gp41 (Wang et al., 2020). The env protein binds to the CD4 surface molecule (Maddon et al., 1986; McDougal et al., 1986) and the chemokine

receptors CXCR4 and CCR5 (Berger et al., 1999). Thus, the primary target for HIV-1 are $CD4^+$ T cells (Engelman and Cherepanov, 2012). The gp120 subunit of the env protein facilitates binding to CD4, leading to the exposure of a chemokine receptor binding site (Kwong et al., 1998; Rizzuto et al., 1998). Chemokine receptor binding induces a conformational change in the gp41 subunit, which inserts into the host cell membrane and causes a fusion of the viral and host cell membrane. This allows the virus to enter the host cell and introduce its capsid core into the host cell cytoplasm (Wang et al., 2020; Wilen et al., 2012). Subsequently, the viral capsid core disassembles and releases the viral RNA as well as the viral enzymes reverse transcriptase and integrase (Engelman and Cherepanov, 2012). The cytoplasm was considered to be the cellular compartment of the capsid core disassembly by early research works. Recently, the nucleus has been suggested to be the location of capsid core disassembly (Toccafondi et al., 2021).

1.2.2 Reverse transcription and proviral integration

The viral reverse transcriptase produces viral dsDNA from the genomic viral RNA. The reverse transcriptase is a heterodimer, comprising an RNA- and a DNA-dependent DNA polymerase that contains Mg^{2+} ions, and an RNase that is specific for RNA-DNA hybrid double strands (Hu and Hughes, 2012).

The reverse transcription process is rather error prone, and the reverse transcriptase lacks a proof reading activity. Determination of the error rate has been proven to be difficult, as it is affected by several conditions *in vitro*, such as Mg^{2+} concentration and pH. Experimental estimates range from 1 to 3 substitutions per 10^5 nucleotides per replication cycle (Abram et al., 2010; Achuthan et al., 2014). If a virus is deficient for the virion infectivity factor (vif) gene, then further hypermutation of the newly generated viral DNA strand by the host cell APOBEC3 proteins can occur (Malim

and Bieniasz, 2012). Mutations of the viral genome can render a provirus to be defective. Such a defective provirus cannot further replicate (Das et al., 2019; Imamichi et al., 2016).

The inaccuracy of viral reverse transcription, together with the high frequency of viral replication and turnover of infected cells in untreated infection leads to a high genetic diversity of viral genomes within a single HIV-infected individual (Coffin, 1995; Hu and Hughes, 2012), despite the initial infection with a single founder virus or small number of founder viruses (Abrahams et al., 2009; Derdeyn et al., 2004; Keele et al., 2008; Sagar et al., 2009; Sagar et al., 2004; Zhu et al., 1996). Over time, mutations accumulate especially in the *env* gene, which encodes the only viral protein that is exposed on the viral surface. This is due to selective pressure from the innate and adaptive immune system (Arrildt et al., 2012; Beretta et al., 2020). The high mutation rate enables the virus to escape from cellular and humoral components of the host immune system and has left efforts to design a vaccine against HIV fruitless (Ganusov et al., 2011; Goonetilleke et al., 2009; Johnson and Desrosiers, 2002; Price et al., 1997).

The viral integrase enzyme enables stable integration of the viral dsDNA into the host cell genome (Craigie and Bushman, 2012). The accuracy of the eukaryotic DNA replication machinery with proof reading mechanisms is much higher compared to the viral reverse transcription, and is estimated to be as low as 1 substitution per 10^{10} nucleotides (Ganai and Johansson, 2016). Therefore, once a provirus is integrated into the host cell genome, it is genetically stable and does not further mutate.

Integration of the HIV provirus into the host cell genome does not occur randomly and is mediated by the host cell protein LEDGF/p75 that can bind to the viral integrase protein. Within the human genome, LEDGF-p75-binding sites are preferentially located in transcription units (Craigie and

Bushman, 2012). Therefore, there is a preference for integration into the introns of actively transcribed host cell genes (Cohn et al., 2015; Craigie and Bushman, 2012; Schroder et al., 2002). *BACH2*, *MKL2*, *DMNT1*, *STAT5B*, and *IL2RB* are among the genes that show enrichment of integrated proviruses (Bedwell et al., 2021; Ikeda et al., 2007; Maldarelli et al., 2014; Wagner et al., 2014).

1.2.3 Proviral transcription, viral assembly and budding

The integrated provirus can be transcribed by hijacking of host cell transcription factors that are abundant in activated $CD4^+$ T cells, such as NF-κB, NFAT, or AP-1 (Pan et al., 2013; Wong and Jiang, 2021). The proviral RNA transcripts comprise spliced mRNAs that are translated into viral proteins, and unspliced genomic viral RNA that is incorporated into a new generation of virions that are released from the host cell membrane (Karn and Stoltzfus, 2012).

The majority of HIV-infected and virus-producing $CD4^+$ T cells dies within one day by apoptosis or pyroptosis, or is cleared by cytotoxic lymphocytes (Doitsh and Greene, 2016; Jones and Walker, 2016; Markowitz et al., 2003). This leads to a progressive loss of $CD4^+$ cells in untreated HIV-1 infection. However, in rare instances, HIV-infected cells do not die but persist. This pool of cells harboring an integrated provirus is persistent over time and constitutes the HIV-1 reservoir (Chun et al., 1997; Finzi et al., 1999; Wong et al., 1997).

1.3 The HIV-1 reservoir

1.3.1 Cellular composition of the HIV-1 reservoir

The HIV-1 reservoir consists of very rare, long-lived cells that harbor an integrated intact provirus that is replication competent and able to produce virions (Cohn et al., 2020; Dufour et al., 2021).

The integrated provirus in these cells can be transcriptionally silent, i.e. latent, or display varying levels of transcription.

The HIV-1 reservoir is established very early in infection (Chun et al., 1998; Colby et al., 2018) and is not eradicated by ART. Thus, even after long time ART, cells with an integrated provirus can be recovered from people living with HIV (Chun et al., 1997; Finzi et al., 1999; Wong et al., 1997). Treatment interruption leads to rebound viraemia in most people living with HIV (Davey et al., 1999; Li et al., 2016). Therefore, the HIV-1 reservoir is considered the main barrier to cure HIV-1 (Cohn et al., 2020; Sengupta and Siliciano, 2018).

Even though myeloid cells, such as monocytes, macrophages, and dendritic cells, can be infected with HIV-1 and therefore potentially contribute to the HIV-1 reservoir (Kandathil et al., 2016; Wong et al., 2019), $CD4^+$ T cells constitute the main cellular compartment of the HIV-1 reservoir (Cohn et al., 2020; Dufour et al., 2021). $CD4^+$ T cells can circulate through the blood, through secondary lymphoid organs such as lymph nodes, the spleen, and the mucosa-associated lymphoid tissue, or home to non-lymphoid tissues (Nguyen et al., 2019). Accordingly, it was shown in a rapid autopsy study that infected cells have a wide anatomical distribution, as cells harboring an integrated provirus were found in all 28 analyzed tissues (Chaillon et al., 2020). Even though follicular helper $CD4^+$ T cells in germinal centers of the lymph nodes (Banga et al., 2016) and $CD4^+$ T cells in the gut-associated lymphoid tissue (Chun et al., 2008; Estes et al., 2017) harbor replication-competent HIV-1 proviruses, circulating blood $CD4^+$ T cells are the most investigated component of the HIV-1 reservoir, due to their less invasive accessibility compared to lymphoid organ and tissue resident $CD4^+$ T cells.

1.3.2 Infected cells harbor defective or intact proviruses

The majority of HIV-1 infected cells in ART-treated individuals harbors a defective provirus. Defective proviruses cannot produce infectious virions, but some can produce viral proteins (Bruner et al., 2016; Cho et al., 2022; Cohn et al., 2015; Einkauf et al., 2022; Eriksson et al., 2013; Ho et al., 2013; Imamichi et al., 2014; Lee et al., 2017; Lorenzi et al., 2016; Peluso et al., 2020; Sanchez et al., 1997). The majority of defects of the HIV genome comprise larger deletions and G-to-A hypermutations. Other, less frequent defects are gene inversions or insertions of genes, smaller deletions and point mutations of the major-splice donor site or the packaging signal (Das et al., 2019; Imamichi et al., 2016).

1.3.3 Transcription of integrated proviruses in HIV-1 infected cells

For a long time, proviruses of the HIV-1 reservoir were thought to be transcriptionally silent (Sengupta and Siliciano, 2018). However, a more differentiated understanding of proviral transcription in cells harboring a provirus is now appreciated.

Early works indicated that proviral transcription is absent or very low (Hermankova et al., 2003). Later studies showed HIV transcription in resting $CD4^+$ T cells from ART-treated, aviraemic individuals (Kearney et al., 2016). However, HIV transcripts are mostly incomplete and therefore not fit to generate infectious particles (Yukl et al., 2018), only present in a very low copy number of 1 to 5 copies per cell (Wiegand et al., 2017), or derived from defective and therefore replication-incompetent proviruses (Imamichi et al., 2016; Pollack et al., 2017). A recent study proposed that clones harboring highly transcribed intact proviruses are slowly eliminated from the HIV-1 reservoir over time, whereas clones with proviruses that display low levels of transcription can be maintained under long-term ART (Einkauf et al., 2022).

Multiple factors can affect proviral transcription, for example the proviral integration site, epigenetic modulation of the proviral integration site, host cell transcription factors, and stochastic events (Sengupta and Siliciano, 2018).

Proviruses in elite controllers who display spontaneous absence of viraemia without ART despite the presence of replication competent integrated proviruses often are distinctly integrated into heterochromatin regions, such as centromeric satellite DNA or in Krüppel-associated box domain-containing zinc finger genes (KRAB-ZNF) on chromosome 19 (Jiang et al., 2020). Huang and colleagues suggest that integration into KRAB-ZNF genes favors a state of latency in a more general sense, not only in elite controllers, as those genes are silenced during T cell activation (Huang et al., 2021). Over time on ART, intact proviruses are more frequently found in non-genic regions and in opposite direction to host genes (Einkauf et al., 2019). Furthermore, histone acetylation and methylation may render the integrated provirus inaccessible to the transcriptional machinery of the host, and thus control proviral transcription (Svensson, 2021).

Transcription factors, such as NF-κB, NFAT, or AP-1 are present in the nucleus of activated $CD4^+$ T cells and can promote proviral transcription. However, in resting $CD4^+$ T cells these transcription factors are restricted, making resting $CD4^+$ T cells less permissive for viral transcription (Pan et al., 2013).

In vitro stimulation of *ex vivo* latently infected $CD4^+$ T cells only reactivates a small fraction of latent proviruses (Ho et al., 2013), and multiple rounds of stimulation are necessary to increase the fraction of HIV-expressing cells (Hosmane et al., 2017). This underscores the stochastic rather than deterministic nature of latency and the reversal of transcriptional quiescence (Hataye et al., 2019).

Since a latently infected cell is characterized by the presence of a transcriptionally silent provirus that has the ability to produce infectious virions, a narrow definition of the latent reservoir only includes cells harboring genetically intact, replication competent proviruses that are not transcribed (Cohn et al., 2020; Sengupta and Siliciano, 2018). In the context of this thesis, "HIV-1 reservoir" includes any cell with an integrated provirus irrespective of its genetic integrity and proviral transcription, whereas "latent reservoir" refers to cells harboring a transcriptionally silent provirus, irrespective of its genetic integrity.

1.3.4 Quantification of the HIV-1 reservoir

The exact quantification of the HIV-1 reservoir is, due to the wide anatomical distribution of cells that harbor an integrated provirus, not feasible (Chaillon et al., 2020). However, infected cells in the blood are a good representation of the proviruses that are present in the overall HIV-1 reservoir (Chaillon et al., 2020; McManus et al., 2019). Approaches to measure HIV-infected cells from blood can be subdivided into viral outgrowth assays and viral DNA amplification assays (Massanella and Richman, 2016).

An early developed method to assess the HIV-1 reservoir is the quantitative viral outgrowth assay (QVOA) . *Ex vivo* $CD4^+$ T cells from blood of a person living with HIV are plated at multiple dilutions in multiwell cell culture plates. Subsequently, cells are cultured in the presence of phytohemagglutinin (PHA) and irradiated peripheral blood mononuclear cells (PBMCs) and $CD4^+$ T cells from an HIV negative donor to propagate viral production. The readout is an ELISA for the viral p24 antigen (Finzi et al., 1997). A further development of the QVOA is the quantitative and qualitative viral outgrowth assays (Q^2VOA) . *Ex vivo* $CD4^+$ T cells of a person living with

HIV are plated at limiting dilution in multiwell cell culture plates and stimulated as in the QVOA. The readout is p24 ELISA, *env* sequencing from viral RNA, and susceptibility of the induced viruses to broadly neutralizing antibodies. Analysis of the genetic diversity of the induced proviruses is possible by *env* sequencing (Lorenzi et al., 2016). The limitation of the viral outgrowth assays is the underestimation of the HIV-1 reservoir. Inducibility of a provirus *in vitro* does not necessarily reflect its inducibility *in vivo*, and has therefore limited biological meaning. Furthermore, only a fraction of the genetically intact proviruses are expressed in viral outgrowth assays after a single round of stimulation (Ho et al., 2013), and even multiple rounds of stimulation do not induce transcription of all genetically intact proviruses (Hosmane et al., 2017).

Viral DNA amplification assays quantify the number of proviral copies independently of *in vitro* inducibility (Massanella and Richman, 2016). Amplification assays can span almost the entire HIV genome (near-full length, NFL) or can span sub-genomic regions of the HIV genome, such as the *env* gene. Only NFL amplification and sequencing can provide information about the genetic intactness of the provirus. Sub-genomic amplification and sequencing assays can overestimate the intact HIV-1 reservoir, since a deleterious mutation may be located outside of the intact amplified sequence, and the provirus may erroneously be considered as intact (Wang et al., 2018b).

The Q4PCR, a nested NFL PCR amplification combined with a four-probe quantitative PCR assay, was developed in the Nussenzweig lab (Gaebler et al., 2019). After limiting dilution of gDNA from $CD4^+$ T cells, a first round of NFL PCR amplification is performed. The PCR product serves as a template for a qPCR with four probe/primer sets that detect different parts of the viral genome. Based on positive qPCR signals, likely intact proviral amplicons are further amplified in a second,

nested NFL PCR. This second NFL PCR product is then sequenced by next generation sequencing which enables definitive identification of genetically intact proviruses.

1.3.5 Dynamics of the HIV-1 reservoir over time

Under ART, the reservoir of cells harboring an integrated provirus has a long half-life time of approximately 4 years. This half-life time of the reservoir is consistent between different assays of HIV-1 reservoir measurement (Antar et al., 2020; Bachmann et al., 2019; Cho et al., 2022; Crooks et al., 2015; Finzi et al., 1999; Peluso et al., 2020; Siliciano et al., 2003). The decay of the intact and the defective HIV-1 reservoir is not uniform. Cells with an intact provirus decay faster than those with a defective provirus, and cells with a defective provirus that can produce viral proteins decay faster than those with a provirus that that cannot produce viral proteins (Antar et al., 2020; Cho et al., 2022; Imamichi et al., 2020; Peluso et al., 2020; Pinzone et al., 2019). At the same time, the diversity of the HIV proviral reservoir decreases overtime, as clones increase in size. This is true for both, defective (Antar et al., 2020; Cohn et al., 2015; Wagner et al., 2013) and intact (Cho et al., 2022) proviruses. All infected cells of an expanded clone share the same provirus, the same unique proviral integration site, and the same unique T cell receptor (TCR) (Cohn et al., 2018; Cohn et al., 2015; Einkauf et al., 2019; Huang et al., 2021; Simonetti et al., 2021). Clonal expansion is not a linear process, as proviral clones can expand and contract over time (Simonetti et al., 2021; Wang et al., 2018a).

Clonal expansion is a multifactorial process (Cohn et al., 2020). Certain proviral integration sites in proximity to genes that can be related to cancer have been proposed to drive clonal expansion (Maldarelli et al., 2014; Wagner et al., 2014). Nevertheless, such an integration may not be sufficient to drive autologous cell proliferation in every case (Wang et al., 2018a). Another factor

of HIV-1 reservoir persistence is cytokine-induced homeostatic proliferation (Chomont et al., 2009). This is a physiologic process of memory $CD4^+$ T cells and not restricted to cells that harbor a provirus (Geginat et al., 2001; Kondrack et al., 2003; Seddon et al., 2003). Importantly, homeostatic proliferation of latently infected cells does not trigger viral replication, thus allowing the latently infected cell to multiply without undergoing cell death (Bosque et al., 2011; Vandergeeten et al., 2013).

Furthermore, antigen-induced proliferation can drive clonal expansion of cells harboring a provirus (Wang et al., 2018a). Infected cells can be specific against HIV-1 (Douek et al., 2002), or antigens of other chronic viral infections, such as CMV or EBV (Henrich et al., 2017; Mendoza et al., 2020; Simonetti et al., 2021).

1.3.6 Markers of the HIV-1 reservoir

Despite years of great research effort to identify biological markers of the HIV-1 reservoir, no such marker or set of markers has been identified. The rationale of discovering markers for infected cells is a specific and selective targeting for eradication of the HIV-1 reservoir (Darcis et al., 2019). Furthermore, since latently infected cells are very rare and can only be detected by viral RNA transcripts or viral proteins upon reactivation of the virus, the transcriptome analysis of latently infected cells that are confirmed to harbor a genetically intact provirus has not been feasible yet. A recent study identified infected $CD4^+$ T cell clones that contain viral RNA expressing cells (Collora et al., 2022). Those clones were enriched in a cytotoxic, Granzyme B-expressing $CD4^+$ T cell subset. However, since the vast majority of integrated proviruses are defective but can produce viral mRNA (Bruner et al., 2016; Cho et al., 2022; Cohn et al., 2015; Einkauf et al., 2022; Eriksson et al., 2013; Ho et al., 2013; Imamichi et al., 2016; Imamichi et al., 2014; Lee et al., 2017; Lorenzi

et al., 2016; Peluso et al., 2020; Pollack et al., 2017; Sanchez et al., 1997), the respective study did not specifically analyze cells harboring an intact provirus, but rather cells with an defective or intact provirus in an unknown ratio.

Even though no definitive marker of infected cells is known, they are enriched in certain $CD4^+$ T cell subsets. Integrated HIV-1 DNA can be found in both, the naïve and the memory $CD4^+$ T cell compartment. However, genetically intact proviruses are enriched in the $CD45RA^-$ memory compartment (Chomont et al., 2009; Duette et al., 2022; Hiener et al., 2017; Lee et al., 2017; Morcilla et al., 2021; Venanzi Rullo et al., 2020; Zerbato et al., 2019).

High levels of CD2 expression (Iglesias-Ussel et al., 2013), and markers of cellular exhaustion, such as PD-1, TIGIT, and LAG-3 (Fromentin et al., 2016) were shown to enrich for HIV DNA, but no discrimination between defective and intact proviruses was made. Therefore, the use of these markers to explicitly enrich cells with an intact integrated provirus remains unclear. In contrast, HLA-DR, which is upregulated in reactivated cells harboring an intact provirus (Cohn et al., 2018) also enriches specifically for intact HIV-1 proviruses in resting infected cells (Horsburgh et al., 2020; Lee et al., 2019).

In vitro reactivation of HIV-1 in *ex vivo* infected cells can induce proviral transcription and translation of viral proteins. This enables identification of infected cells and is an experimental approach in a number of studies (Baxter et al., 2016; Cohn et al., 2018; Grau-Exposito et al., 2019; Liu et al., 2020; Martins et al., 2016; Neidleman et al., 2020; Pardons et al., 2019a; Takata et al., 2019; Telwatte et al., 2019). However, the analysis of resting *ex vivo* infected cells that harbor an intact provirus without T cell stimulation on a single cell level has not been feasible yet and thus, the cellular changes that are induced by latency reversal remain unclear.

Experimental models of latency comprise cell line models and *in vitro* HIV-1 infection of primary $CD4^+$ T cells (Bradley et al., 2018; Cameron et al., 2010; Iglesias-Ussel et al., 2013; Krishnan and Zeichner, 2004; Lassen et al., 2012; Marini et al., 2008; Telwatte et al., 2019). To which degree these models reflect the biology of *ex vivo* infected cells is not known.

1.4 T cell receptor generation and T cell differentiation

1.4.1 Rearrangement of the T cell receptor locus

The adaptive immune system can react to a virtually unlimited number of different antigens. More than 10^8 unique TCRs are estimated to be present in a single individual (Qi et al., 2014). However, the human genome does not encode an equal number of complete unique TCRs. Rather, the TCR is composed of multiple gene sets that are rearranged during T cell development (Nikolich-Zugich et al., 2004).

The TCR locus contains a TCRα locus and a TCRβ locus, each encoding a respective TCR chain. Rearrangement of the TCRβ locus, that comprises approximately 48 functional variable genes (TRBV), 2 functional diversity genes (TRBD), approximately 13 functional joining genes (TRBJ), and 2 functional constant genes (TRBC), occurs first (Figure 1a) (IMGT, 2022). One TRBV, one TRBD, and one TRBJ gene is recombined with one TRBC gene (Figure 1b) (Nikolich-Zugich et al., 2004). Then the TCRα locus, that comprises approximately 45 functional variable genes (TRAV), 50 functional joining genes (TRAJ), and 1 functional constant gene (TRAC), is rearranged (Figure 1d) (IMGT, 2022). One TRAV and one TRAJ gene is recombined with the TRAC gene (Figure 1e) (Nikolich-Zugich et al., 2004). Due to allelic exclusion, a given lymphocyte expresses only a single immune receptor (Nussenzweig et al., 1987). Thus, most T cells express only one TCRα and only one TCRβ chain (Figure 1c and f). However, allelic

exclusion of the TCRα locus is less stringent, and $CD4^+$ T cells can therefore express 2 TCRα chains in some instances (Bassing et al., 2002; Dupic et al., 2019; Nikolich-Zugich et al., 2004). The V(D)J gene rearrangement accounts for a recombinatorial diversity of the TCR. However, the TRAV and TRAJ, and the TRBV, TRBD, and TRBJ genes, respectively, are not joined precisely. Rather, random nucleotides can be added or removed at the gene junctions, accounting for junctional diversity (Figure 1) (Nikolich-Zugich et al., 2004). The recombinatorial and junctional diversity of each TCR chain, together with the diverse joining of an TCRα and a TCRβ chain leads to the vast TCR diversity of more than 10^{20} possible recombinations (Zarnitsyna et al., 2013).

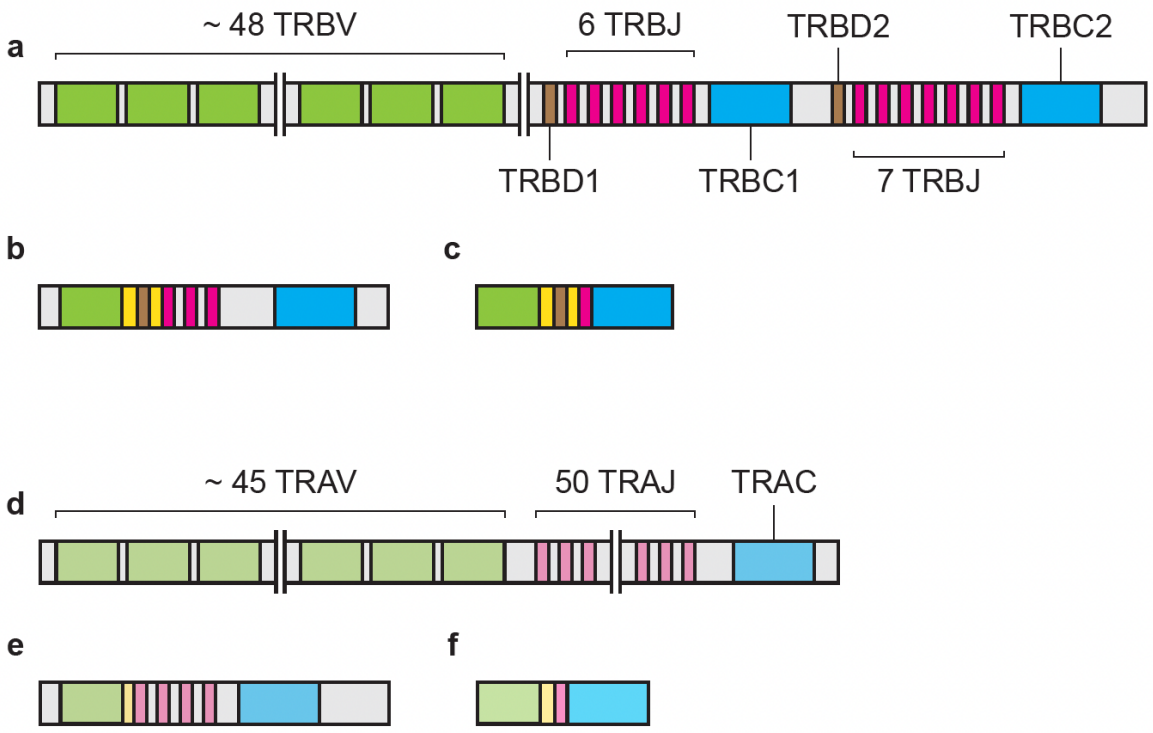


Figure 1 CD4$^+$ T cells express unique αβ T cell receptors.
The scheme represents the rearrangement of the TCRαβ locus. The genes, intergenic regions (depicted in grey) and unique gene junctions (depicted in yellow) are not drawn to scale. The figure is based on Nikolich-Zugich (2004) and the IMGT database (https://www.imgt.org).
a) The TCRβ locus contains approx. 48 TRBV genes, 2 TRBD genes, 13 TRBJ genes, and two TRBC genes.
b) In the re-arranged TCRβ locus, one TRBV gene is recombined with one TRBD gene, one TRBJ gene, and one TRBC gene. The junctions between the TRBV, TRBD, and TRBJ genes are unique.
c) Spliced mRNA of the TCRβ locus.
d) The TCRα locus contains approx. 45 TRAV genes, 50 TRAJ genes, and one TRAC gene.
e) In the re-arranged TCRα locus, one TRAV gene is recombined with one TRAJ gene, and the TRAC gene. The junction between the TRAV and TRAJ genes is unique.
f) Spliced mRNA of the TCRα locus.

1.4.2 $CD4^+$ T cell differentiation

Upon encountering a cognate antigen that is presented by an antigen-presenting cell, naïve $CD4^+$ T cells can undergo clonal expansion and further differentiation (Ruterbusch et al., 2020).

All cells of an expanded clone express the same unique TCR as the founding cell of that clone. Thus, all members of a T cell clone can be identified by their unique TCRαβ sequence (Qi et al., 2014).

Differentiation into T helper cell subsets such as Th1, Th2, and Th17, or Tregs is mostly shaped by the cytokine stimulation that a $CD4^+$ T cell receives from the antigen presenting cell. Subsequent induction of one or several transcription factors determines the subset a $CD4^+$ T cells commits to and the cytokines it produces. Committed Th cells may either join the germinal center reaction as follicular Th cells, or interact with innate and adaptive immune cells outside of germinal centers. There is a notable plasticity, and a $CD4^+$ T cell of one subset can, upon different stimuli, further commit to a different subset (Ruterbusch et al., 2020; Sallusto, 2016). Importantly, members of a single T cell clone can commit to different T helper cell subsets by undergoing asymmetric cell division and by exposure to different cytokine environments and antigen concentrations (Chang et al., 2007; Hale and Ahmed, 2015; Lee et al., 2017).

After a phase of clonal expansion, most cells of a clone undergo apoptosis. However, a small number of cells survive this phase of clonal contraction as $CD4^+$ memory cells that downregulate the expression of CD45RA (MacLeod et al., 2009). Within the memory compartment, several subsets exist.

$CD4^+CD45RA^-CD62L^+CCR7^+$ central memory T (TCM) cells can traffic through the blood and secondary lymphoid organs. TCM are uncommitted to any specific cytokine-producing subset.

Rather, they secrete IL-2, have a high proliferative capacity upon antigen-stimulation and can re-expand a given clone (Chtanova et al., 2005; Gray et al., 2018; MacLeod et al., 2009; MacLeod et al., 2010; Riou et al., 2007; Schiott et al., 2004; Weng et al., 2012).

$CD4^+CD45RA^-CD62L^-CCR7^-$ effector memory T cells (TEM) are able to infiltrate inflamed tissues, are more differentiated then TCM and able to secrete effector cytokines upon antigen stimulation (Farber et al., 2014; Gray et al., 2018; MacLeod et al., 2009; Riou et al., 2007; Weng et al., 2012). Compared to TCM cells, TEM cells are reported to have a higher proliferation and turnover rate in steady state (Macallan et al., 2004).

A subset of the effector memory cell compartment are cytotoxic $CD4^+$ T cells (CTL) that arise in the context of chronic antigen stimulation and inflammation, such as chronic viral infection with HIV-1, CMV, or EBV, or malignancies (Abana et al., 2017; Appay et al., 2002; Juno et al., 2017; Oh and Fong, 2021; Zaunders et al., 2004). $CD4^+$ CTLs can arise from various Th subsets but are most closely related to Th1 cells. They express cytotoxic molecules such as granzymes, perforins, and granulysin, that enable killing of a target cell in a MHCII-restricted manner (Appay et al., 2002; Hashimoto et al., 2019; Takeuchi and Saito, 2017).

1.5 Approaches to HIV cure

Very few cases of HIV remission are reported. In three cases, a person living with HIV received an allogenic bone marrow transplant with a mutation in the CCR5 receptor for an malignant hematological disorder (Gupta et al., 2019; Hutter et al., 2009). Recently, HIV-1 remission after a haplo-cord transplant with a mutation in the CCR5 receptor for a malignant hematological disorder was reported (Marley et al., 2022). However, due to the high risk of a bone marrow ablation

transplantation (Copelan, 2006), this approach is restricted for the indication of hematological disorders.

Proviral gene editing to render genetically intact proviruses replication-incompetent is in its early stages. CRISPR/Cas-mediated disruption of integrated HIV-1 proviruses was performed *in vitro* (Wang et al., 2016) and *in vivo* in animal models of HIV-1 infection (Dash et al., 2019; Kaminski et al., 2016).

Two concepts of altering the proviral expression of the HIV-1 reservoir exist: “Block and lock”, as well as “shock and kill”. The “block and lock” strategy aims to permanently silence proviral transcription and thus to functionally cure HIV, for example by epigenetic modification of the provirus (Ahlenstiel et al., 2020). In contrast, the “shock and kill” approach aims to activate proviral transcription by latency reversing agents (LRAs) in the present of antiretroviral drugs. Subsequently, HIV-expressing cells would either die due to cytopathic effects of the virus, be eliminated by the immune system or could be targeted by broadly neutralizing antibodies, achieving a sterile cure of HIV. However, the overall effect of latency-reversing agents *in vivo* is very low and no significant reduction of the HIV-1 reservoir has been observed (Ait-Ammar et al., 2019).

1.6 Rationale

In order to tailor a cure for HIV, deeper understanding of the HIV-1 reservoir is needed. In particular, characterization of cells with an intact integrated provirus on a single cell level for gene expression and phenotypic profiling without prior T cell stimulation and induction of proviral transcription has not been achieved yet. This is due to the very low frequency of infected cells

harboring an intact provirus and the lack of a marker to detect infected cells harboring an intact provirus without reactivation of HIV.

In this thesis, we identified expanded $CD4^+$ T cell clones harboring an intact provirus from six individuals living with HIV under ART. Based on the unique TCR of each expanded infected $CD4^+$ T cell clone of interest that harbors an intact provirus, we devised a method to enrich and identify each respective infected clone without the necessity to reactivate HIV transcription.

The results presented in this thesis have been accepted for publication. The experiments were conceived and designed together with Yotam Bar-On and Michel C. Nussenzweig, and the input from Mila Jankovic and Lilian B. Cohn. Christian Gaebler and Marina Caskey recruited study participants, supervised sample collection, and collected clinical data. Data analysis was performed together with Michel C. Nussenzweig and the input from Mila Jankovic and Lilian B. Cohn. Thiago Y. Oliveira and Victor Ramos performed the bioinformatic analysis, generated the UMAP figures, and wrote the bioinformatic method sections. The manuscript for publication was written together with Michel C. Nussenzweig, and the input from Mila Jankovic and Lilian B. Cohn, as well as all other co-authors.

CHAPTER 2. Enrichment strategy

We aimed to enrich expanded $CD4^+$ T cell clones harboring an intact provirus, since cells of a very expanded clone are more abundant than single or oligoclonal infected cells.

Each infected cell of a clone harboring a provirus shares three unique molecular identifiers: The integrated provirus, the proviral integration site, and the TCR. The provirus can be identified by PCR amplification and sequencing. Proviral transcription is only detectable in approximately 30 % of cell carrying a provirus (Einkauf et al., 2022), occurs on a very low level in resting $CD4^+$ T cells (Wiegand et al., 2017), and can consist of mostly incomplete proviral transcripts that do not provide insight on the genetic intactness of the respective provirus (Einkauf et al., 2022; Yukl et al., 2018). Therefore, proviral transcripts are unsuitable to identify infected $CD4^+$ T cell clones harboring an intact provirus. In contrast, the TCR is highly transcribed and expressed on the surface, even in resting $CD4^+$ T cells. Thus, if the TCR of a $CD4^+$ T cell clone with an intact integrated provirus is known, this infected clone can be identified based on its unique TCR irrespective of proviral transcription.

In the following work, we stained and sorted $CD4^+$ T cells from ART-treated individuals living with HIV for expression of CD45RA and the TCRβ variable and constant domain. Subsequent gDNA extraction from the sorted populations, limiting dilution, and proviral PCR amplification and sequencing identified the respective markers that enriched the expanded infected clone of interest that harbors an intact provirus.

2.1 Clinical characteristics of the study participants

For this study, six participants living with HIV-1 were included. At the time point of sample collection, all individuals were under ART and aviraemic and had not undergone experimental treatment interventions (Table 1).

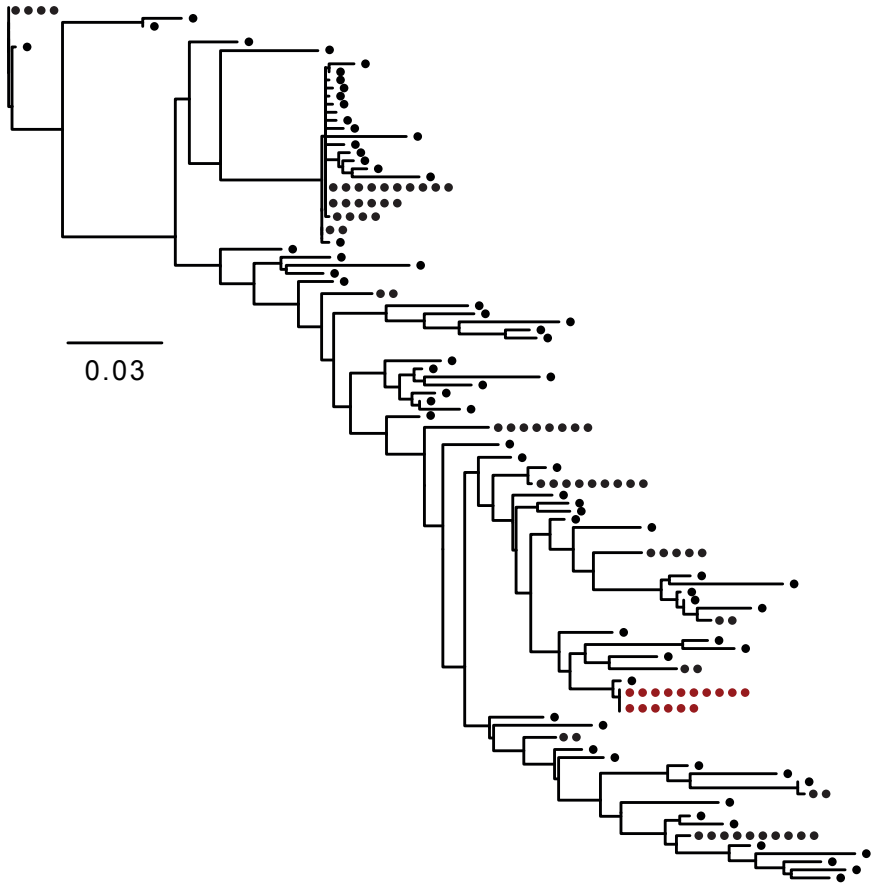
The HIV-1 reservoir of each individual had previously been extensively studied and sequenced by the Nussenzweig laboratory. In each individual, there was an expanded infected clone that harbored an intact provirus and that we aimed to enrich. This infected clone of interest was able to be identified based on its *env* sequence in each individual (Figure 2). In addition, for individuals 603, 605, and B207, the unique TCRαβ sequence of the expanded infected clone of interest was known (Cohn et al., 2018).

Before enrichment, the frequency of the infected clone of interest ranged from 13 to 431 cells per 10^6 $CD4^+$ T cells, based on limiting dilution *env* PCR (individuals 603, 605, and B207), or based on limiting dilution Q4PCR (individuals 5104, 5125, and 9247) (Gaebler et al., 2021; Gaebler et al., 2022) (Table 2).

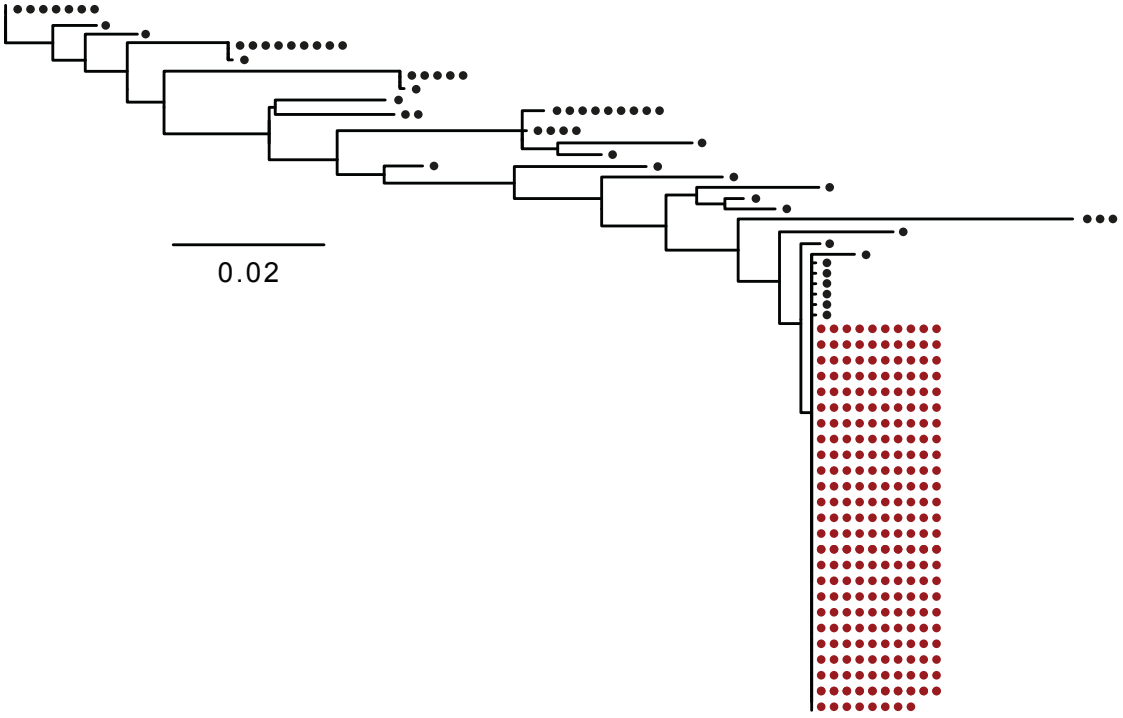
Table 1 Clinical characteristics of study participants.
EVG: elvitegravir; Cobi: cobicistat; TDF: tenofovir disoproxil fumarate; FTC: emtricitabine; RPV: rilpivirine; TAF: tenofovir alafenamide fumurate; BIC: bictegravir.

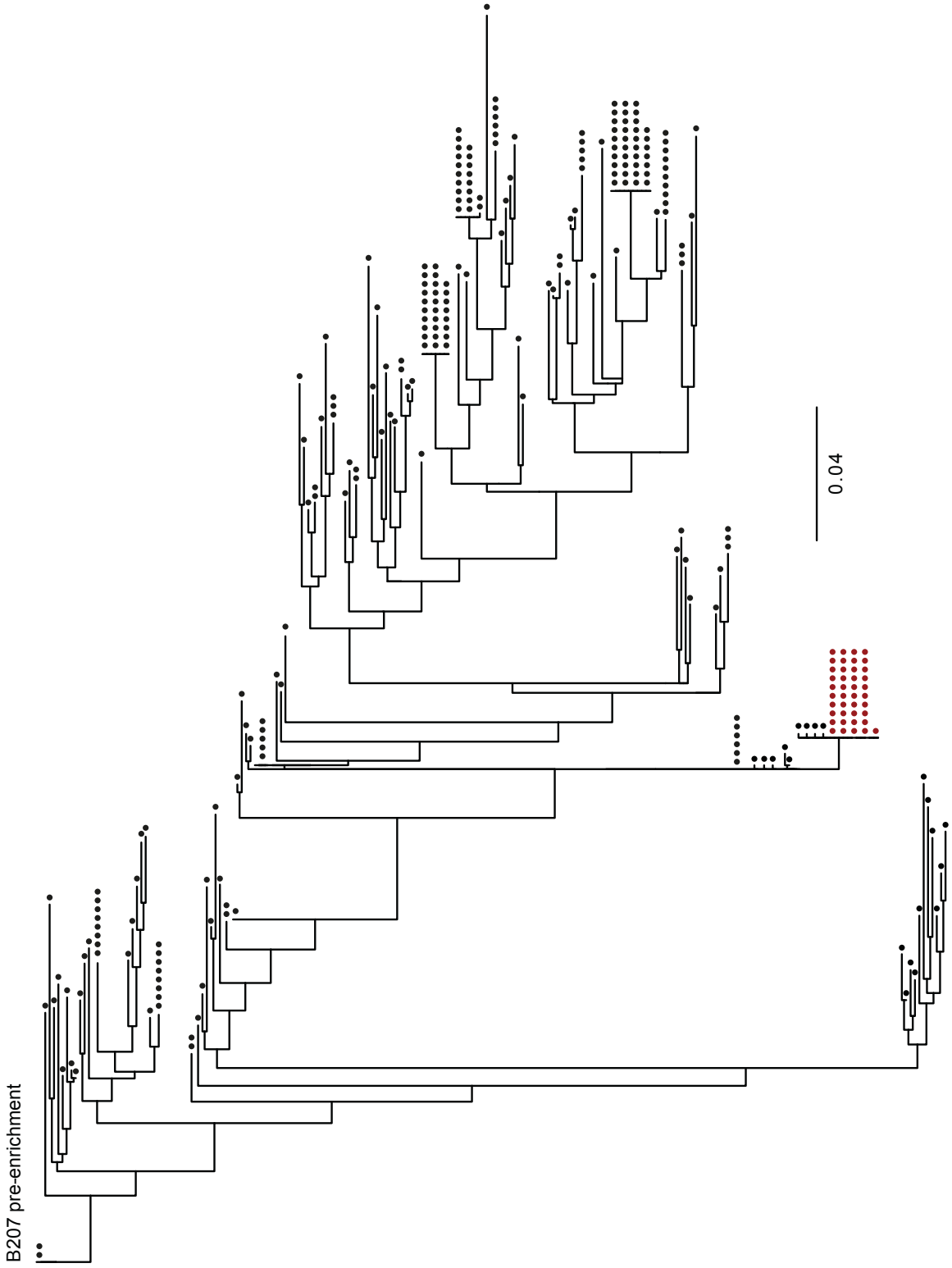
Participant ID	**603**	**605**	**B207**	**5104**	**5125**	**9247**
Age	43	36	47	35	34	31
Sex	M	M	M	M	M	M
Race	White/ Hispanic	White/ Hispanic	White/ Hispanic	Black	Black	Black
Years since HIV-1 diagnosis	12	15	10	7	11	6
Years since ART initiation	10	14	10	7	10	6
Years of un-interrupted ART	10	2	9	7	2	6
Viral load (copies/mL)	< 20	< 20	< 20	< 20	< 20	< 20
$CD4^+$ T cell count	300	371	724	400	450	728
Reported nadir	693	524	50	606	1006	600
ART regimen	EFV/TDF/ FTC	RPV/TDF/ FTC	EFV/TDF/ FTC	BIC/TAF/ FTC	DTG/TDF/ FTC	EVG/Cobi/ TAF/FTC

603 pre-enrichment



605 pre-enrichment




B207 pre-enrichment
0.04

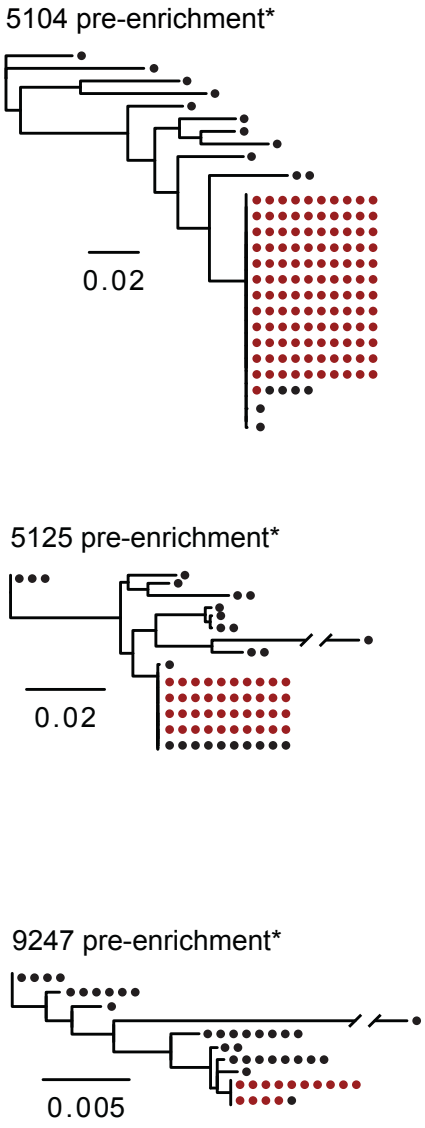


Figure 2 The HIV-1 reservoir in each individual contains an expanded infected clone harboring an intact provirus.
Maximum-likelihood phylogenetic trees show *env* genes from $CD4^+$ T cells. Each dot on the maximum-likelihood phylogenetic trees represents one *env* sequence. The *env* gene that identifies the infected clone of interest is marked red. The scale bars indicate the number of substitutions per site.
For individuals B207, 603, and 605, the *env* gene of the infected clone of interest was previously found in intact NFL sequences from viral outgrowth assays (Cohn et al., 2018). The *env* genes in the trees were recovered by Env PCR from genomic DNA.
For individuals 5104, 5125, and 9247, an expanded infected clone of interest was identified by Q4PCR (marked with the asterisk) from genomic DNA (Gaebler et al., 2021; Gaebler et al., 2022). *Env* sequences were extracted from near-full-length proviral sequences.

Table 2 Frequency, inducibility, and integration site of the infected clone of interest in CD4$^+$ T cells.

CUPM: Clonal units of clone of interest per million CD4$^+$ T cells. [1] (Gaebler et al., 2022), [2] (Gaebler et al., 2021), [3] (Cohn et al., 2018), [4] (Huang et al., 2021)

	603	**605**	**B207**	**5104**	**5125**	**9247**
CUPM Q4PCR	ND	ND	ND	210[1]	28[1]	15[2]
CUPM Env PCR	13	431	164	ND	ND	ND
Infected clone inducible *in vitro*	Yes[3]	Yes[3]	Yes[3,4]	Yes[4]	ND	No
Integration site	ZNF486[4]	ND	ZPF30[4]	ATP2B4, DUXL6[4]	ND	ND

2.2 Enrichment based on CD45RA

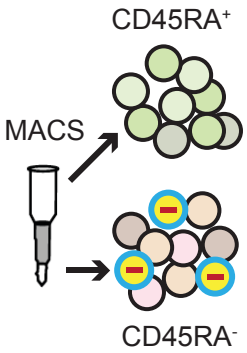
Integrated HIV-1 DNA can be found in both, the naïve and the memory $CD4^+$ T cell compartment. The naïve $CD4^+$ T cell compartment contains integrated proviruses at a low frequency, but with a high proportion of intact proviruses (Venanzi Rullo et al., 2020). However, genetically intact proviruses are overall enriched in the memory compartment (Chomont et al., 2009; Duette et al., 2022; Hiener et al., 2017; Lee et al., 2017; Morcilla et al., 2021; Venanzi Rullo et al., 2020; Zerbato et al., 2019). Naïve $CD4^+$ T cells express CD45RA, which is downregulated upon differentiation into $CD4^+$ T cell memory subsets, and can be re-expressed by terminally differentiated $CD4^+$ T cells (Sallusto and Lanzavecchia, 2011).

To identify if the infected clone of interest in each individual was enriched in the $CD45RA^-$ memory compartment, we magnetically sorted $CD4^+$ T cells into a $CD45RA^+$ and $CD45RA^-$ population, performed limiting dilution of the gDNA of each population, and PCR amplified and sequenced the respective *env* genes (Figure 3a).

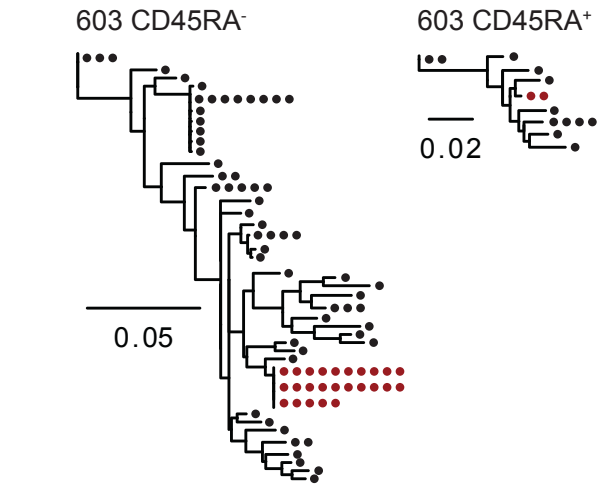
In each individual, the infected clone of interest was almost exclusively found in the $CD4^+CD45RA^-$ memory T cell compartment, in accordance with previous studies (Figure 3b).

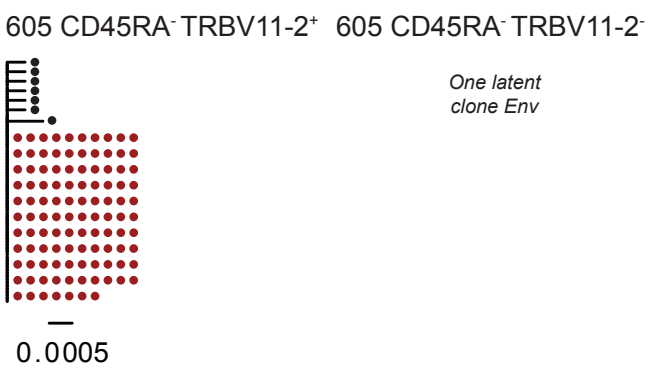
To quantify the relative enrichment of the infected clone of interest, $CD4^+$ T cells from five individuals were stained with an anti-CD45RA antibody and analyzed by flow cytometry (Figure 3c). The fraction of $CD45RA^-$ cells among $CD4^+$ T cells ranged from 42 % to 68 %, depending on the individual, yielding a 1.5- to 2.4-fold enrichment of the clone of interest (Table 4). In conclusion, a very modest enrichment of the infected clones of interest based on CD45RA sorting was achieved.

a

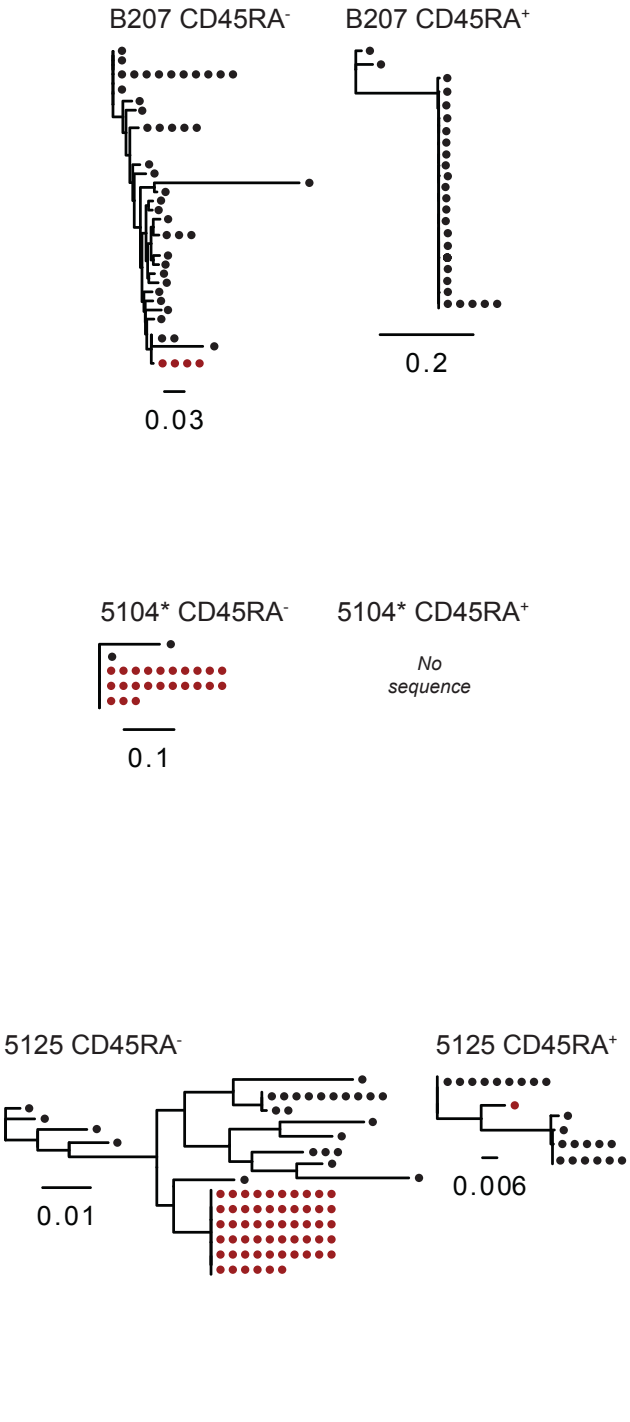

CD45RA+
MACS
CD45RA-

b


603 CD45RA-
603 CD45RA+
0.02
0.05


605 CD45RA- TRBV11-2+
605 CD45RA- TRBV11-2-
One latent clone Env
0.0005

b (cont.)


B207 CD45RA-
B207 CD45RA+
0.2
0.03
5104* CD45RA-
5104* CD45RA+
No sequence
0.1
5125 CD45RA-
5125 CD45RA+
0.01
0.006
9247 CD45RA-
9247 CD45RA+
One non-clonal Env
0.01

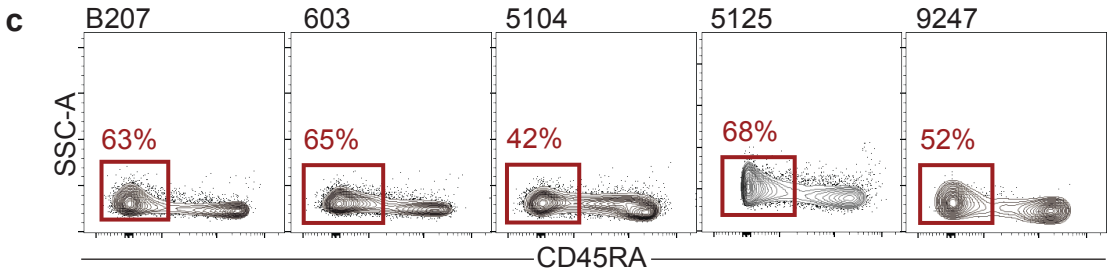


Figure 3 Screening for enrichment of the infected clone of interest by CD45RA sorting.
(a) $CD4^+$ T cells from each individual were magnetically sorted based on CD45RA expression. The gDNA of the $CD45RA^+$ and $CD45RA^-$ population was extracted, limitingly diluted, and Q4PCR or *env* PCR was performed for *env* sequencing.
(b) Maximum-likelihood phylogenetic trees show *env* genes from $CD4^+CD45RA^-$ and $CD4^+CD45RA^+$ T cells. Each dot on the maximum-likelihood phylogenetic trees represents one *env* sequence. The *env* gene of the infected clone of interest is marked red. The scale bars indicate the number of substitutions per site. The asterisk marks maximum-likelihood phylogenetic trees that are based on *env* genes extracted from NFL sequences that were produced by Q4PCR. All other maximum-likelihood phylogenetic trees are based on *env* PCR.
The infected clone of interest in individual 605 expresses TRBV11-2 (Cohn et al., 2018). For this individual, the trees show *env* sequences from the $CD45RA^-TRBV11\text{-}2^+$ and the $CD45RA^+TRBV11\text{-}2^+$ $CD4^+$ T cell populations.
(c) $CD4^+$ T cells were stained for CD45RA. The flowcytometry plots show the expression of CD45RA by $CD3^+CD4^+$ T cells. The red gates show the $CD45RA^-$ population, and the number above each gate shows the fraction of cells within the gate.

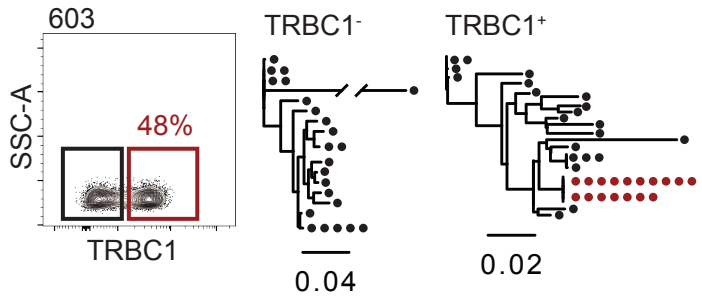
2.3 Enrichment based on TRBC

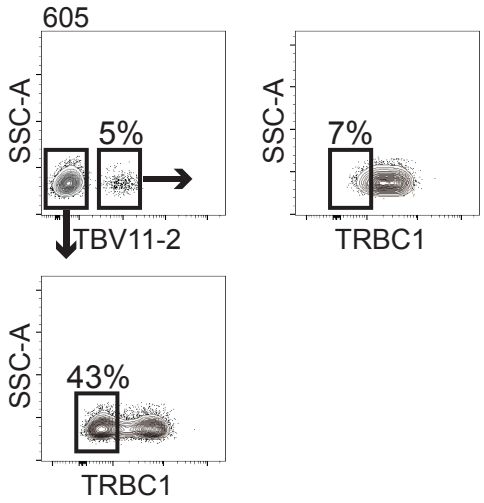
The TCRβ constant domain (TRBC) is encoded by two genes, TRBC1 and TRBC2 (IMGT, 2022). Due to allelic exclusion, only one of the two genes is expressed by a given T cell or T cell clone (Bassing et al., 2002; Nikolich-Zugich et al., 2004; Nussenzweig et al., 1987).

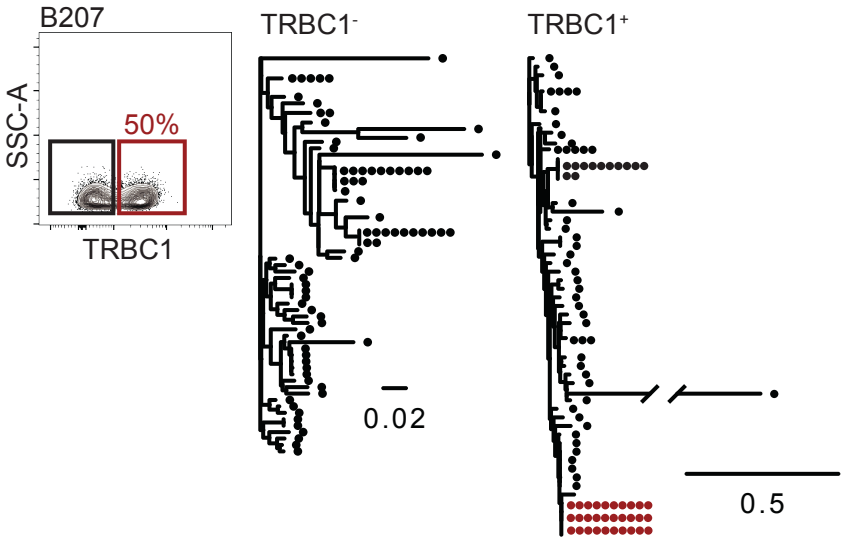
To identify the TRBC gene that is expressed by the infected clone of interest in each individual, $CD4^+$ T cells were stained with an anti-TRBC1 antibody and sorted into a $TRBC1^+$ and $TRBC1^-$ (and thus TRBC2-expressing) population (Figure 4). The gDNA was extracted from each population and subjected to limiting dilution, followed by *env* PCR amplification and sequencing to detect the infected clone of interest.

In individuals B207, 603, 5104, and 5125, the infected clone of interest was found to express TRBC1. In individual 9247, the infected clone of interest expressed TRBC2.

The infected clone of interest in individual 605 expresses TRBV11-2 (Cohn et al., 2018). More than 90 % of $CD4^+TRBV11\text{-}2^+$ T cells in this individual expressed TRBC1 and did not present a clear $TRBC1^-$ population (Figure 4). The infected clone of interest in individual 605 was therefore determined to express TRBC1.


603
SSC-A
48%
TRBC1
TRBC1-
TRBC1+
0.04
0.02


605
SSC-A
5%
TBV11-2
SSC-A
7%
TRBC1
SSC-A
43%
TRBC1


B207
SSC-A
50%
TRBC1
TRBC1-
TRBC1+
0.02
0.5

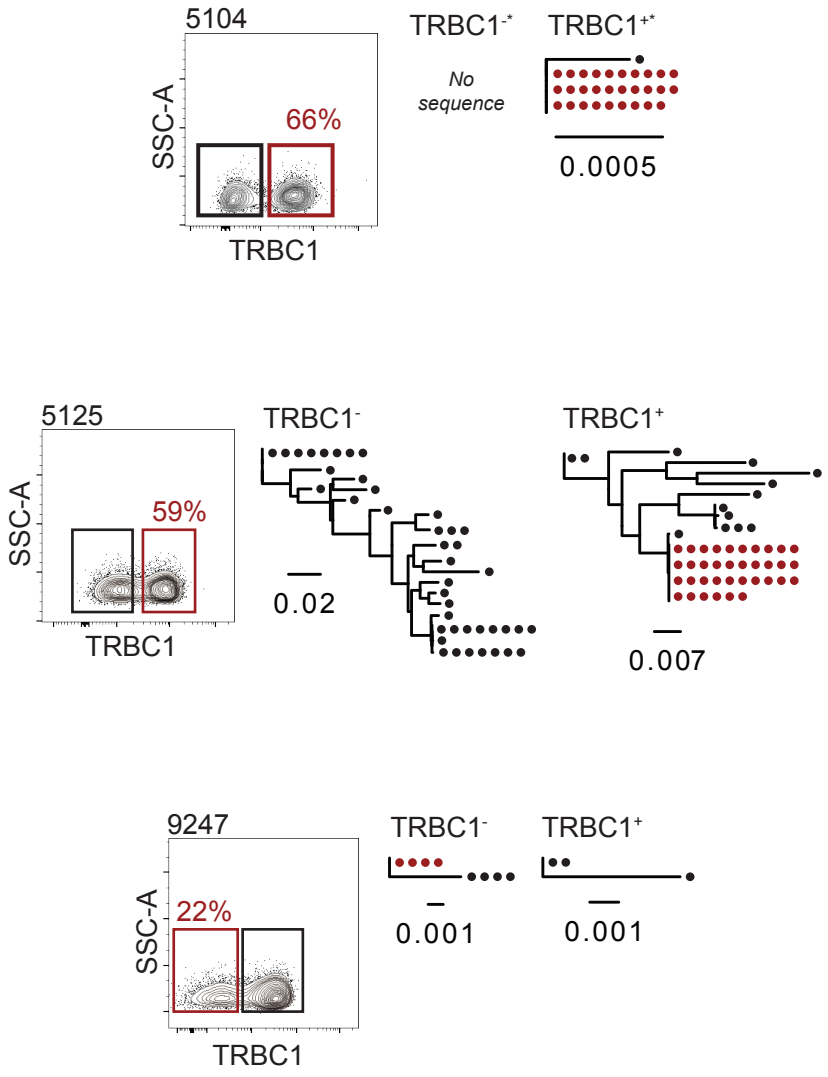


Figure 4 Screening for enrichment of the infected clone of interest by TRBC sorting.
Red outlines indicate the population containing the infected clone of interest. The number above the gate shows the fraction of cells within the gate. In the maximum-likelihood phylogenetic trees, each dot represents a recovered *env* sequence from the respective subpopulation of $CD4^+$ T cells. The *env* gene of the infected clone of interest is marked in red. The scale bars indicate the number of substitutions per site. The asterisk marks maximum-likelihood phylogenetic trees that are based on *env* genes extracted from NFL sequences that were produced by limiting dilution Q4PCR. All other maximum-likelihood phylogenetic trees are based on limiting dilution *env* PCR.

2.4 Enrichment based on TRBV

The TCRβ variable domain (TRBV) is encoded by 48 genes (IMGT, 2022). Due to allelic exclusion, only one TRBV gene is expressed by a given T cell or T cell clone (Bassing et al., 2002; Nikolich-Zugich et al., 2004; Nussenzweig et al., 1987). The exact TCRαβ sequences of the infected clones of interest in individuals 603, 605, and B207 had been sequenced previously and the respective TRBV genes identified to be TRBV-19, TRBV11-2, and TRBV7-8 (Cohn et al., 2018). In contrast, the exact TCRαβ sequences of the infected clones of interest in individuals 5104, 5125, and 9247 were unknown and the respective TRBV genes remained to be identified.

For identification of the TRBV gene that was expressed by the infected clone of interest, the Beta Mark TCR Vbeta Repertoire Kit (Beckman Coulter) was used. The kit comprises 24 anti-TRBV antibodies that are distributed over 8 vials (A – H). In each vial, there are three antibodies, each conjugated with either phycoerythrin (PE), fluorescein isothiocyanate (FITC), or PE-FITC (Table 3).

Table 3 Anti-TRBV antibody cocktails of the Beta Mark TCR Vbeta Repertoire Kit (Beckman Coulter).

Vial	Target TRBV	Fluorochrome
A	TRBV5-5	PE
	TRBV4-1, TRBV4-2, TRBV4-3	PE-FITC
	TRBV28	FITC
B	TRBV3-1	PE
	TRBV19	PE-FITC
	TRBV14	FITC
C	TRBV18	PE
	TRBV5-1	PE-FITC
	TRBV30	FITC
D	TRBV6-5, TRBV6-6, TRBV6-9	PE
	TRBV6-6	PE-FITC
	TRBV12-3, TRBV12-4	FITC
E	TRBV5-6	PE
	TRBV20-1	PE-FITC
	TRBV10-3	FITC
F	TRBV13	PE
	TRBV9	PE-FITC
	TRBV11-2	FITC
G	TRBV25-1	PE
	TRBV2	PE-FITC
	TRBV27	FITC
H	TRBV6-2	PE
	TRBV29-1	PE-FITC
	TRBV4-3	FITC

2.4.1 Enrichment based on TRBV in individual 5014

To identify if the infected clone of interest expressed a TRBV domain that could be stained with the TRBV antibody kit, all 24 anti-TRBV antibodies were applied to $CD4^+$ T cells. Subsequently, a $TRBV^+$ and a $TRBV^-$ population was sorted. Limiting dilution of gDNA from each population and *env* PCR amplification and sequencing indicated that the infected clone of interest was present in the $TRBV^-$ population and could therefore be enriched 2-fold (Figure 5). Thus, the infected clone of interest in this individual expressed one of the 28 TRBV domains that cannot be stained with the anti-TRBV antibody kit.

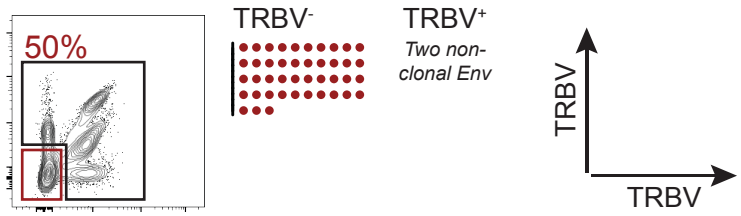


Figure 5 Screening for enrichment of the infected clone of interest by TRBV sorting in individual 5104.
Red outlines indicate the population containing the infected clone of interest. The number above the gate shows the fraction of cells within the red gate. Maximum-likelihood phylogenetic trees of *env* sequences are derived from each sorted population, as indicated. Each dot represents a single *env* gene. The *env* gene of the infected clone of interest is marked in red.

2.4.2 Enrichment based on TRBV in individual 5125

In individual 5125, sorting of $CD4^+$ T cells with all 24 TRBV antibodies and subsequent *env* PCR amplification and sequencing revealed that the infected clone of interest was present in the $TRBV^+$ population and therefore amenable to further enrichment (Figure 6a).

Next, $CD4^+$ T cells were stained with either of 8 sets of 3 antibodies each (vials A – H). The infected clone of interest was only present in population $TRBV^G$, comprising three distinct TRBV domains, based on *env* PCR amplification and sequencing (Figure 6b). The respective three antibodies were again applied to $CD4^+$ T cells and each TRBV domain was sorted individually. This revealed that the infected clone of interest in individual 5125 expressed TRBV2 (Figure 6c).

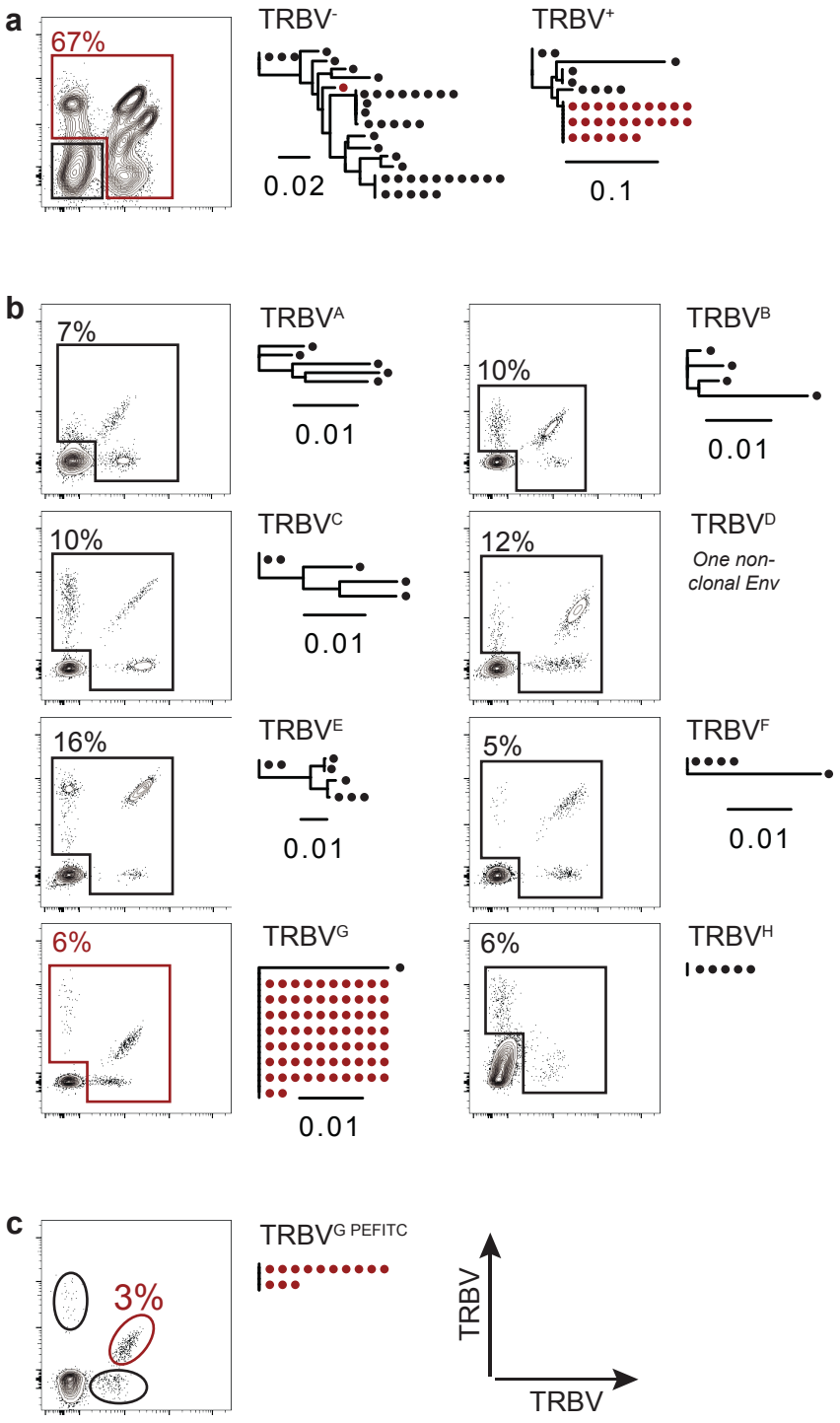

a
67%
TRBV-
0.02
TRBV+
0.1
b
7%
TRBVA
0.01
10%
TRBVB
0.01
10%
TRBVC
0.01
12%
TRBVD
One non-clonal Env
16%
TRBVE
0.01
5%
TRBVF
0.01
6%
TRBVG
0.01
6%
TRBVH
c
3%
TRBVG PEFITC
TRBV
TRBV

Figure 6 Screening for enrichment of the infected clone of interest by TRBV sorting in individual 5125.

Red outlines indicate the population containing the infected clone of interest. The number above each gate shows the fraction of cells within the gate. Maximum-likelihood phylogenetic trees of *env* sequences are derived from each sorted population, as indicated. Each dot represents a single *env* gene. The *env* gene of the infected clone of interest is marked in red. The scale bars indicate the number of substitutions per site.

2.4.3 Enrichment based on TRBV in individual 9247

In individual 9247, initial screening with all 24 anti-TRBV antibodies indicated that the infected clone of interest was enriched in the $TRBV^+$ population, based on *env* PCR amplification and sequencing (Figure 7a). Sorting with two sets of 12 antibodies each (vials A – D and vials E – H) further narrowed down the TRBV domains in question, as the infected clone of interest was only found in one population comprising 12 TRBV domains (Figure 7b). The respective 12 antibodies were divided into four groups of three antibodies. Only one group of three anti-TRBV antibodies contained the infected clone of interest (Figure 7c). The respective three antibodies were again applied to $CD4^+$ T cells and each TRBV domain was sorted individually. This revealed that the infected clone of interest in individual 9247 expressed TRBV4-3 (Figure 7d).

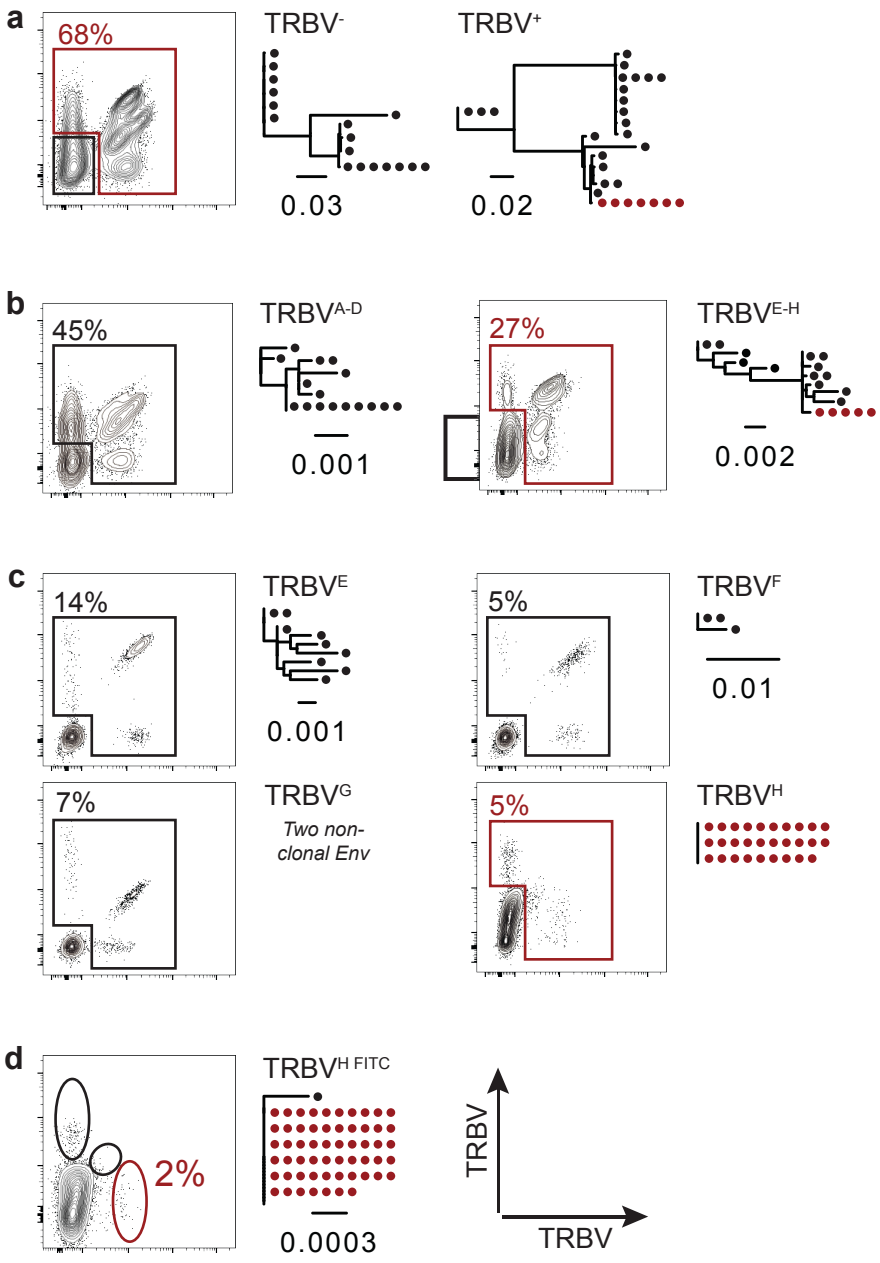


Figure 7 Screening for enrichment of the infected clone of interest by TRBV sorting in individual 9247.

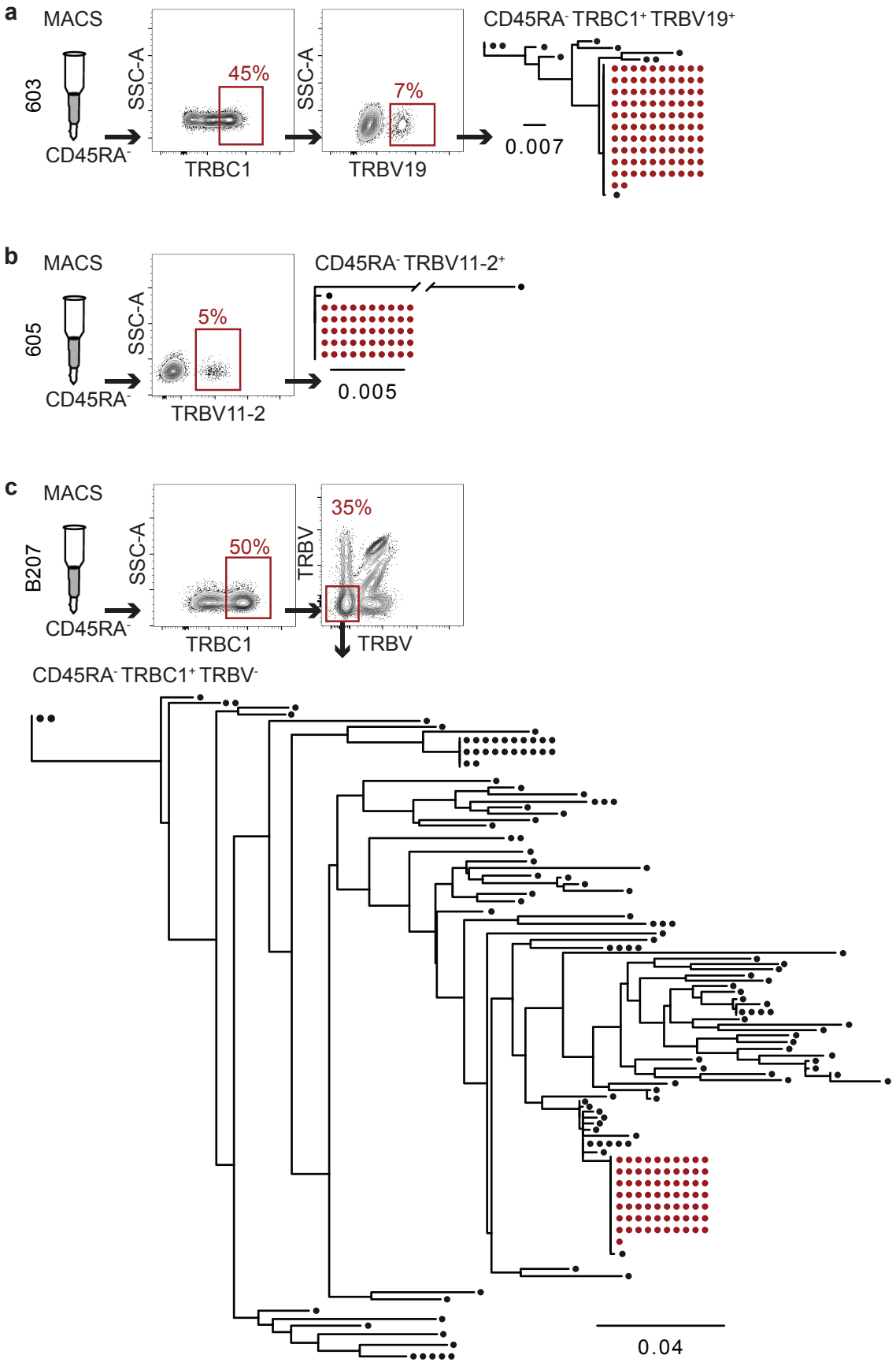
Red outlines indicate the population containing the infected clone of interest. The number above each gate shows the fraction of cells within the gate. Maximum-likelihood phylogenetic trees of *env* sequences are derived from each sorted population, as indicated. Each dot represents a single *env* gene. The *env* gene of the infected clone of interest is marked in red. The scale bars indicate the number of substitutions per site.

2.5 Combined enrichment based on CD45RA, TRBC, and TRBV

To maximize enrichment, the sorting steps based on CD45RA, TRBC, and TRBV were combined for each individual. For individual 603, $CD4^+CD45RA^-TRBC1^+TRBV19^+$ cells were sorted, yielding and overall 47-fold enrichment of the infected clone of interest, based on flowcytometry analysis (Table 4, Figure 8). For individual 605, sorting of $CD4^+CD45RA^-TRBV11\text{-}2^+$ cells yielded an estimated 40-fold enrichment. Since no antibody was available for the TRBV domain that was expressed by the infected clone of interest in individual B207, cells were sorted into a $CD4^+CD3^+CD45RA^-TRBC1^+TRBV^-$ population, with a moderate 9-fold enrichment. Similarly, sorting for the $CD4^+CD3^+CD45RA^-TRBC1^+TRBV^-$ population in individual 5104 led to a moderate 11-fold enrichment. In individual 5125, the infected clone of interest was enriched 54-fold in the $CD4^+CD45RA^-TRBC1^+TRBV2^+$ population. In individual 9247, sorting for the $CD4^+CD45RA^-TRBC1^-TRBV4\text{-}3^+$ population enriched the infected clone of interest 675-fold, based on the flowcytometry analysis (Table 4, Figure 8).

Table 4 Relative enrichment of the infected clone of interest per marker based on flow cytometry.
For individual 605, staining of $CD4^+$ T cells for CD45RA was not performed, and the frequency of memory cells among $CD4^+$ T cells was estimated 50 %.

	CD45RA	TRBC	TRBV	combined
603	1.5	2.2	14.3	47.2
605	~ 2 (estimate)	-	20	~ 40
B207	1.6	2	2.9	9.3
5104	2.4	1.6	2.9	11.1
5125	1.5	1.8	20	54
9247	1.9	5	71	674.5


a
MACS
603
CD45RA-
SSC-A
45%
TRBC1
SSC-A
7%
TRBV19
CD45RA- TRBC1+ TRBV19+
0.007
b
MACS
605
CD45RA-
SSC-A
5%
TRBV11-2
CD45RA- TRBV11-2+
0.005
c
MACS
B207
CD45RA-
SSC-A
50%
TRBC1
TRBV
35%
TRBV
CD45RA- TRBC1+ TRBV-
0.04

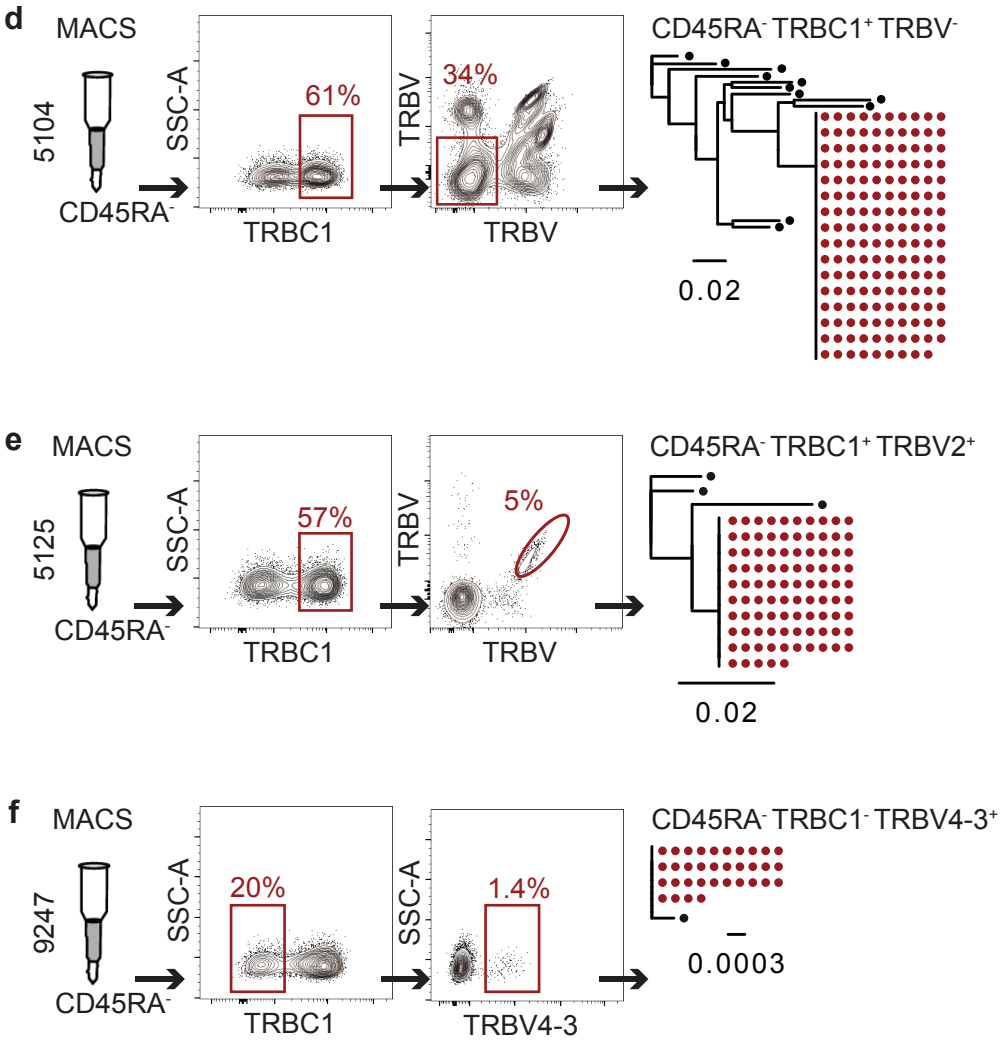

d
MACS
5104
CD45RA-
SSC-A
61%
TRBC1
TRBV
34%
TRBV
CD45RA- TRBC1+ TRBV-
0.02
e
MACS
5125
CD45RA-
SSC-A
57%
TRBC1
TRBV
5%
TRBV
CD45RA- TRBC1+ TRBV2+
0.02
f
MACS
9247
CD45RA-
SSC-A
20%
TRBC1
SSC-A
1.4%
TRBV4-3
CD45RA- TRBC1- TRBV4-3+
0.0003

Figure 8 Combined enrichment of the infected clone of interest based on CD45RA, TRBC, and TRBV expression in individuals 603, 605, B207, 5104, 5125, and 9247.

Red outlines indicate the population containing the infected clone of interest. The number above each gate shows the fraction of cells within the gate. Maximum-likelihood phylogenetic trees of *env* sequences are derived from each sorted population, as indicated. Each dot represents a single *env* gene. The *env* gene of the infected clone of interest is marked in red. The scale bars indicate the number of substitutions per site.

(a) Enrichment in individual 603 after purification of $CD45RA^-$ memory cells followed by staining and sorting for $TRBC1^+$ and $TRBV19^+$ cells. Flow cytometry plots show TRBC1 and TRBV19 staining. Percentages of $CD45RA^-TRBC1^+$, and $CD45RA^-TRBC1^+TRBV19^+$ cells are indicated.

(b) Enrichment in individual 605 after purification of $CD45RA^-$ memory cells followed by staining $TRBV11\text{-}2^+$ cells, which were > 90 % $TRBC1^+$ (Figure 4). Flow cytometry plot shows TRBV11-2 staining. Percentage of $CD45RA^-TRBV11\text{-}2^+$ cells is indicated.

(c) Enrichment in individual B207 after purification of $CD45RA^-$ memory cells followed by staining and sorting for $TRBC1^+$ and $TRBV^-$ cells. Flow cytometry plots show TRBC1 and combined 24 TRBV staining. Percentages of $CD45RA^-TRBC1^+$, and $CD45RA^-TRBC1^+TRBV^-$ cells are indicated.

(d) Enrichment in individual 5104 after purification of $CD45RA^-$ memory cells followed by staining and sorting for $TRBC1^+$ and $TRBV^-$ cells. Flow cytometry plots show TRBC1 and TRBV staining. Percentages of $CD45RA^-TRBC1^+$, and $CD45RA^-TRBC1^+TRBV^-$ cells are indicated.

(e) Enrichment in individual 5125 after purification of $CD45RA^-$ memory cells followed by staining and sorting for $TRBC1^+$ and $TRBV2^+$ cells. Flow cytometry plots show TRBC1 and TRBV2 staining. Percentages of $CD45RA^-TRBC1^+$, and $CD45RA^-TRBC1^+TRBV2^+$ cells are indicated.

(f) Enrichment in individual 9247 after purification of $CD45RA^-$ memory cells followed by staining and sorting for $TRBC1^-$ and $TRBV4\text{-}3^+$ cells. Flow cytometry plots show TRBC1 and TRBV4-3 staining. Percentages of $CD45RA^-TRBC1^-$, and $CD45RA^-TRBC1^-TRBV4\text{-}3^+$ cells are indicated.

CHAPTER 3. TCR identification of infected clones of interest.

3.1 Combined *env* and TCR sequencing

TRBC- and TRBV- based sorting of $CD4^+$ T cells of individuals 5104, 5125, and 9247 enabled identification of the respective gene that is expressed by the infected clone of interest in each individual, whereas the exact TCRαβ sequences remained unknown.

To obtain the TCRαβ sequence of the infected clone of interest, enrichment based on CD45RA, TRBC, and TRBV was performed as described above. The enriched sample was bulk sorted for 10x Genomics TCR sequencing and, in addition, was sorted at a frequency of 5 cells per well into multiwell plates (Figure 9a).

The 10x Genomics TCR sequencing revealed the TCRαβ sequences (10x TCR clonotypes) of all the expanded T cell clones and single T cells that were present in the enriched sample. One of the expanded T cell clones in each sample was presumably the infected clone of interest.

Amplification and sequencing of *env* genes from the multiwell plates revealed which wells contained a cell of the infected clone of interest (Env^+ wells). Presumably, Env^+ wells would contain four random TCRs and one TCR that pertains to the infected clone of interest. This infected clone of interest TCR would be associated with the majority of or all Env^+ wells. In contrast, wells in which no cell of the infected clone of interest was detected (Env^- wells) would contain five random TCRs and would not be associated with the infected clone of interest TCR.

TCR amplification and sequencing was performed on all Env^+ wells and a random selection of Env^- wells. The frequency of each 10x TCR clonotype from the 10x Genomics TCR sequencing data was determined in the Env^+ and Env^- wells. A 10x TCR clonotype was considered to be present in a well if its unique TCRα, or unique TCRβ, or both, were recovered from a well.

3.2 Verification of the approach

The TCRαβ sequences of the infected clones of interest in individuals 603 and 605 were already known (Cohn et al., 2018). The enriched samples of these two individuals were therefore chosen to verify the approach.

For individual 603, the TCRs of 8 Env^+ and 77 Env^- wells were amplified and sequenced. The infected clone of interest TCR (10x TCR clonotype 43582) was present in 62.5 % of Env^+ wells, but in only 4.2 % of Env^- wells. In contrast, TCRs that were not associated with the infected clone of interest were not enriched in Env^+ wells. For example, the next frequent 10x TCR clonotypes 3122, 63760, and 77395 were each present in 1 out of 8 Env^+ wells (12,5 %) (Figure 9b).

For individual 605, the TCRs of 21 Env^+ wells and 62 Env^- wells were amplified and sequenced. The infected clone of interest TCR (10x TCR clonotype 21327) was present in 90.5 % of Env^+ wells, but in only 6.5 % of Env^- wells. No other 10x TCR clonotype was enriched in Env^+ wells, as the next frequent 10x TCR clonotypes in Env^+ wells were clonotypes 1580 and 52471 (9.5 % each). In conclusion, the combination of 10x Genomics TCR sequencing together with *env* and TCR sequencing from multiwell plates is a reliable method to identify the TCR of expanded infected clones.

3.3 TCR identification of infected clones of interest

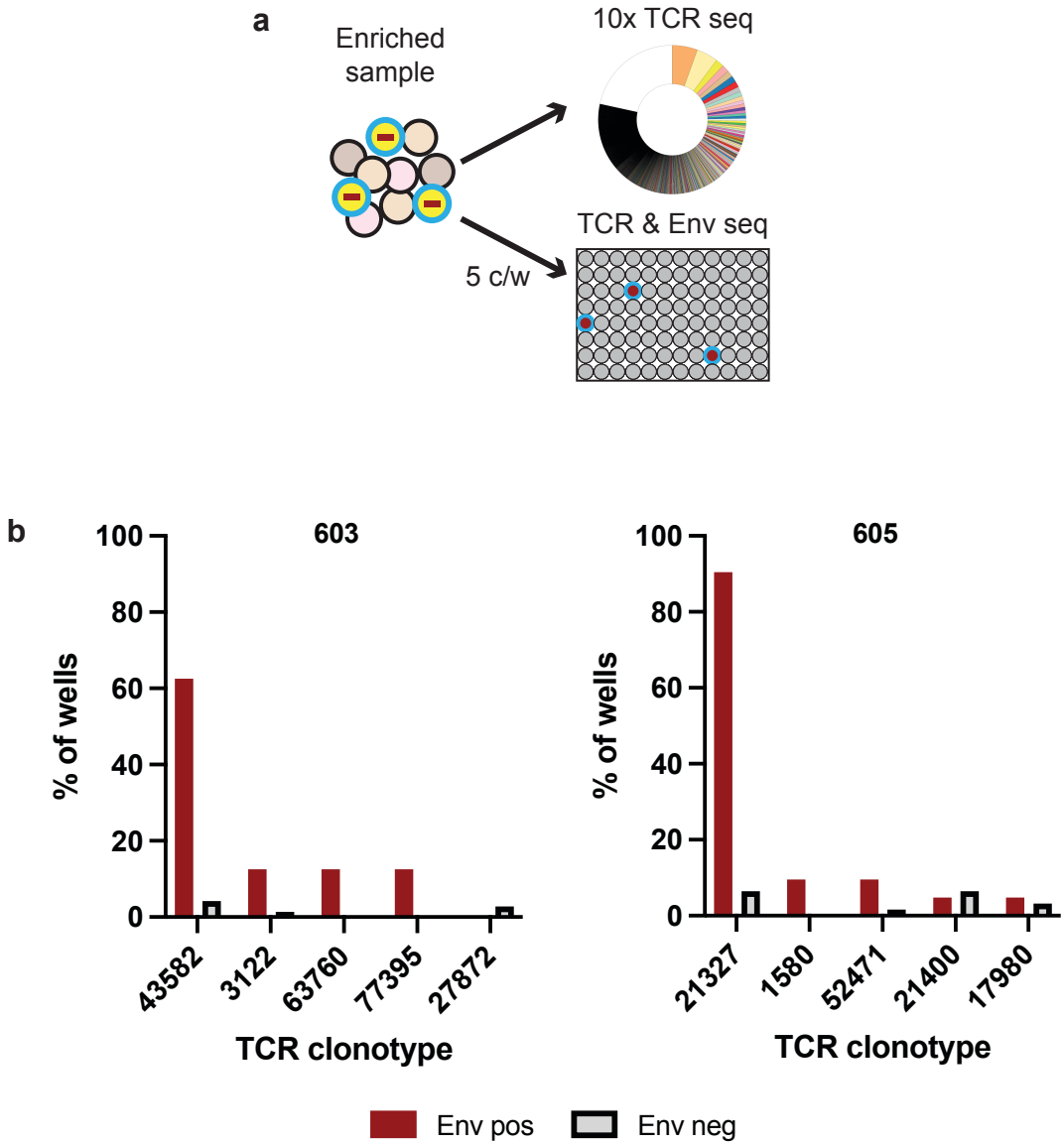
For individual 5104, the TCRs of 7 Env^+ wells and 36 Env^- wells were amplified and sequenced. Only 10x TCR clonotype 10133 (TRAV12-3/J9/TRBV5-4/J1-1) was present in all Env^+ wells and absent in all Env^- negative wells and was therefore identified as the infected clone of interest TCR in this individual (Figure 9c). No other 10x TCR clonotype was exclusively enriched in Env^+ wells

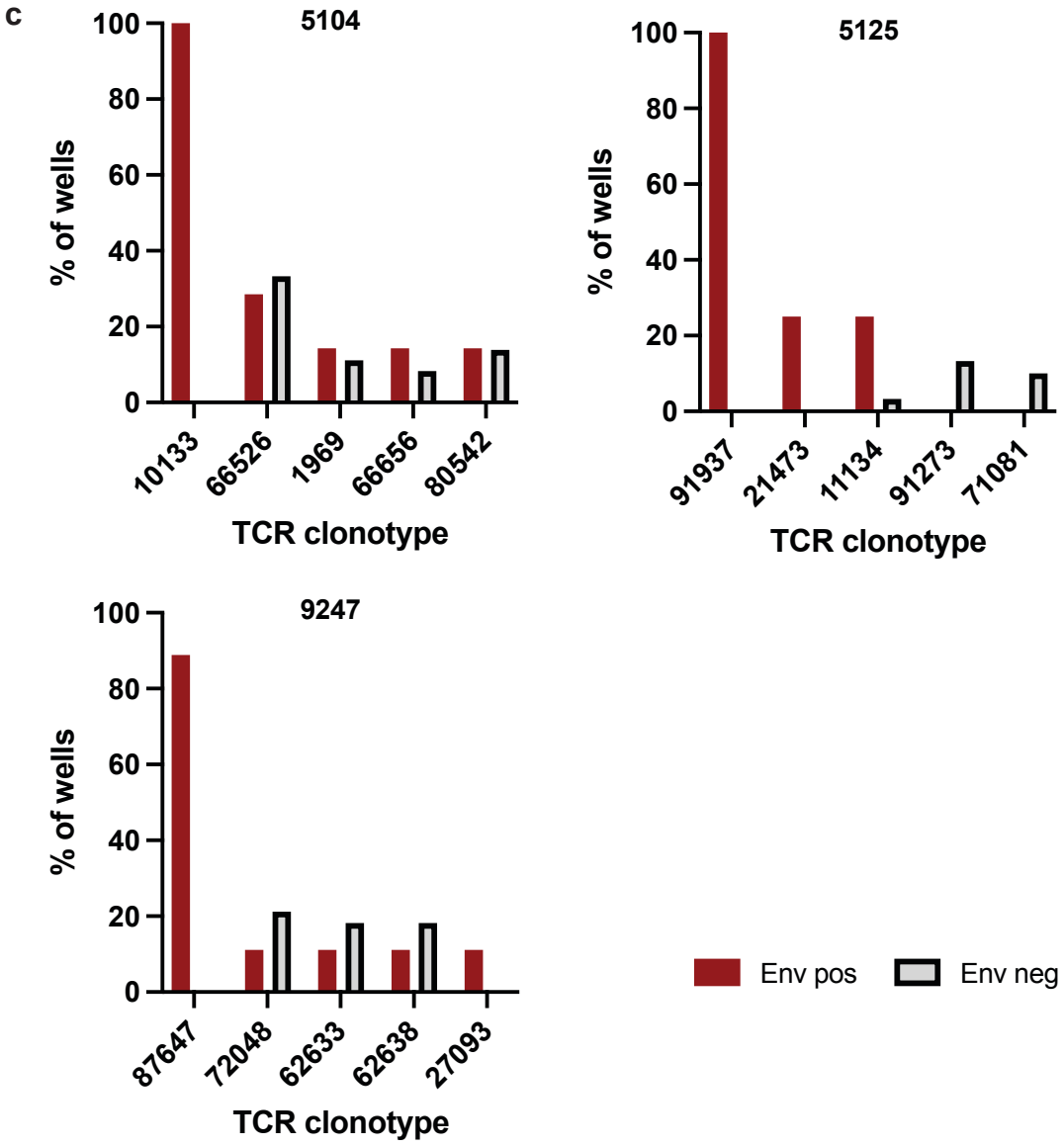
to the same extent, as the next frequent 10x TCR clonotype 66526 was found in only 28.6 % of Env^+ wells, but also in 33.3 % of Env^- wells (Figure 9c).

For individual 5125, the TCRs of 4 Env^+ and 31 Env^- wells were amplified and sequenced. 10x TCR clonotype 91937 (TRAV26-2/J32/TRBV2/J1-1) was present in all Env^+ wells and absent in all Env^- wells and therefore identified as the infected clone of interest TCR. The next frequent 10x TCR clonotypes 21473 and 11134 were present in only one Env^+ well each (25 %) (Figure 9c).

For individual 9247, the TCRs of 9 Env^+ and 42 Env^- wells were amplified and sequenced. 10x TCR clonotype 87647 (TRAV38-1/J33/TRBV4-3/J2-3) was found in 88.9 % of Env^+ wells, but in no Env^- well. 10x TCR clonotype 87647 was therefore identified as the infected clone of interest TCR in individual 9247. The next frequent 10x TCR clonotypes 72048, 62633, 62638, and 27093 were present in only in one Env^+ well each (11 %) (Figure 9c).

Subsequently, the infected clone of interest was identified among all T cell clones and single cells in the 10x Genomics TCR sequencing data of each individual based on its unique TCR (Figure 9d).


a
Enriched sample
10x TCR seq
TCR & Env seq
5 c/w
b
603
605
% of wells
100
80
60
40
20
0
43582
3122
63760
77395
27872
21327
1580
52471
21400
17980
TCR clonotype
Env pos
Env neg


c
5104
% of wells
100
80
60
40
20
0
10133
66526
1969
66656
80542
TCR clonotype
5125
% of wells
100
80
60
40
20
0
91937
21473
11134
91273
71081
TCR clonotype
9247
% of wells
100
80
60
40
20
0
87647
72048
62633
62638
27093
TCR clonotype
Env pos
Env neg

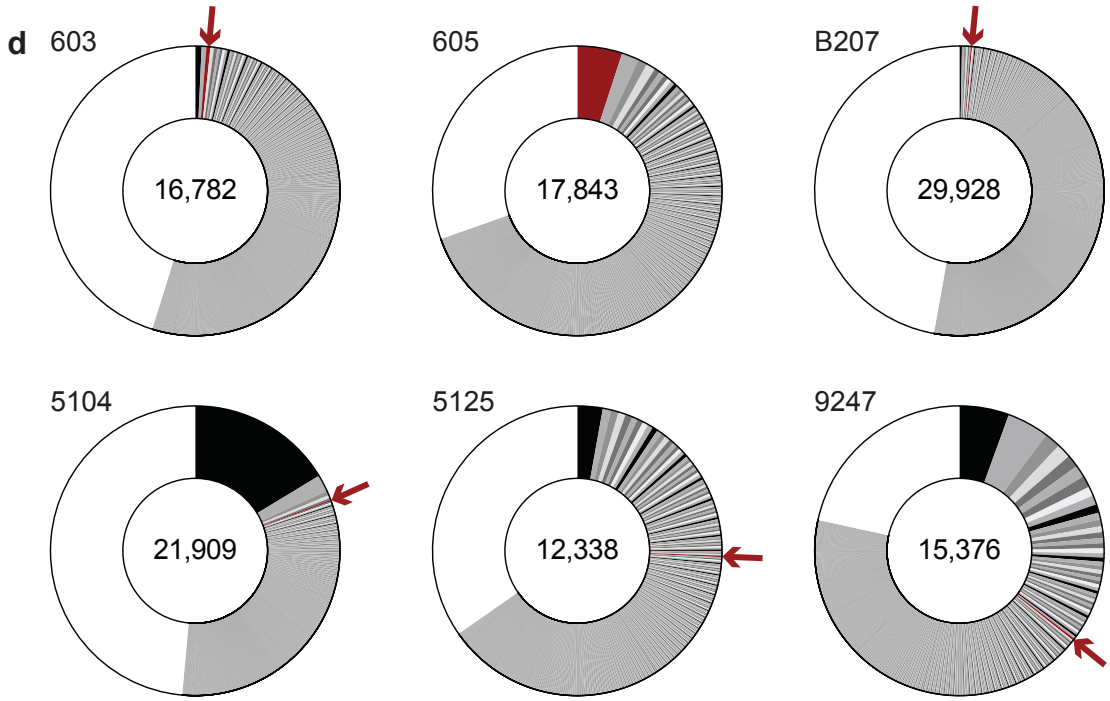


Figure 9 TCR identification of infected clones of interest.
(a) After enrichment based on CD45RA, TRBC, and TRBV, the sample was subjected to 10x Genomics TCR sequencing and sorted at a frequency of 5 cells per well into multiwell plates. Env^+ wells were identified based on *env* PCR amplification and sequencing. All Env^+ and a random selection of Env^- wells were selected for TCR sequencing.
(b) The number above each graph is the study participant ID. The x-axis shows the 5 most frequent 10x TCR clonotypes in Env^+ wells for the respective individual. For each 10x TCR clonotype, its frequency in Env^+ wells (red bars) and in Env^- wells (gray bars) is shown. For individual 603, clonotype 43582 was previously identified as the infected clone of interest TCR. For individual 605, clonotype 21327 was previously identified as the infected clone of interest TCR (Cohn et al., 2018).
(c) The number above each graph is the study participant ID. The x-axis shows the 5 most frequent 10x TCR clonotypes in Env^+ wells for the respective individual. For each 10x TCR clonotype, its frequency in Env^+ wells (red bars) and in Env^- wells (gray bars) is shown. For each individual, the most enriched 10x TCR clonotype was identified as the infected clone of interest TCR.
(d) Pie charts show the relative size of TCR clones as slices. The areas indicated in white represent unique TCR sequences of single cells. The number on the left above the pie chart is the donor ID for each individual. The number in the center of the pie chart represents the number of cells assayed for each individual. The infected clone of interest is indicated by a red arrow and pie slice.

3.4 Fraction of infected cells within a T cell clone

Infection of a $CD4^+$ T cell can occur at different time points during clonal expansion. If a $CD4^+$ T cell becomes infected with HIV-1 and subsequently undergoes clonal expansion, all members of the T cell clone harbor the same HIV-1 provirus and express the same unique TCR. Thus, the number of cells carrying the provirus equals the number of cells expressing the infected clone TCR ($n_{env} = n_{TCR}$) (Figure 10a).

In contrast, if a $CD4^+$ T cells undergoes clonal expansion and only one member of that clone becomes infected with HIV-1 and then further expands, only a fraction of the T cell clone harbors the HIV-1 provirus. In this case, the number of cells carrying the provirus is smaller than the number of cells expressing the TCR that is expressed by the infected clone ($n_{env} < n_{TCR}$) (Figure 10b).

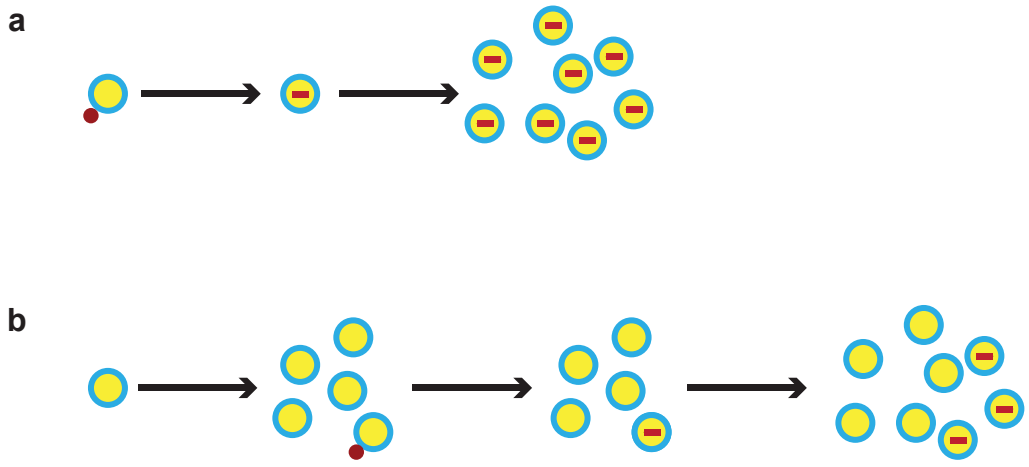


Figure 10 Fraction of infected cells within a clone.
(a) Infection of a $CD4^+$ T cell before clonal expansion results in a homogeneous clone in which all members harbor a provirus.
(b) Infection of a single $CD4^+$ T cell of an already expanded clone results in a heterogeneous clone in which only some members harbor a provirus.

To estimate the fraction of infected cells within the T cell clone of interest, we asked if the frequency of the infected clone of interest provirus matched the frequency of T cells expressing the infected clone of interest TCR. The infected clone of interest was enriched based on CD45RA, TRBC, and TRBV. Subsequent limiting dilution of gDNA and *env* PCR amplification and sequencing quantified the number of proviral copies per 10^4 $CD4^+$ T cells. We compared this proviral copy number with the frequency of cells expressing the infected clone of interest TCR per 10^4 $CD4^+$ T cells, based on the 10x Genomics TCR sequencing.

After enrichment, the frequency of cells pertaining to the infected clone of interest for individuals 603, 605, B207, 5104, 5125, and 9247 was 20, 163, 6, 14, 15, and 18 cells per 10^4 $CD4^+$ T cells, respectively, based on limiting dilution *env* PCR and sequencing (Table 5). The frequency of cells pertaining to the infected clone of interest for individuals 603, 605, B207, 5104, 5125, and 9247 was 45, 502, 14, 16, 17, and 23 cells per 10^4 $CD4^+$ T cells, respectively, based on 10x Genomics TCR sequencing (Table 5).

When analyzing the ratio of cells expressing the infected clone of interest TCR to the number of cells harboring the infected clone of interest provirus per 10^4 $CD4^+$ T cells, it became apparent that in individuals B207, 603, and 605, only a fraction of the T cell clone is in fact infected ($n_{env} < n_{TCR}$). This corresponds to a HIV-1 infection of the T cell clone after the onset of clonal expansion (Figure 10b). In individuals 5104, 5125, and 9247, the number of proviral copies per 10^4 $CD4^+$ T cells was approximately equal to the number of cells expressing the infected clone of interest TCR per 10^4 $CD4^+$ T cells ($n_{env} = n_{TCR}$). This indicates HIV-1 infection of the $CD4^+$ T cell before the onset of or very early in clonal expansion (Figure 10a).

Table 5 Comparison between the number of *env* copies per 10^4 $CD4^+$ T cells (based on *env* PCR and sequencing) and the frequency of the infected clone TCR (based on 10x Genomics TCR sequencing) after enrichment.

	Infected clone of interest *env* copies per 10^4 $CD4^+$ T cells	Cells expressing the infected clone of interest TCR per 10^4 $CD4^+$ T cells	Ratio TCR/*env*
603	20	45	2.3
605	163	502	3.1
B207	6	14	2.3
5104	14	16	1.1
5125	15	17	1.1
9247	18	23	1.3

CHAPTER 4. Transcriptome analysis of infected clones of interest.

4.1 10x Genomics platform

The 10x Genomics platform provides gene expression analysis and TCR sequencing on a single cell level. Reverse transcription of mRNA generates cDNA that is labelled with a unique barcode per cell. From the barcoded cDNA, a gene expression library and a TCR library can be amplified and sequenced. The barcode links the gene expression data to the TCR data of each cell.

HIV-1 is only transcribed in a fraction of infected cells (Baxter et al., 2016; Collora et al., 2022; Einkauf et al., 2022) on a very low level (Hermankova et al., 2003; Wiegand et al., 2017). Since the 10x Genomics platform is limited in its sequencing depth to highly expressed genes, HIV-1 expression was not expected to be reliably identified within the infected clones of interest. Rather, identification of the infected clone of interest in each individual was based on its unique TCR, which is highly expressed even in resting $CD4^+$ T cells.

4.2 Uniform Manifold Approximation Projection analysis of infected clones of interest

The transcriptome data of 109,217 single cells from enriched samples of all six individuals was used to generate a Uniform Manifold Approximation Projection (UMAP) with 15 clusters. Enrichment based on CD45RA, TRBC, and TRBV resulted in a rather homogenous sample per individual, consisting of $CD4^+$ memory cells that expressed the same TRBC domain, and the same TRBV domain (individuals 603, 605, 5125, and 9247) or the same 28 TRBV domains (individuals B207 and 5104). Since the 10x Genomics gene expression data of all individuals was pooled for the UMAP, it represents a heterogenous set of memory $CD4^+$ T cells.

All members of a T cell clone are similar in their gene expression profile in that they express the same TCR genes, which leads to TCR-biased clustering of cells of one clone. To prevent TCR-

biased clustering, the TCR sequencing data was excluded for UMAP generation. However, 10x Genomics TCR sequencing data was used to identify T cell clones and single cells and the infected clone of interest in each individual. In total, 1,050 cells were identified to belong to the infected clones of interest, based on the unique respective TCR. The highest number of infected clone cells belonged to individual 605 (844 cells), and the lowest number of infected clone cells belonged to individual 5125 (21 cells) (Table 6).

Visual inspection of the UMAP showed that cells of infected clones that harbor an intact HIV-1 provirus seemed to preferentially cluster in cluster 7. However, some members of the respective infected clones of interest were also distributed across other clusters of the UMAP (Figure 11). Quantification of the distribution of infected cells across the gene expression clusters revealed the presence of infected cells in 10 out of the 15 clusters. (Figure 11, Table 6). In five of the six individuals, the majority of cells pertaining to the infected clone of interest were located in cluster 7. In individual 9247, the largest fraction of cells pertaining to the infected clone of interest was found in cluster 6, and the second largest fraction in cluster 7.

Nonetheless, smaller fractions of cells pertaining to the infected clones of interest also belonged to clusters 0, 1, 2, 3, 4, 5, 6, 8, and 11.

Table 6 Upper part: Number of cells analyzed by 10x Genomics per individual. Lower part: Distribution [%] of the infected clone of interest, the next smaller, and the next bigger clone per individual over the 15 gene expression clusters.

Number of cells analyzed by 10x genomics single cell gene expression							
Individual	**603**	**605**	**B207**	**5104**	**5125**	**9247**	**Sum**
# of cells	16,224	17,357	27,472	21,143	12,173	14,848	109,217

Individual	**603**			**605**			**B207**		
Clone	smaller (n=64)	infected (n=75)	bigger (n=84)	smaller (n=330)	infected (n=844)	bigger	smaller (n=40)	infected (n=41)	bigger (n=42)
Cluster 0	17.2	18.7	34.5	8.5	7.8	-	0.0	12.2	35.7
Cluster 1	54.7	0.0	4.8	68.2	2.7	-	5.0	0.0	26.2
Cluster 2	1.6	0.0	0.0	0.0	0.0	-	0.0	0.0	11.9
Cluster 3	1.6	0.0	0.0	1.5	0.6	-	5.0	4.9	4.8
Cluster 4	3.1	2.7	7.1	4.5	1.2	-	25.0	2.4	0.0
Cluster 5	4.7	10.7	10.7	11.5	1.7	-	0.0	0.0	4.8
Cluster 6	4.7	0.0	0.0	0.0	0.7	-	0.0	0.0	2.4
Cluster 7	0.0	60.0	23.8	2.1	63.7	-	0.0	73.2	4.8
Cluster 8	0.0	8.0	0.0	0.0	3.8	-	0.0	7.3	0.0
Cluster 9	0.0	0.0	0.0	0.0	0.5	-	2.5	0.0	0.0
Cluster 10	0.0	0.0	0.0	0.3	0.2	-	62.5	0.0	0.0
Cluster 11	7.8	0.0	16.7	1.8	16.1	-	0.0	0.0	9.5
Cluster 12	0.0	0.0	2.4	0.0	0.4	-	0.0	0.0	0.0
Cluster 13	4.7	0.0	0.0	0.6	0.2	-	0.0	0.0	0.0
Cluster 14	0.0	0.0	0.0	0.9	0.4	-	0.0	0.0	0.0
Sum	100	100	100	100	100	-	100	100	100

Individual	5104			5125			9247		
Clone	smaller (n=32)	infected (n=35)	bigger (n=38)	smaller (n=21)	infected (n=21)	bigger (n=22)	smaller (n=34)	infected (n=34)	bigger (n=35)
Cluster 0	43.8	17.1	0.0	33.3	9.5	9.1	8.8	0.0	0.0
Cluster 1	0.0	5.7	0.0	19.0	14.3	36.4	2.9	2.9	5.7
Cluster 2	40.6	8.6	0.0	28.6	4.8	27.3	20.6	23.5	34.3
Cluster 3	3.1	0.0	0.0	9.5	0.0	9.1	0.0	0.0	0.0
Cluster 4	3.1	2.9	0.0	0.0	4.8	13.6	35.3	8.8	31.4
Cluster 5	6.3	0.0	0.0	4.8	4.8	0.0	5.9	0.0	5.7
Cluster 6	0.0	0.0	0.0	4.8	0.0	4.5	23.5	35.3	22.9
Cluster 7	3.1	65.7	0.0	0.0	47.6	0.0	0.0	29.4	0.0
Cluster 8	0.0	0.0	100.0	0.0	9.5	0.0	0.0	0.0	0.0
Cluster 9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Cluster 10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Cluster 11	0.0	0.0	0.0	0.0	4.8	0.0	0.0	0.0	0.0
Cluster 12	0.0	0.0	0.0	0.0	0.0	0.0	2.9	0.0	0.0
Cluster 13	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Cluster 14	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Sum	100	100	100	100	100	100	100	100	100

Our analysis was focused on clonally expanded cells. More than half of the cells in each enriched sample belonged to clonally expanded cells (Figure 9d). We therefore asked if clonal expansion was a hallmark of cluster 7.

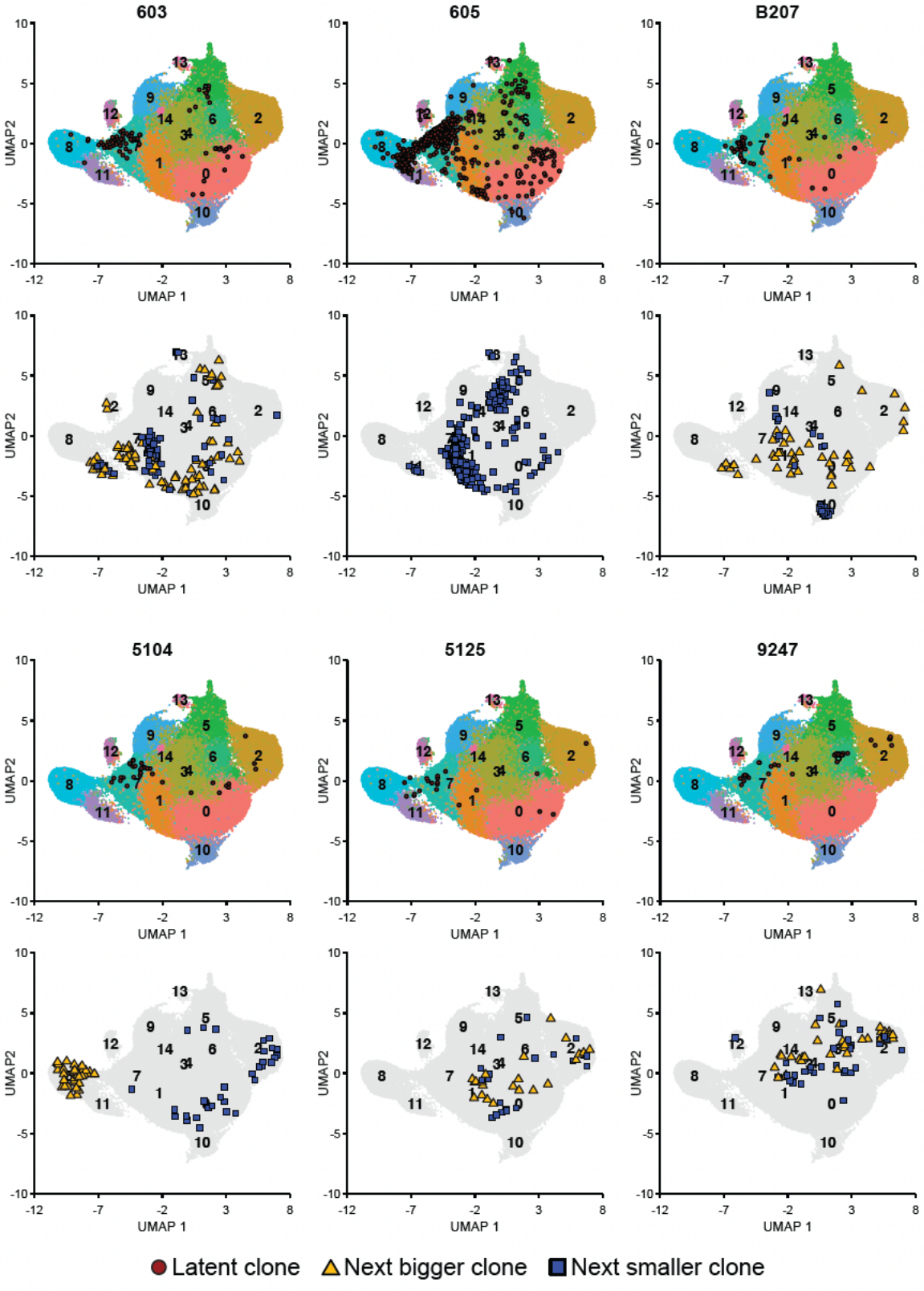
When analyzing cluster 7 alone, clonally expanded, uninfected cells that did not belong to the infected clone of interest contributed 58 – 89 % of cells. However, 8 – 29 % of cells in cluster 7 were not clonally expanded (Table 7). Thus, clonal expansion was not a prerequisite for a $CD4^+$ T cell to fall into cluster 7.

Table 7 Contribution of uninfected clones and single cells to cluster 7.

	Cluster 7					
	# of cells	# of cells of infected clone of interest	# *[%]* of uninfected cells	# *[%]* of uninfected cells in clones	# of uninfected clones	# *[%]* of singlet T cells
603	887	45	842 *[94.9]*	581 *[65.5]*	330	261 *[29.4]*
605	1,989	538	1,451 *[73]*	1,156 *[58.1]*	498	295 *[14.8]*
B207	1,295	30	1,265 *[97.7]*	890 *[68.7]*	526	375 *[29]*
5104	960	23	937 *[97.6]*	655 *[68.2]*	383	282 *[29.4]*
5125	297	10	287 *[96.6]*	223 *[75.1]*	111	64 *[21.5]*
9247	358	10	348 *[97.2]*	318 *[88.8]*	100	30 *[8.4]*

Next, we mapped the next larger T cell clone and the next smaller T cell clone, compared to the infected clone of interest, on the UMAP. Of note, in individual 605, the infected clone of interest was the largest clone, and therefore only the next smaller clone was mapped. Visual inspection of the UMAP with the neighboring clones showed that there did not seem to be a preferential clustering (Figure 11). Quantification of the distribution of T cell clones that were comparable in size to the infected clones of interest revealed that those T cell clones were found in 14 out of the clusters 15 (Figure 11, Table 6). Some of the uninfected clones displayed a preferential clustering, such as the next smaller T cell clone of individual 605 in cluster 1 or the next larger T cell clone of individual 5104 in cluster 8. In contrast, other clones, such as the next larger and smaller T cell clone in individual 5125 were rather randomly distributed over a wide range of clusters.

We concluded that expanded infected clones that harbor an intact HIV-1 provirus preferentially cluster in cluster 7. This preferential clustering was not based on clonal expansion. Furthermore, cluster 7 did not only contain infected cells, but also a high number of uninfected cells. We concluded that the gene expression profile in cluster 7 is not unique for cells pertaining to infected clones that harbor an intact provirus.


603
605
B207
5104
5125
9247
UMAP 1
UMAP2
Latent clone
Next bigger clone
Next smaller clone

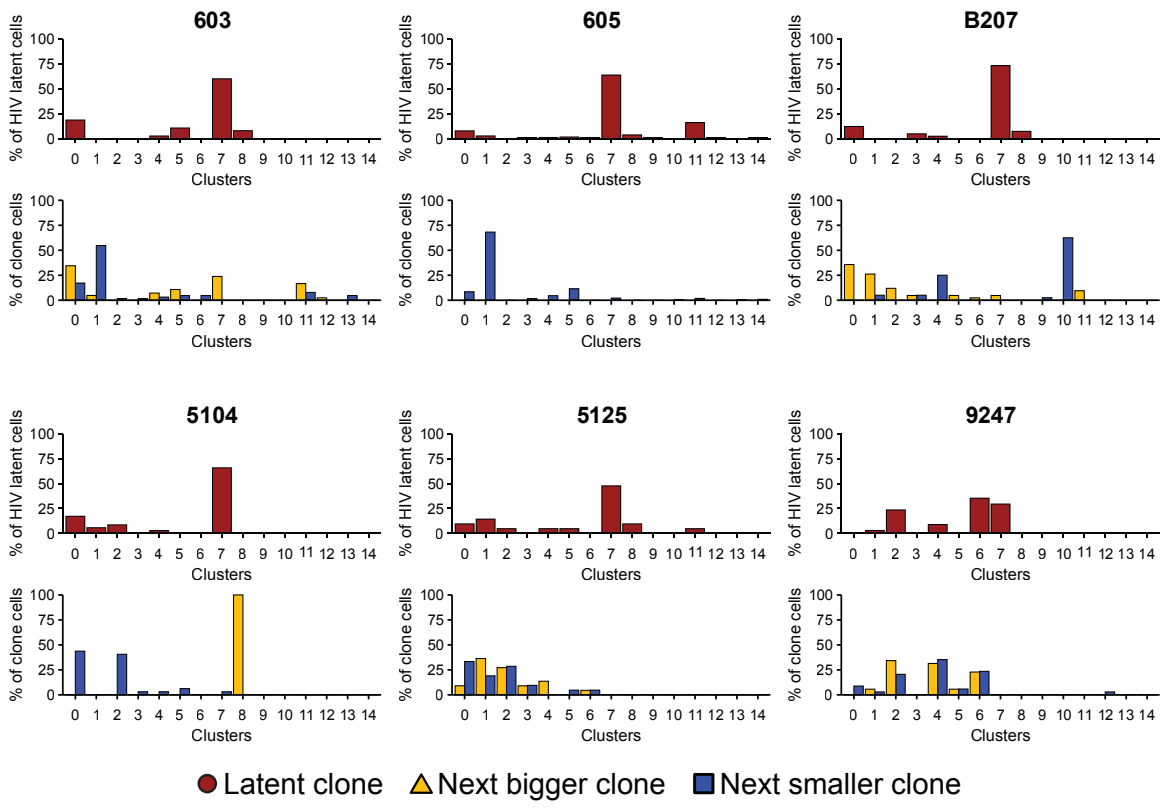


Figure 11 Uniform Manifold Approximation and Projection (UMAP) of 10x Genomics gene expression data.

Data representing gene expression by 109,217 individual cells is shown. The infected clone of interest (red circles) as well as the next bigger (yellow triangles) and next smaller T cell clone (blue squares) in size were located on the UMAP by their TCR sequence. For individual 605, the infected clone of interest was the biggest clone, and only the next smaller clone is shown.

The bar graphs show the distribution of the infected clone of interest (red bars) over the 15 UMAP clusters, and the fraction of cells of the next bigger (yellow bars) and the next smaller (blue bars) T cell clone in size to the infected clone of interest over the 15 UMAP clusters (Table 7).

4.3 Differentially expressed genes in cluster 7

We asked which genes were upregulated or downregulated in cluster 7, which contained the majority of cells pertaining to $CD4^+$ T cell clones harboring an intact provirus, compared to all other clusters. We found a set of 100 genes that were differentially expressed (either up- or down-regulated by an average log2-fold change) in cluster 7 compared to all others clusters (Table 8).

A heatmap of up to ten differentially upregulated genes that characterize each cluster was generated by supervised clustering (Figure 12). Cluster 7 displayed upregulation of the MHCII class molecules HLA-DRB1, HLA-DPA1, and HLA-DPB1, as well as of the MHCII chaperone protein CD74 (Fortin et al., 2013). Further upregulated genes in cluster 7 comprised the granzyme molecules GZMA and GZMK and the cystatin CST7, a granzyme-regulating protein (Perisic Nanut et al., 2014; Perisic Nanut et al., 2017), as well as the nuclear proteins LYAR and DUSP2. In addition, the chemokine ligand CCL5, a potent inhibitor of HIV-1 entry to host cells, was upregulated in cluster 7 (Cocchi et al., 1995; Dragic et al., 1996). Of particular note, except for GZMK, all genes that were upregulated in cluster 7 were also upregulated in cluster 8. However, cluster 8 featured additional upregulated genes that are characteristic of cytotoxic $CD4^+$ T cells: The granzymes GZMB and GZMH, perforin 1 (PRF1) and granulysin (GNLY), and the Natural Killer Cell Granule Protein 7 (NKG7) (Appay et al., 2002; Hashimoto et al., 2019). We concluded that cells in cluster 7 upregulated a distinct set of genes that was mostly shared with cluster 8. However, cluster 8 displayed additional upregulated genes that are characteristic of cytotoxic $CD4^+$ T cells and that were not highly expressed in cluster 7.

We asked if sorting for $GZMA^+$, $GZMK^+$, or $CCL5^+$ cells would specifically enrich the infected clones of interest. Sorting for $GZMA^+$ or $GZMK^+$ $CD4^+$ T cells failed to specifically enrich the infected clone of interest. This in accordance with the finding that the majority, but not all infected

cells were found in cluster 7, and that cluster 7 also contained uninfected cells. Intracellular staining for CCL5 did not provide a clear population of CCL5-expressing $CD4^+$ T cells.

Table 8 Differentially expressed genes (average log2fold < -2.5; average log2fold > 2.5) in cluster 7 compared to all other clusters.

Ensembl gene ID	External gene name	p value	Average log2 fold change
ENSG00000115687	PASK	2.5074E-273	-0.937004582
ENSG00000188404	SELL	1.1432E-303	-0.757375402
ENSG00000168209	DDIT4	4.2236E-135	-0.753863566
ENSG00000070756	PABPC1	1.3615E-275	-0.577338107
ENSG00000074800	ENO1	3.9911E-72	-0.552495574
ENSG00000273149	AL138963.4	1.69448E-43	-0.503468146
ENSG00000111669	TPI1	3.75841E-56	-0.502939955
ENSG00000095794	CREM	7.5629E-48	-0.48733596
ENSG00000157601	MX1	3.44639E-38	-0.463685933
ENSG00000149212	SESN3	2.09806E-96	-0.458160858
ENSG00000126353	CCR7	5.6185E-113	-0.438022627
ENSG00000177410	ZFAS1	9.1716E-145	-0.432744809
ENSG00000173762	CD7	1.2733E-119	-0.422179842
ENSG00000059804	SLC2A3	4.68348E-68	-0.401753273
ENSG00000197061	HIST1H4C	1.74666E-54	-0.367052221
ENSG00000234741	GAS5	1.8049E-124	-0.365481063
ENSG00000134333	LDHA	3.62429E-18	-0.356852098
ENSG00000197989	SNHG12	2.6524E-59	-0.354230806
ENSG00000105220	GPI	2.30295E-38	-0.33651571
ENSG00000146278	PNRC1	2.56242E-66	-0.33232545
ENSG00000196352	CD55	3.58249E-86	-0.331613182
ENSG00000269028	MTRNR2L12	9.28887E-42	-0.326076812
ENSG00000251562	MALAT1	4.23243E-61	-0.326073295
ENSG00000240972	MIF	1.16951E-48	-0.325935464
ENSG00000067225	PKM	3.0003E-53	-0.324450491
ENSG00000166012	TAF1D	4.1997E-62	-0.323635813
ENSG00000130066	SAT1	6.88644E-24	-0.320711104
ENSG00000161011	SQSTM1	3.25091E-23	-0.307785538
ENSG00000181163	NPM1	1.7052E-145	-0.295019474
ENSG00000081059	TCF7	2.53957E-59	-0.293559954
ENSG00000100906	NFKBIA	6.90414E-22	-0.286544704

ENSG00000138795	LEF1	4.65739E-63	-0.283635236
ENSG00000102144	PGK1	3.47758E-18	-0.283588302
ENSG00000104765	BNIP3L	4.68824E-42	-0.280401542
ENSG00000144381	HSPD1	1.33312E-51	-0.27754675
ENSG00000096384	HSP90AB1	6.19504E-51	-0.27513593
ENSG00000114023	FAM162A	3.98958E-38	-0.271026449
ENSG00000167658	EEF2	2.0542E-116	-0.263184273
ENSG00000081320	STK17B	1.01417E-40	-0.263109056
ENSG00000105193	RPS16	7.12703E-82	-0.25357902
ENSG00000213145	CRIP1	1.84831E-74	0.251381711
ENSG00000204642	HLA-F	1.67539E-32	0.252987681
ENSG00000064666	CNN2	7.14478E-34	0.254603703
ENSG00000136167	LCP1	4.88255E-38	0.255115632
ENSG00000145247	OCIAD2	1.35036E-33	0.255136956
ENSG00000135441	BLOC1S1	8.55057E-29	0.256708541
ENSG00000217555	CKLF	6.26258E-37	0.268552743
ENSG00000179218	CALR	5.15079E-26	0.269118004
ENSG00000166710	B2M	0	0.270596986
ENSG00000034713	GABARAPL2	7.97742E-47	0.271789708
ENSG00000213626	LBH	1.70427E-35	0.275234928
ENSG00000158062	UBXN11	2.96782E-31	0.277861974
ENSG00000135046	ANXA1	6.7754E-104	0.277916253
ENSG00000240065	PSMB9	9.1741E-63	0.278633877
ENSG00000165929	TC2N	3.38931E-34	0.282431558
ENSG00000136810	TXN	4.66479E-33	0.284410494
ENSG00000115232	ITGA4	4.07453E-43	0.296034223
ENSG00000234745	HLA-B	2.442E-301	0.299434559
ENSG00000182718	ANXA2	1.2129E-48	0.300086815
ENSG00000105404	RABAC1	3.41798E-59	0.300856113
ENSG00000108518	PFN1	5.1347E-192	0.308425928
ENSG00000075624	ACTB	3.1245E-173	0.312823868
ENSG00000163191	S100A11	5.8168E-123	0.313585328
ENSG00000213719	CLIC1	3.2501E-80	0.321146717
ENSG00000126246	IGFLR1	1.12795E-32	0.321234367
ENSG00000027869	SH2D2A	1.2519E-59	0.325040074
ENSG00000008517	IL32	1.885E-203	0.326877861
ENSG00000100300	TSPO	4.23815E-58	0.328869744
ENSG00000130592	LSP1	5.13406E-85	0.335795569
ENSG00000170571	EMB	1.30243E-48	0.340520299
ENSG00000110324	IL10RA	7.30307E-71	0.345878196
ENSG00000092841	MYL6	1.8186E-212	0.349021829

ENSG00000160255	ITGB2	1.63101E-55	0.376681806
ENSG00000197747	S100A10	5.2906E-220	0.381136641
ENSG00000169442	CD52	2.9441E-231	0.386540002
ENSG00000132965	ALOX5AP	2.30411E-89	0.390330172
ENSG00000148362	PAXX	1.9717E-94	0.391613282
ENSG00000197540	GZMM	5.469E-80	0.411312119
ENSG00000002586	CD99	1.1703E-191	0.411377898
ENSG00000111796	KLRB1	8.73173E-56	0.434616046
ENSG00000196154	S100A4	9.403E-205	0.446707794
ENSG00000142669	SH3BGRL3	0	0.473220735
ENSG00000126264	HCST	3.4505E-135	0.484457968
ENSG00000103187	COTL1	8.1033E-136	0.491151274
ENSG00000100097	LGALS1	3.0001E-118	0.520481509
ENSG00000116824	CD2	7.0921E-219	0.527921045
ENSG00000051523	CYBA	0	0.535424541
ENSG00000133321	PLAAT4	1.0403E-293	0.6130956
ENSG00000235576	LINC01871	6.8684E-252	0.712737658
ENSG00000186810	CXCR3	0	0.863806843
ENSG00000145220	LYAR	0	0.91690397
ENSG00000223865	HLA-DPB1	0	1.005834612
ENSG00000077984	CST7	0	1.00668931
ENSG00000231389	HLA-DPA1	0	1.0517126
ENSG00000145649	GZMA	0	1.159266639
ENSG00000196126	HLA-DRB1	0	1.189048765
ENSG00000158050	DUSP2	0	1.377812382
ENSG00000271503	CCL5	0	1.450021207
ENSG00000019582	CD74	0	1.478166065
ENSG00000113088	GZMK	0	2.055910352

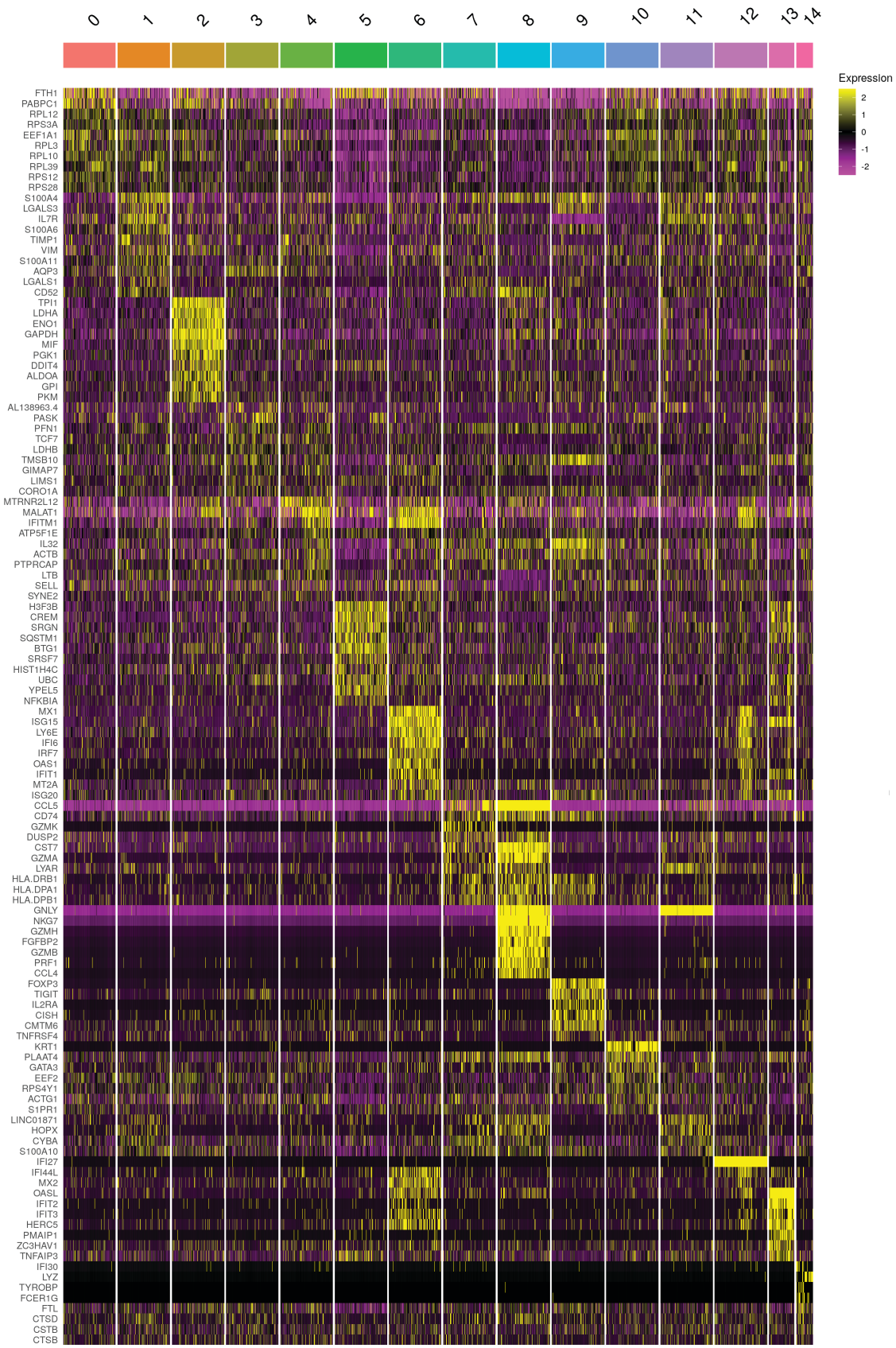

0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Expression
2
1
0
-1
-2
FTH1
PABPC1
RPL12
RPS3A
EEF1A1
RPL3
RPL10
RPL39
RPS12
RPS28
S100A4
LGALS3
IL7R
S100A6
TIMP1
VIM
S100A11
AQP3
LGALS1
CD52
TPI1
LDHA
ENO1
GAPDH
MIF
PGK1
DDIT4
ALDOA
GPI
PKM
AL138963.4
PASK
PFN1
TCF7
LDHB
TMSB10
GIMAP7
LIMS1
CORO1A
MTRNR2L12
MALAT1
IFITM1
ATP5F1E
IL32
ACTB
PTPRCAP
LTB
SELL
SYNE2
H3F3B
CREM
SRGN
SQSTM1
BTG1
SRSF7
HIST1H4C
UBC
YPEL5
NFKBIA
MX1
ISG15
LY6E
IFI6
IRF7
OAS1
IFIT1
MT2A
ISG20
CCL5
CD74
GZMK
DUSP2
CST7
GZMA
LYAR
HLA.DRB1
HLA.DPA1
HLA.DPB1
GNLY
NKG7
GZMH
FGFBP2
GZMB
PRF1
CCL4
FOXP3
TIGIT
IL2RA
CISH
CMTM6
TNFRSF4
KRT1
PLAAT4
GATA3
EEF2
RPS4Y1
ACTG1
S1PR1
LINC01871
HOPX
CYBA
S100A10
IFI27
IFI44L
MX2
OASL
IFIT2
IFIT3
HERC5
PMAIP1
ZC3HAV1
TNFAIP3
IFI30
LYZ
TYROBP
FCER1G
FTL
CTSD
CSTB
CTSB

Figure 12 Heat map of cluster-defining, differentially expressed genes.
Heat map shows up to 10 of the most up-regulated genes (average log2fold > 2.5) per UMAP cluster compared to all other clusters. Genes are indicated on the left and clusters above. Yellow shows relatively highly expressed genes and purple relatively downregulated genes.

4.4 Comparison of gene expression cluster 7 to cytotoxic $CD4^+$ T cells

Cluster 7 and cluster 8 seemed closely related in the expression of genes that are associated with cytotoxic $CD4^+$ T cells. We aimed to further elucidate the similarities and differences between the two clusters. Based on literature research, a set of hallmark genes that are differentially expressed in cytotoxic $CD4^+$ T cells was generated for unsupervised clustering.

Cytotoxic $CD4^+$ cells are expanded in acute, chronic, and ART-treated HIV-1 infection (Appay et al., 2002). They express the cytotoxic molecules perforin, granzymes, and granulysin, and surface molecules such as NKG7, Killer Cell Lectin Like Receptor G1 and B1 (KLRG1 and KLRB1), and Integrin Subunit α L (ITGAL), but lack expression of CD45RA, CD27, CD28, and CCR7 (Appay et al., 2002; Hashimoto et al., 2019). Cytotoxic $CD4^+$ T cells can develop from various different T cell lineages under the regulation of transcription factors such as T-bet, EOMES, Runx3, Blimp-1, ThPOK, Bcl6, and TCF1 (Cachot et al., 2021; Juno et al., 2017; Takeuchi and Saito, 2017; Zaunders et al., 2004).

The unsupervised clustering of the UMAP gene expression clusters for hallmark genes of cytotoxic $CD4^+$ T cells confirmed that cluster 7 is most closely related to cluster 8 (Figure 13). Yet, the two clusters were very distinct in their gene expression. Cluster 8 expressed a host of genes that are required for cytotoxicity, such as EOMES, PRF1, GZMA, GZMB, and GNLY, but lacked the upregulation of CD27, CD28, and CCR7. Thus, cluster 8 represented $CD4^+$ CTLs. In contrast, in cluster 7 GZMK, and to a lower extent GZMA were the only upregulated cytotoxic molecules. The expression level of EOMES was much lower in cluster 7 compared to cluster 8.

We conclude that cluster 7 was closely related to cluster 8, which represented cytotoxic $CD4^+$ T cells, but that a majority of genes that are required for cytotoxicity was not expressed in cluster 7.

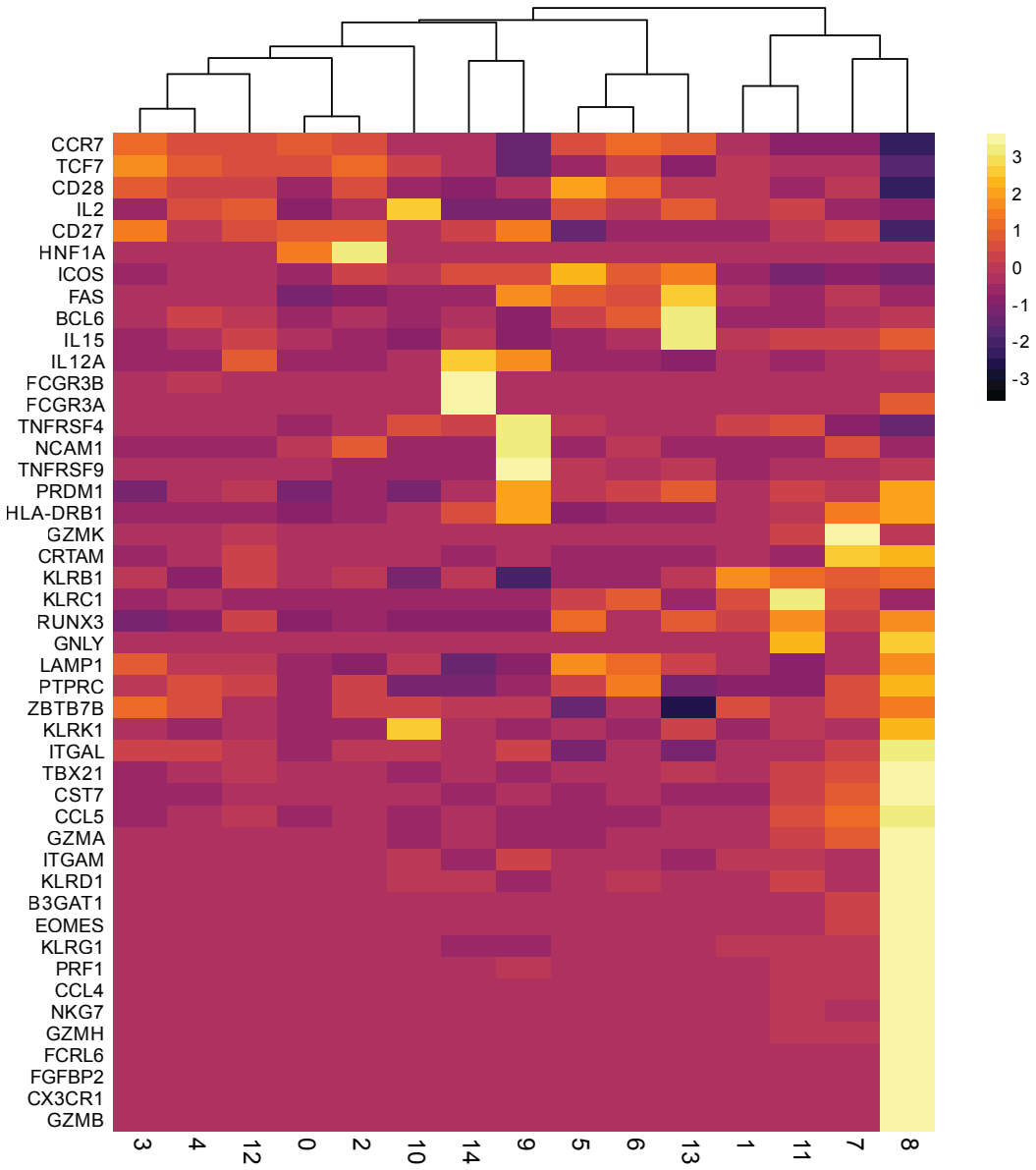


Figure 13 Heat map of differentially expressed genes in cytotoxic CD4$^+$ T cells.
The heat map shows unsupervised clustering based on differentially expressed genes in cytotoxic CD4$^+$ T cells (Appay et al., 2002; Cachot et al., 2021; Hashimoto et al., 2019; Juno et al., 2017; Takeuchi and Saito, 2017; Zaunders et al., 2004). Genes are indicated on the left and clusters are indicated below. Yellow shows relatively highly expressed genes and black shows relatively downregulated genes.

4.5 Analysis of infected clones of interest with a multimodal reference data set

We asked if the infected clones of interest would also display a preferential clustering when mapped on an external $CD4^+$ T cell data set as a reference. Furthermore, we asked if infected cells would be enriched in a specific $CD4^+$ T cell memory population.

The reference data set was based on combined gene expression sequencing and surface marker evaluation (Hao et al., 2021). Mapping of our 10x Genomics gene expression data identified naïve, regulatory (Treg), central memory (TCM), effector memory (TEM), and cytotoxic (CTL) $CD4^+$ T cells, as well as proliferating $CD4^+$ T cells (Figure 14a). Because the samples were $CD45RA^-$ memory $CD4^+$ T cells, the $CD45RA^+$ naïve $CD4^+$ T cell compartment was only marginally represented. The projection of the gene expression clusters on the reference map showed that cluster 7 mostly coincided with the TEM compartment, whereas cluster 8 fell into the CTL compartment (Figure 14b).

The infected clone of interest of each individual was identified based on its unique TCR and mapped on the reference UMAP. Analysis of the distribution of the infected clones of interest over the $CD4^+$ memory T cell subpopulations showed that cells that pertained to the infected clones of interest were present in the TCM, TEM, and CTL compartment. In four out of six individuals, the infected clone of interest was predominantly found in the TEM compartment. In individuals 5104 and 9247, the majority of cells that pertained to the infected clones of interest was in the TCM compartment, and the remaining cells that pertained to the infected clones of interest were in the TEM compartment. In individuals, 603, 605, and 5125, a small fraction of cells that pertained to the infected clones of interest formed part of the cytotoxic $CD4^+$ T cell compartment (Figure 14). The TCM compartment was represented at much higher cell numbers than the TEM compartment. Thus, cells that pertained to the infected clones of interest in the TCM compartment might have

been overrepresented compared to cells that pertained to the infected clones of interest in the TEM compartment. We asked what the relative contribution of the infected clones of interest to each $CD4^+$ memory T cell subset was.

In all individuals but individual 605, the contribution of infected clones of interest to the TEM compartment was the highest (Table 9). In individual 605, the infected clone of interest contributed the most to the CTL compartment, and the second largest contribution was to the TEM compartment. In individuals 603 and 5125, the infected clone of interest also contributed minutely to the CTL compartment.

We concluded that the infected clones of interest were preferentially found in the TEM compartment, but contributed to the TCM and CTL compartment as well.

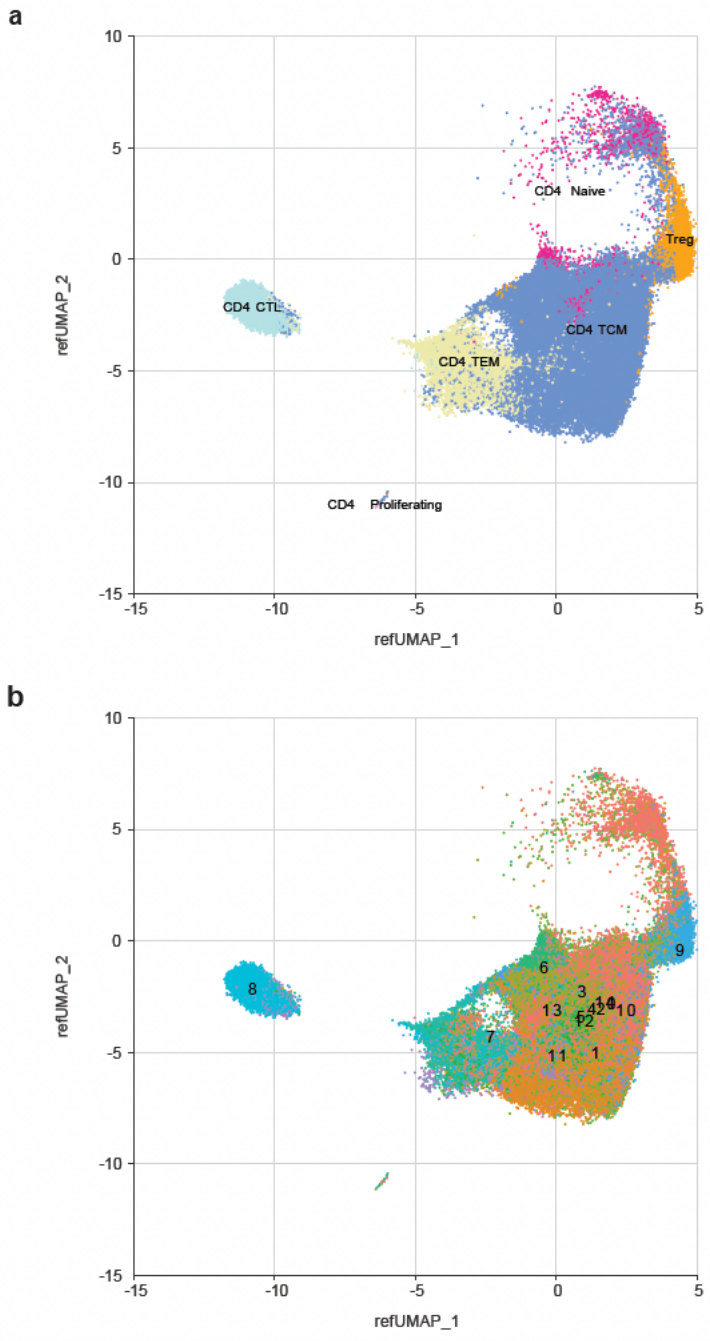

a
CD4 Naive
Treg
CD4 CTL
CD4 TCM
CD4 TEM
CD4 Proliferating
refUMAP_2
refUMAP_1
b
refUMAP_2
refUMAP_1

Figure 14 Projection of 10x Genomics gene expression data on a reference data set of $CD4^+$ T cells from multimodal single cell sequencing (Hao et al., 2021).

Projection of data representing mRNA expression by 109,217 individual cells on a multimodal UMAP of $CD4^+$ T cells from HIV-negative individuals. Treg: regulatory T cell. TCM: central memory. TEM: Effector memory. CTL: Cytotoxic lymphocyte.

a) Clusters are labeled according to Hao et al., 2021.

b) Clusters are labeled according to UMAPs based on 10x Genomics gene expression data in Figure 10. Cluster 7 overlaps mainly with the TEM compartment, and cluster 8 represents the CTL compartment.

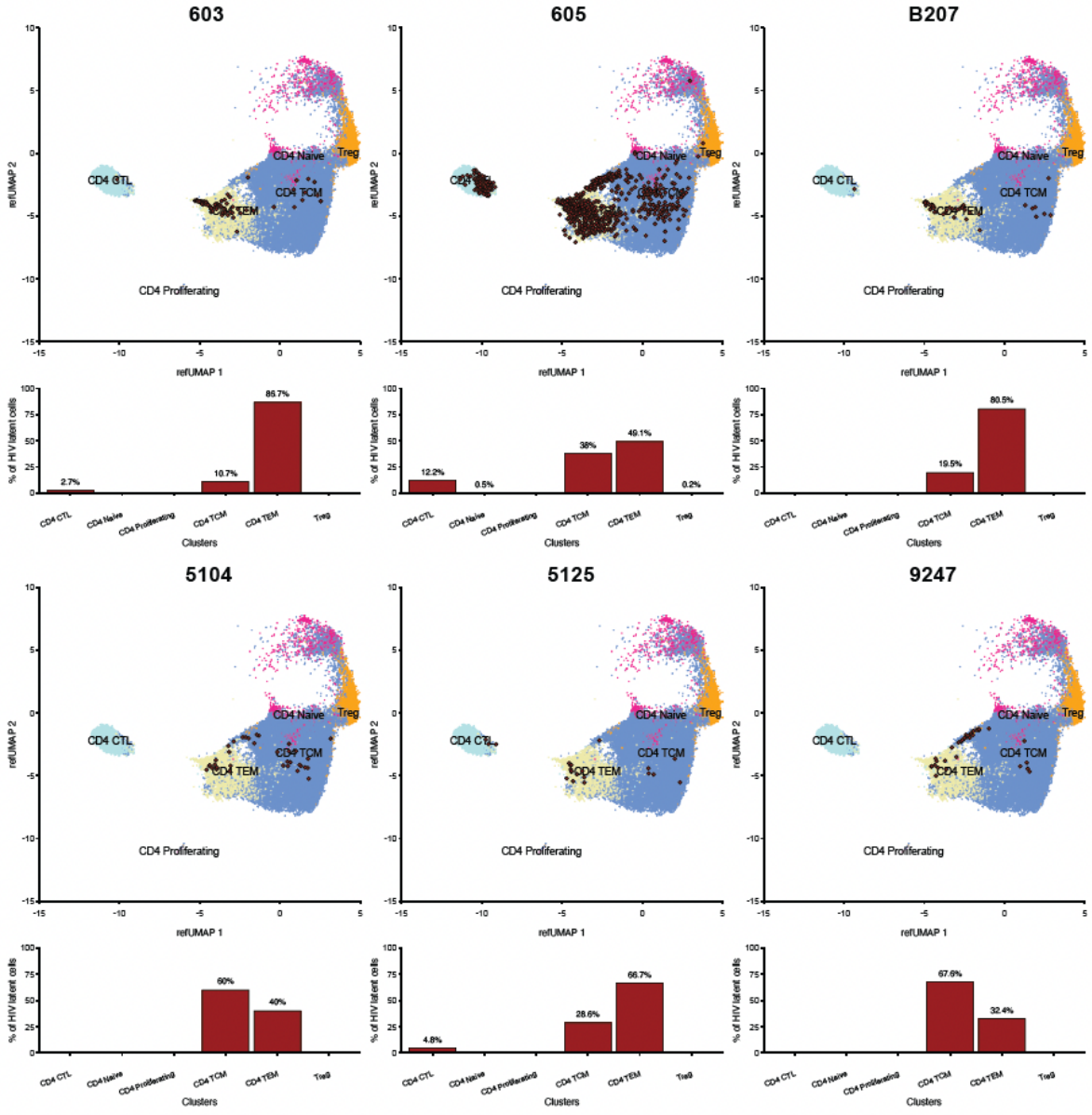


Figure 15 Projection of infected clones of interest from 10x Genomics gene expression data on Uniform Manifold Approximation and Projection (UMAP) based on a CD4$^+$ T cell reference data set (Hao et al., 2021).

Projection of data representing mRNA expression by 109,217 individual cells on a multimodal UMAP of CD4$^+$ T cells from HIV-negative individuals. The infected clone of interest in each individual was located in the UMAP by its TCR sequence and is represented as red dots. Underneath each UMAP, the bar graph shows the fraction of cells of the infected clone of interest in each T cell subpopulation as indicated by the number above each bar.

Table 9 Percentage of cells pertaining to the infected clones of interest relative to cluster size.

	603	**605**	**B207**	**5104**	**5125**	**9247**
CD4 CTL	1.6 %	58.5 %	0.0 %	0.0 %	0.9 %	0.0 %
CD4 Naïve	0.0 %	1.0 %	0.0 %	0.0 %	0.0 %	0.0 %
CD4 Proliferating	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
CD4 TCM	0.1 %	2.2 %	0.0 %	0.1 %	0.1 %	0.2 %
CD4 TEM	6.1 %	21.9 %	2.4 %	1.1 %	3.8 %	2.0 %
Treg	0.0 %	0. 4%	0.0 %	0.0 %	0.0 %	0.0 %

4.6 Detection of HIV transcripts

After enrichment based on CD45RA, TRBV, and TRBC, the majority of infected cells belonged to the infected clones of interest. However, *env* sequences that did not pertain to the infected clone of interest were enriched and recovered by *env* PCR amplification and sequencing as well (Figure 7).

We asked if the 10x gene expression data sets contained sequences from HIV transcripts.

In total, 109,217 cells were analyzed, of which 1,050 belonged to the intact clones of interest. HIV transcripts were detected in only 58 cells total, that derived from individuals 603, 605, B207, 5104, and 9247 (Table 10). No HIV reads were detected for individual 5125.

Of the 58 cells with HIV reads:

- 6 cells belonged to the infected clone of interest in individual 605. No HIV expression was detected for the infected clones of interest from the other individuals.
- 26 cells belonged to T cell clones other than the infected clones of interest
- 11 cells were single T cells
- 15 cells had no TCR sequence recovered and the respective clonal status was therefore not defined (ND).

The 58 cells with HIV reads were present in 14 out of 15 gene expression clusters. Only one such cell was found in gene expression cluster 8, which coincides with the CTL compartment. Per cell, 1 to 34 HIV reads were detected (mean 4.9). Based on the unique molecular identifiers (UMI), these reads were based on 1 to 4 HIV transcripts per cell (mean 1.2).

The 10x Genomics platform preferentially captures highly expressed genes. HIV transcription in unstimulated infected cells is silent or occurs only on a very low level (Baxter et al., 2016; Collora

et al., 2022; Einkauf et al., 2022; Hermankova et al., 2003; Wiegand et al., 2017). In accordance with those previous findings, in the vast majority of cells that belonged to an infected clone of interest, no HIV reads were found.

HIV expression was detected based on a very low number of HIV reads that derived from a very low number of HIV transcripts. Cells without detectable HIV transcripts could either be transcriptionally silent or express HIV on such a low level that it was below the sensitivity of the 10x Genomics platform and therefore not captured. As a result, the distinction between transcriptionally active infected cells and transcriptionally silent, latently infected cells was not possible.

Table 10 HIV transcription in resting infected cells.
When a TCR of a given cell with HIV transcripts was not captured by the 10x Genomics platform, the clonal status indicates "ND".

Individual	Clonal status	Cluster	# of HIV reads	# of UMIs
603	clonal	1	14	2
	clonal	0	7	2
	clonal	5	6	1
	clonal	7	1	1
	clonal	12	11	1
	ND	13	6	1
	clonal	1	6	1
	clonal	3	9	2
	clonal	0	3	1
	single	2	10	1
605	ND	5	34	4
	infected clone of interest	7	5	1
	clonal	12	6	1
	clonal	0	4	1
	infected clone of interest	7	3	1
	clonal	0	1	1
	ND	0	3	1
	ND	1	1	1
	infected clone of interest	7	4	1
	ND	11	16	1
	infected clone of interest	7	13	2
	infected clone of interest	7	3	1
	infected clone of interest	7	5	1
	ND	7	4	1
B207	single	5	20	2
	single	12	14	1
	single	12	10	1
	clonal	7	13	1
	single	7	9	1

	ND	7	4	1
	clonal	5	32	4
	ND	0	6	1
	single	9	1	1
	clonal	1	6	1
	clonal	4	3	1
	single	1	3	1
	clonal	0	2	1
	clonal	5	7	1
	clonal	1	3	1
	clonal	6	1	1
	clonal	2	9	1
	clonal	7	13	1
	single	3	7	2
	clonal	0	18	3
	clonal	0	1	1
	ND	8	1	1
	clonal	0	1	1
	single	0	17	1
	ND	12	4	1
	ND	3	19	4
	single	3	11	1
5104	clonal	3	4	1
	single	2	16	1
	clonal	10	12	1
9247	ND	5	1	1
	ND	7	1	1
	ND	2	1	1
	ND	2	1	1

CHAPTER 5. Discussion.

Infected cells that carry an intact HIV-1 provirus are exceedingly scarce (Cho et al., 2022; Cohn et al., 2020; Dufour et al., 2021) and do not express any known definitive marker that would identify these cells or enable their enrichment (Darcis et al., 2019). Therefore, a single cell analysis of infected cells harboring an intact provirus in their resting state has not been feasible so far. Gene expression analysis of resting infected cells that harbor an intact provirus on a single cell level could elucidate whether these cells display a unique and characteristic gene expression profile with distinct pathways that facilitate proviral silencing and persistence of HIV-1 infected cells.

Proviral reactivation was previously employed to identify infected cells. Reactivated cells that harbor an intact provirus display a different gene expression profile than reactivated uninfected cells (Cohn et al., 2018). However, whether resting infected cells that harbor an intact provirus differ from uninfected cells has not been investigated yet. It is therefore unknown if integration of an intact provirus shapes infected cells in a way that enables control of proviral transcription and thus renders infect cells of the HIV-1 reservoir fundamentally different from uninfected $CD4^+$ T cells.

In this study, the unique TCR of expanded infected clones that harbor an intact provirus was leveraged to enrich and identify those clones within 10x Genomics gene expression and TCR sequencing data. Thus, infected $CD4^+$ T cell clones that carried an intact provirus were analyzed without prior latency reversal, in their resting state.

5.1 Expanded infected clones preferentially, but not exclusively display a T effector memory cell gene expression profile

Over time on ART, the HIV-1 reservoir is maintained by proliferation of infected cells (Antar et al., 2020; Cho et al., 2022; Cohn et al., 2018; Cohn et al., 2015; Maldarelli et al., 2014; Wagner et al., 2013). How proviruses can persist in infected cells and how they can remain latent in proliferating cells has been a highly investigated subject in HIV research. Several non-exclusive mechanism of HIV-1 reservoir maintenance have been proposed: Homeostatic proliferation, the proviral integration site, and antigen-induced proliferation (Cohn et al., 2020).

Homeostatic proliferation of infected cells can be induced by IL7 and IL15 (Chomont et al., 2009). This cytokine-dependent but antigen-independent proliferation is a physiologic mechanism of $CD4^+$ memory cell homeostasis and not exclusive of infected cells (Geginat et al., 2001; Kondrack et al., 2003), and infected $CD4^+$ T cells proliferate at the same rate as uninfected $CD4^+$ T cells (Vandergeeten et al., 2013).

The cells pertaining to infected clones that harbor an intact provirus in this study preferentially, but not exclusively, displayed a gene expression profile of $CD4^+$ T effector memory cells. This is in accordance with previous findings that showed enrichment of genetically intact proviruses in the TEM compartment (Duette et al., 2022; Hiener et al., 2017; Morcilla et al., 2021). TEM cells are highly responsive to cytokine-induced proliferation, which can compensate for their high turnover rate (Geginat et al., 2001; Macallan et al., 2004; MacLeod et al., 2010). IL7-induced proliferation in the absence of viral activation was first described in an *in vitro* model of HIV latency (Bosque et al., 2011). Vandergeeten and colleagues found different effects of IL7 on productively infected and latently infected cells: Whereas productively infected *ex vivo* $CD4^+$ T

cells from viremic individuals increase viral production upon IL7 stimulation, latently infected *ex vivo* $CD4^+$ T cells from ART treated individuals proliferate without proviral activation (Vandergeeten et al., 2013). IL7 was furthermore shown to reduce spontaneous apoptosis in $CD4^+$ T cells from people living with HIV (Vassena et al., 2007). Musick and colleagues reported on one individual living with HIV under ART whose HIV-1 reservoir contains one highly expanded clone harboring an intact provirus (Musick et al., 2019). Notably, most cells of that clone pertained to the TEM compartment, despite the fact that infected TEM cells displayed the highest levels of HIV transcription. Thus, IL7 allows latently infected cells to proliferate without HIV-reactivation and seems to enable tolerance of a certain level of proviral transcription in infected cells.

The integration site can impact both, infected cell proliferation and proviral persistence. An *in vitro* model of HIV infection showed that proviral integration occurs preferentially into *Alu* repeats and actively transcribed genes (Schroder et al., 2002). Studies on *ex vivo* $CD4^+$ T cells from people living with HIV-1 under ART confirmed preferential proviral integration into introns of highly expressed genes and near *Alu* repeats in genic and non-genic regions (Cohn et al., 2015; Han et al., 2004). Integration into *BACH2*, *MKL1, MKL2, IL2RB* (Bedwell et al., 2021; Ikeda et al., 2007; Maldarelli et al., 2014; Wagner et al., 2014), as well as into cancer-related genes such as *MDC1*, *STAT5B*, and *MAPK1* (Bedwell et al., 2021; Wagner et al., 2014) has been associated with infected cell persistence and proliferation. In general, however, selection against proviruses that are integrated into genic regions occurs over time on ART (Cohn et al., 2015), and intact proviruses in people living with HIV on ART are more frequently found in non-genic regions and in opposite direction to host genes (Einkauf et al., 2019). In elite controllers who are aviraemic without ART, intact proviruses that are integrated into inaccessible heterochromatin regions such as centromeric

satellite DNA or into Krüppel-associated box domain-containing zinc finger genes (KRAB-ZNF) on chromosome 19 remain silent over time (Jiang et al., 2020). In general, inaccessibility of the proviral integration site due to histone acetylation and methylation can favor latency (Svensson, 2021). Huang and colleagues suggested that integration into KRAB-ZNF genes enables intact proviruses to remain latent during cell proliferation, since those genes are silenced during T cell activation (Huang et al., 2021).

Lastly, proviral integration can affect the transcription of genes that are down-stream of the integration site (Liu et al., 2020). However, since the integration site is unique for each single infected cell and infected clones, its impact on gene expression and possible effects on the transcriptional state of the cell cannot be generalized. Therefore, the proviral integration site is not always sufficient to explain clonal expansion of infected cells (Wang et al., 2018a).

Antigen-induced expansion of infected cells is yet another mechanism of HIV-1 reservoir maintenance (Cohn et al., 2020; Wang et al., 2018a). TCM cells can undergo rapid proliferation upon re-encounter with their cognate antigen (Gray et al., 2018; Riou et al., 2007; Weng et al., 2012). The combination of the cytokines IL-7, IL-15, TNF-α, IL-6, and IL-10 can induce TCM expansion and differentiation into TEM cells. Thus, the TCM compartment can replenish the TEM compartment (Geginat et al., 2001). The finding that a fraction of infected clones harboring an intact provirus in our study was found in the TCM compartment, and that infected $CD4^+$ T cells can be specific against antigens of chronic infections, such as HIV-1 (Douek et al., 2002), CMV, or EBV (Henrich et al., 2017; Mendoza et al., 2020; Simonetti et al., 2021) is in accordance with the notion that antigen exposure can drive proliferation and clonal expansion of infected clones.

Prolonged antigen exposure in the context of chronic viral infection or malignancies can induce the differentiation of $CD4^+$ CTL (Abana et al., 2017; Appay et al., 2002; Juno et al., 2017; Oh and Fong, 2021; Zaunders et al., 2004). The CTL compartment contained a small fraction of cells pertaining to infected clones that harbored an intact provirus in three individuals (603, 605, and 5125). CTL cells can rise from different T cell subsets, but are most closely related to T_H1 cells (Takeuchi and Saito, 2017). T_H1 cells are important for immune reactions against viral and bacterial intracellular pathogens (Annunziato et al., 2014), and intact proviruses are clonally expanded in T_H1-polarized cells (Lee et al., 2017).

How can infected cells be expanded in an antigen-dependent fashion if TCR stimulation, for example by PHA, cannot only lead to $CD4^+$ T cell activation and proliferation, but also to subsequent reactivation of HIV in latently infected cells (Cohn et al., 2018; Laird et al., 2013; Spina et al., 2013)?

One possible explanation that reconciles the seemingly contradictory findings that infected cells can undergo cytokine- and antigen-induced proliferation without proviral reactivation is a characteristic latently infected cell gene expression program that facilitates repression of proviral transcription and is possibly induced by proviral integration.

HIV-1 transcription relies on a plethora of host cell factors, such as the transcription factors NF-κB, NFAT, or AP-1, that are abundant in the nucleus of activated CD4 T^+ cells, but restricted in resting CD4 T^+ cells (Pan et al., 2013; Wong and Jiang, 2021). Further host cell proteins are implied in the control of transcription factors and HIV latency. TRIM32 is a protein that can induce NF-κB translocation into the nucleus and thus promote proviral transcription. It is negatively regulated by MicroRNA155 (MIR155). Thus, MIR155 can facilitate proviral latency (Ruelas et

al., 2015). In cluster 7, TRIM32 was not downregulated, compared to all other clusters. Since mature microRNAs are not polyadenylated (O'Brien et al., 2018), they cannot be reverse transcribed by the 10x Genomics platform. Therefore, differential transcription of MIR155 in cluster 7 could not be assessed. The *PRDM1* gene encodes the Blimp-1 protein, and is expressed in $CD4^+$ memory T cells, including TCM cells, and highly expressed in TEM cells. Blimp-1 represses HIV transcription (Kaczmarek Michaels et al., 2015). However, *PRDM1* expression was not differentially upregulated in cluster 7, which coincides mainly with the TEM compartment. Antigen-stimulation and TCR signaling can induce T cell proliferation and HIV reactivation (Cohn et al., 2018; Laird et al., 2013; Spina et al., 2013). However, chronic infection, including HIV-1, can also lead to T cell exhaustion, which is characterized by a state of unresponsiveness towards antigen stimulation. Exhausted T cells display surface molecules such as PD-1, LAG3, and TIGIT, that subsequently suppress NF-κB, NFAT, or AP-1 (Wherry and Kurachi, 2015). Infected cells are enriched among $CD4^+$ T cells that express PD-1, LAG3, and TIGIT (Fromentin et al., 2016; Hatano et al., 2013). However, it is not known if those exhaustion markers enrich specifically infected cells that harbor an intact provirus, and cluster 7 did not display a distinct upregulation of these markers. Neither did cluster 7 show upregulation of EOMES, a transcription factor that is highly expressed at an advanced stage of exhaustion (Wherry and Kurachi, 2015).

In conclusion, the infected clones of interest that harbored an intact provirus were enriched in cluster 7. This cluster did not display upregulation of exemplary genes that are known to repress proviral transcription or that are upregulated in T cell exhaustion and thus would favor proviral latency and persistence of infected cells. This speaks against the hypothesis that proviral integration induces a “latently infected cell gene expression program”.

The search for unique biomarkers of infected cells for purification or therapeutic targeting is based on the assumption that infected cells are profoundly different from uninfected $CD4^+$ T cells, and that infected cells are uniform. However, no such biomarker has been discovered.

Our data suggests that, in general, proviral integration does not shape infected cells. Therefore, infected resting $CD4^+$ T cells do not have a uniform transcriptome. The infected clones of interest were distributed over 10 out of 15 gene expression clusters, and formed part of the TEM, TCM, and CTL compartment.

If proviral integration does not induce a transcriptional program that enables latency, and if infected cells are not otherwise uniformly shaped by latent infection, how can latently infected cells arise and persist?

One hypothesis of latently infected cell generation is that if an activated $CD4^+$ T cell becomes infected with HIV-1 and subsequently transitions into a resting memory cell, it can become a latently infected cell. Thus, latency would be a coincidental proviral transcription state that is the consequence of a physiological $CD4^+$ T cell process, rather than a virally induces cellular program. An infected cell would just have to survive long enough to enter the memory compartment before cytopathic effects of the virus or the immune system can kill the infected cell, or a cell could be infected just before becoming a memory cell (Mbonye and Karn, 2017; Murray et al., 2016). Our data supports this hypothesis, since we did not find a distinct gene expression cluster of infected clones that harbored an intact provirus. Rather, these clones of interest clustered together with uninfected memory cells that were contained in the enriched sample. Cells pertaining to the infected clones of interest were found in 10 out of 15 gene expression clusters, representing diverse transcriptional programs.

We projected our gene expression data on a reference map of $CD4^+$ T cells from HIV-negative donors. The infected clones of interest and uninfected $CD4^+$ T cells from people living with HIV did not cluster apart from $CD4^+$ T cells from HIV-negative donors. Rather, all three cell types, irrespective of proviral integration, were distributed over the TEM, TCM, and CTL compartment. This is in accordance with the notion that proviral integration does not induce a "latently infected cell phenotype", but can be maintained silently in several resting $CD4^+$ T cell memory subsets. The diversity within a single expanded infected T cell clone that we observed is a physiological finding and is accounted for by asymmetric cell division and different cytokine stimuli and antigen concentrations that members of a clone encounter (Chang et al., 2007; Hale and Ahmed, 2015; Lee et al., 2017).

In three infected clones of interest (individuals 5104, 5125, and 9247) analyzed in our work, all members carried an intact provirus. In the other three intact infected clones of interest (individuals 603, 605, and B207), only a part of the $CD4^+$ T cell clone harbored an intact provirus, while other members of the $CD4^+$ T cell clone were uninfected. However, no difference between the fully and partially infected clones of interest in terms of distribution over gene expression clusters or preferential clustering in gene expression cluster 7 was apparent.

In addition to the infected clones of interest, we detected 52 cells with viral RNA expression that did not belong to any infected clone of interest. These cells, that harbored proviruses that were not analyzed for genetic integrity and could therefore be intact or defective, were widely distributed over 14 out of 15 gene expression clusters. This further corroborates that proviral integration does not impact the fate of infected clones.

Persistence of intact proviruses in different T cell subsets with very distinct gene expression profiles allows the assumption that there is no one mechanism that enables latency. Rather, latency

can be seen as a multifactorial state in permissive cells that can be impacted by the proviral integration, the host cell differentiation, cytokine and antigen exposures subsequent to proviral integration, and stochastic events.

The clones that we analyzed in the present work were confirmed to be genetically intact. Infected cells that harbor an intact provirus are a rare fraction within the HIV-1 reservoir (Bruner et al., 2016; Cho et al., 2022; Cohn et al., 2015; Einkauf et al., 2022; Eriksson et al., 2013; Ho et al., 2013; Imamichi et al., 2014; Lee et al., 2017; Lorenzi et al., 2016; Peluso et al., 2020; Sanchez et al., 1997). Collora and colleagues identified infected clones based on viral RNA expression, but did not confirm the genetic integrity of the proviruses (Collora et al., 2022). Thus, by default, most of those clones are likely to harbor a defective provirus, and only a small fraction of those clones would be intact. Those infected clones were not preferentially found in the TEM compartment, but in the CTL compartment. This could be a hallmark of clones that express viral RNA, or a hallmark of infected clones that harbor a defective provirus.

The infected clone of interest in individual 605 contained a small number of cells that expressed viral RNA. Most cells of that clone were found in the TEM, and the TCM compartment, and only a small fraction was present in the CTL compartment. The 10x Genomics gene expression platform did not reliably identify viral RNA^+ cells. Therefore, whether the other five infected clones of interest expressed viral RNA or were completely latently infected cannot be assessed. One can hypothesize, however, that the integration of a defective provirus impacts a host cell differently than the integration of an intact provirus, leading to enrichment of infected clones that harbor a defective provirus in the CTL compartment.

In summary, the findings in our study imply that there is no unifying trait within the gene expression profile of expanded infected clones harboring an intact provirus that would provide a single mechanism of proviral latency and persistence. Proviral integration does not seem to shape the transcriptome of the host cell. Rather, infected clones can be permissive to latency in a number of ways and in different T cell subsets. Thus, a single curative strategy that would target the entire HIV-1 reservoir does not appear feasible.

5.2 Resistance to cytotoxic lymphocytes as a possible mechanism of HIV-1 reservoir maintenance

In addition to the mechanisms of proviral persistence that are described above, proviral escape variants from CTL are selected and persist over time, whereas CTL sensitive viruses are more often eliminated from the HIV-1 reservoir (Ganusov et al., 2011; Goonetilleke et al., 2009; Price et al., 1997). In general, resistance to $CD8^+$ T cell-mediated killing is inherent to cells harboring an intact provirus and is associated with high expression of the prosurvival factor BCL-2 (Huang et al., 2018; Ren et al., 2020).

Duette and colleagues observed enrichment of cells harboring intact proviruses in the TEM compartment. This enrichment in the TEM compartment was not based on escape variants that evade $CD8^+$ CTL. They proposed that the viral Nef protein, which can decrease the expression of MHCI and CD4 by the host cell and therefore confer immune evasion, was suggested to favor survival of infected cells harboring an intact provirus cells in the TEM compartment (Basmaciogullari and Pizzato, 2014; Duette et al., 2022). However, the expression of the *nef* gene in infected cells harboring an intact provirus was not directly quantified in that study. Therefore, the contribution of *nef* expression to immune evasion from cytotoxic lymphocytes remains unclear.

Furthermore, Nef-induced downregulation of MHCI molecules can prevent killing by MHCI-restricted cytotoxic lymphocytes, but not by MHCII-restricted lymphocytes, such as $CD4^+$ CTLs that are expanded in HIV infection and can be HIV-specific (Appay et al., 2002; Johnson et al., 2015; Soghoian et al., 2012).

The gene expression cluster that contained most cells of the infected clones of interest in our study was marked by the increased expression of cystatin F (CST7). Cystatin F is considered to be a key regulator of granzymes and thus cytotoxicity (Perisic Nanut et al., 2014). Cells that secrete cystatin F can negatively regulate cytotoxicity and therefore evade CTLs (Perisic Nanut et al., 2017; Prunk et al., 2020). Thus, upregulation of cystatin F is a potential mechanism of persistence of the HIV-1 reservoir. Similarly, Collora and colleagues found that a subset of infected cells with proviral transcription had increased expression of Serpin B9, a protease inhibitor that can also regulate cytotoxic molecules (Collora et al., 2022). However, the increase in Serpin B9 expression was not significant, and whether the infected cells in that study carried an intact HIV-1 provirus was not confirmed.

5.3 Comparison of *ex vivo* resting infected clones of interest with *in vitro* reactivated infected cells and *in vitro* models of latency

Cohn and colleagues performed *in vitro* reactivation of *ex vivo* cells that harbored an intact provirus. Subsequently, the gene expression of *env*-expressing reactivated $CD4^+$ T cells was analyzed on a single cell basis (Cohn et al., 2018). Three individuals (603, 605, and B207) were also included in our present study. Cohn and colleagues compared the gene expression profile of reactivated infected cells that harbored an intact provirus to reactivated uninfected cells. HLA-DR,

which enriches for intact HIV-1 proviruses (Horsburgh et al., 2020; Lee et al., 2019) is up-regulated in reactivated infected cells. It was also upregulated in gene expression cluster 7 that contained the majority of cells pertaining to the infected clones of interest. In contrast, CCL5 was down-regulated in reactivated infected cells but up-regulated in our cluster 7. Other genes that Cohn and colleagues found up-regulated in reactivated infected cells were TIGIT, CCL3, PRDM1, MAF, IL21-AS1, or BATF3. Some of those genes are associated with HIV-1 latency and suppression of viral transcription, but were not differently expressed cluster 7 in our analysis. Furthermore, reactivated infected cells and reactivated uninfected cells were clustering apart from each other. *In vitro* reactivation of latent cell lines suggests that the difference in gene expression profiles between reactivated infected cells and reactivated uninfected cells is induced by the presence of viral RNA and active viral production (Krishnan and Zeichner, 2004). This could explain the distinct gene expression profiles of reactivated infected and reactivated uninfected *ex vivo* $CD4^+$ T cells. In contrast, our study highlights that, in their resting state, infected cells harboring an intact provirus were found in the same clusters as uninfected $CD4^+$ T cells and share similar gene expression profiles.

Importantly, the LRA that is chosen for reactivation of latent cells impacts which latent cells are reactivated, since not all LRAs reactivate each T cell subset in a similar manner. For example, Panobinostat and Ingenol have a higher latency reversing effect on TCM cells than on TEM cells, whereas Romidepsin preferentially reactivates latent cells in the TEM compartment (Grau-Exposito et al., 2019; Pardons et al., 2019b). Thus, identification of infected cells by proviral reactivation with an LRA is biased towards the specific T cell subset that is preferentially reactivated by the respective LRA. In contrast, our study analyzed entire resting infected clones harboring intact proviruses.

Iglesias-Ussel and colleagues used a model of HIV latency with primary $CD4^+$ T cells (Iglesias-Ussel et al., 2013). A total RNA microarray found distinct features of *in vitro* generated latent cells, compared to uninfected cells. Especially high expression of CD2 was characteristic of latent cells in the model, and sorting of *ex vivo* $CD4^+$ T cells from ART-treated people living with HIV verified enrichment of HIV-1 DNA in $CD2^{high}$ cells. However, whether $CD2^{high}$ cells were enriched for intact proviruses was not analyzed.

CD2 was among the 15 most up-regulated genes in cluster 7 in our data set. In contrast to the *in vitro* model, we did not observe a distinct gene expression profile of resting infected clones harboring an intact provirus that would set them apart from uninfected cells, since cluster 7 did not only contain the majority of cells pertaining to the infected clones of interest, but also a large number of uninfected cells.

In another latency model with primary $CD4^+$ T cells, Bradley and colleagues observed overlapping transcriptomes between latently infected and uninfected cells (Bradley et al., 2018), which is more consistent with our findings that resting infected and uninfected $CD4^+$ T cells can share similar gene expression profiles. Nonetheless, 17 genes were differentially expressed in *in vitro* latently infected $CD4^+$ T cells, compared to co-cultured uninfected $CD4^+$ T cells. Only three of those 17 genes were also differentially expressed in our gene expression cluster 7. However, only *CRIP1* was upregulated in both, *in vitro* generated latent cells and gene expression cluster 7. In contrast, *MIF* and *MTRNR2L12* were upregulated in *in vitro* generated latently infected cells, but downregulated in gene expression cluster 7.

In conclusion, reactivated *ex vivo* infected cells harboring an intact provirus and *in vitro* infected primary $CD4^+$ T cells as models of latency differ substantially from *ex vivo* resting infected cells harboring an intact provirus. This highlights the importance of the present work, which did not rely on latency reversal or modelling latency *in vitro* but characterized *ex vivo* resting infected cells harboring an intact provirus.

5.4 Limitations of the study and future prospects

The present study is limited to expanded infected clones harboring an intact provirus from aviraemic individuals on long-time ART. The clonality of the HIV-1 reservoir increases over time (Antar et al., 2020; Cho et al., 2022; Cohn et al., 2015; Wagner et al., 2013), and there is an inverse correlation between the size of an infected clone and the probability that this respective infected clone is reactivated and expresses HIV proteins (Lorenzi et al., 2016). The majority of the HIV-1 proviral reservoir is defective and especially selected for clones that harbor a defective provirus that cannot produce viral proteins (Bruner et al., 2016; Cho et al., 2022; Cohn et al., 2015; Einkauf et al., 2022; Eriksson et al., 2013; Ho et al., 2013; Imamichi et al., 2014; Lee et al., 2017; Lorenzi et al., 2016; Peluso et al., 2020; Sanchez et al., 1997). Only a small fraction of the HIV-1 reservoir harbors an intact provirus, and absence or low level proviral transcription is selected over intact clones with higher proviral transcription (Einkauf et al., 2022). Therefore, the infected clones of interest in this study were selected over time under ART. The transcriptional difference of less expanded infected clones or single infected cells harboring an intact provirus that over time become activated and eliminated from the HIV-1 reservoir, opposed to expanded infected clones that are less likely to be reactivated and therefore persist, could provide a deeper inside into mechanisms of latency and proviral persistence. This could be achieved by the longitudinal

collection of blood samples from people living with HIV, and the enrichment and gene expression analysis of various infected clones over time.

Duette and colleagues found differences in the distribution of infected cells harboring intact or defective proviruses over $CD4^+$ T cell naïve and memory compartments, implying that the provirus can affect the fate of the host $CD4^+$ T cell (Duette et al., 2022). Enrichment and gene expression analysis of infected clones with various categories of proviral defects (for example, deletion of certain genes, mutations in the major splice donor site or packaging signal element) could elucidate effects of the provirus and viral genes on the fate of the host $CD4^+$ T cell.

A technical limitation of the present study is the limited sequencing depth of the 10x Genomics platform. Deeper sequencing may find differences in gene expression profiles between infected clones harboring an intact provirus and uninfected cells. It may further be able to reliably distinguish infected cells with proviral transcription from completely transcriptionally silent latent cells.

In our study, no differences between completely and partially infected clones harboring an intact provirus were apparent. However, within partially infected clones, we were unable to distinguish on a single cell level cells that harbored a provirus from cells that did not harbor a provirus within the respective $CD4^+$ T cell clone. Combined proviral sequencing of gDNA and gene expression analysis on a single cell level could highlight differences of infected and uninfected cells within a $CD4^+$ T cell clone.

The samples studied in this work were derived from leukapheresis and therefore represented infected cells in the blood circulation. Proviruses in infected cells from blood are representative of those found in other anatomical compartments (Chaillon et al., 2020; McManus et al., 2019). However, the transcriptional differences of infected cells in different anatomical compartments remain unclear and are difficult to assess, given the inaccessibility of tissue-resident $CD4^+$ T cells and $CD4^+$ T cells in secondary lymphoid organs.

The impact of the biological sex on the course of HIV-1 infection, latency, and curative strategies has become an aspect of HIV research (Ait-Ammar et al., 2019; Falcinelli et al., 2020; Moran et al., 2022). The biological sex of the six study participants was male, and inclusion of participants of non-male biological sex may extend the understanding of latency and proviral persistence in people living with HIV.

CHAPTER 6. APPENDIX

6.1 Data availability

The data reported in this thesis is archived at the following databases: Single cell RNA-Seq and TCR data is available at NCBI GEO (GSE204756); Envelope sequences deposited into the Genbank database (ON662322-ON664914); raw FASTQ sequences used to identify the TCR of the infected clones of interest are available at NCBI SRA (SRR19524296-SRR19524298).

The code for single-cell analysis was released at https://doi.org/10.5281/zenodo.6950427.

The code to identify the TCR of the infected clones of interest is available on github (https://github.com/victor-ramos/demultiplex_and_assembly_TCR) and at https://doi.org/10.5281/zenodo.6954076.

6.2 Participant cohort

Study participants were recruited at the Rockefeller University hospital and gave informed written consent before participation in the study. The study protocols and procedures met the standards of Good Clinical Practice and were approved by the institutional review board of the Rockefeller University. The participants' characteristics, including age and sex, are available in Table 1.

After leukapheresis, peripheral blood mononuclear cells (PBMCs) were isolated by Ficoll separation and stored in aliquots in liquid nitrogen. At the time of sample collection, all study participants were receiving ART, were aviremic and had not undergone experimental treatment regimes.

6.3 Methods

6.3.1 Cell sorting

All procedures were performed while maintaining cells at 4 °C. $CD4^+$ T cells were negatively selected from PBMCs with the $CD4^+$ T cell isolation kit (Miltenyi, cat. 130-059-901). For memory cell separation, $CD45RA^-$ T cells were negatively selected using magnetic CD45RA MicroBeads (Miltenyi, cat. 130-045-901).

$CD4^+CD45RA^-$ T cells were incubated with Fc-blocking reagent for 10 min (Miltenyi, cat. 130-059-901). Fixable Viability Dye eFluor 780 (Invitrogen, cat. 65-0865-14) was used for live/dead cell staining (dilution 1:1,000). The following antibodies were used for surface staining (dilution 1:100): PerCP/Cy5.5 anti-human CD4 (BioLegend, cat. 317428), PacificBlue anti-human CD3 (BioLegend, cat. 300431), Brilliant Violet 605 anti-human TCR Cβ1 (BD, cat. 747979), FITC anti-human TCR Vβ17 (Beckman Coulter, cat. IM1234), FITC anti-human TCR Vβ21.3 (Backman Coulter, cat. IM1483), FITC anti-human TCR Vβ7.2 (Beckman Coulter, cat. B06666), and Beta Mark TCR Vbeta Repertoire Kit (Beckman Coulter, cat. IM3497). The Vbeta Repertoire Kit contains 24 different anti-TRBV antibodies that come in 8 vials (A – H). In each vial, there are three antibodies conjugated with either PE, or FITC, or PE-FITC.

For enrichment, the following population was sorted: individuals B207 and 5104 $CD3^+CD4^+TRBC1^+TRBV^-$ lymphocytes; individual 603 $CD4^+TRBC1^+TRBV19^+$ lymphocytes; individual 605 $CD4^+TRBV11\text{-}2^+$ lymphocytes; individual 5125 $CD4^+TRBC1^+TRBV2^+$ lymphocytes; individual 9247 $CD4^+TRBC1^-TRBV4\text{-}3^+$ lymphocytes. Sorts were performed on BD FACS Aria IIu and BD FACS Aria III using FACSDiva software (version 2.0.2, BD). Flowcytometry data was analyzed with FlowJo (version 10.8.1, BD)

6.3.2 gDNA extraction and quantification

Sorted cells were incubated with Proteinase K buffer (10 mM Tris, pH 8, 0.5% SDS, 100 mM NaCl, and 1 mM EDTA) and 250 μg/mL Proteinase K (ThermoFisher, cat. AM2548) at 50 °C for 6 hours. An equal volume of phenol-chloroform-isoamyl (Sigma-Aldrich, cat. 77617) was added and gDNA was extracted with Phase Lock Gel light tubes (QuantaBio, cat. 10847-800). gDNA was precipitated with ethanol and GlycoBlue Coprecipitant (ThermoFisher, cat. AM9515), washed twice with 70 % ethanol and resuspended in TE buffer. gDNA concentration was measured by Qubit dsDNA BR Assay Kit (ThermoFisher, cat. Q32853).

6.3.3 Near full-length proviral amplification and Q4PCR

Near full-length proviral (NFL) amplification was performed as previously described (Gaebler et al., 2019).

The first round of near full-length proviral amplification (NFL1 PCR) was run in a 20 μL reaction volume with 0.5 U of Platinum Taq DNA Polymerase High Fidelity (Fisher Scientific, cat. 11-304-029) and outer PCR primers BLOuterF (5′-AAATCTCTAGCAGTGGCGCCCGAACAG-3′) and BLOuterR (5′-TGAGGGATCTCTAGTTACCAGAGTC-3′) (0.2 μM each) (Li et al., 2007), 2 mM $MgSO_4$ and 0.2 mM dNTPs at 94 °C for 2 min; (94 °C for 30 s, 64 °C for 30 s, and 68 °C for 10 min) for 3 cycles; (94 °C for 30 s, 61 °C for 30 s, and 68 °C for 10 min) for 3 cycles; (94 °C for 30 s, 58 °C for 30 s, and 68 °C for 10 min) for 3 cycles; (94 °C for 30 s, 55 °C for 30 s, and 68 °C for 10 min) for 41 cycles; and then 68 °C for 10 min as a touch down PCR.

Next, a quadruplex qPCR in a total reaction volume of 10 μL was performed. To 5 μL of TaqMan Universal PCR Master Mix, (ThermoFisher, cat. 4304437), 1 μL of the NFL1 PCR product and 4 μL of a master mix containing the following primer pairs and internal probes were added (final

concentration in 10 µL reaction volume in brackets): PS F (5′-TCTCTCGACGCAGGACTC-3′) and PS R (5′-TCTAGCCTCCGCTAGTCAAA-3′) (675 nM each), PS probe (5′/Cy5/TTTGGCGTA/TAO/CTCACCAGTCGCC/3′/IAbRQSp (187.5 nM) (Bruner, 2018); env F (5′-AGTGGTGCAGAGAGAAAAAAGAGC-3′) and env R (5′-GTCTGGCCTGTACCGTCAGC-3′) (90 nM each), env probe (5′/VIC/CCTTGGGTTCTTGGGA/3′/MGB) (25 nM) (Bruner et al., 2019);

gag F (5′-ATGTTTTCAGCATTATCAGAAGGA-3′) and gag R (5′-TGCTTGATGTCCCCCCACT-3′) (337.5 nM each), gag probe (5′/6-FAM/CCACCCCAC/ZEN/AAGATTTAAACACCATGCTAA/3′/IABkFQ) (93.75 nM) (Palmer et al., 2003);

pol F (5′-GCACTTTAAATTTTCCCATTAGTCCTA-3′) and pol R (5′-CAAATTTCTACTAATGCTTTTATTTTTTC-3′) (675 nM each), pol probe (5′/NED/AAGCCAGGAATGGATGGCC/3′/MGB) (187.5 nM) (Schmid et al., 2010). qPCR was performed using the Applied Biosystem QuantStudio 6 Flex Real-Time PCR System with the following conditions:

94 °C for 10 min; (94 °C for 15 s, 60 °C for 1 min) for 40 cycles.

QuantStudio Real-Time PCR Software version 1.3 (ThermoFisher) was used for qPCR data analysis. Baseline correction (start cycle 3, end cycle 10) and normalized reported signal threshold (ΔRn threshold = 0.025) was set for all targets/probes. A fluorescent signal above the threshold between cycle value 10 and 40 of any two or more probes was identified and the corresponding NFL1 PCR samples were selected for NFL2 PCR.

NFL2 PCR was run in a 20 µL reaction volume with 1 µL of NFL1 PCR product, 0.5 U of Platinum Taq DNA Polymerase High Fidelity (Fisher Scientific, cat. 11-304-029) and inner PCR primers

275F (5′-ACAGGGACCTGAAAGCGAAAG-3′) and 280R (5′-CTAGTTACCAGAGTCACA CAACAGACG-3) (0.2 µM each) (Ho et al., 2013), 2 mM $MgSO_4$ and 0.2 mM dNTPs at 94 °C for 2 min; (94 °C for 30 s, 64 °C for 30 s, and 68 °C for 10 min) for 3 cycles; (94 °C for 30 s, 61 °C for 30 s, and 68 °C for 10 min) for 3 cycles; (94 °C for 30 s, 58 °C for 30 s, and 68 °C for 10 min) for 3 cycles; (94 °C for 30 s, 55 °C for 30 s, and 68 °C for 10 min) for 41 cycles; and then 68 °C for 10 min as a touch down PCR. NFL2 PCR products were then sequenced.

For each sample, a serial dilution of gDNA was performed to achieve a fluorescent signal above the threshold between cycle value 10 and 40 of any two or more probes in < 30 % of wells.

6.3.4 Env PCR

Envelope gene amplification was performed as previously described (Salazar-Gonzalez et al., 2008; Salazar-Gonzalez et al., 2009).

The first round of *env* amplification (Env1 PCR) for individuals 603, B207, 5104, 5125, and 9247 was run in a 20 µL reaction volume with 0.5 U of Platinum Taq DNA Polymerase High Fidelity (Fisher Scientific, cat. 11-304-029) and outer PCR primers Env5out (5′-TAGAGCCCTGGAAGCATCCAGGAAG-3′) and Env3out (5′-TTGCTACTTGTGATTGCTCCATGT-3′) (0.2 µM each) (Salazar-Gonzalez et al., 2008), 2 mM $MgSO_4$ and 0.2 mM dNTPs at 94 °C for 2 min; (94 °C for 15 s, 58.5 °C for 30 s, and 68 °C for 3 min) for 35 cycles; then 68 °C for 15 min. Due to a primer mismatch, the first round of *env* amplification (Env1 PCR) for individual 605 was run in a 20 µL reaction volume with 0.5 U of Platinum Taq DNA Polymerase High Fidelity (Fisher Scientific, cat. 11-304-029) and outer PCR primers B3F3 (5′-TGGAAAGGTGAAGGGGCAGT-AGTAATAC-3′) (Salazar-Gonzalez et al., 2009) and Env3out (5′-TTGCTACTTGTGATTGCTCCATGT-3′) (0.2 µM each), 2 mM $MgSO_4$

and 0.2 mM dNTPs at 94 °C for 2 min; (94 °C for 15 s, 60.4 °C for 30 s, and 68 °C for 6 min) for 40 cycles; then 68 °C for 15 min.

The second round of *env* amplification (Env2 PCR) for all individuals was run in a 20 µL reaction volume with 1 µL of Env1 PCR product, 0.5 U of Platinum Taq DNA Polymerase High Fidelity (Fisher Scientific, cat. 11-304-029) and inner PCR primers Env5in (5′-TTAGGCATCTCCTATGGCAGGAAGAAG-3′) and Env3in (5′-GTCTCGAGATACTGCTCCCACCC-3′) (0.2 µM each) (Salazar-Gonzalez et al., 2008), 2 mM $MgSO_4$ and 0.2 mM dNTPs at 94 °C for 2 min; (94 °C for 15 s, 61 °C for 30 s, and 68 °C for 3 min) for 35 cycles; then 68 °C for 15 min.

4 µL aliquots of Env2 PCR products were added to 16 µL aliquots of nuclease-free water and run on E-Gel 96 Agarose Gels with SYBR Safe DNA Gel Stain, 1% (ThermoFisher, cat. G720841) for visualization. Samples with a band size of ~ 2.5 kb were selected for sequencing. For each sample, a serial dilution of gDNA was performed to achieve a PCR product band on agarose gel for < 30 % of wells.

For *env* gene sequencing for the infected clone of interest TCR identification in individual 605, a shortened nested PCR was performed to amplify a ~ 500 bp amplicon of the *env* gene in which the expanded infected clone of interest differed from other proviruses that were enriched. The first round of *env* amplification (Env1short PCR) for individual 605 was run in a 20 µL reaction volume with 0.5 U of Platinum Taq DNA Polymerase High Fidelity (Fisher Scientific, cat. 11-304-029) and outer PCR primers Env 1,133 F1 (5'- GAGGGGAATTTTTCTACTGTAACAC-3') and Env 1,956 R1 (5'- GTTCTGCGAATTTT-CAATTAAGGTG -3') (0.2 µM each), 2 mM $MgSO_4$ and 0.2 mM dNTPs at 94 °C for 2min; (94 °C for 15 s, 56 °C for 30 s, and 68 °C for 1 min) for 35

cycles; and 68 °C for 15 min. The second round of *env* amplification (Env2short PCR) for individual 605 was run in a 20 μL reaction volume with 1 μL of Env1short PCR product, 0.5 U of Platinum Taq DNA Polymerase High Fidelity (Fisher Scientific, cat. 11-304-029) and inner PCR primers Env 1,240 F2 (5'- ATCACACTCCGATGC-AGAATAAAAC-3') and Env 1,765 R2 (5'-TTAGGTATCT-TTCCACAGCCAGTAC-3') (0.2 μM each), 2 mM $MgSO_4$ and 0.2 mM dNTPs at 94 °C for 2 min; (94 °C for 15 s, 58 °C for 30 s, and 68 °C for 1 min) for 35 cycles; and 68 °C for 15 min. Env2short PCR products were either sequenced by Sanger sequencing (GeneWiz, Azenta Life Sciences) with primers 1,240 F2 (5'- ATCACACTCCGATGC-AGAATAAAAC-3') and Env 1,765 R2 (5'-TTAGGTATCT-TTCCACAGCCAGTAC-3') or on the Illumina MiSeq platform as described below.

6.3.5 Env2 PCR product and NFL2 PCR product sequencing

DNA concentrations of second round PCR products were measured with Qubit dsDNA BR Assay Kit (ThermoFisher, cat. Q32853). Samples were diluted to a concentration of 10 – 20 ng/μL. For tagmentation, 1 μL of diluted second round PCR product was added to 0.25 μL Nextera TDE1 Tagment DNA enzyme and 1.25 μL of TD Tagmentation buffer (Illumina Tagment DNA TDE1 Enzyme and Buffer, Illumina cat. 20034198). Subsequently, DNA fragments were ligated to i5/i7 barcoded primers from the Illumina Nextera XT Index Kit v2 A – D (Nextera XT Index Kit v2 Set A, Illumina, cat. FC-131-2001 – FC-131-2004) using KAPA HiFi HotStart ReadyMix PCR Kit (Roche, cat. 07958935001). The ligated DNA fragments were pooled and purified with Agencourt RNAClean XP magnetic beads (Beckman Coulter, cat. A63987) for paired-end sequencing at a concentration of 12 pM with MiSeq Reagent Kit v2 (300-cycles) (Illumina, cat. MS-102-2002). Sequences were assembled with the Defective and Intact HIV Genome Assembler.

6.3.6 Phylogenetic trees

Env sequences were aligned to the HIV HXB2CG *env* sequence using Geneious Prime software (version 11.0.12, Biomatters). Maximum-likelihood phylogenetic trees were built with PHYML, substitution model HKY85, without bootstrapping, to identify identical *env* sequences by clustering.

6.3.7 10x Genomics

10x Genomics gene expression and V(D)J libraries were generated with the Chromium Single Cell 5' Library & Gel Bead Kit (10x Genomics, cat. PN-1000014) and Chromium Single Cell V(D)J Enrichment Kit, Human T Cell (10x Genomics, cat. PN-1000005) as described in the 10x Genomics protocol. The 5' expression library was sequenced with NovaSeq 6000 S1 (100 cycles) (Illumina, cat. 20012865) and the V(D)J library was sequenced with NextSeq 500/550 Mid Output Kit v2.5 (300 Cycles) (Illumina, cat. 20024905).

6.3.8 Infected clone of interest TCR identification

Infected cells pertaining to the clones of interest were enriched based on CD45RA, TRBC, and TRBV, and 5 cells per well were sorted into 96-well plates containing TCL buffer (Qiagen, cat. 1031576) with 1 % beta-mercaptethanol and snap frozen on dry ice. Plates were stored at - 80 °C until further use. After thawing on ice, magnetic bead clean-up was performed with Agencourt RNAClean XP magnetic beads (Beckman Coulter, cat. A63987). TCR mRNA was reverse transcribed into cDNA in a reaction volume of 20 µL with 200 U of SuperScript III Reverse Transcriptase (Invitrogen, cat. 18080044), 40 U of RNasin Plus Ribonuclease Inhibitor (Promega,

cat. N2615), 0.5 mM dNTPs, 5 mM DTT, and the following primers: AC1R (5'-ACACATCAGAATCCTTACTTTG-3'), BC1R (5'-CAGTATCTGGAGTCATTGA-3') (0.125 μM each) (Mamedov et al., 2013). After a second magnetic bead clean-up, binding TCR cDNA and gDNA, the first round of *env* amplification PCR from gDNA was performed as described above. After this, each well contained multiple copies of TCR cDNA and, if an infected cell was present, multiple copies of the *env* gene. A 1 μL aliquot from each well was taken for the second round of *env* amplification and sequencing to identify wells that contained a cell of the infected clone of interest (Env^+ wells). From all Env^+ wells and a random selection of Env^- wells, an aliquot was taken to amplify the TCRα and TCRβ chain separately in duplicates with primers as previously described (Han et al., 2014). The first round of PCR amplification was performed with 2 μL aliquots of cDNA in a total reaction volume of 25 μL, using 0.75 U of HotStarTaq DNA polymerase (Qiagen, cat. 203209), the forward primers TRAV Ph1 (0.06 μM each) and the reverse primer AlphaPhase1 (0.03 μM) for TCRα amplification; the forward primers TRBV Ph1 (0.06 μM each) and the reverse primer BetaPhase1 (0.03 μM) for TCRβ amplification, and 0.2 mM dNTPs at 95 °C for 15 min; (94 °C for 30 s, 62 °C for 1 min, and 72 °C for 1 min) for 25 cycles; and 72 °C for 5 min. The second round of PCR amplification was performed with 1 μL aliquots of first round PCR amplification products in a total reaction volume of 25 μL, using 0.75 U of HotStarTaq DNA polymerase (Qiagen, cat. 203209), the forward primers hTRAV (0.06 μM each) and the reverse primer AlphaPhase2 (0.03 μM) for TCRα amplification, the forward primers hTRBV (each at 0.06 μM) and the reverse primer BetaPhase2 (0.03 μM) for TCRβ amplification, and 0.2 mM dNTPs at 95 °C for 15 min; (94 °C for 30 s, 64 °C for 1 min, and 72 °C for 1 min) for 25 cycles; and 72 °C for 5 min. The third round of PCR amplification was performed with 1 μL aliquots of second round PCR amplification products in a total reaction volume of 25 μL, using 0.75 U of

HotStarTaq DNA polymerase (Qiagen, cat. 203209), the AlphaBC primers (0.05 μM) for TCRα amplification, the BetaBC primers (0.05 μM) for TCRβ amplification, PlateNN primers (0.05 μM), PE1 and PE2 primers (0.05 μM each), and 0.2 mM dNTPs at 95 °C for 15 min; (94 °C for 30 s, 64 °C for 1 min, and 72 °C for 1 min) for 36 cycles; and 72 °C for 5 min. The barcoded PCR amplification products were pooled, concentrated with Agencourt RNAClean XP magnetic beads (Beckman Coulter, cat. A63987), and run on an 1.5 % agarose gel. The band between 300 and 400 bp was excised and purified with the NucleoSpin Gel and PCR Clean-up kit (Quiagen, cat. 740609.50). The purified product was subjected to paired-end sequencing at a concentration of 20 pM with MiSeq Reagent Kit v3 (600 cycle) (Illumina, cat. MS-102-3003).

PhiX-derived reads were removed from downstream analysis using a k-mer based approach implemented by bbduk.sh from BBTools v38.72 (https://sourceforge.net/projects/bbmap/). Samples were demultiplexed according to the combination of oligos used to uniquely identify the plate, row, and column they were placed (Han et al., 2014). The quality control check was performed with Trim Galore package v0.6.7 (https://github.com/FelixKrueger/TrimGalore) to trim Illumina adapters and low-quality bases. Paired-overlapping reads were exported into a single read by BBMerge. Subsequently, TRUST4 (Song et al., 2021) was used to reconstruct and annotate the T cell receptor (TCR) sequences. There is a chance of incorrect TCR assignment if the sequencing errors occur at the barcodes present in reads. A threshold was established based on the number of reads used to assemble a specific TCR contig assigned to an empty well. In most empty wells, less than 20 reads for a single TCR chain sequence were detected. An individual threshold was established for each sequencing run. We obtained the putative clonotypes, defined by the 10x Genomics V(D)J single-cell sequencing, that shared either the α or β chain found in that well. The

most frequent clonotype associated with Env^+ wells in each individual was selected as the clonotype associated with proviral integration.

The bottleneck in this method is the *env* amplification from one single copy. The efficiency of the TCR amplification is documented as 88 % for TCR α and 93 % for TCR β based on published literature (Han et al., 2014). This is because TCR is abundantly expressed in T cells. Since either a unique TCR α or β chain is sufficient to identify a TCR of interest from the 10x TCR sequencing data, the detection rate for the specific TCR in Env^+ wells is rather high.

6.3.9 Single-cell RNA-seq and single-cell TCR-seq processing

Single-cell RNA-seq binary base call (BCL) files were demultiplexed and converted into FASTQ files using BCLtoFastq prior to alignment to hg38 with CellRanger (v4.0.0) and analyzed in R studio with Seurat (v4). Cells with a mitochondrial proportion greater than 5 % and/or a feature count <200 or >2,500 were discarded. Sample batches were combined, normalized and scaled with SCTransform. Uniform Manifold Approximation and Projection (UMAP) clustering was performed selecting the first thirty principal components. Single-cell TCR-seq FASTQs were aligned with CellRanger (v4.0.0) to the default CellRanger VDJ reference. Output contig annotations were filtered and analyzed in R studio with Seurat.

6.3.10 Mapping scRNA-seq to $CD4^+$ T cells reference

$CD4^+$ T cell population was extracted from published human peripheral blood cells multimodal annotated reference (Hao et al., 2021). The UMAP reference from extracted $CD4^+$ T population was recreated using the first 50 principal components of the RNA expression slot and the cells from each individual were anchored and mapped utilizing the FindTransferAnchors and

MapQuery functions from Seurat (reference.reduction = "pca", dims = 1:50, reduction.model = umap).

6.3.11 HIV-1 transcript detection using 10x data

Samtools was used to extract reads that cellranger failed to align to the human reference. Sequences extracted from the BAM file generated by cellranger contain identified cell barcodes and UMI in the sequence header. We used bbduk.sh from the BBtools package to search for reads containing HIV-1 k-mers (k=31) from intact HIV-1 genome sequences obtained from the Los Alamos HIV database. Cell barcode and UMI were extracted from sequences containing HIV-1 k-mers and used to calculate HIV-1 expression.

6.4 Number of cells analyzed before and after enrichment

Table 11 Number of cells analyzed and assay type before and after enrichment

Individual	Sample	Cells assayed	Sequences of clone of interest	Sequences of clone of interest per 10^6 $CD4^+$ T cells	Assay (Ref.)
603	$CD4^+$	1,228,800	16	13.0	Env PCR
	enriched	49,920	102	2,043.0	Env PCR
605	$CD4^+$	576,000	248	430.5	Env PCR
	enriched	3,072	50	16,276.0	Env PCR
B207	$CD4^+$	249,600	41	164.3	Env PCR
	enriched	115,200	71	616.3	Env PCR
5104	$CD4^+$	576,000	121	210.0	Q4PCR (Gaebler et al., 2022)
	enriched	115,200	159	1,380.2	Env PCR
5125	$CD4^+$	1,440,000	40	27.8	Q4PCR (Gaebler et al., 2022)
	enriched	64,512	95	1,472.6	Env PCR
9247	$CD4^+$	960,000	14	14.6	Q4PCR (Gaebler et al., 2021)
	enriched	19,200	34	1,770.8	Env PCR

CHAPTER 7. References

Abana, C.O., Pilkinton, M.A., Gaudieri, S., Chopra, A., McDonnell, W.J., Wanjalla, C., Barnett, L., Gangula, R., Hager, C., Jung, D.K., *et al.* (2017). Cytomegalovirus (CMV) Epitope-Specific CD4(+) T Cells Are Inflated in HIV(+) CMV(+) Subjects. J Immunol *199*, 3187-3201.

Abrahams, M.R., Anderson, J.A., Giorgi, E.E., Seoighe, C., Mlisana, K., Ping, L.H., Athreya, G.S., Treurnicht, F.K., Keele, B.F., Wood, N., *et al.* (2009). Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-poisson distribution of transmitted variants. J Virol *83*, 3556-3567.

Abram, M.E., Ferris, A.L., Shao, W., Alvord, W.G., and Hughes, S.H. (2010). Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. J Virol *84*, 9864-9878.

Achuthan, V., Keith, B.J., Connolly, B.A., and DeStefano, J.J. (2014). Human immunodeficiency virus reverse transcriptase displays dramatically higher fidelity under physiological magnesium conditions in vitro. J Virol *88*, 8514-8527.

Ahlenstiel, C.L., Symonds, G., Kent, S.J., and Kelleher, A.D. (2020). Block and Lock HIV Cure Strategies to Control the Latent Reservoir. Front Cell Infect Microbiol *10*, 424.

Ait-Ammar, A., Kula, A., Darcis, G., Verdikt, R., De Wit, S., Gautier, V., Mallon, P.W.G., Marcello, A., Rohr, O., and Van Lint, C. (2019). Current Status of Latency Reversing Agents Facing the Heterogeneity of HIV-1 Cellular and Tissue Reservoirs. Front Microbiol *10*, 3060.

Altman, L.K. (1981). RARE CANCER SEEN IN 41 HOMOSEXUALS. In The New York Times, pp. 20.

Annunziato, F., Cosmi, L., Liotta, F., Maggi, E., and Romagnani, S. (2014). Human Th1 dichotomy: origin, phenotype and biologic activities. Immunology.

Antar, A.A., Jenike, K.M., Jang, S., Rigau, D.N., Reeves, D.B., Hoh, R., Krone, M.R., Keruly, J.C., Moore, R.D., Schiffer, J.T., *et al.* (2020). Longitudinal study reveals HIV-1-infected CD4+ T cell dynamics during long-term antiretroviral therapy. J Clin Invest *130*, 3543-3559.

Appay, V., Zaunders, J.J., Papagno, L., Sutton, J., Jaramillo, A., Waters, A., Easterbrook, P., Grey, P., Smith, D., McMichael, A.J., *et al.* (2002). Characterization of CD4(+) CTLs ex vivo. J Immunol *168*, 5954-5958.

Arrildt, K.T., Joseph, S.B., and Swanstrom, R. (2012). The HIV-1 env protein: a coat of many colors. Curr HIV/AIDS Rep *9*, 52-63.

Arts, E.J., and Hazuda, D.J. (2012). HIV-1 antiretroviral drug therapy. Cold Spring Harb Perspect Med *2*, a007161.

Bachmann, N., von Siebenthal, C., Vongrad, V., Turk, T., Neumann, K., Beerenwinkel, N., Bogojeska, J., Fellay, J., Roth, V., Kok, Y.L., *et al.* (2019). Determinants of HIV-1 reservoir size and long-term dynamics during suppressive ART. Nat Commun *10*, 3193.

Banga, R., Procopio, F.A., Noto, A., Pollakis, G., Cavassini, M., Ohmiti, K., Corpataux, J.M., de Leval, L., Pantaleo, G., and Perreau, M. (2016). PD-1(+) and follicular helper T cells are responsible for persistent HIV-1 transcription in treated aviremic individuals. Nat Med *22*, 754-761.

Basmaciogullari, S., and Pizzato, M. (2014). The activity of Nef on HIV-1 infectivity. Front Microbiol *5*, 232.

Bassing, C.H., Swat, W., and Alt, F.W. (2002). The mechanism and regulation of chromosomal V(D)J recombination. Cell *109 Suppl*, S45-55.

Bavinton, B.R., Pinto, A.N., Phanuphak, N., Grinsztejn, B., Prestage, G.P., Zablotska-Manos, I.B., Jin, F., Fairley, C.K., Moore, R., Roth, N.*, et al.* (2018). Viral suppression and HIV transmission in serodiscordant male couples: an international, prospective, observational, cohort study. Lancet HIV *5*, e438-e447.

Baxter, A.E., Niessl, J., Fromentin, R., Richard, J., Porichis, F., Charlebois, R., Massanella, M., Brassard, N., Alsahafi, N., Delgado, G.G.*, et al.* (2016). Single-Cell Characterization of Viral Translation-Competent Reservoirs in HIV-Infected Individuals. Cell Host Microbe *20*, 368-380.

Beckman Coulter. https://www.beckman.com/reagents/coulter-flow-cytometry/antibodies-and-kits/clinical-research-systems-and-kits/im3497.

Bedwell, G.J., Jang, S., Li, W., Singh, P.K., and Engelman, A.N. (2021). rigrag: high-resolution mapping of genic targeting preferences during HIV-1 integration in vitro and in vivo. Nucleic Acids Res *49*, 7330-7346.

Beretta, M., Migraine, J., Moreau, A., Essat, A., Goujard, C., Chaix, M.L., Drouin, A., Bouvin-Pley, M., Meyer, L., Barin, F.*, et al.* (2020). Common evolutionary features of the envelope glycoprotein of HIV-1 in patients belonging to a transmission chain. Sci Rep *10*, 16744.

Berger, E.A., Murphy, P.M., and Farber, J.M. (1999). Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. Annu Rev Immunol *17*, 657-700.

Bosque, A., Famiglietti, M., Weyrich, A.S., Goulston, C., and Planelles, V. (2011). Homeostatic proliferation fails to efficiently reactivate HIV-1 latently infected central memory CD4+ T cells. PLoS Pathog *7*, e1002288.

Bradley, T., Ferrari, G., Haynes, B.F., Margolis, D.M., and Browne, E.P. (2018). Single-Cell Analysis of Quiescent HIV Infection Reveals Host Transcriptional Profiles that Regulate Proviral Latency. Cell Rep *25*, 107-117 e103.

Bruner, K.M., Murray, A.J., Pollack, R.A., Soliman, M.G., Laskey, S.B., Capoferri, A.A., Lai, J., Strain, M.C., Lada, S.M., Hoh, R.*, et al.* (2016). Defective proviruses rapidly accumulate during acute HIV-1 infection. Nat Med *22*, 1043-1049.

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

Bruner, K.S., Robert (2018). Compositions and methods related to characterizing proviral reservoirs. United States Patent Application US15/568,893, filed April 22, 2016.

Cachot, A., Bilous, M., Liu, Y.C., Li, X., Saillard, M., Cenerenti, M., Rockinger, G.A., Wyss, T., Guillaume, P., Schmidt, J.*, et al.* (2021). Tumor-specific cytolytic CD4 T cells mediate immunity against human cancer. Sci Adv *7*.

Cameron, P.U., Saleh, S., Sallmann, G., Solomon, A., Wightman, F., Evans, V.A., Boucher, G., Haddad, E.K., Sekaly, R.P., Harman, A.N.*, et al.* (2010). Establishment of HIV-1 latency in resting CD4+ T cells depends on chemokine-induced changes in the actin cytoskeleton. Proc Natl Acad Sci U S A *107*, 16934-16939.

CDC (1981). Pneumocystis Pneumonia --- Los Angeles. In Epidemiologic Notes and Reports (Center of Disease Control and Prevention).

Chaillon, A., Gianella, S., Dellicour, S., Rawlings, S.A., Schlub, T.E., De Oliveira, M.F., Ignacio, C., Porrachia, M., Vrancken, B., and Smith, D.M. (2020). HIV persists throughout deep tissues with repopulation from multiple anatomical sources. J Clin Invest *130*, 1699-1712.

Chang, J.T., Palanivel, V.R., Kinjyo, I., Schambach, F., Intlekofer, A.M., Banerjee, A., Longworth, S.A., Vinup, K.E., Mrass, P., Oliaro, J.*, et al.* (2007). Asymmetric T lymphocyte division in the initiation of adaptive immune responses. Science *315*, 1687-1691.

Cho, A., Gaebler, C., Olveira, T., Ramos, V., Saad, M., Lorenzi, J.C.C., Gazumyan, A., Moir, S., Caskey, M., Chun, T.W.*, et al.* (2022). Longitudinal clonal dynamics of HIV-1 latent reservoirs measured by combination quadruplex polymerase chain reaction and sequencing. Proc Natl Acad Sci U S A *119*.

Chomont, N., El-Far, M., Ancuta, P., Trautmann, L., Procopio, F.A., Yassine-Diab, B., Boucher, G., Boulassel, M.R., Ghattas, G., Brenchley, J.M.*, et al.* (2009). HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. Nat Med *15*, 893-900.

Chtanova, T., Newton, R., Liu, S.M., Weininger, L., Young, T.R., Silva, D.G., Bertoni, F., Rinaldi, A., Chappaz, S., Sallusto, F.*, et al.* (2005). Identification of T cell-restricted genes, and signatures for different T cell responses, using a comprehensive collection of microarray datasets. J Immunol *175*, 7837-7847.

Chun, T.W., Engel, D., Berrey, M.M., Shea, T., Corey, L., and Fauci, A.S. (1998). Early establishment of a pool of latently infected, resting CD4(+) T cells during primary HIV-1 infection. Proc Natl Acad Sci U S A *95*, 8869-8873.

Chun, T.W., Nickle, D.C., Justement, J.S., Meyers, J.H., Roby, G., Hallahan, C.W., Kottilil, S., Moir, S., Mican, J.M., Mullins, J.I.*, et al.* (2008). Persistence of HIV in gut-associated lymphoid tissue despite long-term antiretroviral therapy. J Infect Dis *197*, 714-720.

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

Cocchi, F., DeVico, A.L., Garzino-Demo, A., Arya, S.K., Gallo, R.C., and Lusso, P. (1995). Identification of RANTES, MIP-1 alpha, and MIP-1 beta as the major HIV-suppressive factors produced by CD8+ T cells. Science *270*, 1811-1815.

Coffin, J.M. (1995). HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. Science *267*, 483-489.

Cohen, M.S., Chen, Y.Q., McCauley, M., Gamble, T., Hosseinipour, M.C., Kumarasamy, N., Hakim, J.G., Kumwenda, J., Grinsztejn, B., Pilotto, J.H.*, et al.* (2016). Antiretroviral Therapy for the Prevention of HIV-1 Transmission. N Engl J Med *375*, 830-839.

Cohn, L.B., Chomont, N., and Deeks, S.G. (2020). The Biology of the HIV-1 Latent Reservoir and Implications for Cure Strategies. Cell Host Microbe *27*, 519-530.

Cohn, L.B., da Silva, I.T., Valieris, R., Huang, A.S., Lorenzi, J.C.C., Cohen, Y.Z., Pai, J.A., Butler, A.L., Caskey, M., Jankovic, M.*, et al.* (2018). Clonal CD4(+) T cells in the HIV-1 latent reservoir display a distinct gene profile upon reactivation. Nat Med *24*, 604-609.

Cohn, L.B., Silva, I.T., Oliveira, T.Y., Rosales, R.A., Parrish, E.H., Learn, G.H., Hahn, B.H., Czartoski, J.L., McElrath, M.J., Lehmann, C.*, et al.* (2015). HIV-1 integration landscape during latent and active infection. Cell *160*, 420-432.

Colby, D.J., Trautmann, L., Pinyakorn, S., Leyre, L., Pagliuzza, A., Kroon, E., Rolland, M., Takata, H., Buranapraditkun, S., Intasan, J.*, et al.* (2018). Rapid HIV RNA rebound after antiretroviral treatment interruption in persons durably suppressed in Fiebig I acute HIV infection. Nat Med *24*, 923-926.

Collora, J.A., Liu, R., Pinto-Santini, D., Ravindra, N., Ganoza, C., Lama, J.R., Alfaro, R., Chiarella, J., Spudich, S., Mounzer, K.*, et al.* (2022). Single-cell multiomics reveals persistence of HIV-1 in expanded cytotoxic T cell clones. Immunity.

Copelan, E.A. (2006). Hematopoietic stem-cell transplantation. N Engl J Med *354*, 1813-1826.

Craigie, R., and Bushman, F.D. (2012). HIV DNA integration. Cold Spring Harb Perspect Med *2*, a006890.

Crooks, A.M., Bateson, R., Cope, A.B., Dahl, N.P., Griggs, M.K., Kuruc, J.D., Gay, C.L., Eron, J.J., Margolis, D.M., Bosch, R.J.*, et al.* (2015). Precise Quantitation of the Latent HIV-1 Reservoir: Implications for Eradication Strategies. J Infect Dis *212*, 1361-1365.

Darcis, G., Berkhout, B., and Pasternak, A.O. (2019). The Quest for Cellular Markers of HIV Reservoirs: Any Color You Like. Front Immunol *10*, 2251.

Das, A.T., Pasternak, A.O., and Berkhout, B. (2019). On the generation of the MSD- class of defective HIV proviruses. Retrovirology *16*, 19.

Dash, P.K., Kaminski, R., Bella, R., Su, H., Mathews, S., Ahooyi, T.M., Chen, C., Mancuso, P., Sariyer, R., Ferrante, P.*, et al.* (2019). Sequential LASER ART and CRISPR Treatments Eliminate HIV-1 in a Subset of Infected Humanized Mice. Nat Commun *10*, 2753.

Davey, R.T., Jr., Bhat, N., Yoder, C., Chun, T.W., Metcalf, J.A., Dewar, R., Natarajan, V., Lempicki, R.A., Adelsberger, J.W., Miller, K.D.*, et al.* (1999). HIV-1 and T cell dynamics after interruption of highly active antiretroviral therapy (HAART) in patients with a history of sustained viral suppression. Proc Natl Acad Sci U S A *96*, 15109-15114.

Deeks, S.G., Overbaugh, J., Phillips, A., and Buchbinder, S. (2015). HIV infection. Nat Rev Dis Primers *1*, 15035.

Derdeyn, C.A., Decker, J.M., Bibollet-Ruche, F., Mokili, J.L., Muldoon, M., Denham, S.A., Heil, M.L., Kasolo, F., Musonda, R., Hahn, B.H.*, et al.* (2004). Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. Science *303*, 2019-2022.

Doitsh, G., and Greene, W.C. (2016). Dissecting How CD4 T Cells Are Lost During HIV Infection. Cell Host Microbe *19*, 280-291.

Douek, D.C., Brenchley, J.M., Betts, M.R., Ambrozak, D.R., Hill, B.J., Okamoto, Y., Casazza, J.P., Kuruppu, J., Kunstman, K., Wolinsky, S.*, et al.* (2002). HIV preferentially infects HIV-specific CD4+ T cells. Nature *417*, 95-98.

Dragic, T., Litwin, V., Allaway, G.P., Martin, S.R., Huang, Y., Nagashima, K.A., Cayanan, C., Maddon, P.J., Koup, R.A., Moore, J.P.*, et al.* (1996). HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. Nature *381*, 667-673.

Duette, G., Hiener, B., Morgan, H., Mazur, F.G., Mathivanan, V., Horsburgh, B.A., Fisher, K., Tong, O., Lee, E., Ahn, H.*, et al.* (2022). The HIV-1 proviral landscape reveals Nef contributes to HIV-1 persistence in effector memory CD4+ T-cells. J Clin Invest.

Dufour, C., Gantner, P., Fromentin, R., and Chomont, N. (2021). The multifaceted nature of HIV latency. J Clin Invest *131*.

Dupic, T., Marcou, Q., Walczak, A.M., and Mora, T. (2019). Genesis of the alphabeta T-cell receptor. PLoS Comput Biol *15*, e1006874.

Einkauf, K.B., Lee, G.Q., Gao, C., Sharaf, R., Sun, X., Hua, S., Chen, S.M., Jiang, C., Lian, X., Chowdhury, F.Z.*, et al.* (2019). Intact HIV-1 proviruses accumulate at distinct chromosomal positions during prolonged antiretroviral therapy. J Clin Invest *129*, 988-998.

Einkauf, K.B., Osborn, M.R., Gao, C., Sun, W., Sun, X., Lian, X., Parsons, E.M., Gladkov, G.T., Seiger, K.W., Blackmer, J.E.*, et al.* (2022). Parallel analysis of transcription, integration, and sequence of single HIV-1 proviruses. Cell *185*, 266-282 e215.

Engelman, A., and Cherepanov, P. (2012). The structural biology of HIV-1: mechanistic and therapeutic insights. Nature Reviews Microbiology *10*, 279-290.

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

Estes, J.D., Kityo, C., Ssali, F., Swainson, L., Makamdop, K.N., Del Prete, G.Q., Deeks, S.G., Luciw, P.A., Chipman, J.G., Beilman, G.J.*, et al.* (2017). Defining total-body AIDS-virus burden with implications for curative strategies. Nat Med *23*, 1271-1276.

Falcinelli, S.D., Shook-Sa, B.E., Dewey, M.G., Sridhar, S., Read, J., Kirchherr, J., James, K.S., Allard, B., Ghofrani, S., Stuelke, E.*, et al.* (2020). Impact of Biological Sex on Immune Activation and Frequency of the Latent HIV Reservoir During Suppressive Antiretroviral Therapy. J Infect Dis *222*, 1843-1852.

Farber, D.L., Yudanin, N.A., and Restifo, N.P. (2014). Human memory T cells: generation, compartmentalization and homeostasis. Nat Rev Immunol *14*, 24-35.

Finzi, D., Blankson, J., Siliciano, J.D., Margolick, J.B., Chadwick, K., Pierson, T., Smith, K., Lisziewicz, J., Lori, F., Flexner, C.*, et al.* (1999). Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. Nat Med *5*, 512-517.

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

Fortin, J.S., Cloutier, M., and Thibodeau, J. (2013). Exposing the Specific Roles of the Invariant Chain Isoforms in Shaping the MHC Class II Peptidome. Front Immunol *4*, 443.

Fromentin, R., Bakeman, W., Lawani, M.B., Khoury, G., Hartogensis, W., DaFonseca, S., Killian, M., Epling, L., Hoh, R., Sinclair, E.*, et al.* (2016). CD4+ T Cells Expressing PD-1, TIGIT and LAG-3 Contribute to HIV Persistence during ART. PLoS Pathog *12*, e1005761.

Gaebler, C., Falcinelli, S.D., Stoffel, E., Read, J., Murtagh, R., Oliveira, T.Y., Ramos, V., Lorenzi, J.C.C., Kirchherr, J., James, K.S.*, et al.* (2021). Sequence Evaluation and Comparative Analysis of Novel Assays for Intact Proviral HIV-1 DNA. J Virol *95*.

Gaebler, C., Lorenzi, J.C.C., Oliveira, T.Y., Nogueira, L., Ramos, V., Lu, C.L., Pai, J.A., Mendoza, P., Jankovic, M., Caskey, M.*, et al.* (2019). Combination of quadruplex qPCR and next-generation sequencing for qualitative and quantitative analysis of the HIV-1 latent reservoir. J Exp Med *216*, 2253-2264.

Gaebler, C., Nogueira, L., Stoffel, E., Oliveira, T.Y., Breton, G., Millard, K.G., Turroja, M., Butler, A., Ramos, V., Seaman, M.S.*, et al.* (2022). Prolonged viral suppression with anti-HIV-1 antibody therapy. Nature.

Ganai, R.A., and Johansson, E. (2016). DNA Replication-A Matter of Fidelity. Mol Cell *62*, 745-755.

Ganusov, V.V., Goonetilleke, N., Liu, M.K., Ferrari, G., Shaw, G.M., McMichael, A.J., Borrow, P., Korber, B.T., and Perelson, A.S. (2011). Fitness costs and diversity of the cytotoxic T lymphocyte (CTL) response determine the rate of CTL escape during acute and chronic phases of HIV infection. J Virol *85*, 10518-10528.

Geginat, J., Sallusto, F., and Lanzavecchia, A. (2001). Cytokine-driven proliferation and differentiation of human naive, central memory, and effector memory CD4(+) T cells. J Exp Med *194*, 1711-1719.

Goonetilleke, N., Liu, M.K., Salazar-Gonzalez, J.F., Ferrari, G., Giorgi, E., Ganusov, V.V., Keele, B.F., Learn, G.H., Turnbull, E.L., Salazar, M.G.*, et al.* (2009). The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. J Exp Med *206*, 1253-1272.

Grau-Exposito, J., Luque-Ballesteros, L., Navarro, J., Curran, A., Burgos, J., Ribera, E., Torrella, A., Planas, B., Badia, R., Martin-Castillo, M.*, et al.* (2019). Latency reversal agents affect differently the latent reservoir present in distinct CD4+ T subpopulations. PLoS Pathog *15*, e1007991.

Gray, J.I., Westerhof, L.M., and MacLeod, M.K.L. (2018). The roles of resident, central and effector memory CD4 T-cells in protective immunity following infection or vaccination. Immunology.

Gupta, R.K., Abdul-Jawad, S., McCoy, L.E., Mok, H.P., Peppa, D., Salgado, M., Martinez-Picado, J., Nijhuis, M., Wensing, A.M.J., Lee, H.*, et al.* (2019). HIV-1 remission following CCR5Delta32/Delta32 haematopoietic stem-cell transplantation. Nature *568*, 244-248.

Hale, J.S., and Ahmed, R. (2015). Memory T follicular helper CD4 T cells. Front Immunol *6*, 16.

Han, A., Glanville, J., Hansmann, L., and Davis, M.M. (2014). Linking T-cell receptor sequence to functional phenotype at the single-cell level. Nat Biotechnol *32*, 684-692.

Han, Y., Lassen, K., Monie, D., Sedaghat, A.R., Shimoji, S., Liu, X., Pierson, T.C., Margolick, J.B., Siliciano, R.F., and Siliciano, J.D. (2004). Resting CD4+ T cells from human immunodeficiency virus type 1 (HIV-1)-infected individuals carry integrated HIV-1 genomes within actively transcribed host genes. J Virol *78*, 6122-6133.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M.*, et al.* (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573-3587 e3529.

Hashimoto, K., Kouno, T., Ikawa, T., Hayatsu, N., Miyajima, Y., Yabukami, H., Terooatea, T., Sasaki, T., Suzuki, T., Valentine, M.*, et al.* (2019). Single-cell transcriptomics reveals expansion of cytotoxic CD4 T cells in supercentenarians. Proc Natl Acad Sci U S A *116*, 24242-24251.

Hatano, H., Jain, V., Hunt, P.W., Lee, T.H., Sinclair, E., Do, T.D., Hoh, R., Martin, J.N., McCune, J.M., Hecht, F.*, et al.* (2013). Cell-based measures of viral persistence are associated with immune activation and programmed cell death protein 1 (PD-1)-expressing CD4+ T cells. J Infect Dis *208*, 50-56.

Hataye, J.M., Casazza, J.P., Best, K., Liang, C.J., Immonen, T.T., Ambrozak, D.R., Darko, S., Henry, A.R., Laboune, F., Maldarelli, F.*, et al.* (2019). Principles Governing Establishment versus Collapse of HIV-1 Cellular Spread. Cell Host Microbe *26*, 748-763 e720.

Henrich, T.J., Hobbs, K.S., Hanhauser, E., Scully, E., Hogan, L.E., Robles, Y.P., Leadabrand, K.S., Marty, F.M., Palmer, C.D., Jost, S.*, et al.* (2017). Human Immunodeficiency Virus Type 1 Persistence Following Systemic Chemotherapy for Malignancy. J Infect Dis *216*, 254-262.

Hermankova, M., Siliciano, J.D., Zhou, Y., Monie, D., Chadwick, K., Margolick, J.B., Quinn, T.C., and Siliciano, R.F. (2003). Analysis of human immunodeficiency virus type 1 gene expression in latently infected resting CD4+ T lymphocytes in vivo. J Virol *77*, 7383-7392.

Hiener, B., Horsburgh, B.A., Eden, J.S., Barton, K., Schlub, T.E., Lee, E., von Stockenstrom, S., Odevall, L., Milush, J.M., Liegler, T.*, et al.* (2017). Identification of Genetically Intact HIV-1 Proviruses in Specific CD4(+) T Cells from Effectively Treated Participants. Cell Rep *21*, 813-822.

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

Horsburgh, B.A., Lee, E., Hiener, B., Eden, J.S., Schlub, T.E., von Stockenstrom, S., Odevall, L., Milush, J.M., Liegler, T., Sinclair, E.*, et al.* (2020). High levels of genetically intact HIV in HLA-DR+ memory T cells indicates their value for reservoir studies. AIDS *34*, 659-668.

Hosmane, N.N., Kwon, K.J., Bruner, K.M., Capoferri, A.A., Beg, S., Rosenbloom, D.I., Keele, B.F., Ho, Y.C., Siliciano, J.D., and Siliciano, R.F. (2017). Proliferation of latently infected CD4(+) T cells carrying replication-competent HIV-1: Potential role in latent reservoir dynamics. J Exp Med *214*, 959-972.

Hu, W.S., and Hughes, S.H. (2012). HIV-1 reverse transcription. Cold Spring Harb Perspect Med *2*.

Huang, A.S., Ramos, V., Oliveira, T.Y., Gaebler, C., Jankovic, M., Nussenzweig, M.C., and Cohn, L.B. (2021). Integration features of intact latent HIV-1 in CD4+ T cell clones contribute to viral persistence. J Exp Med *218*.

Huang, S.H., Ren, Y., Thomas, A.S., Chan, D., Mueller, S., Ward, A.R., Patel, S., Bollard, C.M., Cruz, C.R., Karandish, S.*, et al.* (2018). Latent HIV reservoirs exhibit inherent resistance to elimination by CD8+ T cells. J Clin Invest *128*, 876-889.

Hutter, G., Nowak, D., Mossner, M., Ganepola, S., Mussig, A., Allers, K., Schneider, T., Hofmann, J., Kucherer, C., Blau, O.*, et al.* (2009). Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. N Engl J Med *360*, 692-698.

Iglesias-Ussel, M., Vandergeeten, C., Marchionni, L., Chomont, N., and Romerio, F. (2013). High levels of CD2 expression identify HIV-1 latently infected resting memory CD4+ T cells in virally suppressed subjects. J Virol *87*, 9148-9158.

Ikeda, T., Shibata, J., Yoshimura, K., Koito, A., and Matsushita, S. (2007). Recurrent HIV-1 integration at the BACH2 locus in resting CD4+ T cell populations during effective highly active antiretroviral therapy. J Infect Dis *195*, 716-725.

Imamichi, H., Dewar, R.L., Adelsberger, J.W., Rehm, C.A., O'Doherty, U., Paxinos, E.E., Fauci, A.S., and Lane, H.C. (2016). Defective HIV-1 proviruses produce novel protein-coding RNA species in HIV-infected patients on combination antiretroviral therapy. Proc Natl Acad Sci U S A *113*, 8783-8788.

Imamichi, H., Natarajan, V., Adelsberger, J.W., Rehm, C.A., Lempicki, R.A., Das, B., Hazen, A., Imamichi, T., and Lane, H.C. (2014). Lifespan of effector memory CD4+ T cells determined by replication-incompetent integrated HIV-1 provirus. AIDS *28*, 1091-1099.

Imamichi, H., Smith, M., Adelsberger, J.W., Izumi, T., Scrimieri, F., Sherman, B.T., Rehm, C.A., Imamichi, T., Pau, A., Catalfamo, M.*, et al.* (2020). Defective HIV-1 proviruses produce viral proteins. Proc Natl Acad Sci U S A *117*, 3704-3710.

IMGT (2022). https://www.imgt.org/IMGTrepertoire/LocusGenes/genetable/human/geneNumber.html#functional.

Jiang, C., Lian, X., Gao, C., Sun, X., Einkauf, K.B., Chevalier, J.M., Chen, S.M.Y., Hua, S., Rhee, B., Chang, K.*, et al.* (2020). Distinct viral reservoirs in individuals with spontaneous control of HIV-1. Nature *585*, 261-267.

Johnson, S., Eller, M., Teigler, J.E., Maloveste, S.M., Schultz, B.T., Soghoian, D.Z., Lu, R., Oster, A.F., Chenine, A.L., Alter, G.*, et al.* (2015). Cooperativity of HIV-Specific Cytolytic CD4 T Cells and CD8 T Cells in Control of HIV Viremia. J Virol *89*, 7494-7505.

Johnson, W.E., and Desrosiers, R.C. (2002). Viral persistence: HIV's strategies of immune system evasion. Annu Rev Med *53*, 499-518.

Jones, R.B., and Walker, B.D. (2016). HIV-specific CD8(+) T cells and HIV eradication. J Clin Invest *126*, 455-463.

Juno, J.A., van Bockel, D., Kent, S.J., Kelleher, A.D., Zaunders, J.J., and Munier, C.M. (2017). Cytotoxic CD4 T Cells-Friend or Foe during Viral Infection? Front Immunol *8*, 19.

Kaczmarek Michaels, K., Natarajan, M., Euler, Z., Alter, G., Viglianti, G., and Henderson, A.J. (2015). Blimp-1, an intrinsic factor that represses HIV-1 proviral transcription in memory CD4+ T cells. J Immunol *194*, 3267-3274.

Kaminski, R., Bella, R., Yin, C., Otte, J., Ferrante, P., Gendelman, H.E., Li, H., Booze, R., Gordon, J., Hu, W.*, et al.* (2016). Excision of HIV-1 DNA by gene editing: a proof-of-concept in vivo study. Gene Ther *23*, 696.

Kandathil, A.J., Sugawara, S., and Balagopal, A. (2016). Are T cells the only HIV-1 reservoir? Retrovirology *13*, 86.

Karim, S.S.A., and Baxter, C. (2019). HIV incidence rates in adolescent girls and young women in sub-Saharan Africa. Lancet Glob Health *7*, e1470-e1471.

Karn, J., and Stoltzfus, C.M. (2012). Transcriptional and posttranscriptional regulation of HIV-1 gene expression. Cold Spring Harb Perspect Med *2*, a006916.

Kearney, M.F., Wiegand, A., Shao, W., Coffin, J.M., Mellors, J.W., Lederman, M., Gandhi, R.T., Keele, B.F., and Li, J.Z. (2016). Origin of Rebound Plasma HIV Includes Cells with Identical Proviruses That Are Transcriptionally Active before Stopping of Antiretroviral Therapy. J Virol *90*, 1369-1376.

Keele, B.F., Giorgi, E.E., Salazar-Gonzalez, J.F., Decker, J.M., Pham, K.T., Salazar, M.G., Sun, C., Grayson, T., Wang, S., Li, H.*, et al.* (2008). Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci U S A *105*, 7552-7557.

Kim, J., Vasan, S., Kim, J.H., and Ake, J.A. (2021). Current approaches to HIV vaccine development: a narrative review. J Int AIDS Soc *24 Suppl 7*, e25793.

Kondrack, R.M., Harbertson, J., Tan, J.T., McBreen, M.E., Surh, C.D., and Bradley, L.M. (2003). Interleukin 7 regulates the survival and generation of memory CD4 cells. J Exp Med *198*, 1797-1806.

Krishnan, V., and Zeichner, S.L. (2004). Host cell gene expression during human immunodeficiency virus type 1 latency and reactivation and effects of targeting genes that are differentially expressed in viral latency. J Virol *78*, 9458-9473.

Krishnaratne, S., Hensen, B., Cordes, J., Enstone, J., and Hargreaves, J.R. (2016). Interventions to strengthen the HIV prevention cascade: a systematic review of reviews. Lancet HIV *3*, e307-317.

Kwong, P.D., Wyatt, R., Robinson, J., Sweet, R.W., Sodroski, J., and Hendrickson, W.A. (1998). Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature *393*, 648-659.

Laird, G.M., Eisele, E.E., Rabi, S.A., Lai, J., Chioma, S., Blankson, J.N., Siliciano, J.D., and Siliciano, R.F. (2013). Rapid quantification of the latent reservoir for HIV-1 using a viral outgrowth assay. PLoS Pathog *9*, e1003398.

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

Lee, E., Bacchetti, P., Milush, J., Shao, W., Boritz, E., Douek, D., Fromentin, R., Liegler, T., Hoh, R., Deeks, S.G.*, et al.* (2019). Memory CD4 + T-Cells Expressing HLA-DR Contribute to HIV Persistence During Prolonged Antiretroviral Therapy. Front Microbiol *10*, 2214.

Lee, G.Q., Orlova-Fink, N., Einkauf, K., Chowdhury, F.Z., Sun, X., Harrington, S., Kuo, H.H., Hua, S., Chen, H.R., Ouyang, Z.*, et al.* (2017). Clonal expansion of genome-intact HIV-1 in functionally polarized Th1 CD4+ T cells. J Clin Invest *127*, 2689-2696.

Li, B., Gladden, A.D., Altfeld, M., Kaldor, J.M., Cooper, D.A., Kelleher, A.D., and Allen, T.M. (2007). Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution. J Virol *81*, 193-201.

Li, J.Z., Etemad, B., Ahmed, H., Aga, E., Bosch, R.J., Mellors, J.W., Kuritzkes, D.R., Lederman, M.M., Para, M., and Gandhi, R.T. (2016). The size of the expressed HIV reservoir predicts timing of viral rebound after treatment interruption. AIDS *30*, 343-353.

Liu, R., Yeh, Y.J., Varabyou, A., Collora, J.A., Sherrill-Mix, S., Talbot, C.C., Jr., Mehta, S., Albrecht, K., Hao, H., Zhang, H.*, et al.* (2020). Single-cell transcriptional landscapes reveal HIV-1-driven aberrant host gene transcription as a potential therapeutic target. Sci Transl Med *12*.

Lorenzi, J.C., Cohen, Y.Z., Cohn, L.B., Kreider, E.F., Barton, J.P., Learn, G.H., Oliveira, T., Lavine, C.L., Horwitz, J.A., Settler, A.*, et al.* (2016). Paired quantitative and qualitative assessment of the replication-competent HIV-1 reservoir and comparison with integrated proviral DNA. Proc Natl Acad Sci U S A *113*, E7908-E7916.

Macallan, D.C., Wallace, D., Zhang, Y., De Lara, C., Worth, A.T., Ghattas, H., Griffin, G.E., Beverley, P.C., and Tough, D.F. (2004). Rapid turnover of effector-memory CD4(+) T cells in healthy humans. J Exp Med *200*, 255-260.

MacLeod, M.K., Clambey, E.T., Kappler, J.W., and Marrack, P. (2009). CD4 memory T cells: what are they and what can they do? Semin Immunol *21*, 53-61.

MacLeod, M.K., Kappler, J.W., and Marrack, P. (2010). Memory CD4 T cells: generation, reactivation and re-assignment. Immunology *130*, 10-15.

Maddon, P.J., Dalgleish, A.G., McDougal, J.S., Clapham, P.R., Weiss, R.A., and Axel, R. (1986). The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. Cell *47*, 333-348.

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

Malim, M.H., and Bieniasz, P.D. (2012). HIV Restriction Factors and Mechanisms of Evasion. Cold Spring Harb Perspect Med *2*, a006940.

Mamedov, I.Z., Britanova, O.V., Zvyagin, I.V., Turchaninova, M.A., Bolotin, D.A., Putintseva, E.V., Lebedev, Y.B., and Chudakov, D.M. (2013). Preparing unbiased T-cell receptor and antibody cDNA libraries for the deep next generation sequencing profiling. Front Immunol *4*, 456.

Marini, A., Harper, J.M., and Romerio, F. (2008). An in vitro system to model the establishment and reactivation of HIV-1 latency. J Immunol *181*, 7713-7720.

Markowitz, M., Louie, M., Hurley, A., Sun, E., Di Mascio, M., Perelson, A.S., and Ho, D.D. (2003). A novel antiviral intervention results in more accurate assessment of human immunodeficiency virus type 1 replication dynamics and T-cell decay in vivo. J Virol *77*, 5037-5038.

Marley, G., Tan, R.K.J., and Tang, W. (2022). Stem cell research finds possible HIV cure with cord blood transplant. Innovation (Camb) *3*, 100238.

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

Massanella, M., and Richman, D.D. (2016). Measuring the latent reservoir in vivo. J Clin Invest *126*, 464-472.

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

McDougal, J.S., Maddon, P.J., Dalgleish, A.G., Clapham, P.R., Littman, D.R., Godfrey, M., Maddon, D.E., Chess, L., Weiss, R.A., and Axel, R. (1986). The T4 glycoprotein is a cell-surface receptor for the AIDS virus. Cold Spring Harb Symp Quant Biol *51 Pt 2*, 703-711.

McManus, W.R., Bale, M.J., Spindler, J., Wiegand, A., Musick, A., Patro, S.C., Sobolewski, M.D., Musick, V.K., Anderson, E.M., Cyktor, J.C*., et al.* (2019). HIV-1 in lymph nodes is maintained by cellular proliferation during antiretroviral therapy. J Clin Invest *129*, 4629-4642.

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

Moore, M.D., and Hu, W.S. (2009). HIV-1 RNA dimerization: It takes two to tango. AIDS Rev *11*, 91-102.

Moran, J.A., Turner, S.R., and Marsden, M.D. (2022). Contribution of Sex Differences to HIV Immunology, Pathogenesis, and Cure Approaches. Front Immunol *13*, 905773.

Morcilla, V., Bacchus-Souffan, C., Fisher, K., Horsburgh, B.A., Hiener, B., Wang, X.Q., Schlub, T.E., Fitch, M., Hoh, R., Hecht, F.M.*, et al.* (2021). HIV-1 Genomes Are Enriched in Memory CD4(+) T-Cells with Short Half-Lives. mBio *12*, e0244721.

Murray, A.J., Kwon, K.J., Farber, D.L., and Siliciano, R.F. (2016). The Latent Reservoir for HIV-1: How Immunologic Memory and Clonal Expansion Contribute to HIV-1 Persistence. J Immunol *197*, 407-417.

Musick, A., Spindler, J., Boritz, E., Perez, L., Crespo-Velez, D., Patro, S.C., Sobolewski, M.D., Bale, M.J., Reid, C., Keele, B.F.*, et al.* (2019). HIV Infected T Cells Can Proliferate in vivo Without Inducing Expression of the Integrated Provirus. Front Microbiol *10*, 2204.

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

Nguyen, Q.P., Deng, T.Z., Witherden, D.A., and Goldrath, A.W. (2019). Origins of CD4(+) circulating and tissue-resident memory T-cells. Immunology *157*, 3-12.

Nikolich-Zugich, J., Slifka, M.K., and Messaoudi, I. (2004). The many important facets of T-cell repertoire diversity. Nat Rev Immunol *4*, 123-132.

Nussenzweig, M.C., Shaw, A.C., Sinn, E., Danner, D.B., Holmes, K.L., Morse, H.C., 3rd, and Leder, P. (1987). Allelic exclusion in transgenic mice that express the membrane form of immunoglobulin mu. Science *236*, 816-819.

O'Brien, J., Hayder, H., Zayed, Y., and Peng, C. (2018). Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. Front Endocrinol (Lausanne) *9*, 402.

Oh, D.Y., and Fong, L. (2021). Cytotoxic CD4(+) T cells in cancer: Expanding the immune effector toolbox. Immunity *54*, 2701-2711.

Palmer, S., Wiegand, A.P., Maldarelli, F., Bazmi, H., Mican, J.M., Polis, M., Dewar, R.L., Planta, A., Liu, S., Metcalf, J.A.*, et al.* (2003). New real-time reverse transcriptase-initiated PCR assay with single-copy sensitivity for human immunodeficiency virus type 1 RNA in plasma. J Clin Microbiol *41*, 4531-4536.

Pan, X., Baldauf, H.M., Keppler, O.T., and Fackler, O.T. (2013). Restrictions to HIV-1 replication in resting CD4+ T lymphocytes. Cell Res *23*, 876-885.

Pardons, M., Baxter, A.E., Massanella, M., Pagliuzza, A., Fromentin, R., Dufour, C., Leyre, L., Routy, J.P., Kaufmann, D.E., and Chomont, N. (2019a). Single-cell characterization and quantification of translation-competent viral reservoirs in treated and untreated HIV infection. PLoS Pathog *15*, e1007619.

Pardons, M., Fromentin, R., Pagliuzza, A., Routy, J.P., and Chomont, N. (2019b). Latency-Reversing Agents Induce Differential Responses in Distinct Memory CD4 T Cell Subsets in Individuals on Antiretroviral Therapy. Cell Rep *29*, 2783-2795 e2785.

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

Perisic Nanut, M., Sabotic, J., Jewett, A., and Kos, J. (2014). Cysteine cathepsins as regulators of the cytotoxicity of NK and T cells. Front Immunol *5*, 616.

Perisic Nanut, M., Sabotic, J., Svajger, U., Jewett, A., and Kos, J. (2017). Cystatin F Affects Natural Killer Cell Cytotoxicity. Front Immunol *8*, 1459.

Pinzone, M.R., VanBelzen, D.J., Weissman, S., Bertuccio, M.P., Cannon, L., Venanzi-Rullo, E., Migueles, S., Jones, R.B., Mota, T., Joseph, S.B.*, et al.* (2019). Longitudinal HIV sequencing reveals reservoir expression leading to decay which is obscured by clonal expansion. Nat Commun *10*, 728.

Pollack, R.A., Jones, R.B., Pertea, M., Bruner, K.M., Martin, A.R., Thomas, A.S., Capoferri, A.A., Beg, S.A., Huang, S.H., Karandish, S.*, et al.* (2017). Defective HIV-1 Proviruses Are Expressed and Can Be Recognized by Cytotoxic T Lymphocytes, which Shape the Proviral Landscape. Cell Host Microbe *21*, 494-506 e494.

Price, D.A., Goulder, P.J., Klenerman, P., Sewell, A.K., Easterbrook, P.J., Troop, M., Bangham, C.R., and Phillips, R.E. (1997). Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection. Proc Natl Acad Sci U S A *94*, 1890-1895.

Prunk, M., Perisic Nanut, M., Jakos, T., Sabotic, J., Svajger, U., and Kos, J. (2020). Extracellular Cystatin F Is Internalised by Cytotoxic T Lymphocytes and Decreases Their Cytotoxicity. Cancers (Basel) *12*.

Qi, Q., Liu, Y., Cheng, Y., Glanville, J., Zhang, D., Lee, J.Y., Olshen, R.A., Weyand, C.M., Boyd, S.D., and Goronzy, J.J. (2014). Diversity and clonal selection in the human T-cell repertoire. Proc Natl Acad Sci U S A *111*, 13139-13144.

Ren, Y., Huang, S.H., Patel, S., Alberto, W.D.C., Magat, D., Ahimovic, D., Macedo, A.B., Durga, R., Chan, D., Zale, E.*, et al.* (2020). BCL-2 antagonism sensitizes cytotoxic T cell-resistant HIV reservoirs to elimination ex vivo. J Clin Invest *130*, 2542-2559.

Riou, C., Yassine-Diab, B., Van grevenynghe, J., Somogyi, R., Greller, L.D., Gagnon, D., Gimmig, S., Wilkinson, P., Shi, Y., Cameron, M.J.*, et al.* (2007). Convergence of TCR and cytokine signaling leads to FOXO3a phosphorylation and drives the survival of CD4+ central memory T cells. J Exp Med *204*, 79-91.

Rizzuto, C.D., Wyatt, R., Hernandez-Ramos, N., Sun, Y., Kwong, P.D., Hendrickson, W.A., and Sodroski, J. (1998). A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding. Science *280*, 1949-1953.

Rodger, A.J., Cambiano, V., Bruun, T., Vernazza, P., Collins, S., Degen, O., Corbelli, G.M., Estrada, V., Geretti, A.M., Beloukas, A.*, et al.* (2019). Risk of HIV transmission through condomless sex in serodifferent gay couples with the HIV-positive partner taking suppressive antiretroviral therapy (PARTNER): final results of a multicentre, prospective, observational study. Lancet *393*, 2428-2438.

Ruelas, D.S., Chan, J.K., Oh, E., Heidersbach, A.J., Hebbeler, A.M., Chavez, L., Verdin, E., Rape, M., and Greene, W.C. (2015). MicroRNA-155 Reinforces HIV Latency. J Biol Chem *290*, 13736-13748.

Ruterbusch, M., Pruner, K.B., Shehata, L., and Pepper, M. (2020). In Vivo CD4(+) T Cell Differentiation and Function: Revisiting the Th1/Th2 Paradigm. Annu Rev Immunol *38*, 705-725.

Saag, M.S., Gandhi, R.T., Hoy, J.F., Landovitz, R.J., Thompson, M.A., Sax, P.E., Smith, D.M., Benson, C.A., Buchbinder, S.P., Del Rio, C.*, et al.* (2020). Antiretroviral Drugs for Treatment and Prevention of HIV Infection in Adults: 2020 Recommendations of the International Antiviral Society-USA Panel. JAMA *324*, 1651-1669.

Sagar, M., Laeyendecker, O., Lee, S., Gamiel, J., Wawer, M.J., Gray, R.H., Serwadda, D., Sewankambo, N.K., Shepherd, J.C., Toma, J.*, et al.* (2009). Selection of HIV variants with signature genotypic characteristics during heterosexual transmission. J Infect Dis *199*, 580-589.

Sagar, M., Lavreys, L., Baeten, J.M., Richardson, B.A., Mandaliya, K., Ndinya-Achola, J.O., Kreiss, J.K., and Overbaugh, J. (2004). Identification of modifiable factors that affect the genetic diversity of the transmitted HIV-1 population. AIDS *18*, 615-619.

Salazar-Gonzalez, J.F., Bailes, E., Pham, K.T., Salazar, M.G., Guffey, M.B., Keele, B.F., Derdeyn, C.A., Farmer, P., Hunter, E., Allen, S.*, et al.* (2008). Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. J Virol *82*, 3952-3970.

Salazar-Gonzalez, J.F., Salazar, M.G., Keele, B.F., Learn, G.H., Giorgi, E.E., Li, H., Decker, J.M., Wang, S., Baalwa, J., Kraus, M.H.*, et al.* (2009). Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. J Exp Med *206*, 1273-1289.

Sallusto, F. (2016). Heterogeneity of Human CD4(+) T Cells Against Microbes. Annu Rev Immunol *34*, 317-334.

Sallusto, F., and Lanzavecchia, A. (2011). Memory in disguise. Nat Med *17*, 1182-1183.

Sanchez, G., Xu, X., Chermann, J.C., and Hirsch, I. (1997). Accumulation of defective viral genomes in peripheral blood mononuclear cells of human immunodeficiency virus type 1-infected individuals. J Virol *71*, 2233-2240.

Schiott, A., Lindstedt, M., Johansson-Lindbom, B., Roggen, E., and Borrebaeck, C.A. (2004). CD27- CD4+ memory T cells define a differentiated memory population at both the functional and transcriptional levels. Immunology *113*, 363-370.

Schmid, A., Gianella, S., von Wyl, V., Metzner, K.J., Scherrer, A.U., Niederost, B., Althaus, C.F., Rieder, P., Grube, C., Joos, B.*, et al.* (2010). Profound depletion of HIV-1 transcription in patients initiating antiretroviral therapy during acute infection. PLoS One *5*, e13310.

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

Seddon, B., Tomlinson, P., and Zamoyska, R. (2003). Interleukin 7 and T cell receptor signals regulate homeostasis of CD4 memory cells. Nat Immunol *4*, 680-686.

Sengupta, S., and Siliciano, R.F. (2018). Targeting the Latent Reservoir for HIV-1. Immunity *48*, 872-895.

Shaw, G.M., and Hunter, E. (2012). HIV transmission. Cold Spring Harb Perspect Med *2*.

Siliciano, J.D., Kajdas, J., Finzi, D., Quinn, T.C., Chadwick, K., Margolick, J.B., Kovacs, C., Gange, S.J., and Siliciano, R.F. (2003). Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. Nat Med *9*, 727-728.

Simonetti, F.R., Zhang, H., Soroosh, G.P., Duan, J., Rhodehouse, K., Hill, A.L., Beg, S.A., McCormick, K., Raymond, H.E., Nobles, C.L.*, et al.* (2021). Antigen-driven clonal selection shapes the persistence of HIV-1-infected CD4+ T cells in vivo. J Clin Invest *131*.

Soghoian, D.Z., Jessen, H., Flanders, M., Sierra-Davidson, K., Cutler, S., Pertel, T., Ranasinghe, S., Lindqvist, M., Davis, I., Lane, K.*, et al.* (2012). HIV-specific cytolytic CD4 T cell responses during acute HIV infection predict disease outcome. Sci Transl Med *4*, 123ra125.

Song, L., Cohen, D., Ouyang, Z., Cao, Y., Hu, X., and Liu, X.S. (2021). TRUST4: immune repertoire reconstruction from bulk and single-cell RNA-seq data. Nat Methods *18*, 627-630.

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

Spinner, C.D., Boesecke, C., Zink, A., Jessen, H., Stellbrink, H.J., Rockstroh, J.K., and Esser, S. (2016). HIV pre-exposure prophylaxis (PrEP): a review of current knowledge of oral systemic HIV PrEP in humans. Infection *44*, 151-158.

Svensson, J.P. (2021). Targeting Epigenetics to Cure HIV-1: Lessons From (and for) Cancer Treatment. Front Cell Infect Microbiol *11*, 668637.

Takata, H., Kessing, C., Sy, A., Lima, N., Sciumbata, J., Mori, L., Jones, R.B., Chomont, N., Michael, N.L., Valente, S.*, et al.* (2019). Modeling HIV-1 Latency Using Primary CD4(+) T Cells from Virally Suppressed HIV-1-Infected Individuals on Antiretroviral Therapy. J Virol *93*.

Takeuchi, A., and Saito, T. (2017). CD4 CTL, a Cytotoxic Subset of CD4(+) T Cells, Their Differentiation and Function. Front Immunol *8*, 194.

Telwatte, S., Moron-Lopez, S., Aran, D., Kim, P., Hsieh, C., Joshi, S., Montano, M., Greene, W.C., Butte, A.J., Wong, J.K.*, et al.* (2019). Heterogeneity in HIV and cellular transcription profiles in cell line models of latent and productive infection: implications for HIV latency. Retrovirology *16*, 32.

Toccafondi, E., Lener, D., and Negroni, M. (2021). HIV-1 Capsid Core: A Bullet to the Heart of the Target Cell. Front Microbiol *12*, 652486.

UNAIDS (2021). UNAIDS DATA 2021 (Geneva: Joint United Nations Programme on HIV/AIDS).

Vandergeeten, C., Fromentin, R., DaFonseca, S., Lawani, M.B., Sereti, I., Lederman, M.M., Ramgopal, M., Routy, J.P., Sekaly, R.P., and Chomont, N. (2013). Interleukin-7 promotes HIV persistence during antiretroviral therapy. Blood *121*, 4321-4329.

Vassena, L., Proschan, M., Fauci, A.S., and Lusso, P. (2007). Interleukin 7 reduces the levels of spontaneous apoptosis in CD4+ and CD8+ T cells from HIV-1-infected individuals. Proc Natl Acad Sci U S A *104*, 2355-2360.

Venanzi Rullo, E., Pinzone, M.R., Cannon, L., Weissman, S., Ceccarelli, M., Zurakowski, R., Nunnari, G., and O'Doherty, U. (2020). Persistence of an intact HIV reservoir in phenotypically naive T cells. JCI Insight *5*.

Wagner, T.A., McKernan, J.L., Tobin, N.H., Tapia, K.A., Mullins, J.I., and Frenkel, L.M. (2013). An increasing proportion of monotypic HIV-1 DNA sequences during antiretroviral treatment suggests proliferation of HIV-infected cells. J Virol *87*, 1770-1778.

Wagner, T.A., McLaughlin, S., Garg, K., Cheung, C.Y., Larsen, B.B., Styrchak, S., Huang, H.C., Edlefsen, P.T., Mullins, J.I., and Frenkel, L.M. (2014). HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. Science *345*, 570-573.

Wang, G., Zhao, N., Berkhout, B., and Das, A.T. (2016). A Combinatorial CRISPR-Cas9 Attack on HIV-1 DNA Extinguishes All Infectious Provirus in Infected T Cell Cultures. Cell Rep *17*, 2819-2826.

Wang, Q., Finzi, A., and Sodroski, J. (2020). The Conformational States of the HIV-1 Envelope Glycoproteins. Trends Microbiol *28*, 655-667.

Wang, Z., Gurule, E.E., Brennan, T.P., Gerold, J.M., Kwon, K.J., Hosmane, N.N., Kumar, M.R., Beg, S.A., Capoferri, A.A., Ray, S.C.*, et al.* (2018a). Expanded cellular clones carrying

replication-competent HIV-1 persist, wax, and wane. Proc Natl Acad Sci U S A *115*, E2575-E2584.

Wang, Z., Simonetti, F.R., Siliciano, R.F., and Laird, G.M. (2018b). Measuring replication competent HIV-1: advances and challenges in defining the latent reservoir. Retrovirology *15*, 21.

Weng, N.P., Araki, Y., and Subedi, K. (2012). The molecular basis of the memory T cell response: differential gene expression and its epigenetic regulation. Nat Rev Immunol *12*, 306-315.

Wherry, E.J., and Kurachi, M. (2015). Molecular and cellular insights into T cell exhaustion. Nat Rev Immunol *15*, 486-499.

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

Wilen, C.B., Tilton, J.C., and Doms, R.W. (2012). HIV: cell binding and entry. Cold Spring Harb Perspect Med *2*.

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

Wong, L.M., and Jiang, G. (2021). NF-kappaB sub-pathways and HIV cure: A revisit. EBioMedicine *63*, 103159.

Wong, M.E., Jaworowski, A., and Hearps, A.C. (2019). The HIV Reservoir in Monocytes and Macrophages. Front Immunol *10*, 1435.

Yukl, S.A., Kaiser, P., Kim, P., Telwatte, S., Joshi, S.K., Vu, M., Lampiris, H., and Wong, J.K. (2018). HIV latency in isolated patient CD4(+) T cells may be due to blocks in HIV transcriptional elongation, completion, and splicing. Sci Transl Med *10*.

Zarnitsyna, V.I., Evavold, B.D., Schoettle, L.N., Blattman, J.N., and Antia, R. (2013). Estimating the diversity, completeness, and cross-reactivity of the T cell repertoire. Front Immunol *4*, 485.

Zaunders, J.J., Dyer, W.B., Wang, B., Munier, M.L., Miranda-Saksena, M., Newton, R., Moore, J., Mackay, C.R., Cooper, D.A., Saksena, N.K*., et al.* (2004). Identification of circulating antigen-specific CD4+ T lymphocytes with a CCR5+, cytotoxic phenotype in an HIV-1 long-term nonprogressor and in CMV infection. Blood *103*, 2238-2247.

Zerbato, J.M., McMahon, D.K., Sobolewski, M.D., Mellors, J.W., and Sluis-Cremer, N. (2019). Naive CD4+ T Cells Harbor a Large Inducible Reservoir of Latent, Replication-competent Human Immunodeficiency Virus Type 1. Clin Infect Dis *69*, 1919-1925.

Zhu, T., Wang, N., Carr, A., Nam, D.S., Moor-Jankowski, R., Cooper, D.A., and Ho, D.D. (1996). Genetic characterization of human immunodeficiency virus type 1 in blood and genital secretions: evidence for viral compartmentalization and selection during sexual transmission. J Virol *70*, 3098-3107.