Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Kelsie Brooks Date

HIV-1 Reservoir Formation in the Context of Pre-Antiretroviral Therapy Infection

By
Kelsie Brooks
Doctor of Philosophy

Graduate Division of Biological and Biomedical Sciences
Microbiology and Molecular Genetics

_____
Eric Hunter, Ph.D.
Advisor

_____
Baek Kim, Ph.D., RPh
Committee Member

_____
Mirko Paiardini, Ph.D.
Committee Member

_____
Guido Silvestri, M.D.
Committee Member

_____
Sam Speck, Ph.D.
Committee Member

Accepted:

_____
Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____
Date

HIV-1 Reservoir Formation in the Context of Pre-Antiretroviral Therapy Infection

By

Kelsie Brooks
B.S., Oklahoma State University, 2013

Advisor: Eric Hunter, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in the Graduate Division of Biological and Biomedical Sciences
Microbiology and Molecular Genetics
2019

Abstract

HIV-1 Reservoir Formation in the Context of Pre-Antiretroviral Therapy Infection

By Kelsie Brooks

While antiretroviral therapy (ART) has changed the landscape of the HIV-1 epidemic, reducing mortality and providing an effective means of preventing infections, it fails to cure HIV-1 infection. ART drugs do not target the reservoir, a persistent population of latently infected cells that remain even with suppression of virus to undetectable levels in the blood.

Here we investigate the establishment of the reservoir in the context of the natural history of HIV-1 infection, sampling virus from seroconversion through chronic infection just before ART initiation, and assess the relationship between reservoir proviral variants and pre-therapy virus. In 13 Zambian individuals, we observed that although the majority of proviral variants in the reservoir are most closely related to virus present shortly before the initiation of treatment, provirus most closely related to very early infection virus is present in five of 13 individuals as well. In two individuals, we detected provirus matching the transmitted/founder (TF) virus, indicating persistence of the variant for over six years of ART-naïve infection in one case. Reservoir proviral variants were also significantly more closely related to the TF virus than virus in the blood just prior to the start of therapy, highlighting the contributions of earlier variants to the reservoir population. Sequential sampling of four individuals at two time points during ART indicated a yet shorter distance from the TF virus with continued time on treatment, or an enrichment in early infection variants.

Taken together, we conclude that these observations indicate the reservoir is seeded throughout ART-naïve infection, and that HIV-1 cure strategies designed to reduce or eliminate the reservoir must account for the presence of virus archived at all stages of the infection.

HIV-1 Reservoir Formation in the Context of Pre-Antiretroviral Therapy Infection

By

Kelsie Brooks
B.S., Oklahoma State University, 2013

Advisor: Eric Hunter, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in the Graduate Division of Biological and Biomedical Sciences
Microbiology and Molecular Genetics
2019

Acknowledgments

There are too many people who've willed me through the long process of getting a Ph.D. for me to thank them all, but the below individuals deserve particular kudos:

- My parents, Bob and Pam, and my siblings, Erin and Colin, for loving and supporting me. Your sympathy and advice—particularly for enduring the difficulties of graduate school, which you've all experienced—have been invaluable.
- Current and former graduate student labmates (Dan, Jess, Debby, Zach, Martin, Elina, Sarah, and Sam) for your thoughtful scientific discussions, patience, and camaraderie as most of us questioned our life choice to pursue a Ph.D. at one time or another. A lab that ziplines together stays together!
- All the Hunter lab members who keep us going in their tireless maintenance of the day-to-day needs: sample procurement, dozens of functioning thermocyclers, and supplies (including tea and lab meeting bagels—thanks, Effie!).
- My MMG and IMP friends for being a welcoming, hilarious, inspiring group of people to work with and get to know, and for becoming my Atlanta family.
- My friends from way back, Kira and Christiona, and the Oklahoma State crew for accepting me in all my weird lizard-loving-ness, and always being there for a good talk.
- My weird pets, the geckos and dragon Smaug, for their reminder to enjoy the little things in life (bugs).
- Eric, for giving me the freedom to explore, providing guidance when those explorations sometimes failed, and for being an ally in my scientific journey.

Table of Contents

Chapter I: Introduction

Current Status of the Epidemic and Impetus for Further Research

The impact of HIV/AIDS on global health is staggering: in over 30 years since the HIV epidemic began, approximately 32 million people have died of AIDS-associated illnesses, and nearly 38 million individuals were living with HIV in 2018.[1] Although HIV/AIDS is no longer among the top ten causes of death worldwide,[2] and the number of new HIV infections globally in 2018 was 40% lower than the highest annual total of nearly three million in 1997,[1] the epidemic is far from over. Worryingly, some countries and regions have experienced considerable increases in the annual number of new HIV infections.[3] The UNAIDS 90-90-90 goals for effectively ending the epidemic by 2030—90% of HIV+ individuals aware of their status, 90% of individuals with known HIV infection receiving continued treatment, and 90% of those on treatment reaching undetectable levels of virus in blood by 2020[4]—will require continued and expanded funding and support from global partners to achieve.

Progress toward ending the HIV/AIDS epidemic has required a uniquely multi-faceted approach in comparison to globally or regionally eliminated viral pathogens, such as variola virus and polio virus, which have benefitted from the deployment of effective vaccines. In absence of a sufficiently protective vaccine, HIV prevention has employed combinatorial approaches of structural, biomedical, and behavioral interventions to avert new infections, such as HIV testing and counseling (behavioral intervention) along with condom provision (biomedical intervention).[5] Key in the reduction of new infections has been the rollout of effective antiretroviral therapy (ART), both for treatment as prevention (TasP) and pre-exposure prophylaxis (PrEP).[6] Taken daily, oral PrEP has demonstrated a 44% reduction in infections among men who have sex with men, and more than 60% reduction of infections in heterosexual

men and women.[7-9] Infections in HIV- partners of HIV+ individuals beginning ART at CD4+ T cell counts of 350 to 500/mm^3 were reduced 89% when compared to infections in partners of HIV+ individuals with delayed initiation of treatment until CD4+ T cell counts were 250 or lower; early treatment was also associated with 41% fewer clinical events than delayed treatment, indicating that early treatment of HIV infected patients not only benefits the individuals, but also provides a considerable community benefit in preventing transmission, the crux of TasP.[10]

Effective ART has not only changed the landscape of HIV prevention, but also considerably reduced mortality associated with HIV/AIDS, which has decreased 33% in the last nine years.[1] Although HIV infected individuals taking ART in developed countries can now experience a life expectancy near that of an uninfected individual,[11] treatment must be maintained throughout life for sustained viral suppression, and it is not curative. Even with very early initiation of therapy at three days following SIV infection in rhesus macaques, and within two days of birth for an infant born to an HIV+ mother who had not taken ART, the infection was nonetheless established and persistent; virus suppressed with ART rebounded with therapy interruption.[12,13] The challenge of HIV cure is further evident in the mere two instances of such to date, both of which were achieved with stem-cell transplantation of cells containing a small deletion in the CCR5 gene (Δ32), rendering the cells non-permissive to infection by HIV utilizing the CCR5 molecule as a co-receptor for entry.[14-16] This approach, which was performed in the context of patient leukemia[14,15] and Hodgkin's lymphoma,[16] is clearly not tractable for broad implementation, but several other strategies for sterilizing cure with no evidence of continued HIV infection, and functional cure in which control can be reliably achieved for considerable periods of time without ART, are actively being pursued.[17]

Achieving a functional or sterilizing HIV cure requires an understanding of the HIV reservoir, which is a persistent population of latently infected cells refractory to treatment, but containing HIV that retains the ability to replicate.[18-20] Although ART can suppress virus in the blood to below the limit of detection in standard assays (50 copies/mL), invariably virus rebounds from the reservoir with treatment interruption.[21-23] The viral life cycle in part permits the existence of latently infected cells, as retroviruses like HIV stably integrate a DNA genome (provirus) into the host cell chromosome.[24] In resting cells with low levels of active transcription factors such as NF-$_K$B and NFAT that are important for provirus expression, it may be sufficiently silenced for the host cell to evade immune system recognition and clearance, and the provirus is therefore maintained in a latent state with the cell.[25] Additional epigenetic silencing and the proviral integration locus also play a role in the establishment of latency.[25] Without active viral replication, antiretrovirals are rendered ineffective against latent proviruses, as the viral proteins the drugs target are not produced. Furthermore, HIV's tropism for CD4+ T cells permits infection of long-lived cells, as latently infected resting memory CD4+ T cells have been shown to have a half-life of approximately 44 months *in vivo*.[26,27] Persistent, latently infected cells with replication-competent provirus are nonetheless rare,[28] posing one of several difficulties in identifying cells of the reservoir for elimination with cure strategies.

While targeting latently infected cells persisting during ART is an approach to reduce the reservoir once established, very early treatment initiation has been investigated as a means of limiting its formation.[23] Treatment even during Fiebig I, the earliest detectable stage of HIV infection lasting approximately one week and preceding seroconversion,[29] does not prevent reservoir establishment.[23] However, individuals treated starting in early HIV infection do exhibit smaller reservoirs than those initiating ART during the later chronic stage of infection,[30,31]

suggesting that the reservoir continues to be formed throughout active, ART-naïve infection. The dynamics of reservoir establishment in HIV infection are highly relevant to the growing cure efforts, yet relatively little research investigating the reservoir in the context of the natural history of the infection has been done. Elucidating the relationship of the reservoir to the course of the pre-therapy infection is the focus of the subsequent chapter and the unifying theme of this work, while this introduction shall provide the context of the research by first detailing the natural history of HIV infection, from acquisition to chronic infection, with an emphasis on the viral evolution and immune dysfunction that occur, and concluding with an overview of the current knowledge of the reservoir and gaps in understanding that the studies detailed herein address.

HIV Acquisition: the Transmission Bottleneck and Features of Acute Infection

HIV infection or acquisition can occur through multiple routes, such as sex, injection drug use, and vertical transmission from HIV infected mothers to infants before, during, or after birth. The risk of infection associated with a given exposure is challenging to quantify, but varies across routes, with vertical mother-to-child transmission (MTCT) having a considerably higher probability of infection than injection drug use or sexual transmission at an estimated 2,260 infections per 10,000 exposures for MTCT versus 63 infections for injection drug use, and four to 138 for sexual exposure.[32] The wide range of estimated infections for sexual exposure relates to distinctly different risks for penile-vaginal and insertive anal intercourse compared to receptive anal intercourse, with the former associated with four to 11 infections per 10,000 exposures, while 138 are estimated for receptive anal intercourse.[32] Despite such disparate risk for HIV acquisition in sexual exposures, interventions such as condom usage and PrEP broadly reduce the likelihood of HIV acquisition,[33] while other factors such as the presence of sexually

transmitted infections increase the risk of an individual acquiring HIV,[34-36] perhaps through the promotion of a genital environment high in chemokines such as macrophage inflammatory protein 1-alpha and 1-beta (MIP-1α and MIP-1β) and interferon gamma-induced protein 10 (IP-10) that may enhance immune cell recruitment and infection.[37] Genital ulcers and inflammation (GUI) likewise affect HIV acquisition, reducing selection at the transmission bottleneck,[38,39] a defining feature associated with acute infection.

Regardless of the route of infection, HIV infection is typically established by a single genetic variant, known as the transmitted/founder (TF) virus, despite the presence of multiple genetically distinct viruses, or a quasispecies, in a transmitting partner or source.[40-47] This marked reduction in genetic diversity with infection is known as the transmission bottleneck,[48] and investigations of the genotypes and phenotypes of TF viruses in comparison to their corresponding source quasispecies have been an important, active area of research. TF studies aim to understand the mechanisms of TF selection and elucidate unique properties of transmitted variants, which could serve as vaccine targets. Accurate means of identifying the TF virus and cloning methods for generating authentic, infectious TF viruses[49,50] have been essential tools for studying transmitted viruses, while paired samples from transmitting (donor) and newly infected (recipient) partners are critical for examining the transmission bottleneck.

With recognition that an extreme genetic bottleneck exists in HIV transmission, several hypotheses for this selection emerged. The selection might be stochastic, representing a mere reduction in numbers of variants in the inoculum to a single infecting variant. In such instance, transmitted variants would be expected to belong to the most prominent population of the donor quasispecies, as they would represent a larger fraction of the variants from which the TF virus could derive. However, in paired samples of the donor genital tract and recipient plasma in eight

heterosexual transmissions, the TF virus derived from minor, rather than major, donor genital tract populations in all cases.[51] This phenomenon argues for non-stochastic selection at the transmission bottleneck, but does not indicate a mechanism by which the selection might occur.

Another hypothesis for TF selection was that TF viruses represent highly replicating viruses capable of out-competing other variants. With six TF virus clones from acute infections and corresponding non-transmitted (NT) variant clones from the donor quasispecies, TF viruses were not found to have a higher or significantly different *in vitro* replicative capacity than the corresponding NT variants.[52] Furthermore, these TF viruses did not demonstrate a marked resistance to viral restriction by interferon-α (IFN-α)[52] or enhanced capability for HLA downregulation that prevents immune system recognition of infected cells[53] in comparison to NT donor variants. However, additional studies of TF viruses have found both increased replication capacity and interferon resistance.[54,55] Such disparities in findings may reflect the differing comparisons made, as in Parrish et al.,[54] TF viruses were compared to chronic infection viruses not derived from the transmitting partner rather than to donor quasispecies variants, as was the case for transmission pairs investigated by Deymier and Ende et al.[52] Likewise the means of isolating virus in Iyer et al.[55] is distinct from that of Deymier and Ende et al.,[52] with Iyer et al. expanding virus in culture while Deymier and Ende cloned variants from PCR-amplified virus. Without a clear indication that TF viruses are consistently selected for higher IFN-α resistance, this suggests that TF viruses may not be selected for improved evasion of targeting by the innate immune system, which is further implied by the lack of significantly different NK cell mediated suppression of viral replication in cells infected by TF versus NT viruses.[53] Additionally, TF viruses in the context of heterosexual transmission exhibit enhanced sensitivity to donor neutralizing antibodies.[42,52] The phenotype is genetically linked through a significant negative

correlation of neutralization sensitivity with the distance to the amino acid sequence of a consensus virus for the viral subtype circulating in the country of infection[52]. TF viruses are correspondingly significantly more closely related to the consensus than NT viruses,[52] and amino acid residues that match the consensus of the cohort studied are also preferentially transmitted, both in heterosexual transmission and MTCT.[39,56] This potentially indicates a genetic sieving effect against transmission of variants that have accumulated the most mutations, perhaps through selection by the immune system for antibody resistance or escape of CD8+ T cell-mediated targeting. Studies indicating that more ancestral virus from within the donor quasispecies is transmitted support this genetic selection.[57,58]

Once infection is established by the TF virus(es), there are predictable dynamics of viral load and host responses. Following the Fiebig I stage of infection in which viral RNA alone is detectable, viral p24 antigen can be detected in Fiebig II, though it may subsequently be lost to detection, and progressive antibody presence defines Fiebig stages III-VI.[29] Fiebig I-II are generally considered to represent acute HIV infection, with up to approximately six months post-infection considered early or recent infection.[59] Although viral RNA is detectable in Fiebig I, it peaks with the appearance of antibodies in Fiebig II-III,[29,48] concomitant with a considerable loss of CD4+ T cells.[60] The subsequent decline of peak viremia is associated with the development and activity of HIV-specific CD8+ T cells,[61,62] as well as a rebound in CD4+ T cells.[60] Decline from peak viremia is typically complete approximately 18 days post-peak, with the lowest viral load at this early stage of infection approximating the set-point of viremia[63] that may be maintained for years even in the absence of ART. Set-point viral load (SPVL) is a significant predictor of disease progression, with higher viral loads linked to a shorter time to AIDS,[64] indicating that a key marker of HIV pathogenesis is established at an early stage of infection.

Transmitted virus phenotype has also been connected to SPVL. TF virus replicative capacity, as assessed by replication of chimeras for the transmitted Gag or Reverse Transcriptase and Integrase proteins in laboratory HIV strain backbones, is significantly positively correlated to SPVL in the recipient,[65,66] and low TF virus chimera replication is predictive of significantly slower CD4+ T cell loss.[66,67] These findings indicate a role for TF virus replication in disease, with highly replicating TF virus chimeras also generating a more perturbed immune system in early infection through higher levels of pro-inflammatory cytokines, as well as more activated and exhausted CD8+ T cells critical for clearance of virally infected cells.[67] The presence of more than one TF virus may likewise heighten immune system dysfunction, as transmission of multiple variants versus a single TF has been linked to significantly higher viral loads,[68] as well as more rapid disease progression.[69] Features of early HIV infection, including the post-peak viral load and phenotype of the TF virus(es), are therefore important indicators of and contributors to the extensive immune system disruption that occurs with HIV infection, which ultimately leads to AIDS in absence of ART.

Chronic Infection: Immune System Dysfunction

HIV causes a persistent infection with continued viral replication and CD4+ T cell decline in the chronic stage, as the immune system is not able to control and clear the virus. This is despite early and considerable induction of numerous antiviral and pro-inflammatory cytokines such as IFN-α, IFN-γ, tumor necrosis factor-alpha (TNF-α), IL-15, and IL-18.[70] Indeed the "cytokine storm" induced by HIV begins earlier in infection and reaches even greater magnitude than corresponding responses against other pathogens such as Hepatitis B and C viruses,[70] which can be cleared. The reason for failure of HIV clearance therefore does not appear to be a lack of immune stimulation or response, and instead likely represents a complex

dynamic between the virus and host, with multiple viral mechanisms of evading host responses that, in perpetually responding to the infection, promote continued immune dysfunction.

As a component of the adaptive immune response, CD8+ T cells provide an antigen-specific means of controlling virally infected cells. However, HIV-specific CD8+ T cells in chronic infection express significantly more of the programmed death 1 (PD-1) receptor than do cells of the total CD8+ population, and are considerably less capable of proliferating with this higher PD-1 expression, indicating functional exhaustion.[71] The level of PD-1 in the total CD8+ T cell population is also significantly positively correlated to viral load, indicating that as the viral antigen persists and increases, host CD8+ T cells broadly become less capable of proliferating and executing effector functions.[71] Furthermore, CD8+ T cell activation (as measured by CD38) is linked to viral load, with a corresponding increase in CD38 with higher viral loads, and higher CD8+ T cell activation in late stages of infection is associated with a shorter survival time.[72] Although the development of HIV-specific CD8+ T cells is initially associated with a drop in viral load,[62] suggesting effective targeting of the virus contributing to its clearance, HIV escapes from CD8+ cytotoxic T lymphocyte (CTL) cells even at a very early stage of infection,[73] and continued escape throughout chronic infection[74] likely contributes to ineffective CTL responses, allowing continued viral replication that fuels CD8+ T cell activation and exhaustion, ultimately with a poorer prognosis for disease progression.

While CD8+ T cell activation may be connected to survival in late stages of infection, CD4+ T cell loss is the defining feature for disease progression. Gut tissue demonstrates considerable depletion of CD4+ T cells with acute SIV infection,[75] and the profound loss is noted in HIV infection as well; reductions in gut CD4+ T cell numbers persist throughout the chronic stages of the infection.[76] Such extensive loss of gut CD4+ T cells appears to be due to

the high proportion of primary targets for HIV infection, CCR5+ CD4+ T cells and proliferating CD4+ T cells, as defined by Ki67, in the gut rather than lymph nodes or the blood.[76] Failure of gut CD4+ T cells to completely reconstitute from the considerable depletion is likely due in part to the overall slower thymic output of CD4+ than CD8+ T cells.[60] This phenomenon does not explain continued CD4+ T cell decline in chronic infection, however, nor the ability of CD4+ T cells to reconstitute with ART, suggesting there are several processes at play affecting CD4+ T cell counts. Indeed, plasma viral load is correlated significantly with CD4+ T cell decline,[77] indicating at least some direct effect of the virus on CD4+ T cell loss, but immune activation and homeostatic proliferation perturbations in the CD4+ T cell compartment are likely contributors as well to the sustained, gradual decline of CD4+ T cells with chronic HIV infection.[78]

Immune cell function and quantity are disturbed with HIV infection, but the immune dysregulation is not restricted to just cells, as tissues and organs are affected as well. Microbial translocation from the gut, as measured by the bacterial membrane molecule LPS, is significantly higher in HIV infected individuals than uninfected, likely indicating gut damage and decreased barrier permeability in this tissue.[79] Furthermore, increased levels of LPS were associated with higher levels of activated CD8+ T cells and IFN-α, indicating a systemic inflammatory effect.[79] Inflammation in the lymph nodes of HIV-infected individuals can result in collagen deposition, which disrupts the lymph node architecture necessary for promoting interactions between various immune cells, and therefore is considered to indicate a deficiency of the lymph nodes to assist in maintaining homeostasis.[80] Levels of inflammation characteristic of chronic HIV infection are further associated with higher morbidity in HIV patients for certain age-related conditions such as cardiovascular disease and cancer, a phenomenon known as “inflammaging,” wherein many organs and systems are affected.[81] With many years’ use of antiretrovirals, the effects due strictly

to infection and those instead resultant from treatment side effects are difficult to untangle,[81] but even fully virally suppressive ART does not reduce all systemic inflammation to pre-infection levels.[82] Consequently, a signature of chronic, ART-naive HIV infection is considerable morbidity and mortality arising from rapid, profound, and sustained effects on the immune system, with the hallmark loss of CD4+ T cells.

Chronic Infection: Evolution of the Quasispecies

Despite the extreme genetic bottleneck of HIV transmission, with most infections initiated by a single variant, HIV rapidly mutates and evolves in a new host to form a population of distinct but closely related variants, or a quasispecies. The quasispecies arises as the highly error-prone HIV reserve transcriptase incorporates approximately one mutation in the HIV genome for every round of viral replication.[83] Acute infection sequences from single-variant infection demonstrate a star-like phylogeny, with random rather than shared mutations coalescing around a consensus,[43] but this phylogeny is quickly broken as mutations are selected by the immune system. Immediately following the rapid rise in viremia and expansion of HIV-specific CD8+ T cells in early infection, the first CTL escape mutations arise.[73,84] Some CTL escape mutations may negatively impact the ability of the virus to replicate,[85] and require the concerted mutation of other regions in the genome for re-gaining appropriate function,[65] or may revert if no longer relevant for the HLA of the infected individual.[86] Thus the evolutionary landscape of HIV is highly dynamic and complex.

In a similar fashion to the mutations permitting CTL escape, drug resistance mutations arise quickly with monotherapy. Several conserved mutations confer resistance to zidovudine (AZT), the first antiretroviral approved for HIV, though differences in the genetic background of the virus may alter the timing and pattern of the emergence of these mutations in a patient.[87]

While non-suppressive ART may lead to the selection of drug resistance, resistance mutations in a patient quasispecies do not necessarily preclude the effective use of a combinatorial drug regimen including the antiretroviral with demonstrated resistance; in part, this may be due to the fact that ancestral, drug sensitive variants persist even with years of sub-optimal treatment.[88]

As CTL-driven viral escape occurs quickly in infection, so immune-mediated selection of the *env* gene is not far behind. Indeed B cell responses to Env begin approximately one week after virus is detectable in the blood, although they are not sufficient for contributing to controlling viral load.[89] As a protein on the surface of the virion, Env is a ready target of antibodies, and has considerable mutational space in which to attain mutations while still remaining functional, with the diversity in *env* the greatest for any HIV gene. Furthermore, HIV *env* is extraordinarily diverse even when compared to other viral pathogens with error-prone replication machinery: *env* diversity in approximately 20 chronically HIV-infected individuals far exceeds that of nearly 100 sequences of the influenza attachment protein, hemagglutinin, in a flu season.[90] Env selection occurs continually throughout the infection, such that many individuals develop broadly neutralizing antibodies (bnAbs) against Env during chronic, ART-naïve infection.[91] Vaccination strategies to recapitulate this progressive generation of bnAbs with serial immunogens have largely proven ineffective. Such difficulty reveals the complex, dynamic interplay of viral evolution and host immune pressure, which is a driving force of the generation of the HIV quasispecies.

The Reservoir: an Enigma

Latently infected cells containing replication-competent provirus that persist during ART comprise the reservoir, the critical barrier to HIV cure. Such cells are difficult to quantify, as they persist within a much larger pool of defective virus.[92,93] Defective proviruses may contain

large deletions or hypermutation resulting from the action of a host protein, APOBEC3G.[94] The canonical G to A mutations of APOBEC3G often result in the generation of nonsense mutations, thus rendering a provirus non-functional, while small mutations in packaging signals or other critical regulatory elements of the proviral genome also result in defective proviruses.[93] With their rarity, latent replication-competent proviruses are difficult to reactivate *in vitro*, requiring large numbers of cells for the gold-standard reactivation assay, the quantitative viral outgrowth assay (QVOA).[95] There are further issues, too, with QVOA, as it fails to induce all replication-competent provirus, requiring repeated rounds of stimulation to reactive a portion of the population each time.[92] However, DNA-based methods of detecting latent provirus may overestimate the true size of the replication-competent population, as many amplify only a portion of the genome, and may do so where there is considerable defect outside the amplified region.[28] New quantitation strategies that more accurately represent the size of the replication-competent reservoir[96,97] have been essential for evaluation of cure strategies.

Means of eliminating the reservoir are diverse, and range from "shock and kill" strategies to genetic targeting.[98] "Shock and kill" refers to the reactivation of provirus with drugs to then permit for clearing of the infected cells, which may occur through cell death or immune-mediated killing.[99] The action of histone deacetylases, which remove acetyl groups from histones and allow tighter wrapping of DNA around chromatin, preventing gene expression, contributes to HIV latency; the activity of the proteins has been targeted with histone deacetylase inhibitors as one means of generating the "shock" to reverse HIV latency. Although latency has been reversed *in vivo* with such inhibitors, this has unfortunately not resulted in a corresponding reduction in latently infected cells *in vivo*,[100] perhaps unsurprising given that active infection does not result in the clearance of all virally-infected cells. Newer strategies for eliminating virus

following its reactivation from latency now include combining a latency reversal agent (LRA) with an immunomodulatory component such as antibody or a therapeutic vaccine to boost responses and kill the actively infected cells.[99]

Genetic strategies to target the reservoir have also been an active area of research, and include modifying the host's cells to render them resistant to infection, as with CCR5Δ32 mutation, or disrupting the proviral genome through DNA modifying systems such as zinc-finger nucleases and CRISPR-Cas9 to prevent viral replication.[98] The latter is a strategy with some promise, as it has been studied in mice with human-like immune cells susceptible to and infected with HIV.[101] Targeting of the proviral long terminal repeat and Gag gene using CRISPR-Cas9 delivered by adeno-associated virus successfully reduced the HIV DNA in multiple tissues, such as gut and spleen, to undetectable levels in two treated animals, and no viral rebound was observed following treatment interruption in these animals.[101] However, this CRISPR-based strategy was successful only a portion of the animals treated, and in combination with long-acting ART,[101] suggesting that the process of excising portions of the proviral genome in latently infected cells might result in low-levels of production of new virus, and therefore ART must be continued during the process.

While research to broadly eliminate latently infected cells is underway, the questions of which cell(s) or anatomic site(s) are responsible for viral rebound with treatment interruption remain. The process of viral recrudescence with treatment interruption appears complex, and indeed one study has indicated multiple foci of infection in the lymph node with treatment interruption, which was hypothesized to explain diverse populations of rebounding virus.[102] This might corroborate the findings of two other studies, which each concluded that rebounding virus appears to represent recombination between pre-rebound viruses, with recombination between

two distinct tissues, blood and lymph nodes, in the work of Vibholm et al.[103,104] Simultaneous reactivation of genetically unique viruses followed by superinfection of the same cell and subsequent replication would provide a mechanism for recombined virus from the same or distinct compartments. Furthermore, Rothenberger et al.[102] described foci of infection in lymph nodes prior to interruption of treatment, suggesting that transcriptionally active cells are present even with viral suppression to <50 copies/mL, which Kearney et al. propose as a source of recrudescent virus.[105] An answer to the question of where HIV rebounds is therefore no simpler than the potential cell types from which it derives, as several—T follicular helper,[106] Th17,[107] central memory,[108] effector memory,[97] and stem cell-like[109]—have been indicated as particularly permissive to infection and/or enriched in the reservoir. Additionally, some groups have observed homogenous reactivated virus, suggesting a single source.[110] The difficulty in determining the multiplicity and source of rebounding virus is likely reflective of the dynamic nature of pre-therapy infection in which infected cells are disseminated throughout tissues, and cell turnover is high.

The answer to where the HIV reservoir resides is not entirely clear, and likewise questions remain in regard to the activity of latently infected cells during treatment. Studies addressing the stability of the reservoir have conflicting findings, in some instances suggesting that there is ongoing evolution despite viral suppression to <50 copies/mL in the blood with ART,[111] while others find a dearth of genetic changes to support continued evolution in the reservoir with effective therapy.[110,112,113] Furthermore, the source of virus contributing to persistent low-level viremia <50 copies/mL is debated, as sustained viral replication could allow for the production of virus in the plasma, yet alternatively, sporadic reactivation and viral production from latently infected cells could also result in low levels of virus. Sporadic

reactivation may be more consistent with the slow decay of latently infected cells, which have a half-life of nearly four years,[27] as continued replication might result in more rapid cell death and turnover. The mechanisms of such long-lived persistence of latently infected cells are less ambiguous than their activity (or lack thereof) during treatment—not only are the latently infected memory CD4+ T cells long-lived by nature, but homeostatic proliferation and clonal expansion[108,114] of these cells permit latently infected daughter cells to persist even as the originally infected cell senesces or reactivates and dies. These cell-controlled processes also explain the genetic signatures of identical proviruses and integration sites in the reservoir, as the provirus is reproduced in its integration locus and sequence with DNA replication of proliferating or expanding cells.

As they are long-lived, latently infected cells could in theory draw from many time points prior to the initiation of therapy. Although the reservoir is certainly established in early infection, the fate of these cells seeded into the reservoir at an early stage of is relatively unknown. Recent work from Brodin et al. has demonstrated that in comparison to pre-therapy sequences, much of the latent provirus remaining during virologically suppressive ART is most closely related to virus circulating in plasma shortly before the initiation of treatment.[115] However, this group also reports an overrepresentation of proviral variants most closely related to those from within approximately six months of infection.[115] These data are quite similar in nature to findings of Abrahams et al., who reactivated latent provirus with QVOA,[116] while the Brodin work utilized a broader PCR-based strategy.[115] The conclusions of Abrahams et al., however, are that late infection variants from one year prior to ART initiation are overrepresented among the reservoir, representing a change in the host immune environment with initiation of ART that results in

extensive seeding at this time.[116] Further investigation into the nature of reservoir formation in the context of pre-therapy infection is therefore a timely and important study to report.

Summary

HIV poses one of the most considerable challenges in infectious disease to date, as there is no vaccine to prevent infection, and no cure for the millions of infected individuals. However, antiretroviral therapy has changed the prognosis of HIV infection from inevitable chronic inflammation and CD4+ T cell loss with eventual progression to AIDS. HIV+ individuals can now expect a life expectancy of nearly that of an uninfected person, but not without the daily regimen of ART. Exploring the complex and sometimes contradictory nature of the reservoir which prevents HIV cure is a critical step to further understanding of this barrier to making HIV infection a temporary, rather than life-long, infection.

Chapter II: HIV-1 variants are archived throughout infection and persist in the reservoir

Kelsie Brooks[1], Bradley R. Jones[2], Dario A. Dilernia[1], Daniel Wilkins[1#a], Daniel T. Claiborne[1#b], Samantha McInally[1], Jill Gilmour[3], William Kilembe[4], Jeffrey B. Joy[2,5], Susan A. Allen[4,6], Zabrina L. Brumme[2,7], and Eric Hunter[1,6]*

[1]Emory Vaccine Center, Emory University, Atlanta, Georgia, United States of America

[2]British Columbia Centre for Excellence in HIV/AIDS, Vancouver, British Columbia, Canada

[3]Human Immunology Lab, International AIDS Vaccine Initiative, London, England, United Kingdom

[4]Zambia-Emory HIV Research Project, Lusaka, Zambia

[5]Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada

[6]Department of Pathology & Laboratory Medicine, Emory University School of Medicine, Atlanta, Georgia, United States of America

[7]Faculty of Health Sciences, Simon Fraser University, Burnaby, British Columbia, Canada

[#a]Current address: Department of Biology, Emory University, Atlanta, Georgia, United States of America

[#b]Current address: Ragon Institute of MGH, MIT & Harvard, Cambridge, Massachusetts, United States of America

*Corresponding author

Manuscript in revision with the journal *PLoS Pathogens*

Abstract

The HIV-1 reservoir consists of latently infected cells that persist despite antiretroviral therapy (ART). Elucidating the proviral genetic composition of the reservoir, particularly in the context of pre-therapy viral diversity, is therefore important to understanding reservoir formation and the persistence of latently infected cells. Here we investigate reservoir proviral variants from 13 Zambian acutely-infected individuals with additional pre-therapy sampling for a unique comparison to the ART-naïve quasispecies. We identified complete transmitted/founder (TF) viruses from seroconversion plasma samples, and additionally amplified and sequenced HIV-1 from plasma obtained one-year post-infection and just prior to ART initiation. While the majority of proviral variants in the reservoir were most closely related to viral variants from the latest pre-therapy time point, we also identified reservoir proviral variants dating to or near the time of infection, and to intermediate time points between infection and treatment initiation. Reservoir proviral variants differing by five or fewer nucleotide changes from the TF virus persisted during treatment in five individuals, including proviral variants that exactly matched the TF in two individuals, one of whom had remained ART-naïve for more than six years. Proviral variants during treatment were significantly less divergent from the TF virus than plasma variants present at the last ART-naïve time point. These findings indicate that reservoir proviral variants are archived throughout infection, recapitulating much of the viral diversity that arises throughout untreated HIV-1 infection, and therefore strategies to target and reduce the reservoir must permit for the clearance of proviruses encompassing this extensive diversity.

Author summary

Despite reducing viremia to levels below the limit of detection in standard assays, effective antiretroviral therapy (ART) does not eradicate cells latently infected with HIV-1.

These cells serve as a reservoir for viral rebound if therapy is interrupted; thus, understanding the composition of the reservoir may yield further targets for HIV-1 cure strategies. We have taken a genetic approach to elucidating the reservoir in 13 Zambian subtype C seroconvertors who were followed longitudinally through ART initiation and virologic suppression. In five of the 13 individuals, provirus sequences identical to or differing by five or fewer nucleotides from the transmitted/founder virus were detected, indicating archiving and persistence of early infection variants for more than six years following infection. While the majority of proviral variants in latently infected cells were most closely related to plasma virus circulating immediately prior to treatment initiation, additional variants dating to intermediate time points in the infection were also observed. These findings demonstrate that virus is archived during all stages of ART-naïve infection, and these variants persist throughout ART. HIV-1 cure strategies to eliminate the reservoir must address the broad genetic diversity of a within-host proviral quasispecies including variants archived from acute through chronic infection.

Introduction

Although over 23 million individuals living with HIV were receiving ART by the end of 2018, only two have been cured following stem-cell transplantation of HIV-1 resistant cells [1-4]. ART alone is not curative due to the persistence and proliferation of latently infected CD4+ T cells harboring intact but quiescent proviruses unaffected by antiretrovirals that target stages of active viral replication [5-11]. This long-lived and potentially self-renewing population of latently infected cells can serve as a reservoir for viral rebound in the event of treatment cessation [11-16], and efforts to understand the reservoir are therefore essential to HIV-1 cure strategies. Genetic approaches investigating the reservoir have sequenced rebounding virus in HIV-1 patients undergoing treatment interruption, reactivated virus from latently infected cells

stimulated *in vitro*, and proviral populations during ART, providing critical insights into the complex nature of provirus remaining during treatment, only a small fraction of which is capable of replicating, and a further subset of which reactivates with treatment interruption [14-26]. The sources and establishment of this reservoir are of considerable interest, and although the reservoir is seeded beginning even in very early stages of infection [14, 20, 27-29], few investigations have explored the relationship of the reservoir to early infection, pre-therapy viral variants. The extent to which these variants may persist in the reservoir during virologically suppressive ART is incompletely understood.

Recent studies examining associations between transmitted virus, its descendent quasispecies in chronic ART-naïve infection, and the reservoir include genetic analyses of amplified virus [30, 31], particularly from quantitative viral outgrowth assay (QVOA) as in Abrahams et al. [31], while additional work from Jones et al. [32] infers age of latent proviral genes in relation to pre-therapy plasma variants. All three groups describe heterogenous populations of proviral sequences that do not indicate ongoing evolution during virologically suppressive ART [30-32]. Furthermore, all three studies observe proviral variants that are inferred to be most closely related to sequences circulating in the plasma immediately prior to the start of treatment as well as variants contemporaneous with the earliest pre-therapy sample [30-32]. The frequencies of proviral variants dating to particular pre-therapy eras are distinct in each study, with Brodin et al. [30] and Abrahams et al. [31] describing 60% and 71%, respectively, of proviral variants during treatment as most closely related to pre-therapy sequences from immediately prior to treatment initiation, while these frequencies are higher than that described in Jones et al. [32]. Given the interpatient variability in proviral population structure present in these studies, it is perhaps unsurprising for discrepancies in findings as well.

In this study, we examined reservoir proviral sequences in the context of pre-therapy plasma HIV-1 RNA diversity and evolution in 13 Zambian seroconvertors. Critically, our reconstruction of within-host HIV-1 evolution includes the inference of the near full-length transmitted/founder virus from single genome amplification (SGA), allowing us to investigate the possible long-term persistence of this sequence within the reservoir. Utilizing the phylogenetic approaches developed by Jones et al. [32] and additional analyses to assess the reservoir during short-term ART, our findings indicate that latent proviral diversity broadly reflects plasma HIV-1 RNA diversity during the period of pre-therapy infection. A majority of variants appear most closely related to those circulating in plasma near the time of ART initiation, but the reservoir quasispecies can in some individuals include variants present at the time of transmission, and demonstrates persistence of variants archived throughout ART-naïve infection.

Results

Participant selection and sampling methods

We identified 13 Zambian seroconvertors from the Zambia-Emory HIV Research Project (ZEHRP) for study according to the following criteria: ART-naïve infection of at least two years, subsequent ART with viremia <50 copies/mL at one or more time points following therapy initiation, and sample availability during treatment (Table 1). All participants received combination ART per country guidelines. We amplified and sequenced a minimum of seven near full-length genomes (NFLGs) by single genome amplification (SGA) from the earliest HIV+ plasma sample available for each participant (seroconversion sample), which was collected a median of 44 days post Estimated Date of Infection (EDI) (Table 1 and Fig 1). Sequencing was performed using Pacific Biosciences Single Molecule, Real-Time (SMRT) sequencing [33], and

the TF virus was inferred as the consensus of the low-diversity NFLGs (S1A Fig). Participant Z1658F was determined to have been infected with two TF viruses (S1 Fig). We additionally utilized SGA and SMRT sequencing to amplify and sequence an approximately 3.6 kb amplicon spanning the *vpu*, *env*, and *nef* genes from plasma samples collected approximately one year following the EDI, as well as from the last pre-therapy time point available (Fig 1). For six individuals, we additionally amplified and sequenced proviral DNA from the last pre-therapy time point to assess the divergence of sequences from the TF virus when collected from the cellular compartment rather than plasma.

For all participants, the *vpu-env-nef* amplicon was also generated by SGA from DNA of peripheral white blood cells collected during treatment; for four individuals, proviral sequences were also sampled at a second, later time point during ART (Table 1 and Figure 1). All sequences analyzed across time points were free of obvious defects such as nonsense mutations, INDELs resulting in disruption of an open reading frame, or in-frame INDELs of more than 90 nucleotides. In total, 1,376 sequences, excluding APOBEC hypermutants, were generated and analyzed for all participants and time points (S1 Table).

Reservoir variants are distributed throughout the phylogenetic trees of viral sequences from each individual

For all 13 participants, a maximum-likelihood phylogeny was inferred from within-host alignments of plasma and proviral DNA *vpu-env-nef* sequences from all time points, where the tree was rooted on the inferred TF virus (Fig 2 and S2 Fig). This permits insight into the divergence of all descendant variants in each population from the TF virus, as well as the inferred evolutionary relationships between them. Divergence of plasma sequences from the TF virus by patristic (root-to-tip) distances increases significantly from one-year post-infection to

the last ART-naïve time point, yet there is a significant decrease in the patristic distances of proviral sequences sampled during treatment compared to those sampled at the last ART-naïve time point (Fig 3). This decrease does not appear to be an artifact of analyzing sequences from cells versus plasma, as there is no significant difference in the patristic distances from the TF virus of variants collected from cells and plasma at the same time point in six participants assessed (S3 Fig). Sequences with minimal divergence from the TF virus are persisting in the reservoir, as can be seen for participants Z1123M and Z326M (Fig 2A, S2F Fig, Fig 3A), where sequences identical to the TF were recovered in the reservoir, having persisted for over two years and six years of ART-naïve infection, respectively.

Although proviral variants closely related to and including the TF virus persist in the reservoir, reservoir variants are distributed throughout the viral phylogenies for each individual. Plasma variants from one year post-EDI and the last pre-therapy time point exhibited a "ladder-like" topology characteristic of within-host phylogenies [34], where plasma sequences from a given time point formed distinct clades in all intrapatient trees, and reservoir variants fell within or between these clades. In our initial analysis, we classified individual reservoir variants as being most closely related to variants of the clade within which they fall, or as intermediate in cases where they do not fall within the one-year or last ART-naïve clades (Fig 3C). Additionally, variants that did not fall within either the one-year post-infection or last ART-naïve clade and were within six nucleotide changes from the TF virus were classified as seroconversion variants, as they exhibit divergence from the TF virus equivalent to the level of diversity among sequences from the serconversion time point. In 10 of 13 individuals, we observed at least one proviral variant classified as most closely related to seroconversion or one-year post-EDI plasma variants, while in all participants we observed proviral variants most closely related to the last

ART-naïve plasma variants, with these sequences making up the greatest portion of the total proviral populations overall (Fig 3C). Taken together, we consider the presence of TF or seroconversion variants, as well as variants classified as intermediate and chronic, to indicate archiving of viral sequences throughout infection.

Dating of provirus integration indicates variant archiving throughout infection

While visualization of within-host phylogenies inferred from pre-therapy plasma sequences and proviruses persisting during ART allowed us to roughly estimate the era in which the latter integrated into the reservoir, more nuanced estimates were needed. To more precisely estimate the age of proviral variants with respect to the ART-naïve infection, we applied the method developed by Jones et al. [32], which utilizes pre-therapy sequences to develop a model of continuous within-host evolution relative to sampling time, and place reservoir variants at a distinct date along the infection history, rather than into a broad category based on relatedness to discrete pre-therapy sampling. For this analysis, we inferred maximum-likelihood phylogenies from pre-therapy plasma and reservoir proviral *env* sequences, where the root was placed at the location that maximized the correlation between the divergence from the root and sample collection date of the pre-therapy plasma sequences (Fig 4A, D). The pre-ART plasma variants were then used to train a linear model that related their divergence from the root and their sample collection dates. The linear model was used to infer the integration date and 95% confidence interval of the age of each proviral variant based upon its divergence from the root (Fig 4B-C, E-F).

Consistent with reservoir proviral variants being seeded at various times spanning infection to treatment initiation, there is considerable discrepancy between proviral sample collection dates and their inferred integration dates, with some variants estimated to have been

integrated near the time of seroconversion. The point estimates of integration dates for Z1165M indicate a variant was archived within one month of the root date (Feb 5, 2006), falling within the date estimates for seroconversion sequences (Jan 24, 2006 to Mar 25, 3006). In contrast, several variants displaying considerably higher divergence from the root were present as well, including those with estimated integration dates consistent with the last ART-naïve plasma variant date estimates. Consistent with our initial analysis, integration date estimates for both Z1165M and Z634F proviral variants supported periodic seeding of variants in the reservoir throughout the infection.

To assess whether the intrahost viral evolution in these participants reliably follows a molecular clock from transmission at a given point in time relative to sample collection, we used Bayesian approaches to infer the root date of the infection directly from sequence data, using pre-therapy *env* and *nef* sequences without evidence of hypermutation or recombination. Bayesian-estimated root dates of Jan 13, 2006 for Z1165M and Jun 29, 2007 for Z634F fit within the corresponding last negative and first positive HIV test dates for both participants, and were within 14 days of the clinically estimated dates of infection (Jan 26, 2006 and Jul 8, 2007, respectively) (Fig 5). These highly concordant estimated dates of infection ensure that the data derived from sampling at discrete time points within the infection nonetheless represent virus clearly evolving in keeping with a molecular clock from a well-defined transmission event.

Repeated sampling during ART and persistence of early infection variants

The dynamics of proviral decay during short-term ART could be considered to bias the results of this investigation, as proviral DNA decay within the three months to approximately two years following treatment initiation sampled here falls within an initial, more rapid phase of decay [35-38].We therefore may be sampling provirus that persists only transiently rather than

comprising the more stable population of latently infected cells with a slower decay rate. To determine if the distribution of sequence variants in these participants was influenced by the relatively short time on treatment, we sampled an additional time point six months to a year further into treatment in four participants. All proviral variants without APOBEC hypermutation were included in phylogenetic trees of the total pre-therapy and during treatment sequences, and phylogenies were again rooted on the TF virus (Fig 6). Early variants were observed throughout the repeated sampling during ART, as proviral variants classified as seroconversion variants were in both the first and second time point during treatment for participant N133M (Fig 6B). For participant Z1165M, a single proviral sequence most closely related to seroconversion sequences was observed in the second time point during treatment (Fig 6C), and in participant Z1788F, seroconversion and early infection variants differing from the TF virus by up to approximately 30 nucleotides were found during the first and second time points following treatment (Fig 6D). Overall, however, proviral sequences from both time points during treatment were intermingled with each other and the sequences from one-year post-infection and the last ART-naïve time point.

To facilitate combining data across participants, root-to-tip distances of each reservoir sequence were normalized to the participant's total tree height. Comparison of these normalized root-to-tip distances by sampling time point during ART revealed significantly shorter patristic or root-to-tip distances for the later reservoir samples compared to the earlier ones (Mann-Whitney, $p=0.0007$). This suggests that, with ongoing treatment, early viral variants are perhaps becoming enriched in the reservoir during ART.

Discussion

We observed that proviral sequences from 13 individuals who had undergone short-term ART were distributed among pre-therapy sequences in phylogenetic trees, with the majority of proviral sequences most closely related to variants from the last ART-naïve time point. However, as analysis of the estimated time of integration for proviral sequences indicates, there is archiving of variants throughout ART-naïve infection, from the earliest time of infection to treatment initiation. This finding is consistent with previous work by Jones et al. [32], but extends the stages of pre-therapy infection explored to acute infection. We identified TF viruses from acute infections with longitudinal follow-up through chronic ART-naïve infection and treatment initiation, while the pre-therapy samples of the two HIV-1 infections investigated in Jones et al. are from chronic infections [32]. Archiving of variants throughout ART-naïve infection is complementary to the observations that the reservoir is smaller and less diverse in individuals beginning treatment early in infection versus during chronic infection [37, 39, 40], since preventing replication with ART ensures a halt in viral evolution and concomitant latent infection with progressively more diverse variants.

Within the diverse populations of proviral sequences we observed, we identified variants that were identical to or contemporaneous with the TF virus after as many as six years of ART-naïve infection and following six to 12 months of ART. These very early, TF-related sequences were observed in 5/13 of the individuals sampled and represented from 2.5-8.0% of all reservoir variants in those individuals. It is clear that these very early viral sequences can persist for several years in the absence of therapy, consistent with their integration in long-lived CD4 T cells. Persistence of ancestral variants is not unprecedented, as several studies assessing drug resistance in patients receiving virologically suppressive ART after a history of non-suppressive

therapy found that both ancestral, drug susceptible virus and variants with resistance mutations persist during years of effective treatment [41-44].

Recent studies have shown that a majority of proviruses persisting during ART exhibit large internal deletions or other defects, such as nonsense mutations resulting from APOBEC-induced hypermutation, which render the provirus defective [22, 24]. Due to sample limitations, we assessed approximately one-third of the genome encompassing the *vpu*, *env*, and *nef* genes, and thus cannot exclude the possibility that sequences we have observed as exact matches to the TF virus in this amplicon might contain differences elsewhere in the genome, including mutations and/or deletions that would prevent viral replication. Nevertheless, all of the sequences used for analysis do represent biologically functional gene regions, since sequences with INDELs and frameshifts were excluded. Unlike Abrahams et al. [31], who used QVOA to characterize reactivatable sequences, we are not exclusively addressing the replication-competent reservoir. However, QVOA are known to underestimate the size of the reservoir, as the bulk of replication-competent proviruses are not induced with single or successive rounds of stimulation [22]. Phylogenetic assessment of HIV-1 DNA during virologically suppressive ART serves to address the broad population of persistent provirus within which the replication-competent reservoir is contained, and address its relationship to pre-therapy virus.

In addition to containing TF virus or very early infection variants in some individuals, reservoir proviral populations were overall less evolved from the TF virus than the sequences at the last ART-naïve time point (Fig 3). This finding may be influenced by the short duration of treatment, as all individuals studied here received ART for less than three years at the time of sample collection during treatment, and three participants were sampled within six months of treatment initiation while the reservoir is less stable. However, we did find that early infection

variants persisted with continued time on treatment in the four individuals sampled twice while receiving ART. Furthermore, sequential sampling indicated that with continued time on treatment, the distance of reservoir variants from the TF virus significantly decreased (Fig 7), indicating an enrichment for variants dating to earlier in the course of the infection. As viremia rapidly declines in a first phase of viral decay following treatment initiation, followed by a second, slower decay phase [45], latently infected cells decay in stages [35-38], perhaps with those infected most recently by variants circulating in the plasma just prior to treatment initiation decaying first. This mechanism would be consistent with the observation that CD4+ central memory T cells from six to 12 years of ART harbor HIV-1 DNA most closely related to early infection sequences [23]. Further studies must address the phylogenetic influence of latently-infected cell decay.

As HIV-1 prevention and treatment efforts are scaled up globally, research efforts to reduce and/or eliminate the reservoir in pursuit of an HIV-1 cure are expanding as well. Towards this goal, it is critical to characterize the genetic diversity of the reservoir to assess the variants that HIV-1 eradication strategies must target. Our findings indicate that virus is archived throughout infection, and cure strategies should therefore address the genetic diversity of reservoir proviral quasispecies with many unique variants, including those dating back to the time of transmission.

Materials and methods

Human subjects

Zambian volunteers were enrolled as heterosexual couples in Couples Voluntary Counseling and Testing (CVCT), with HIV testing and counseling of both partners conducted upon enrollment. Follow-up HIV testing was conducted approximately every three months for

the negative partners of serodiscordant couples, and blood samples were collected from both partners in the event of a positive test as a component of the Zambia-Emory HIV Research Project (ZEHRP). Human subjects protocols for ZEHRP were approved by the University Teaching Hospital Ethics Committee in Lusaka, Zambia, while additional approval for sample or data use was granted by the Institutional Review Boards of Emory University, Simon Fraser University, and Providence Health Care/University of British Columbia. Informed consent for sample collection was obtained for each volunteer upon enrollment in CVCT. Dates for ART initiation were self-reported, and clinically-based estimated dates of infection were calculated with the appropriate formula of the three following: 1. midpoint of dates for the last antibody negative and first antibody positive test; 2. Fourteen days prior to the first p24 antigen positive, antibody negative test; 3. Ten days prior to the first viral load >1600 copies/mL, antibody negative test. Participants Z1123M, Z1047M, and Z1808F had first HIV+ tests that were not antibody positive, instead VL >1600 copies/mL while p24 negative and antibody negative for Z1123M, and p24 positive while antibody negative for participants Z1047M and Z1808F. All other participants had antibody positive tests for their first HIV+ test.

Nucleic acid extraction and cDNA synthesis

Viral RNA in plasma samples was extracted using the QIAamp Viral RNA Mini Kit (QIAGEN) or E.Z.N.A Viral RNA Kit (Omega Bio-Tek) according to the manufacturer's instructions. Briefly, 150 uL plasma was lysed with buffer and centrifuged through a silica column, which was then washed with appropriate buffers. RNA was eluted in >60 uL Buffer AVE (QIAGEN) or DEPC H2O (Omega). RNA served as template in cDNA synthesis reactions described below.

Eleven microliters of viral RNA were used in each 20 uL reverse-transcriptase reaction for cDNA synthesis utilizing SuperScript III or IV Reverse Transcriptase (Invitrogen) according to the manufacturer's instructions, but with an extension time of up to one hour. SuperScript III protocols additionally included a 4°C pause following the one hour extension time for addition of 200 Units RT enzyme proceeding a second extension for two hours at 55°C. Both SuperScriot III and IV protocols included RNase H digestion of RNA-DNA heteroduplexes with 20 min incubations at 20°C. Oligo dT (5'-TTTTTTTTTTTTTTTTT-3') or 1.3'3'PlCb (5'-ACTACTTAGAGCACTCAAGGCAAGCTTTATTG-3') primers were used as anchors in the reactions, and cDNA was directly used in PCR or frozen for subsequent use.

Nucleic acids were extracted from cells using the QIAamp DNA Blood Mini or Midi Kit (QIAGEN) according to the manufacturer's instructions. Briefly, white cell pellets of total white blood cells in RNAlater were processed for lysis in QIAGEN Protease or Proteinase K and lysis buffer. Following addition of 100% ethanol, lysate was applied to a silica column and centrifuged, following by washing of the column. Samples were eluted in QIAGEN buffer AVE and used directly in PCR or frozen for subsequent use.

PCR and amplicon purification

PCR for single genome amplification (SGA) of near full-length genomes (NFLGs) consisted of two rounds of PCR utilizing appropriate template for ≤40% positive reactions of approximately nine kilobases as visualized by gel electrophoresis. Each round of PCR consisted of 25 uL reactions with 0.5 Units Q5 Hot Start High-Fidelity Enzyme (NEB), 1x Q5 Reaction Buffer, 1x Q5 High GC Enhancer, 350 μM each dNTP, 500 nM each primer, template, and nuclease-free H2O to reach 25 uL. PCR primers are described in Rousseau 2006 [39] for both first and second rounds, and first round primers are as follows: 1.U5Cc (5'-

CCTTGAGTGCTCTAAGTAGTGTGTGCCCGTCTGT-3', forward primer) and 1.3'3'PlCb (5'-ACTACTTAGAGCACTCAAGGCAAGCTTTATTG-3', reverse primer). Second round PCR primers are as follows, with 1 uL of first round PCR product used as template in the second round PCR: 2.U5Cd (5'-AGTAGTGTGTGCCCGTCTGTTGTGTGACTC-3', forward primer) and 2.3'3'plCb (5'-TAGAGCACTCAAGGCAAGCTTTATTGAGGCTTA-3', reverse primer). Both first and second round PCR utilized the following program: 98°C for 30 sec, 35 cycles of 98°C for 10 sec and 72°C for 7:30 sec, 72°C for 10:00, and 4°C forever (end). Amplicons were purified with the Wizard SV Gel and PCR Clean-Up System (Promega) according to manufacturere's instructions, eluting in H2O.

For amplification of vpu, env, and nef gene amplicons, two rounds of PCR were used as above for NFLG amplification. Reactions were 20 uL with first round primers: Vif1 KB (5'-GGGTTTATTACAGRGACAGCAGAG-3', forward primer) and Ofm19 (5'-GCACTCAAGGCAAGCTTTATTGAGGCTTA-3', reverse primer). First round PCR product (0.8 uL) was used a template for second round PCR with the following primers: EA1F KB (5'-GCTTAGGCATYTCMTATGGCAGGAAGAAG-3', forward primer) and O1R (5'-AAAGCAGCTGCTTATATGCAGCWTC-3', reverse primer). Amplicons were purified with the NucleoSpin Gel and PCR Clean-Up (Takara), eluting in Elution Buffer NE or H2O.

Next-generation sequencing

All sequencing was performed with Pacific Biosciences SMRTbell sequencing on the RS II, with individual DNA libraries run on a single SMRT cell. Libraries were generated with 30-60 NFLG amplicons combined at eqiumolar concentrations and identified by nucleotide barcode following reamplification of first round PCR products with barcoded primers, or by barcoded adapter from the SMRTbell Barcoded Adapter Complete Prep Kit-96 (Pacific Biosciences).

Libraries for vpu, env, and nef amplicons were generated with 80-100 amplicons per library and identified with barcoded adapters. Libraries were made according to manufacturer's instruction, followed by appropriate size selection with the BluePippin (Sage Science). We are greatly appreciative of library size selection and quality control, as well as sample run of libraries on the RS II performed at the University of Delaware Sequencing and Genotyping Center.

Amplicon reads generated from the SMRT sequencing were analyzed with a unique algorithm to perform read phasing and error correction in generation of final sequences [33]. Libraries generated using Pacific Biosciences barcoded adapters were first analyzed with PacBio SMRT analysis software PB Barcode to separate reads by barcoded adapter prior to additional read phasing and error correction with the algorithm described in Dilernia et al.

Phylogenetic trees and reservoir variant dating

Maximum-likelihood phylogenetic trees for complete amplicons lacking frameshifting INDELs, APOBEC hypermutation, or other deleterious mutations were made with the PhyML plugin [47] of Geneious software v9.0.4 [48] using a general time reversible model with six nucleotide substitution categories and gamma distribution parameter with 100 bootstraps. Trees were rooted on the appropriate transmitted/founder virus sequence trimmed to the *vpu*, *env*, and *nef* gene amplicon. Patristic distances from the TF virus were extracted from the distance matrix. Trees were edited for visualization with MEGA v7.0.26 [49]. Statistics for patristic distance from phylogenetic trees were performed using Prism. APOBEC hypermutants were first removed from analysis with the LANL Hypermut v2.0 tool [50] with appropriate TF virus as the reference sequence, and all sequences of $p<0.05$ considered hypermutated.

For two participants, we estimated the root date of their within-host plasma HIV-1 RNA sequences using established Bayesian methods. Briefly, within-host pre-therapy plasma HIV-1

nef and *env* sequences were first screened for hypermutation (using Hypermut v2.0) and recombination (RDP v4.95 [51]) and any hypermutated or within-host recombinant sequences were removed. Remaining within-host HIV-1 *env* and *nef* sequences from the same original template were then concatenated. We ran two parallel 100,000,000 length chains sampling every 10,000 states in the software package BEAST v1.10.4 [52] for each participant with two partitions (*nef* and *env*). Posterior distributions for the root date were estimated using unlinked SRD06 substitution models [53], unlinked uncorrelated relaxed lognormal clock models [54] and a linked coalescent Bayesian skyline tree model [55] with two groups using a piecewise-constant model. After discarding up to 33% of the initial run as burn-in, the chains from parallel runs were combined with LogCombiner v2.5.2 [56] and analyzed in Tracer v1.7.1 [57] to ensure convergence and verify that effective sample size values were >200 for all parameters.
Proviral variant integration dates were estimated as previously described [32]. Briefly, *env* genes were trimmed from seroconversion and all other pre-therapy variants, as well as proviral variants from samples collected during treatment. Any sequences demonstrating hypermutation or recombination were excluded from analysis, and only the earliest variant of duplicate sequences was kept for analysis. Maximum-likelihood trees were generated with RAxML v8.2.12 [58] and trees were rooted with root-to-tip regression (RTT) using the R package ape v5.3 [59] to maximize the correlation between the divergence from root and the sample collection date of the pre-therapy sequences. The pre-therapy variants were used to train a linear model of the divergence from root and the sample collection date. Finally, the integration dates date and confidence intervals of the proviral variants were estimated from this model.

Acknowledgements

We would like to thank all the ZEHRP study participants, as this work would not have been possible without them. Additionally, we are grateful for the support of Jon Allen and Charlott Morel Sanchez at the Emory Vaccine Center for sample management, and Dr. Paul Farmer at the Emory Vaccine Center for technical assistance regarding the ZEHRP cohort. Furthermore, we are indebted to the efforts of Olga Shevchenko at the University of Delaware Sequencing and Genotyping Center for all Pacific Biosciences sequencing.

Table 1. Participant details.

Participant ID	Last HIV- Test Date	First HIV+ Test Date	Estimated Date of Infection (EDI)[a]	EDI to SC Sample (days)	EDI to ART Start (yrs)	ART Regimen	Time on ART to ART Sample 1, 2 (yrs)
Z1094F	08-Apr-2008	04-Jul-2008	21-May-2008	44	2.75	EFV, TDF, FTC	0.48
Z1123M	14-Mar-2008	12-Apr-2008	02-Apr-2008	10	2.86	EFV, TDF, FTC	0.91
N133M	16-May-2007	16-Aug-2007	01-Jul-2007	46	2.87	EFV, TDF, FTC	0.56, 1.27
Z1788F	10-Nov-2006	10-Feb-2007	26-Dec-2006	46	2.98	EFV, TDF, FTC	0.72, 2.09
Z634F	22-Jun-2007	24-Jul-2007	08-Jul-2007	16	3.11	EFV, TDF, FTC	0.75, 1.40
Z2006M	09-Apr-2009	13-May-2009	26-Apr-2009	17	3.32	NVP, AZT, 3TC/EFV, TDF, FTC	0.32
Z1047M	24-May-2007	14-Aug-2007	31-Jul-2007	24	3.66	EFV, TDF, FTC	0.5
Z1165M	08-Dec-2005	16-Mar-2006	26-Jan-2006	49	3.82	ABC, 3TC, EFV	1.36, 2.14
Z1658F	11-Mar-2006	08-Jun-2006	24-Apr-2006	46	4.63	ABC, 3TC, NVP	0.89
Z1808F	18-Dec-2007	24-Jan-2008	10-Jan-2008	14	4.68	EFV, TDF, FTC	0.31
Z1124F	15-Feb-2006	18-May-2006	02-Apr-2006	46	5.22	ABC, 3TC, NVP	0.54
Z1044M	29-Nov-2005	25-Mar-2006	26-Jan-2006	58	5.23	AZT, 3TC, EFV	0.33
Z326M	21-Nov-2006	17-Feb-2007	04-Jan-2007	44	6.22	EFV, TDF, FTC	0.48
MEDIAN	**NA**	**NA**	**NA**	**44**	**3.66**	**NA**	**0.54, 1.75**

SC=Seroconversion; N=Ndola, Zambia; Z=Lusaka, Zambia; M=Male; F=Female

[a]See materials and methods for calculation of EDI

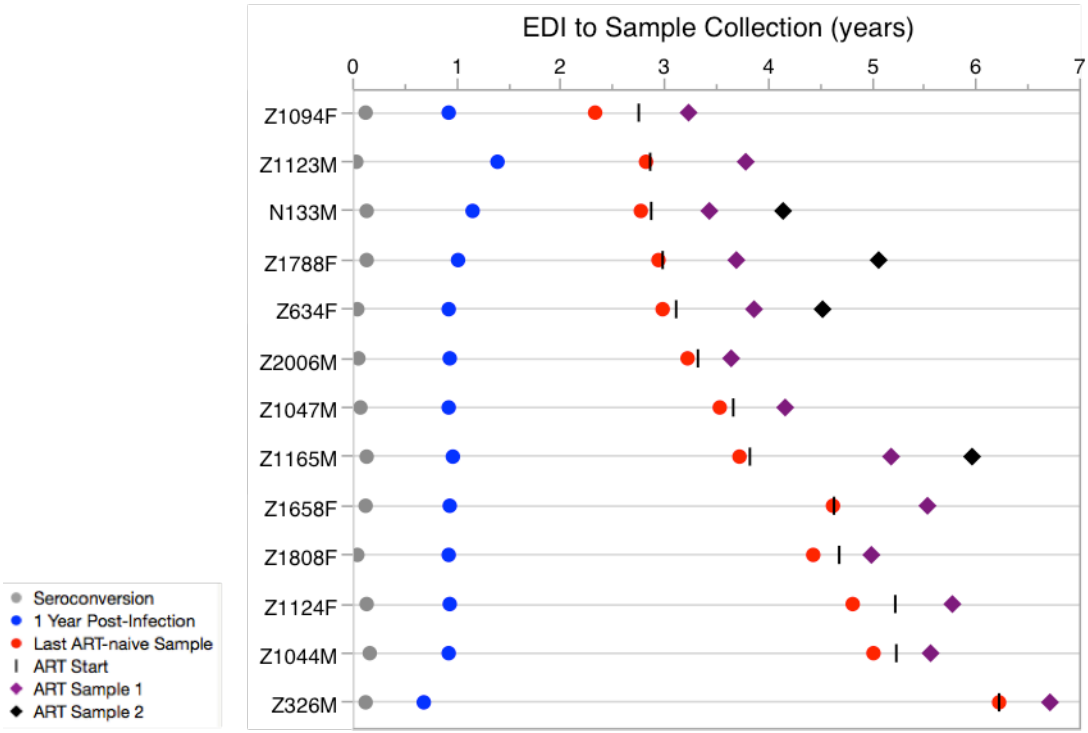


Fig 1. Sampling strategy for study participants. Viral RNA was reverse transcribed, amplified, and sequenced from pre-therapy plasma samples, while proviral DNA was amplified from cells collected during treatment. For six individuals (IDs Z1094F, Z1123M, Z1788F, Z1047M, Z1165M, and Z1658F), we additionally amplified and sequenced proviral DNA from cells at the last pre-therapy time point to assess divergence from the TF virus of contemporaneous sequences from plasma vs. cells. EDI=Estimated Date of Infection

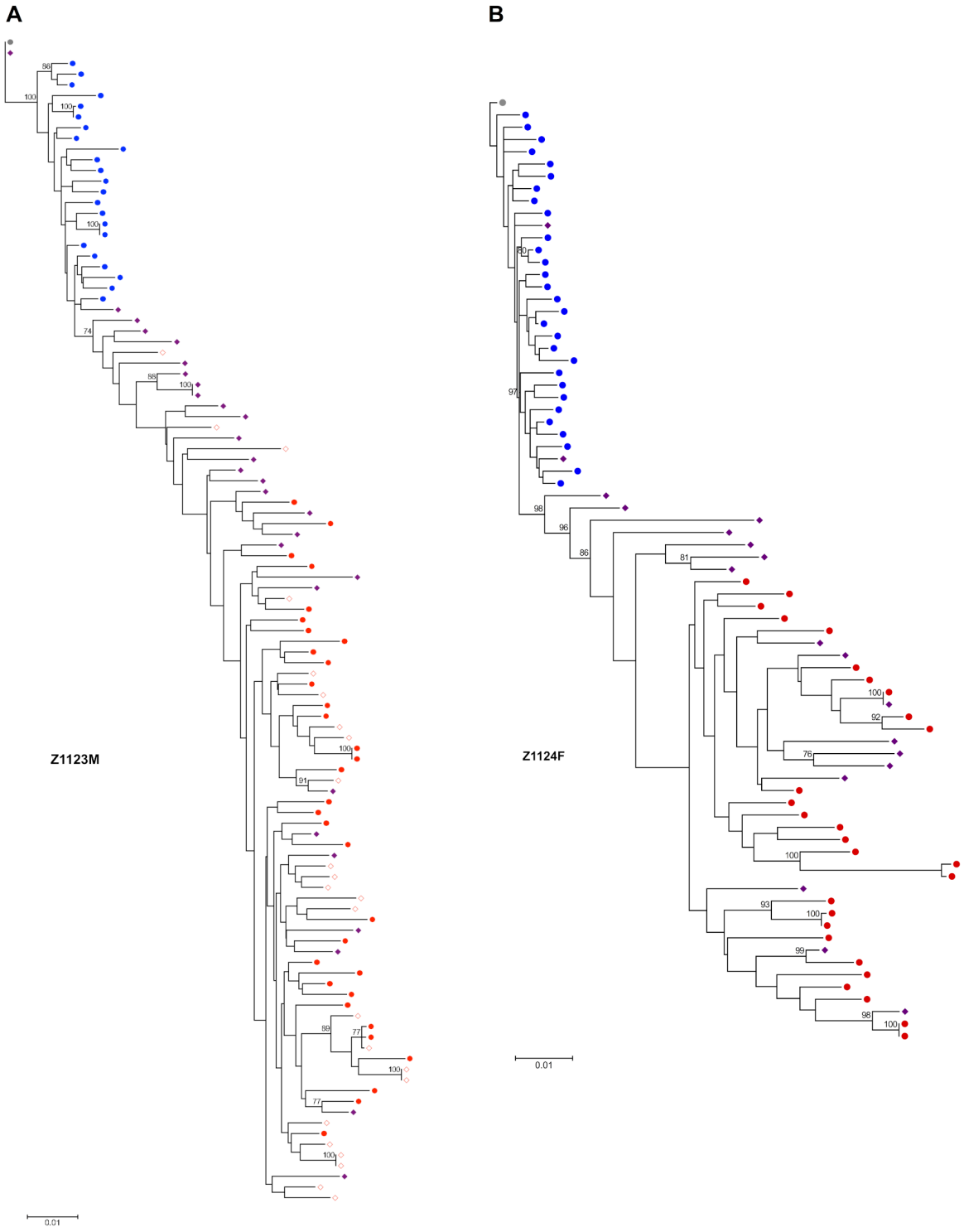


Fig 2. Maximum-likelihood (ML) trees for viral and proviral variants in two individuals.

Representative ML phylogenetic trees for participants Z1123M (A) and Z1124F (B) rooted on

the respective TF virus (grey) identified from the seroconversion sample and depicting all viral variants from one year post infection (blue), the last ART-naïve sample (red), and during treatment (purple diamonds). Variants from cells collected at the last ART-naïve time point are shown in open red diamonds, while all plasma variants are in filled circles. Nodes with bootstrap support values >70 are shown in text.

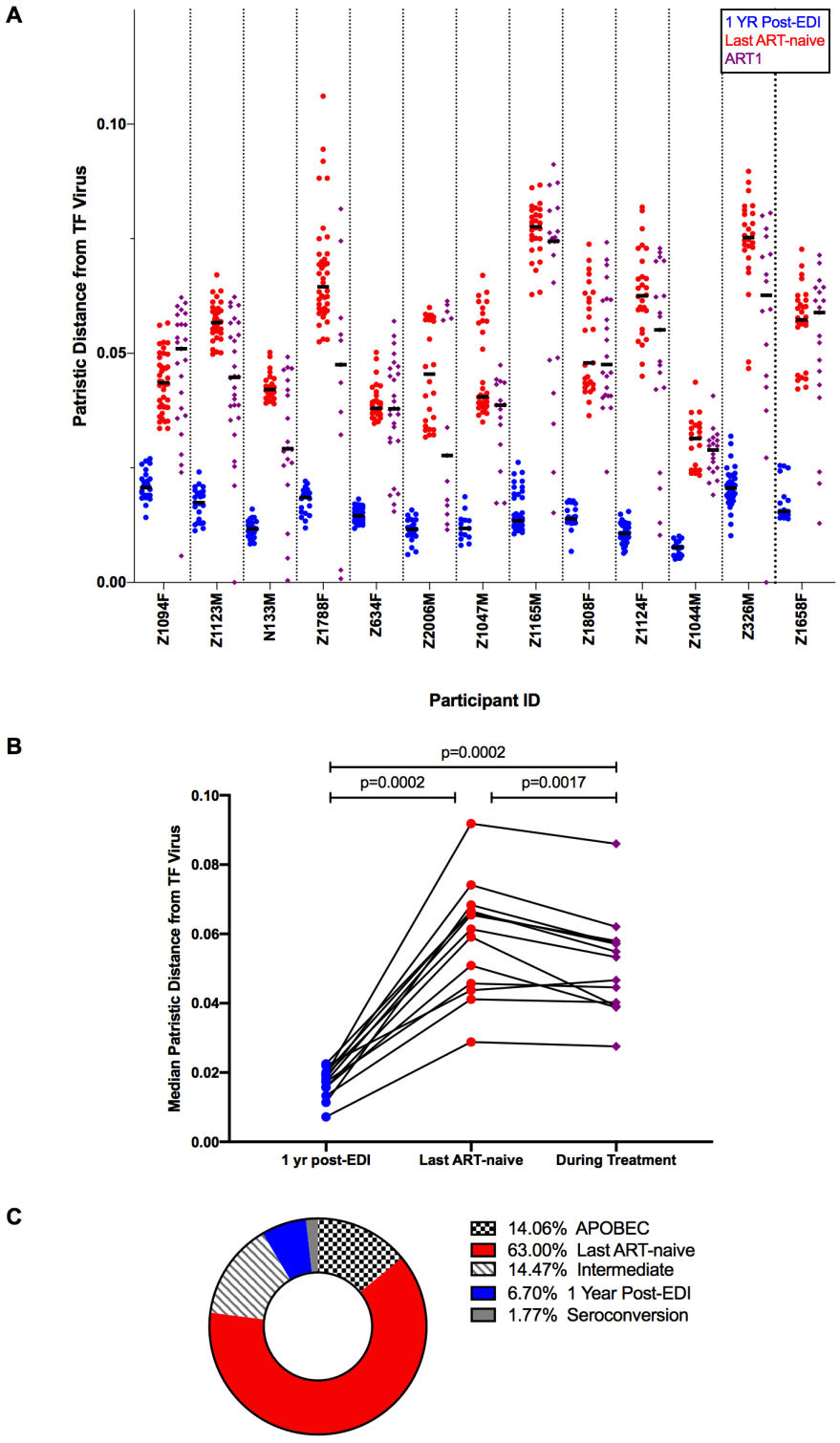


Fig 3. Divergence from TF increases during ART-naïve infection and decreases on treatment.

(A) Patristic distances from the TF virus, or root-to-tip values, from the ML trees are shown for

all 13 participants, with the single instance of multivariant infection in participant Z1658F shown at the far right. Where two samples were assessed during treatment, only the first is shown here. (B) Summary of the median intrapatient patristic distances, where the median distance is significantly different between each time point assessed (Wilcoxon matched-pairs signed rank test). (C) Proportion of proviruses seeded into the reservoir, by era, as estimated from the placement of reservoir sequence in the phylogeny (values represent the median of all participants). A majority of variants are most closely related to variants present at the last ART-naïve time point. The percentage of variants demonstrating APOBEC hypermutation is also indicated, though these sequences were excluded from analysis.

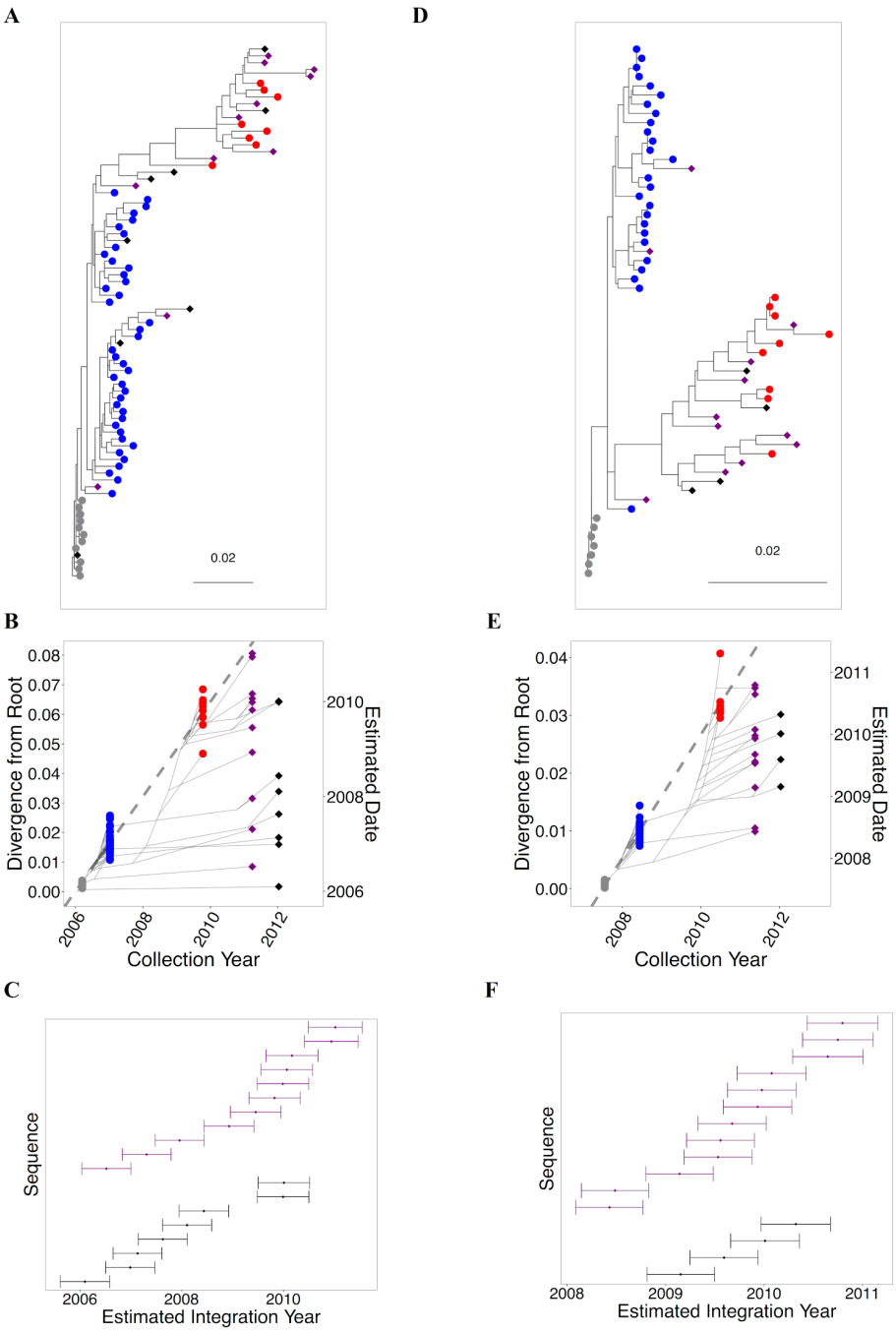


Fig 4. Regression-based inference of time of provirus integration. Figures for participant Z1165M are shown in A-C, and D-F represent participant Z634F. Maximum-likelihood trees of

the env gene for pre-therapy variants (circles), including individual seroconversion variants (grey), one-year sequences (blue), last pre-therapy sequences (red), and proviral variants (diamonds) in A and D. Two samples during treatment were assessed for both participants, with the first in purple, and subsequent in black. Trees were rooted to optimize the correlation between root-to-tip distance and sampling time for all pre-therapy plasma variants. The linear model relating root-to-tip distances to sampling time is shown in the dashed lines of figures B and E, with the pre-therapy variants denoted as colored dots, and the phylogenetic relationships between them denoted as faint grey lines. Proviral variants from samples collected during treatment are shown in diamonds in the same manner. The estimated integration dates of these proviral variants and 95% confidence intervals are shown in the plots C and F.

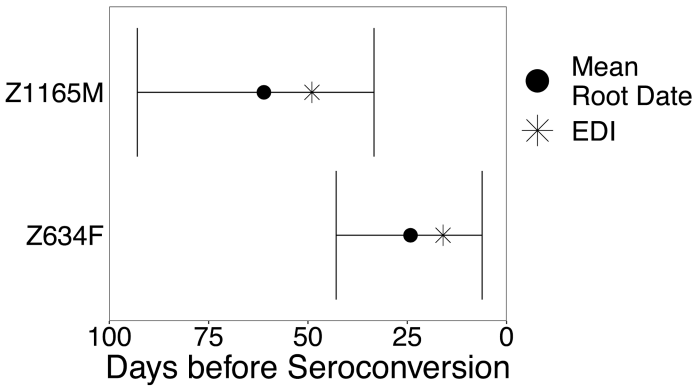


Fig 5. Bayesian root date estimation. Bayesian inference was used to estimate the root date, or time of infection, from pre-therapy plasma env and nef sequences. The 95% highest posterior density intervals surround the mean estimated root dates of the Bayesian inference (circles). The clinically estimated infection dates (EDIs) are depicted as stars. Data are shown for Z1165M and Z634F. The seroconversion dates correspond to the seroconversion sample dates of March 16, 2006 for Z1165M, and July 24, 2007 for Z634F.

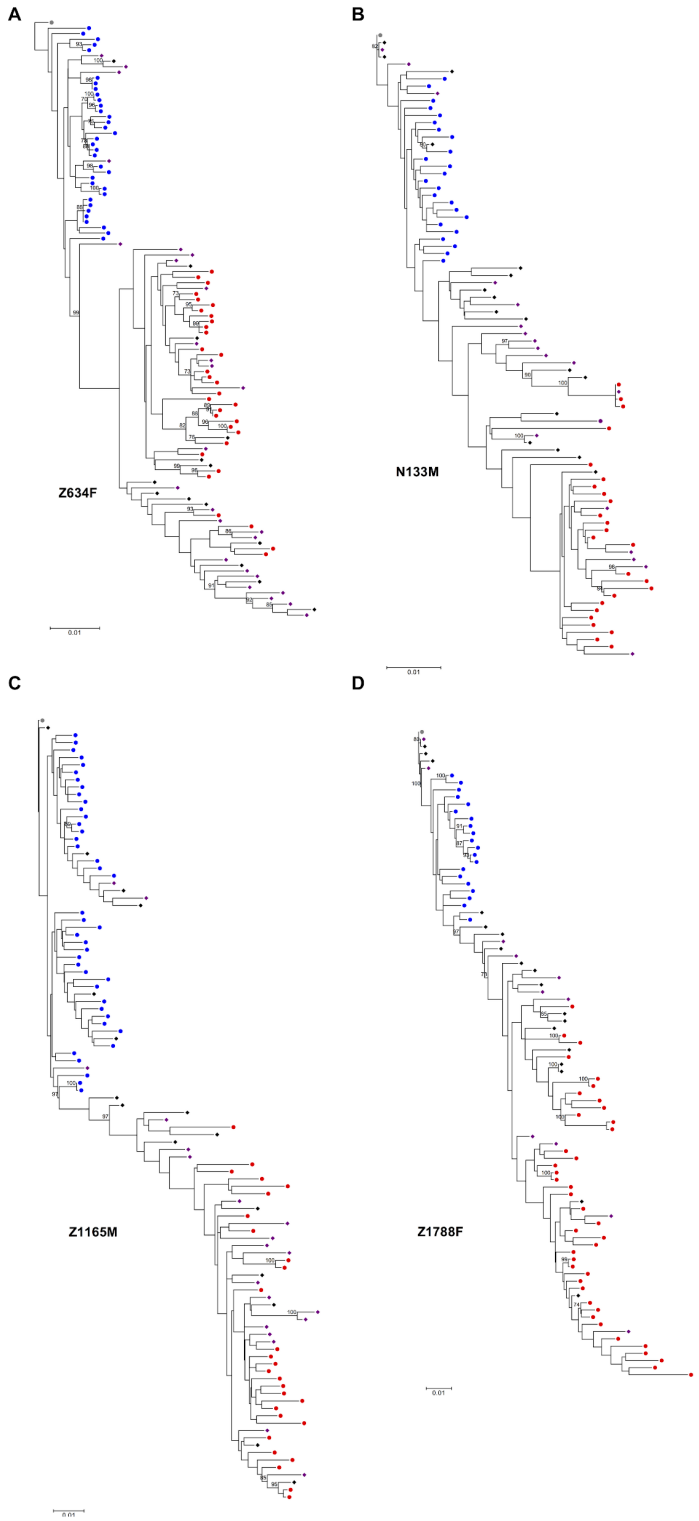


Fig 6. ML trees of all intrahost variants with two samples during treatment. ML phylogenetic trees for all four participants: Z634F (A), N133M (B), Z1165M (C), and Z1788F (D) rooted on

the respective TF virus (grey) and depicting all viral variants from one year post- infection (blue), the last ART-naïve sample (red), and during treatment (purple and black diamonds, with second sample in black). Bootstraps >70 are shown in text.

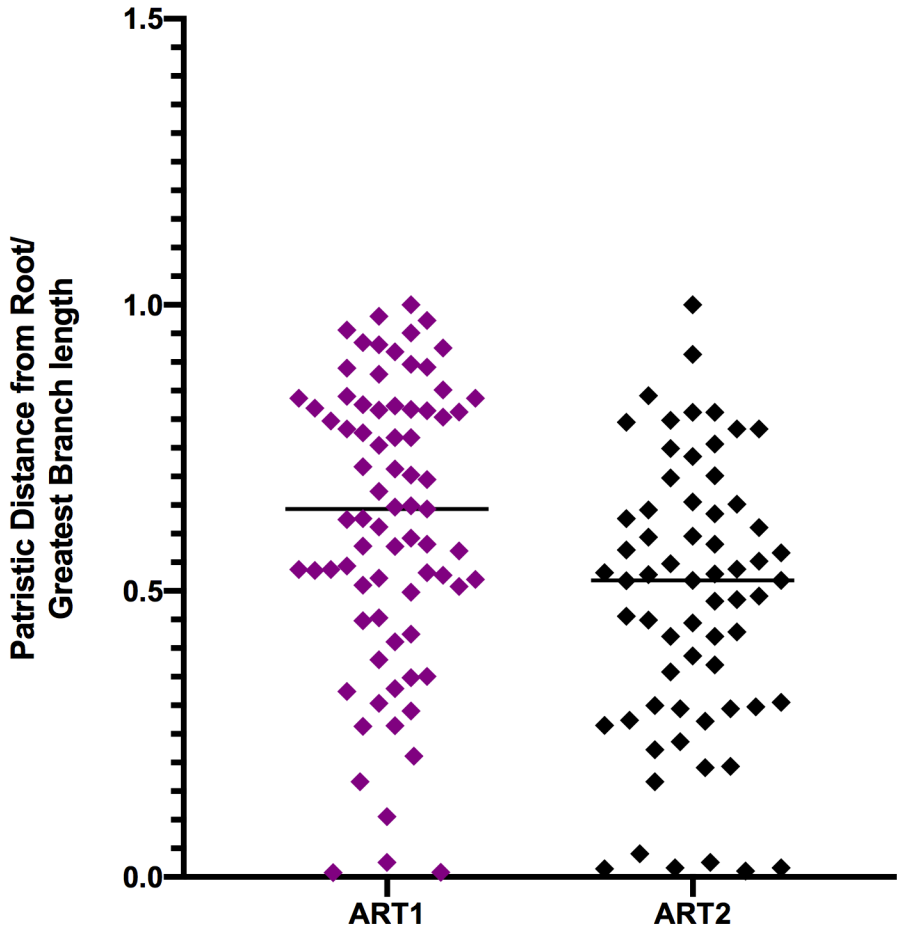
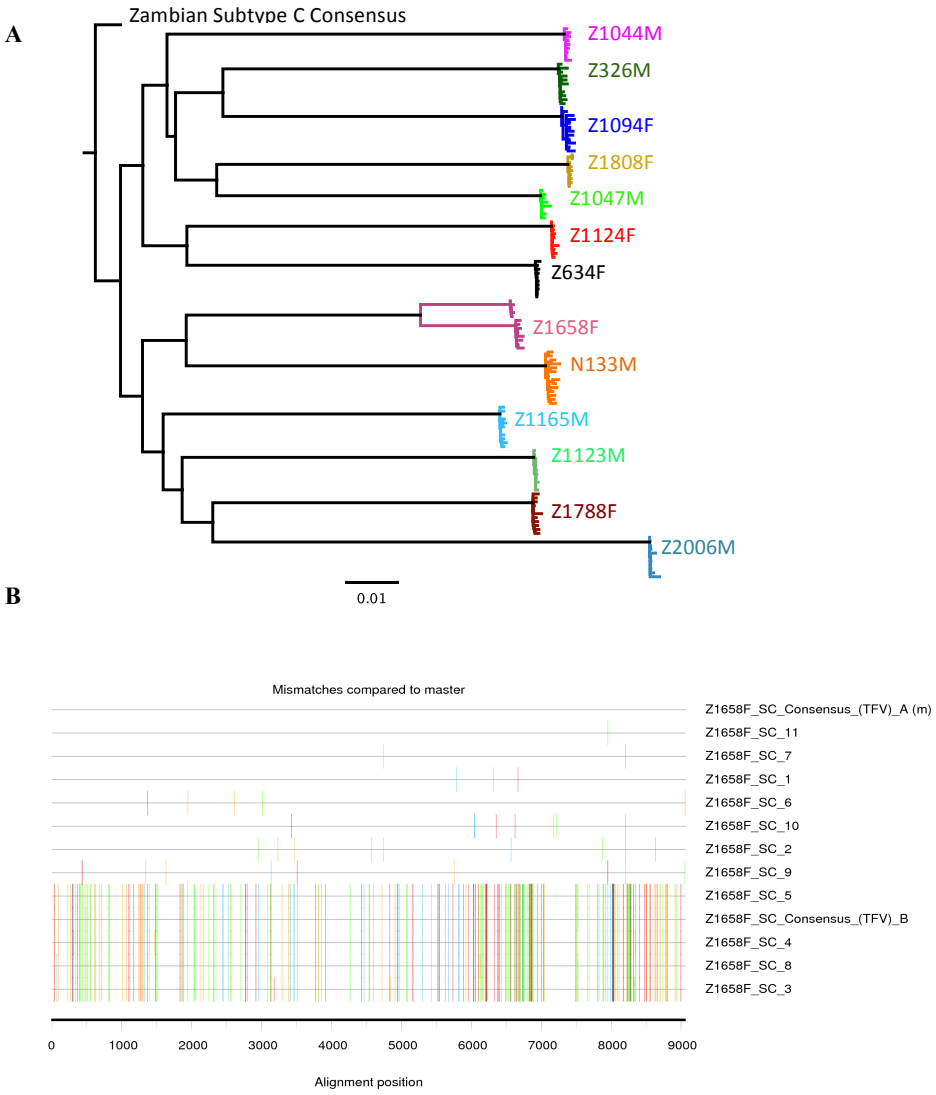


Fig 7. Distance from TF virus decreases with subsequent sampling during treatment. Patristic distances from the TF virus for reservoir variants as a proportion of the greatest intrapatient patristic distance. Distances are significantly lower for variants sampled at the second time point during treatment compared to the first (Mann-Whitney, p=0.0007).

S1 Table. Extended sampling details.

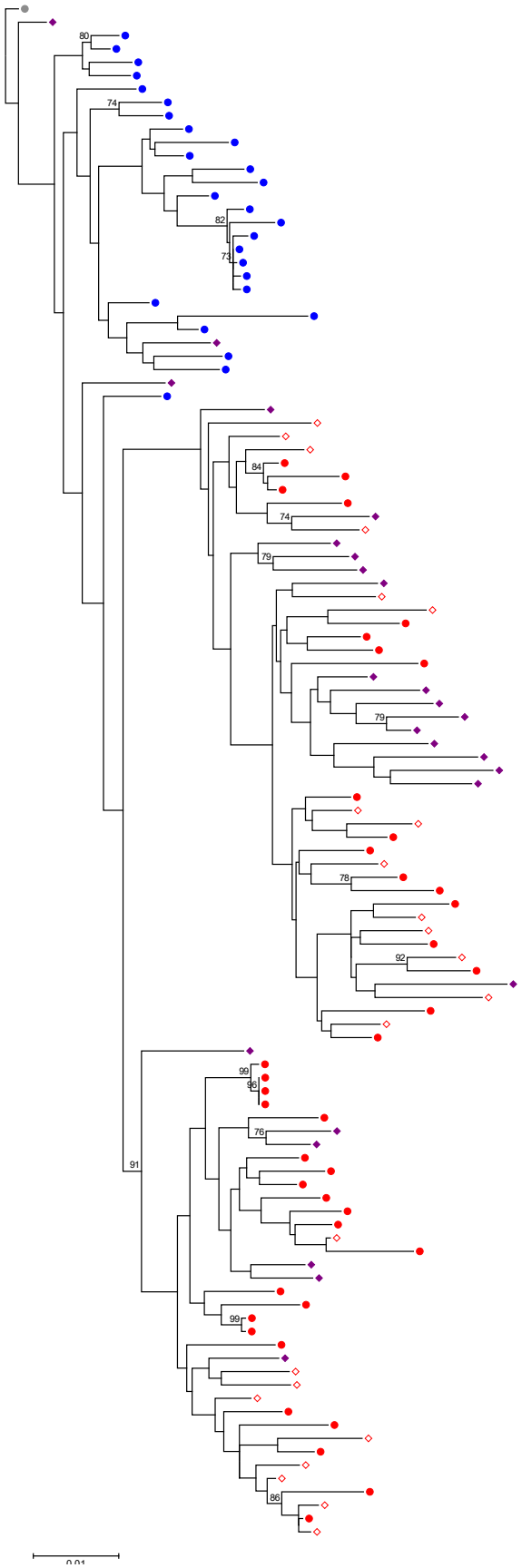
Participant ID	Estimated Date of Infection (EDI)	First HIV+ Test Date	vRNA/ p24/ Ab at First HIV+ Test	SC Sample VL (copies/ mL)[a]	SC Sample # Seqs (AH)	EDI to 1 Yr Sample (Yrs)	1 Yr Sample VL (copies/ mL)	1 Yr Sample # Seqs (AH)	EDI to Last ART-naïve Sample (Yrs)	Last ART-naïve Sample VL (copies/ mL)	Last ART-naïve Sample # Seqs (AH) Plasma-Cells	ART Sample 1 VL (copies/ mL)	ART Sample 1 # Seqs (AH)	ART Sample 2 VL (copies/ mL)	ART Sample 2 # Seqs (AH)
Z1094F	21-May-2008	04-Jul-2008	+/NA/+	550,000	12 (0)	0.92	104,000	26 (0)	2.33	391,000	40 (0)-25 (3)	<50	25 (3)	NA	NA
Z1123M	02-Apr-2008	12-Apr-2008	+/-/-	37,068	8 (0)	1.01	40,500	23 (0)	2.82	110,259	33 (0)-24 (2)	<50	28 (4)	NA	NA
N133M	01-Jul-2007	16-Aug-2007	+/NA/+	180,220	14 (0)	0.68	100,001	24 (0)	2.77	186,000	25 (0)	<50	19 (1)	<50	17 (2)
Z1788F	26-Dec-2006	10-Feb-2007	+/NA/+	1,015,368	10 (0)	0.92	68,000	27 (0)	2.94	18,136	40 (0)-25 (1)	<50	20 (4)	<50	18 (1)
Z634F	08-Jul-2007	24-Jul-2007	+/NA/+	120,756	14 (0)	0.93	64,600	34 (0)	2.98	58,652	32 (1)	<50	27 (1)	<50	17 (6)
Z2006M	26-Apr-2009	13-May-2009	+/NA/+	27,200	13 (0)	0.93	10,900	22 (1)	3.22	109,070	26 (3)	<50	13 (8)	NA	NA
Z1047M	31-Jul-2007	14-Aug-2007	+/+/-	17,300	8 (1)	0.92	538,000	23 (0)	3.53	305,564	35 (1)-25 (1)	<50	21 (3)	NA	NA
Z1165M	26-Jan-2006	16-Mar-2006	+/NA/+	750,000	13 (0)	0.96	122,000	41 (0)	3.72	24,864	28 (0)-24 (2)	<50	20 (6)	<50	16 (3)
Z1658F	24-Apr-2006	08-Jun-2006	+/NA/+	219,111	14 (0)	0.93	44,500	19 (11)	4.62	34,370	30 (1)-25 (2)	<50	19 (9)	NA	NA
Z1808F	10-Jan-2008	24-Jan-2008	+/+/-	1,742,240	10 (0)	0.92	126,248	25 (0)	4.43	57,460	27 (15)	<50	22 (8)	NA	NA
Z1124F	02-Apr-2006	18-May-2006	+/NA/+	333,392	11 (0)	0.92	15,100	29 (0)	4.81	1,072	28 (0)	<50	19 (5)	NA	NA
Z1044M	26-Jan-2006	25-Mar-2006	+/NA/+	119,075	10 (0)	1.15	6,020	21 (0)	5.01	5,287	23 (0)	<50	19 (2)	NA	NA
Z326M	04-Jan-2007	17-Feb-2007	+/NA/+	44,800	12 (0)	1.39	32,700	38 (1)	6.22	9,100	26 (2)	<50	17 (10)	NA	NA
MEAN	**NA**	**NA**	**NA**	**396,656**	**11.5 (0)**	**0.97**	**97,890**	**27.0 (1)**	**3.80**	**100,833**	**30.2 (1.8)**	**<50**	**20.6 (4.9)**	**<50**	**17.0 (3.0)**
MEDIAN	**NA**	**NA**	**NA**	**180,220**	**12.0 (0)**	**0.93**	**64,600**	**24 (0)**	**3.53**	**57,460**	**28 (0)**	**<50**	**19 (4)**	**<50**	**16.5 (2.5)**

SC=Seroconversion
Seqs=Sequences
AH=additional APOBEC Hypermutant sequences excluded from analysis
[a]VL within seven days of sample

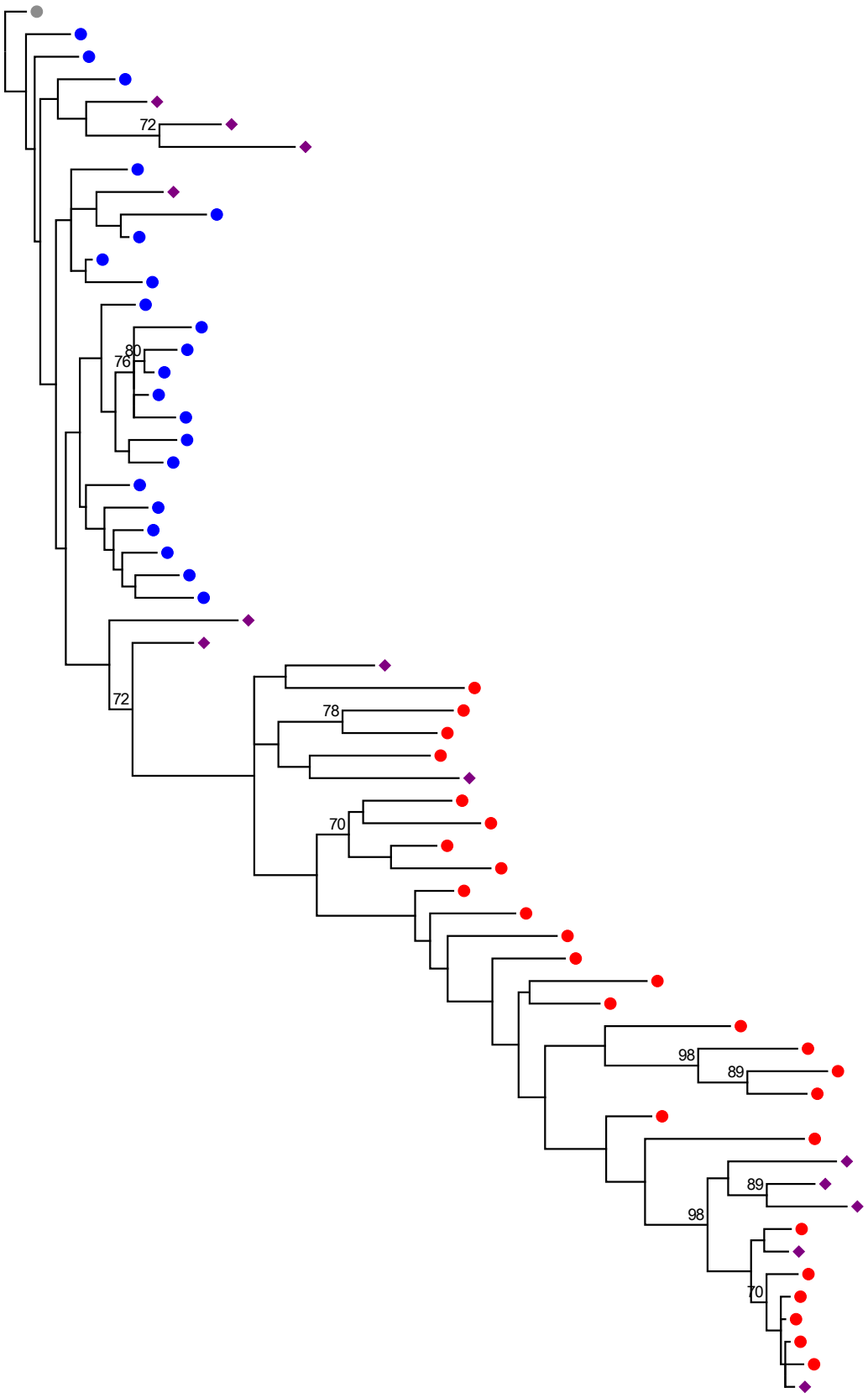


S1 Fig. Seroconversion sequences. (A) Maximum-likelihood phylogenetic tree of the near full-length HIV-1 genomes from all 13 study individuals with color-coding of sequences from each participant. A single distinct clade and very short branch lengths within each participant viral population are indicative of the low diversity at the seroconversion sample time points, except in the case of participant Z1658F (in rose), where two clades are present. The sequences within each clade for Z1658F are low-diversity, consistent with infection being established by two TF

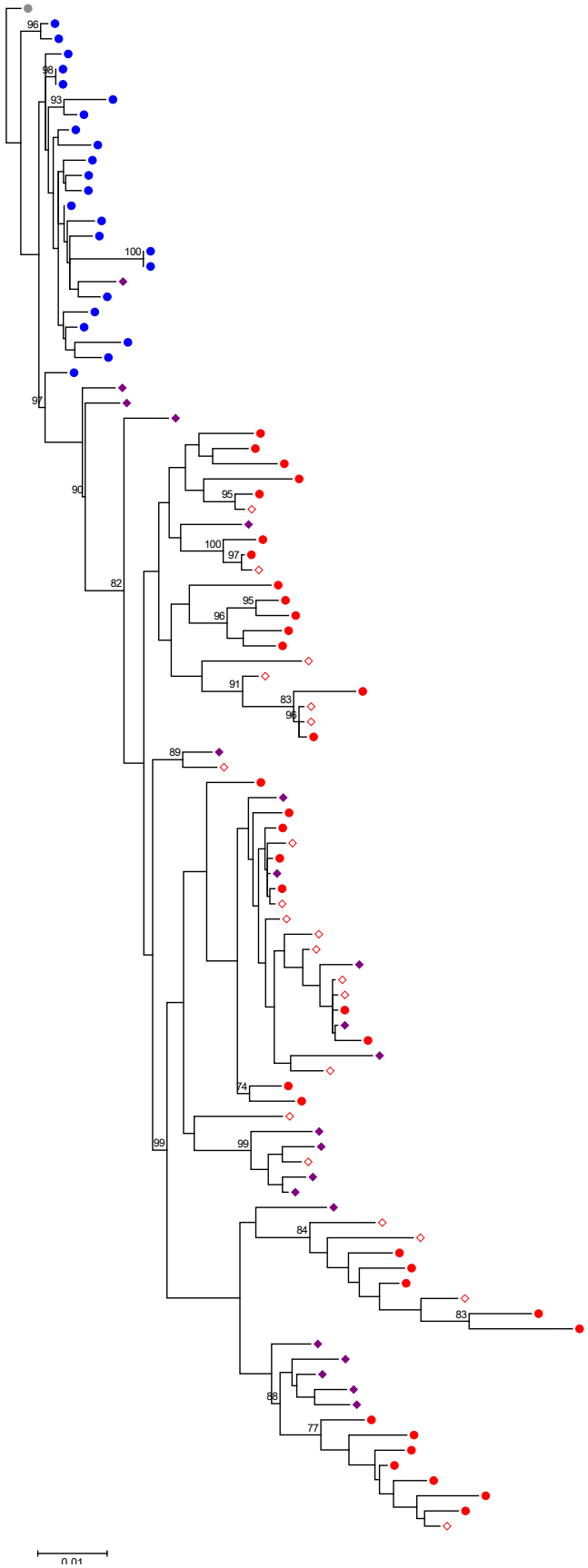
viruses. (B) Highlighter plot of the two viral populations for Z1658F, with one TF virus as the master sequence.



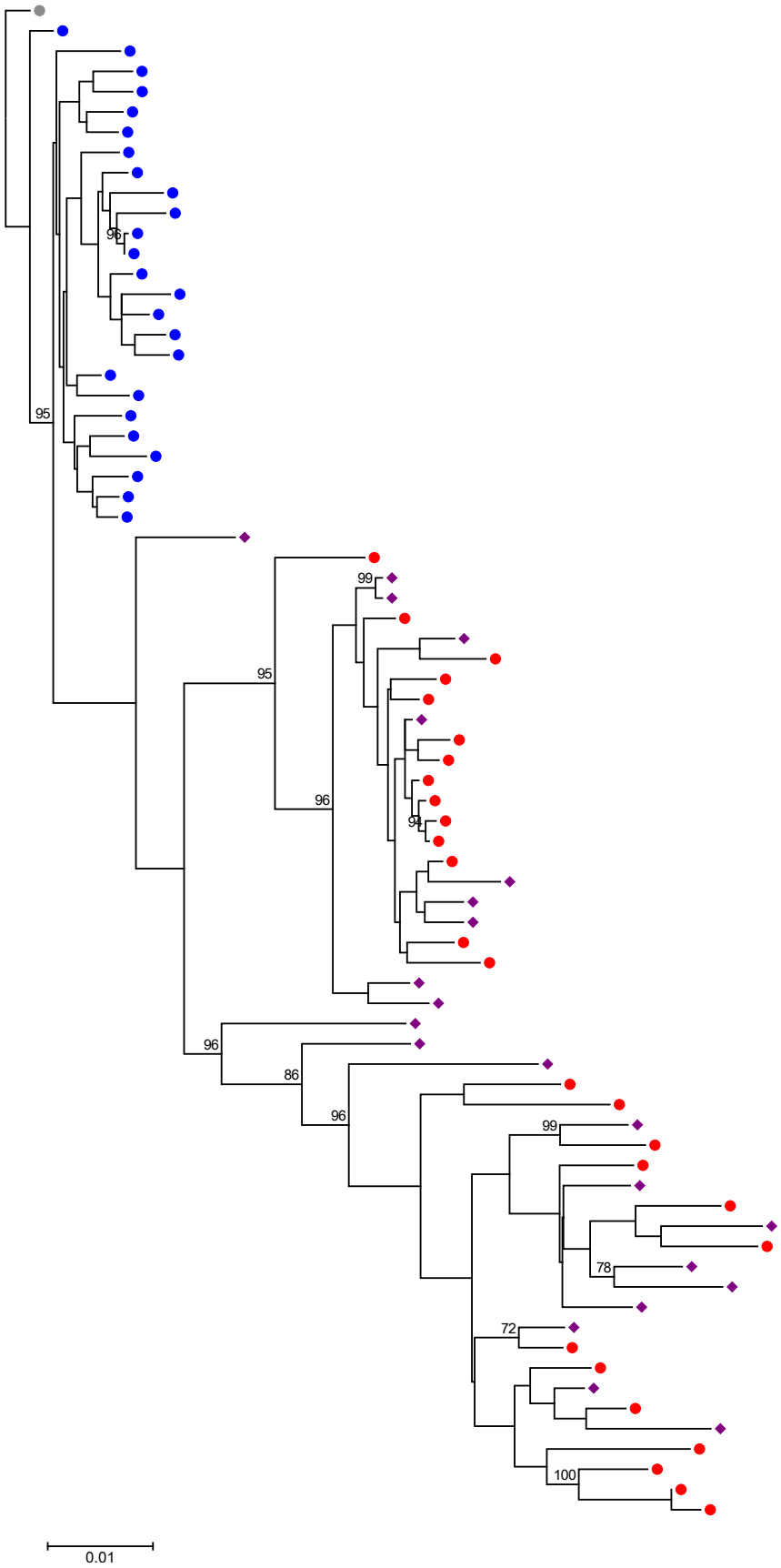
Z1094F



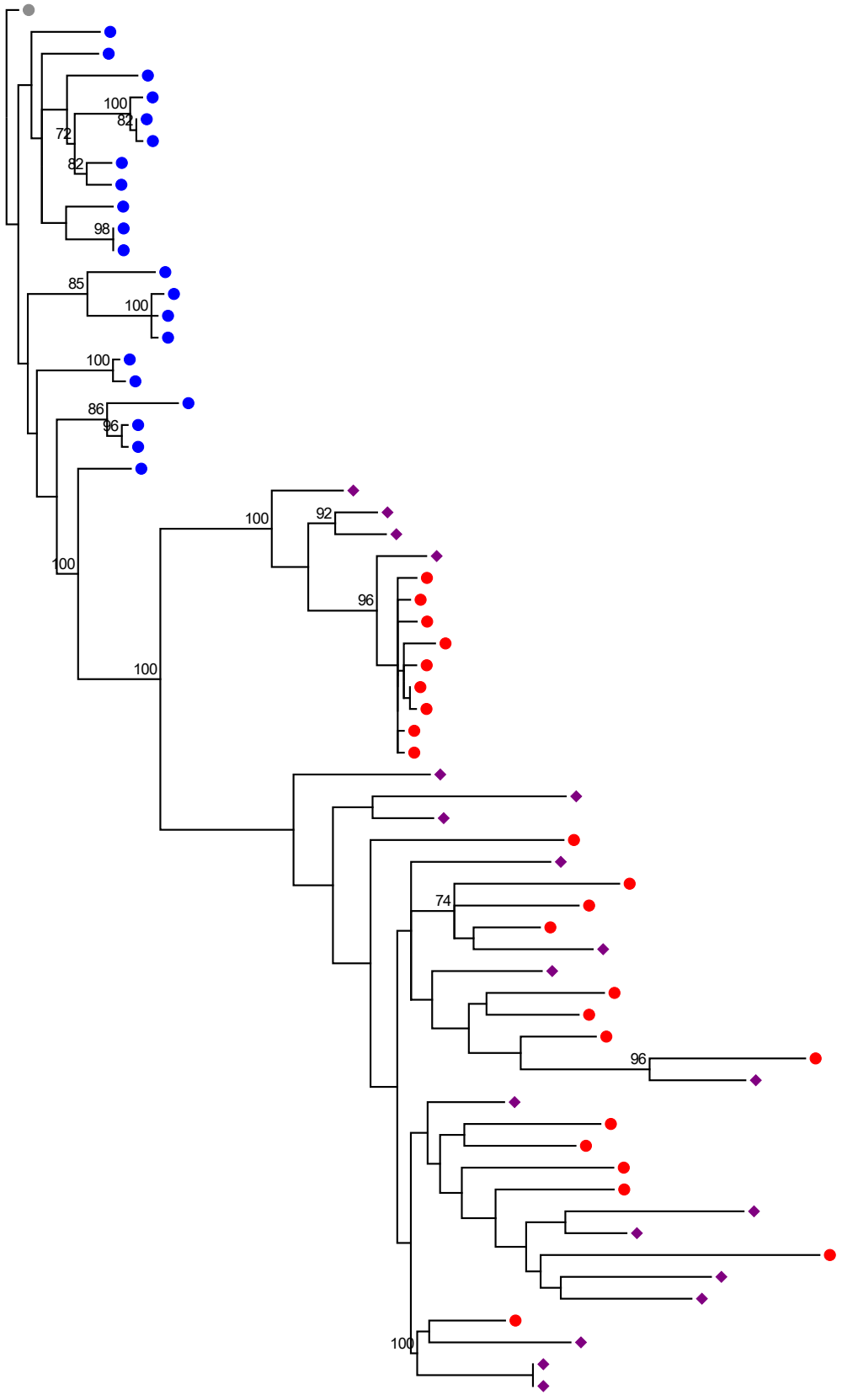
Z2006M


96
98
93
100
97
90
95
100
97
82
95
96
91
83
96
89
74
99
99
84
83
88
77
0.01

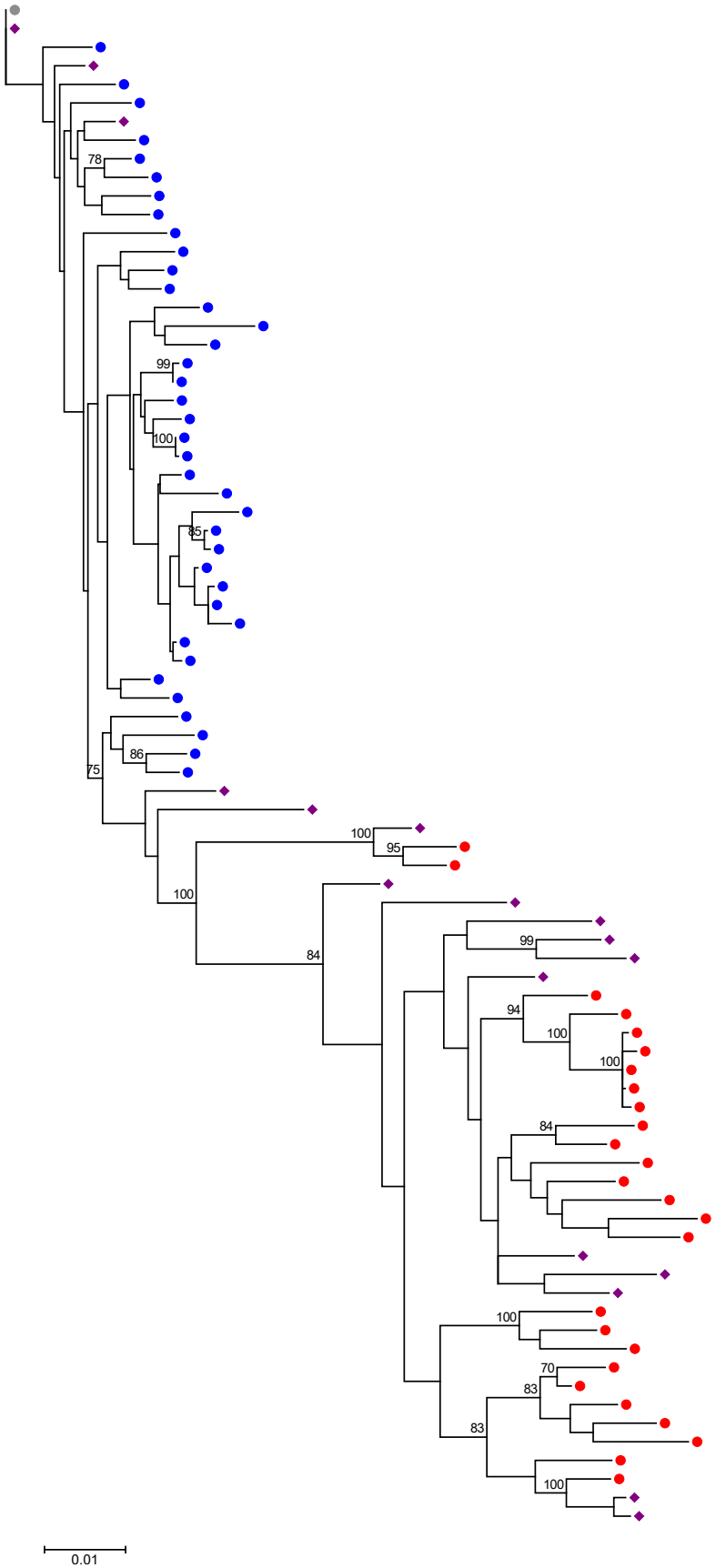
Z1047M



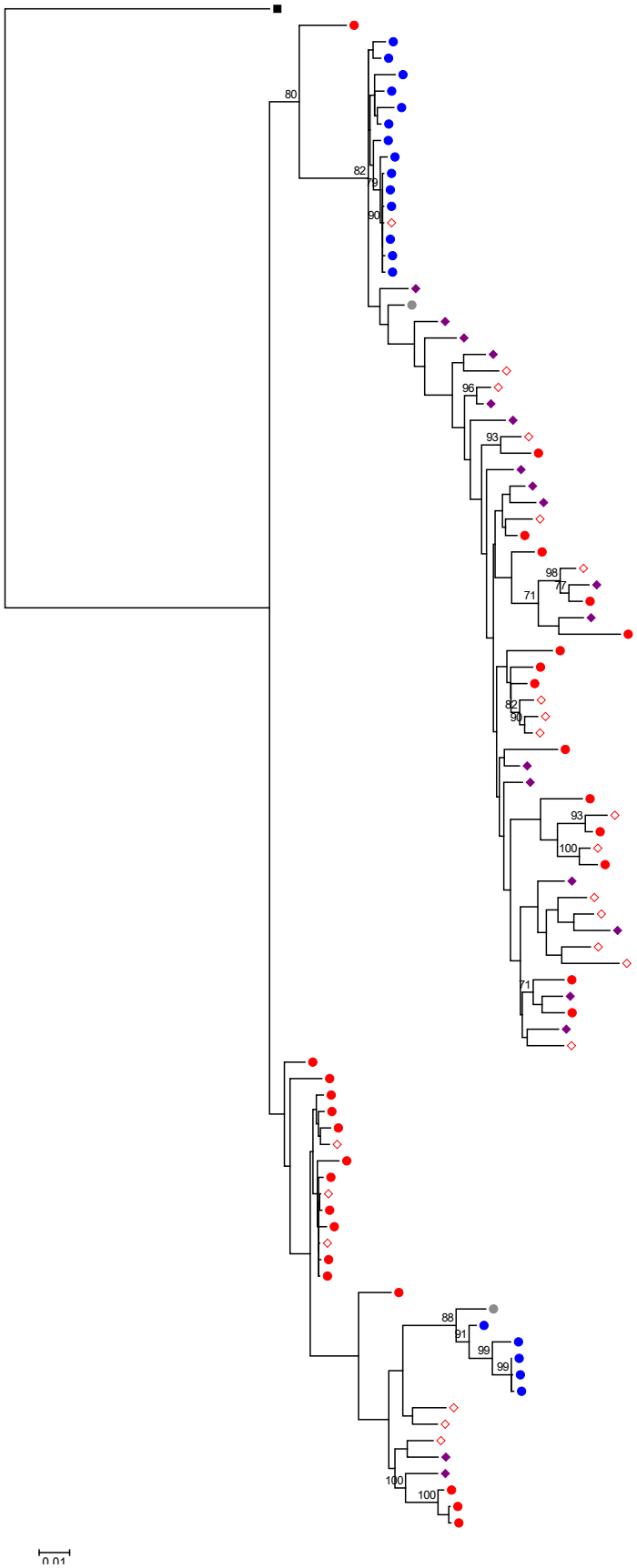
Z1808F



Z1044M

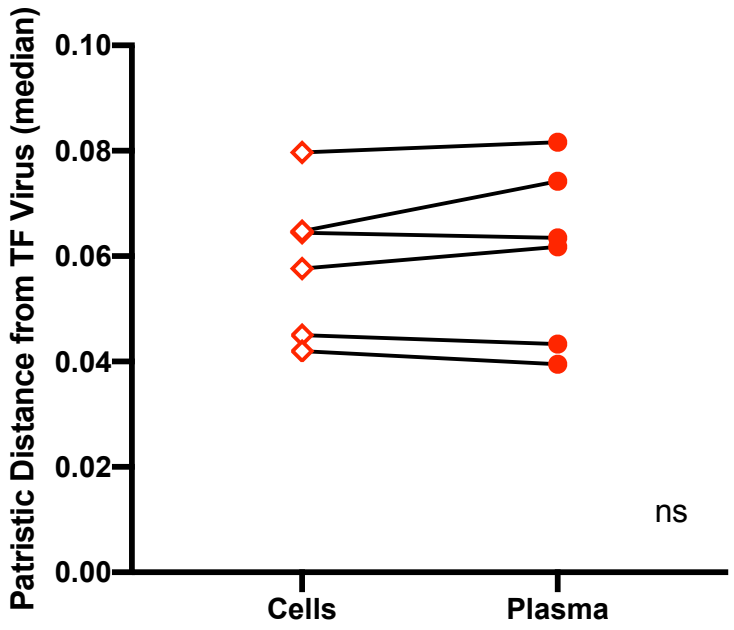

78
99
100
85
86
75
100
95
100
84
99
94
100
100
84
100
70
83
83
100
0.01

Z326M



Z1658F

S2 Fig. Maximum-likelihood tree for seven participants with one sample available during treatment. (A) Z1094F (B) Z2006M (C) Z1047M (D) Z1808F (E) Z1044M (F) Z326M (G) Z1658F. Participant Z1658F was infected with two TF viruses, both included in the ML tree, which is rooted on a Zambian subtype C consensus sequence (black square). All other trees are rooted on the respective TF virus (grey) identified from the seroconversion sample and depict all viral variants from one-year post infection (blue), the last ART-naïve sample (red), and during treatment (purple diamonds). Sequences from cells collected at the last ART-naïve time point are shown in open red diamonds, while all plasma variants are in filled circles. Nodes with bootstrap support values >70 are shown in text.



S3 Fig. Sequences from cells are not significantly less divergent from TF virus than plasma variants at same time point. The median distance from the TF virus for sequences from cells and plasma collected at the last ART-naïve time point is not significantly different (Wilcoxon matched-pairs signed rank test, p=0.5625)

References

1. UNAIDS. Fact sheet: global AIDS update 2019. 2019.

2. Hütter G, Nowak D, Mossner M, Ganepola S, Müßig A, Allers K, et al. Long-Term Control of HIV by CCR5 Delta32/Delta32 Stem-Cell Transplantation. New England Journal of Medicine. 2009;360(7):692-8. doi: 10.1056/NEJMoa0802905.

3. Allers K, Hütter G, Hofmann J, Loddenkemper C, Rieger K, Thiel E, et al. Evidence for the cure of HIV infection by CCR5Δ32/Δ32 stem cell transplantation. Blood. 2011;117(10):2791. doi: 10.1182/blood-2010-09-309591.

4. Gupta RK, Abdul-Jawad S, McCoy LE, Mok HP, Peppa D, Salgado M, et al. HIV-1 remission following CCR5Δ32/Δ32 haematopoietic stem-cell transplantation. Nature. 2019;568(7751):244-8. doi: 10.1038/s41586-019-1027-4.

5. Finzi D, Hermankova M, Pierson T, Carruth LM, Buck C, Chaisson RE, et al. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. Science (New York, NY). 1997;278(5341):1295-300. Epub 1997/11/21. PubMed PMID: 9360927.

6. Wong JK, Hezareh M, Günthard HF, Havlir DV, Ignacio CC, Spina CA, et al. Recovery of Replication-Competent HIV Despite Prolonged Suppression of Plasma Viremia. Science (New York, NY). 1997;278(5341):1291. doi: 10.1126/science.278.5341.1291.

7. Chun T-W, Stuyver L, Mizell SB, Ehler LA, Mican JAM, Baseler M, et al. Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. Proceedings of the National Academy of Sciences. 1997;94(24):13193. doi: 10.1073/pnas.94.24.13193.

8. Chomont N, El-Far M, Ancuta P, Trautmann L, Procopio FA, Yassine-Diab B, et al. HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. Nature

medicine. 2009;15(8):893-900. Epub 2009/06/23. doi: 10.1038/nm.1972. PubMed PMID: 19543283; PubMed Central PMCID: PMCPmc2859814.

9. Maldarelli F, Wu X, Su L, Simonetti FR, Shao W, Hill S, et al. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. Science (New York, NY). 2014;345(6193):179. doi: 10.1126/science.1254194.

10. Wagner TA, McLaughlin S, Garg K, Cheung CYK, Larsen BB, Styrchak S, et al. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. Science (New York, NY). 2014;345(6196):570. doi: 10.1126/science.1256304.

11. Wang Z, Gurule EE, Brennan TP, Gerold JM, Kwon KJ, Hosmane NN, et al. Expanded cellular clones carrying replication-competent HIV-1 persist, wax, and wane. Proceedings of the National Academy of Sciences. 2018;115(11):E2575. doi: 10.1073/pnas.1720665115.

12. Finzi D, Blankson J, Siliciano JD, Margolick JB, Chadwick K, Pierson T, et al. Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. Nature medicine. 1999;5(5):512-7. doi: 10.1038/8394.

13. Siliciano JD, Kajdas J, Finzi D, Quinn TC, Chadwick K, Margolick JB, et al. Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. Nature medicine. 2003;9(6):727-8. Epub 2003/05/20. doi: 10.1038/nm880. PubMed PMID: 12754504.

14. Zhang L, Chung C, Hu B-S, He T, Guo Y, Kim AJ, et al. Genetic characterization of rebounding HIV-1 after cessation of highly active antiretroviral therapy. The Journal of Clinical Investigation. 2000;106(7):839-45. doi: 10.1172/JCI10565.

15. Joos B, Fischer M, Kuster H, Pillai SK, Wong JK, Böni J, et al. HIV rebounds from latently infected cells, rather than from continuing low-level replication. Proceedings of the National Academy of Sciences. 2008;105(43):16725. doi: 10.1073/pnas.0804192105.

16. Kearney MF, Spindler J, Shao W, Yu S, Anderson EM, O'Shea A, et al. Lack of Detectable HIV-1 Molecular Evolution during Suppressive Antiretroviral Therapy. PLoS pathogens. 2014;10(3):e1004010. doi: 10.1371/journal.ppat.1004010.

17. Kearney MF, Wiegand A, Shao W, Coffin JM, Mellors JW, Lederman M, et al. Origin of Rebound Plasma HIV Includes Cells with Identical Proviruses That Are Transcriptionally Active before Stopping of Antiretroviral Therapy. Journal of virology. 2016;90(3):1369. doi: 10.1128/JVI.02139-15.

18. Cohen YZ, Lorenzi JCC, Krassnig L, Barton JP, Burke L, Pai J, et al. Relationship between latent and rebound viruses in a clinical trial of anti–HIV-1 antibody 3BNC117. The Journal of experimental medicine. 2018;215(9):2311. doi: 10.1084/jem.20180936.

19. Palich R, Ghosn J, Chaillon A, Boilet V, Nere ML, Chaix ML, et al. Viral rebound in semen after antiretroviral treatment interruption in an HIV therapeutic vaccine double-blind trial. 2019;(1473-5571 (Electronic)).

20. Colby DJ, Trautmann L, Pinyakorn S, Leyre L, Pagliuzza A, Kroon E, et al. Rapid HIV RNA rebound after antiretroviral treatment interruption in persons durably suppressed in Fiebig I acute HIV infection. Nature medicine. 2018;24(7):923-6. doi: 10.1038/s41591-018-0026-6.

21. Vibholm LK, Lorenzi JCC, Pai JA, Cohen YZ, Oliveira TY, Barton JP, et al. Characterization of Intact Proviruses in Blood and Lymph Node from HIV-Infected Individuals Undergoing Analytical Treatment Interruption. Journal of virology. 2019;93(8):e01920-18. doi: 10.1128/JVI.01920-18.

22. Ho YC, Shan L, Hosmane NN, Wang J, Laskey SB, Rosenbloom DI, et al. Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. Cell. 2013;155(3):540-51. Epub 2013/11/19. doi: 10.1016/j.cell.2013.09.020. PubMed PMID: 24243014; PubMed Central PMCID: PMCPmc3896327.

23. Buzon MJ, Sun H, Li C, Shaw A, Seiss K, Ouyang Z, et al. HIV-1 persistence in CD4+ T cells with stem cell–like properties. Nature medicine. 2014;20:139. doi: 10.1038/nm.3445 https://www.nature.com/articles/nm.3445 - supplementary-information.

24. Bruner KM, Murray AJ, Pollack RA, Soliman MG, Laskey SB, Capoferri AA, et al. Defective proviruses rapidly accumulate during acute HIV-1 infection. Nature medicine. 2016;22:1043. doi: 10.1038/nm.4156 https://www.nature.com/articles/nm.4156 - supplementary-information.

25. Hiener B, Horsburgh BA, Eden JS, Barton K, Schlub TE, Lee E, et al. Identification of Genetically Intact HIV-1 Proviruses in Specific CD4(+) T Cells from Effectively Treated Participants. 2017;(2211-1247 (Electronic)).

26. Bruner KM, Wang Z, Simonetti FR, Bender AM, Kwon KJ, Sengupta S, et al. A quantitative approach for measuring the reservoir of latent HIV-1 proviruses. Nature. 2019;566(7742):120-5. doi: 10.1038/s41586-019-0898-8.

27. Daar ES, Bai J, Hausner MA, Majchrowicz M, Tamaddon M, Giorgi JV. Acute HIV Syndrome after Discontinuation of Antiretroviral Therapy in a Patient Treated before Seroconversion. Annals of Internal Medicine. 1998;128(10):827-9. doi: 10.7326/0003-4819-128-10-199805150-00005.

28. Whitney JB, Hill AL, Sanisetty S, Penaloza-MacMaster P, Liu J, Shetty M, et al. Rapid seeding of the viral reservoir prior to SIV viraemia in rhesus monkeys. Nature.

2014;512(7512):74-7. Epub 2014/07/22. doi: 10.1038/nature13594. PubMed PMID: 25042999; PubMed Central PMCID: PMCPmc4126858.

29. Henrich TJ, Hatano H, Bacon O, Hogan LE, Rutishauser R, Hill A, et al. HIV-1 persistence following extremely early initiation of antiretroviral therapy (ART) during acute HIV-1 infection: An observational study. PLOS Medicine. 2017;14(11):e1002417. doi: 10.1371/journal.pmed.1002417.

30. Brodin J, Zanini FA-O, Thebo L, Lanz C, Bratt G, Neher RA-O, et al. Establishment and stability of the latent HIV-1 DNA reservoir. LID - 10.7554/eLife.18889 [doi] LID - e18889 [pii]. 2016;(2050-084X (Electronic)).

31. Abrahams M-R, Joseph SB, Garrett N, Tyers L, Moeser M, Archin N, et al. The replication-competent HIV-1 latent reservoir is primarily established near the time of therapy initiation. Science Translational Medicine. 2019;11(513):eaaw5589. doi: 10.1126/scitranslmed.aaw5589.

32. Jones BR, Kinloch NN, Horacsek J, Ganase B, Harris M, Harrigan PR, et al. Phylogenetic approach to recover integration dates of latent HIV sequences within-host. 2018;(1091-6490 (Electronic)).

33. Dilernia DA, Chien JT, Monaco DC, Brown MP, Ende Z, Deymier MJ, et al. Multiplexed highly-accurate DNA sequencing of closely-related HIV-1 variants using continuous long reads from single molecule, real-time sequencing. Nucleic acids research. 2015;43(20):e129. Epub 2015/06/24. doi: 10.1093/nar/gkv630. PubMed PMID: 26101252.

34. Theys K, Libin P, Pineda-Peña A-C, Nowé A, Vandamme A-M, Abecasis AB. The impact of HIV-1 within-host evolution on transmission dynamics. Current Opinion in Virology. 2018;28:92-101. doi: https://doi.org/10.1016/j.coviro.2017.12.001.

35. Yerly S, Perneger TV, Vora S, Hirschel B, Perrin L. Decay of cell-associated HIV-1 DNA correlates with residual replication in patients treated during acute HIV-1 infection. AIDS (London, England). 2000;14(18):2805-12. Epub 2001/01/12. doi: 10.1097/00002030-200012220-00001. PubMed PMID: 11153661.

36. Blankson JN, Finzi D, Pierson TC, Sabundayo BP, Chadwick K, Margolick JB, et al. Biphasic decay of latently infected CD4+ T cells in acute human immunodeficiency virus type 1 infection. The Journal of infectious diseases. 2000;182(6):1636-42. Epub 2000/11/09. doi: 10.1086/317615. PubMed PMID: 11069234.

37. Hocqueloux L, Avettand-Fènoël V, Jacquot S, Prazuck T, Legac E, Mélard A, et al. Long-term antiretroviral therapy initiated during primary HIV-1 infection is key to achieving both low HIV reservoirs and normal T cell counts. Journal of Antimicrobial Chemotherapy. 2013;68(5):1169-78. doi: 10.1093/jac/dks533.

38. Besson GJ, Lalama CM, Bosch RJ, Gandhi RT, Bedison MA, Aga E, et al. HIV-1 DNA Decay Dynamics in Blood During More Than a Decade of Suppressive Antiretroviral Therapy. Clinical Infectious Diseases. 2014;59(9):1312-21. doi: 10.1093/cid/ciu585. PubMed PMID: WOS:000344647100018.

39. Archin NM, Vaidya NK, Kuruc JD, Liberty AL, Wiegand A, Kearney MF, et al. Immediate antiviral therapy appears to restrict resting CD4+ cell HIV-1 infection without accelerating the decay of latent infection. Proceedings of the National Academy of Sciences. 2012;109(24):9523. doi: 10.1073/pnas.1120248109.

40. Buzon MJ, Martin-Gayo E, Pereyra F, Ouyang Z, Sun H, Li JZ, et al. Long-term antiretroviral treatment initiated at primary HIV-1 infection affects the size, composition, and decay kinetics of the reservoir of HIV-1-infected CD4 T cells. Journal of virology.

2014;88(17):10056-65. Epub 2014/06/27. doi: 10.1128/jvi.01046-14. PubMed PMID: 24965451; PubMed Central PMCID: PMCPmc4136362.

41. Ruff CT, Ray SC, Kwon P, Zinn R, Pendleton A, Hutton N, et al. Persistence of Wild-Type Virus and Lack of Temporal Structure in the Latent Reservoir for Human Immunodeficiency Virus Type 1 in Pediatric Patients with Extensive Antiretroviral Exposure. Journal of virology. 2002;76(18):9481. doi: 10.1128/JVI.76.18.9481-9492.2002.

42. Verhofstede C, Noe A, Demecheleer E, De Cabooter N, Van Wanzeele F, Van Der Gucht B, et al. Drug-resistant variants that evolve during nonsuppressive therapy persist in HIV-1-infected peripheral blood mononuclear cells after long-term highly active antiretroviral therapy. Jaids-Journal of Acquired Immune Deficiency Syndromes. 2004;35(5):473-83. doi: Doi 10.1097/00126334-200404150-00005. PubMed PMID: WOS:000220632000005.

43. Kieffer TL, Finucane MM, Nettles RE, Quinn TC, Broman KW, Ray SC, et al. Genotypic Analysis of HIV-1 Drug Resistance at the Limit of Detection: Virus Production without Evolution in Treated Adults with Undetectable HIV Loads. The Journal of infectious diseases. 2004;189(8):1452-65.

44. Bailey JR, Sedaghat AR, Kieffer T, Brennan T, Lee PK, Wind-Rotolo M, et al. Residual Human Immunodeficiency Virus Type 1 Viremia in Some Patients on Antiretroviral Therapy Is Dominated by a Small Number of Invariant Clones Rarely Found in Circulating CD4+ T Cells. Journal of virology. 2006;80(13):6441. doi: 10.1128/JVI.00591-06.

45. Perelson AS, Essunger P, Cao Y, Vesanen M, Hurley A, Saksela K, et al. Decay characteristics of HIV-1-infected compartments during combination therapy. Nature. 1997;387(6629):188-91. doi: 10.1038/387188a0.

46. Rousseau CM, Birditt BA, McKay AR, Stoddard JN, Lee TC, McLaughlin S, et al. Large-scale amplification, cloning and sequencing of near full-length HIV-1 subtype C genomes. Journal of virological methods. 2006;136(1-2):118-25. Epub 2006/05/17. doi: 10.1016/j.jviromet.2006.04.009. PubMed PMID: 16701907.

47. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Systematic Biology. 2010;59(3):307-21. doi: 10.1093/sysbio/syq010.

48. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647-9. doi: 10.1093/bioinformatics/bts199.

49. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular Biology and Evolution. 2016;33(7):1870-4. doi: 10.1093/molbev/msw054.

50. Rose PP, Korber BT. Detecting hypermutations in viral sequences with an emphasis on G → A hypermutation. Bioinformatics. 2000;16(4):400-1. doi: 10.1093/bioinformatics/16.4.400.

51. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol. 2015;1(1):vev003. doi: 10.1093/ve/vev003. PubMed PMID: 27774277; PubMed Central PMCID: PMCPMC5014473.

52. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. 2018;4(1):vey016. doi: 10.1093/ve/vey016. PubMed PMID: 29942656; PubMed Central PMCID: PMCPMC6007674.

53. Shapiro B, Rambaut A, Drummond AJ. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. Mol Biol Evol. 2006;23(1):7-9. doi: 10.1093/molbev/msj021. PubMed PMID: 16177232.

54. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. PLoS Biol. 2006;4(5):e88. doi: 10.1371/journal.pbio.0040088. PubMed PMID: 16683862; PubMed Central PMCID: PMCPMC1395354.

55. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol. 2005;22(5):1185-92. doi: 10.1093/molbev/msi103. PubMed PMID: 15703244.

56. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchene S, Fourment M, Gavryushkina A, et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 2019;15(4):e1006650. doi: 10.1371/journal.pcbi.1006650. PubMed PMID: 30958812; PubMed Central PMCID: PMCPMC6472827.

57. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst Biol. 2018;67(5):901-4. doi: 10.1093/sysbio/syy032. PubMed PMID: 29718447; PubMed Central PMCID: PMCPMC6101584.

58. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312-3. doi: 10.1093/bioinformatics/btu033. PubMed PMID: 24451623; PubMed Central PMCID: PMCPMC3998144.

59. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 2019;35(3):526-8. doi: 10.1093/bioinformatics/bty633. PubMed PMID: 30016406.

Chapter III: Discussion

Connections to Infection Dynamics and Future Directions

Investigating the natural history of HIV infection as reported here for 13 Zambian seroconvertors followed longitudinally through chronic infection and treatment initiation is a unique opportunity to assess the timescales for persistence of T cells or their descendant daughter clones in the reservoir. As the virus evolves so rapidly upon infection of a new host, the individual variants remaining in the proviral population during treatment are a timestamp for their entry into the reservoir. We were particularly surprised by the presence of TF virus in the reservoir after up to six years of ART-naïve infection while T cell turnover is rapid. The cell itself may not have persisted for such an extended period of time, and instead may be an expanded clone or daughter cell of the originally infected cell, but the findings nonetheless corroborate the long half-life of central memory CD4+ T cells.[27] Although the amplicon in which we observed exact proviral matches to the TF virus for participants Z1123M and Z326M does not cover the entire genome, and it is possible there are mutations elsewhere in the proviruses from which the amplicons derive, from their similarity in the highly diverse *env* gene alone we conclude that these variants date to the very earliest stages of the infection before considerable mutation took place.

We cannot distinguish if the cell itself is as old as the infection, or if the provirus is contained within a daughter cell, which is an important distinction. Homeostatic proliferation may permit the expansion of infected clones and the generation of daughter cells with both the same proviral genome and integration site as the original infected cell. Where proviruses are identical, integration site sequencing can provide an indication if the cells represent unique infected cells (different integration sites), or expanded clones (same integration site), as the

likelihood of a single viral variant infecting the same locus in different cells is infinitesimally small.

The persistence of such early viral variants in the reservoir brings into question whether the dynamics of CD4+ T cell depletion and expansion in acute infection are contributing to the maintenance of very early virus. As viral load reaches its peak in Fiebig II-III[29] approximately as CD4+ T cells have reached the inflection point of their depletion, there is a considerably high ratio of virus to target cells. With this, and the extraordinarily inflammatory environment of the "cytokine storm"[70] in acute infection, extreme cellular activation and high levels of virus could lead to extremely high rates of infected cells, some of which will persist. This follows the hypothesis of Brodin et al., [115] who considered overrepresentation of early infection sequences in their data to reflect the very high level of viral replication at this time. Following acute infections in the absence of therapy is no longer an ethical study framework, as treatment should be offered as soon as an individual is known to be HIV+, but the study of SIV-infected rhesus macaques might provide a model in which to assess the effects of early infection replication dynamics on viral persistence through chronic infection. Tracking individual T cells with amplification of integration sites and/or T cell receptors in concert with the provirus would indicate the persistence of individual T cells or their daughter cells versus different cells infected with the same virus.

While not exact matches, we did also observe provrial variants very closely related to the TF virus (less than six nucleotides different), which we do consider to date to very early in the infection, likely at the time of the seroconversion sample. We observed these very early variants in five of the 13 individuals, an approximately 40% frequency, which is not unlike the 33% (three of nine participants) in which early infection sequences were observed by the Abrahams

et. al. group.[116] With similar findings of early variant presence, the potential mechanisms for the prolonged persistence of the virus are of interest. We have not seen evidence of a connection to TF virus replicative capacity as assessed by transmitted Gag chimeras, as those individuals with early virus persisting in the proviral population were not clearly discriminated from the remainder of the group by replicative capacity (data not shown). We had hypothesized that higher replicative capacity would lead to a higher DNA burden in central and effector memory cells at an early time point, as has been shown for other Zambian volunteers,[67] and that this might coincide with the persistence of early infection variants at a much later stage of infection. However, in assessing a subset of 10 of the 13 individuals investigated in this study for whom samples and data were available, we did not observe a higher HIV DNA burden in central memory, effector memory, or naïve CD4+ T cells at early time points post-infection in relation to the replicative capacity of the TF virus (data not shown).

In absence of a clear indication of an influence of viral replication on the persistence of early infection variants in latent provirus, we nonetheless wish to consider the overall dynamics of the infection. With slower CD4+ T cell loss, and concomitant reduced disruption of CD4+ T cell homeostasis and slower cell turnover, we might expect a greater likelihood of observing early viral variants in the reservoir. These questions are under investigation through the analysis of pre-therapy CD4+ T cell count and viral load measurements for all study participants.

Additionally, it is interesting to contemplate that as TF viruses have considerably distinct replicative capacities *in vitro*, perhaps they also have distinct capacities to establish latency in host cells. Though less diverse than other regions of the genome, the HIV long terminal repeats (LTRs) are unique to viruses isolated from different individuals, and contain a number of regulatory elements that play a role in proviral transcription or latency. We have cloned authentic

TF viruses for the 13 study individuals, and with these, modified latency reversal assays could provide insight into the capability of the variants to establish and reemerge from latency, perhaps shedding light on the presence of early variants in the proviral populations of some individuals and not others.

While not all individuals exhibited presence of early infection variants, across all individuals, the proviral variants during treatment were significantly closer to the TF virus by patristic distance than the plasma virus at the last ART-naïve timepoint. We consider this to result from the considerable influence of the very early proviral variants in the population on the patristic distances, even though these early infection variants are far less numerous than those most closely related to the last ART-naïve plasma virus. This appears to represent a shift back toward ancestral virus in the reservoir, consistent with our conclusion that the reservoir is seeded throughout the ART-naïve infection. With a representation of earlier variants, even at lower frequency than those most closely related to the last ART-naïve virus, it is clear to us that the reservoir represents an archive of much of the ART-naïve infection.

A limitation of our study is the sampling of provirus from cells collected during treatment at early times post treatment-initiation, while the reservoir is undergoing an early and more rapid phase of decay.[117] Therefore, we may not be sampling provirus that is likely to form the very stable reservoir. As we sampled four individuals with an initial and subsequent time point during ART, we found that the subsequent variants on treatment were closer to the TF virus (as expressed by individual variant distance as a proportion of the total intrapatient distance). We consider this an intriguing indication that as time on treatment continues, early infection variants are not lost, but perhaps enriched in the reservoir population. As we observe a decreased distance from the TF virus in the reservoir proviral population from an initial to subsequent sampling on

treatment, the variants most evolved from the TF virus that are circulating just before treatment initiation may in fact be those which decay most rapidly. Therefore, although predominant in the reservoir at time points shortly after the initiation of treatment, the proviral variants that are most closely related to the virus circulating in plasma just prior to initiation of therapy might represent the least stable of the proviral population.

Significance for the Field and Implications for HIV Cure

The findings detailed herein that both early variants and those most closely related to virus circulating at the last ART-naïve time point are persisting in the reservoir are in fact remarkably consistent with the data of both the Brodin and Abrahams groups. Despite the differences in methods, as Brodin et al.[115] amplified a portion of the *gag* gene rather than the predominantly *env* amplicon we selected, and Abrahams et al.[116] assessed the replication-competent reservoir with QVOA, the structures of the phylogenetic trees between the three studies demonstrate a marked degree of similarity. Each study, ours included, observed variable reservoir proviral populations with considerable interpatient heterogeneity. Indeed, although Abrahams et al. generally observe that the majority of replication-competent provirus is most closely related to the virus at the last ART-naïve time point, in two individuals approximately 20% of the latent virus dated to within a year of infection.[116] In the opposite scenario, though we see provirus dating to approximately the first year of infection or earlier in the majority of study participants (10 of 13), we do observe three individuals where there simply does not appear to be early virus archived, or at least not at the level of detection of our study. Again, the mechanisms for persistence of early virus in some participants but not others is particularly intriguing, and merits investigation of the history of infection with regard to viral load and CD4+ T cell decline for any indications as to the cause.

As the reservoir appears to be seeded throughout the course of ART-naïve infection, with TF virus or very early variants represented in some individuals, but predominantly provirus most closely related to the virus at the last ART-naïve time point, the goal to link HIV+ patients to care as soon as infection is identified should be further supported. Though ART does come with side effects related to drug toxicity, these appear overshadowed in comparison to the benefits of preventing the continued loss of CD4+ T cells and extreme disruption of the immune system evidenced in ART-naïve infection. Furthermore, in agreement with previous studies, there is no indication of ongoing viral evolution during ART in our data. In fact, there appears to be a shift back to more ancestral virus with continued time on treatment, rather than the increased distance that would be expected with evolution and accumulation of new mutations.

For strategies in which suppression or cure of HIV is the goal, individuals starting ART sooner in infection may prove better candidates, as they have smaller, less diverse reservoirs than individuals who begin treatment during chronic infection.[31] These findings led to a hypothesis that the reservoir is seeded throughout infection, which is consistent with our data that both early and chronic infection viruses are persisting in the proviral population. We do also agree with the proposal from Abrahams et al. that administration of a means of limiting reservoir formation at the time of ART initiation could lead to a smaller reservoir, as a considerable fraction appears to date from the time just prior to therapy initiation.

Summary

We have provided the field with an additional analysis that details the formation of the reservoir in the context of the natural history of the infection. Such studies will become increasingly unlikely to occur, as the advent of ART initiation in early infection renders the several years of ART-naïve infection investigated here unethical. Our conclusions, though

markedly different than those of Abrahams et al.,[116] who focused on the majority of reservoir variants that are most closely related to those at the last pre-ART time point, are supported by similar data. We have not identified a feature of the virus or host that explains persistence of early infection variants in some, but not all, participants in our study, and will investigate potential causes related to the capability of the TF virus to persist in a latent state *in vitro*. All the studies to date investigating the formation of the reservoir in the context of the early through chronic ART-naïve infection should encourage the initiation of treatment in early infection, as it is clear that beginning therapy while there is a viral population of limited diversity would lead to formation of a more genetically homogenous reservoir.

References

1. UNAIDS. Fact sheet: global AIDS update 2019. 2019.

2. WHO. Global Health Estimates 2016: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2016. 2018.

3. UNAIDS. Miles to Go—Closing Gaps, Breaking Barriers, Righting Injustices. 2018.

4. UNAIDS. 90-90-90: An ambitious treatment target to help end the AIDS epidemic. 2014.

5. UNAIDS. Combination HIV Prevention:
Tailoring and Coordinating Biomedical,
Behavioural and Structural Strategies to Reduce New HIV Infections—a UNAIDS Discussion Paper. 2010.

6. Karim SSA, Karim QA. Antiretroviral prophylaxis: a defining moment in HIV control. *The Lancet* 2011; **378**(9809): e23-e5.

7. Grant RM, Lama JR, Anderson PL, et al. Preexposure Chemoprophylaxis for HIV Prevention in Men Who Have Sex with Men. *New England Journal of Medicine* 2010; **363**(27): 2587-99.

8. Thigpen MC, Kebaabetswe PM, Paxton LA, et al. Antiretroviral Preexposure Prophylaxis for Heterosexual HIV Transmission in Botswana. *New England Journal of Medicine* 2012; **367**(5): 423-34.

9. Baeten JM, Donnell D, Ndase P, et al. Antiretroviral Prophylaxis for HIV Prevention in Heterosexual Men and Women. *New England Journal of Medicine* 2012; **367**(5): 399-410.

10. Cohen MS, Chen YQ, McCauley M, et al. Antiretroviral Therapy for the Prevention of HIV-1 Transmission. *New England Journal of Medicine* 2016; **375**(9): 830-9.

11. Samji H, Cescon A, Hogg RS, et al. Closing the Gap: Increases in Life Expectancy among Treated HIV-Positive Individuals in the United States and Canada. *PLOS ONE* 2013; **8**(12): e81355.

12. Whitney JB, Hill AL, Sanisetty S, et al. Rapid seeding of the viral reservoir prior to SIV viraemia in rhesus monkeys. *Nature* 2014; **512**(7512): 74-7.

13. Ledford H. HIV rebound dashes hope of 'Mississippi baby' cure. *Nature News* 2014.

14. Hütter G, Nowak D, Mossner M, et al. Long-Term Control of HIV by CCR5 Delta32/Delta32 Stem-Cell Transplantation. *New England Journal of Medicine* 2009; **360**(7): 692-8.

15. Allers K, Hütter G, Hofmann J, et al. Evidence for the cure of HIV infection by CCR5Δ32/Δ32 stem cell transplantation. *Blood* 2011; **117**(10): 2791.

16. Gupta RK, Abdul-Jawad S, McCoy LE, et al. HIV-1 remission following CCR5Δ32/Δ32 haematopoietic stem-cell transplantation. *Nature* 2019; **568**(7751): 244-8.

17. Barré-Sinoussi F, Ross AL, Delfraissy J-F. Past, present and future: 30 years of HIV research. *Nature Reviews Microbiology* 2013; **11**: 877.

18. Finzi D, Hermankova M, Pierson T, et al. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science (New York, NY)* 1997; **278**(5341): 1295-300.

19. Chun TW, Carruth L, Finzi D, et al. Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature* 1997; **387**(6629): 183-8.

20. Wong JK, Hezareh M, Günthard HF, et al. Recovery of Replication-Competent HIV Despite Prolonged Suppression of Plasma Viremia. *Science (New York, NY)* 1997; **278**(5341): 1291.

21. Chun TW, Davey RT, Jr., Engel D, Lane HC, Fauci AS. Re-emergence of HIV after stopping therapy. *Nature* 1999; **401**(6756): 874-5.

22. Perrin L, Rakik A, Yerly S, et al. Combined therapy with zidovudine and L-697,661 in primary HIV infection. *AIDS (London, England)* 1996; **10**(11): 1233-7.

23. Colby DJ, Trautmann L, Pinyakorn S, et al. Rapid HIV RNA rebound after antiretroviral treatment interruption in persons durably suppressed in Fiebig I acute HIV infection. *Nature medicine* 2018; **24**(7): 923-6.

24. Craigie R, Bushman FD. HIV DNA integration. *Cold Spring Harbor perspectives in medicine* 2012; **2**(7): a006890-a.

25. Dahl V, Josefsson L, Palmer S. HIV reservoirs, latency, and reactivation: prospects for eradication. *Antiviral Res* 2010; **85**(1): 286-94.

26. Finzi D, Blankson J, Siliciano JD, et al. Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nature medicine* 1999; **5**(5): 512-7.

27. Siliciano JD, Kajdas J, Finzi D, et al. Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nature medicine* 2003; **9**(6): 727-8.

28. Eriksson S, Graf EH, Dahl V, et al. Comparative Analysis of Measures of Viral Reservoirs in HIV-1 Eradication Studies. *PLoS pathogens* 2013; **9**(2): e1003174.

29. Fiebig EW, Wright DJ, Rawal BD, et al. Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. *AIDS (London, England)* 2003; **17**(13): 1871-9.

30. Hocqueloux L, Avettand-Fènoël V, Jacquot S, et al. Long-term antiretroviral therapy initiated during primary HIV-1 infection is key to achieving both low HIV reservoirs and normal T cell counts. *Journal of Antimicrobial Chemotherapy* 2013; **68**(5): 1169-78.

31. Buzon MJ, Martin-Gayo E, Pereyra F, et al. Long-term antiretroviral treatment initiated at primary HIV-1 infection affects the size, composition, and decay kinetics of the reservoir of HIV-1-infected CD4 T cells. *Journal of virology* 2014; **88**(17): 10056-65.

32. Patel P, Borkowf CB, Brooks JT, Lasry A, Lansky A, Mermin J. Estimating per-act HIV transmission risk: a systematic review. *AIDS (London, England)* 2014; **28**(10): 1509-19.

33. CDC. Effectiveness of Prevention Strategies to Reduce the Risk of Acquiring or Transmitting HIV. 2019.

34. Boily M-C, Baggaley RF, Wang L, et al. Heterosexual risk of HIV-1 infection per sexual act: systematic review and meta-analysis of observational studies. *The Lancet Infectious Diseases* 2009; **9**(2): 118-29.

35. Venkatesh KK, van der Straten A, Cheng H, et al. The relative contribution of viral and bacterial sexually transmitted infections on HIV acquisition in southern African women in the Methods for Improving Reproductive Health in Africa study. *International Journal of STD & AIDS* 2011; **22**(4): 218-24.

36. Jones J, Weiss K, Mermin J, et al. Proportion of Incident Human Immunodeficiency Virus Cases Among Men Who Have Sex With Men Attributable to Gonorrhea and Chlamydia: A Modeling Analysis. *Sexually Transmitted Diseases* 2019; **46**(6).

37. Masson L, Passmore J-AS, Liebenberg LJ, et al. Genital Inflammation and the Risk of HIV Acquisition in Women. *Clinical Infectious Diseases* 2015; **61**(2): 260-9.

38. Haaland RE, Hawkins PA, Salazar-Gonzalez J, et al. Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1. *PLoS pathogens* 2009; **5**(1): e1000274.

39. Carlson JM, Schaefer M, Monaco DC, et al. HIV transmission. Selection bias at the heterosexual HIV-1 transmission bottleneck. *Science (New York, NY)* 2014; **345**(6193): 1254031.

40. Zhu T, Mo H, Wang N, et al. Genotypic and phenotypic characterization of HIV-1 patients with primary infection. *Science (New York, NY)* 1993; **261**(5125): 1179-81.

41. Scarlatti G, Leitner T, Halapi E, et al. Comparison of variable region 3 sequences of human immunodeficiency virus type 1 from infected children with the RNA and DNA sequences of the virus populations of their mothers. *Proceedings of the National Academy of Sciences* 1993; **90**(5): 1721.

42. Derdeyn CA, Decker JM, Bibollet-Ruche F, et al. Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. *Science (New York, NY)* 2004; **303**(5666): 2019-22.

43. Keele BF, Giorgi EE, Salazar-Gonzalez JF, et al. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proceedings of the National Academy of Sciences of the United States of America* 2008; **105**(21): 7552-7.

44. Salazar-Gonzalez JF, Salazar MG, Keele BF, et al. Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. *The Journal of experimental medicine* 2009; **206**(6): 1273-89.

45. Masharsky AE, Dukhovlinova EN, Verevochkin SV, et al. A Substantial Transmission Bottleneck among Newly and Recently HIV-1-Infected Injection Drug Users in St Petersburg, Russia. *The Journal of infectious diseases* 2010; **201**(11): 1697-702.

46. Chaillon A, Gianella S, Little SJ, et al. Characterizing the multiplicity of HIV founder variants during sexual transmission among MSM. *Virus evolution* 2016; **2**(1): vew012-vew.

47. Tully DC, Ogilvie CB, Batorsky RE, et al. Differences in the Selection Bottleneck between Modes of Sexual Transmission Influence the Genetic Composition of the HIV-1 Founder Virus. *PLoS pathogens* 2016; **12**(5): e1005619.

48. Shaw GM, Hunter E. HIV transmission. *Cold Spring Harbor perspectives in medicine* 2012; **2**(11): a006965.

49. Rousseau CM, Birditt BA, McKay AR, et al. Large-scale amplification, cloning and sequencing of near full-length HIV-1 subtype C genomes. *Journal of virological methods* 2006; **136**(1-2): 118-25.

50. Deymier MJ, Claiborne DT, Ende Z, et al. Particle infectivity of HIV-1 full-length genome infectious molecular clones in a subtype C heterosexual transmission pair following high fidelity amplification and unbiased cloning. *Virology* 2014; **468-470c**: 454-61.

51. Boeras DI, Hraber PT, Hurlston M, et al. Role of donor genital tract HIV-1 diversity in the transmission bottleneck. *Proceedings of the National Academy of Sciences* 2011; **108**(46): E1156.

52. Deymier MJ, Ende Z, Fenton-May AE, et al. Heterosexual Transmission of Subtype C HIV-1 Selects Consensus-Like Variants without Increased Replicative Capacity or Interferon-α Resistance. *PLoS pathogens* 2015; **11**(9): e1005154.

53. Ende Z, Deymier MJ, Claiborne DT, et al. HLA Class I Downregulation by HIV-1 Variants from Subtype C Transmission Pairs. *Journal of virology* 2018; **92**(7): e01633-17.

54. Parrish NF, Gao F, Li H, et al. Phenotypic properties of transmitted founder HIV-1. *Proceedings of the National Academy of Sciences* 2013; **110**(17): 6626.

55. Iyer SS, Bibollet-Ruche F, Sherrill-Mix S, et al. Resistance to type 1 interferons is a major determinant of HIV-1 transmission fitness. *Proceedings of the National Academy of Sciences* 2017; **114**(4): E590.

56. Naidoo VL, Mann JK, Noble C, et al. Mother-to-Child HIV Transmission Bottleneck Selects for Consensus Virus with Lower Gag-Protease-Driven Replication Capacity. *Journal of virology* 2017; **91**(17): e00518-17.

57. Sagar M, Laeyendecker O, Lee S, et al. Selection of HIV variants with signature genotypic characteristics during heterosexual transmission. *The Journal of infectious diseases* 2009; **199**(4): 580-9.

58. Redd AD, Collinson-Streng AN, Chatziandreou N, et al. Previously transmitted HIV-1 strains are preferentially selected during subsequent sexual transmissions. *The Journal of infectious diseases* 2012; **206**(9): 1433-42.

59. DHSS. US. Guidelines for the Use of Antiretroviral Agents in Adults and Adolescents with HIV. 2018.

60. Douek DC, Picker LJ, Koup RA. T Cell Dynamics in HIV-1 Infection. *Annual Review of Immunology* 2003; **21**(1): 265-304.

61. Borrow P, Lewicki H, Hahn BH, Shaw GM, Oldstone MB. Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection. *Journal of virology* 1994; **68**(9): 6103.

62. Koup RA, Safrit JT, Cao Y, et al. Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. *Journal of virology* 1994; **68**(7): 4650.

63. Robb ML, Eller LA, Kibuuka H, et al. Prospective Study of Acute HIV-1 Infection in Adults in East Africa and Thailand. *New England Journal of Medicine* 2016; **374**(22): 2120-30.

64. Mellors JW, Rinaldo CR, Gupta P, White RM, Todd JA, Kingsley LA. Prognosis in HIV-1 Infection Predicted by the Quantity of Virus in Plasma. *Science (New York, NY)* 1996; **272**(5265): 1167-70.

65. Prince JL, Claiborne DT, Carlson JM, et al. Role of transmitted Gag CTL polymorphisms in defining replicative capacity and early HIV-1 pathogenesis. *PLoS pathogens* 2012; **8**(11): e1003041.

66. Ojwach DBA, MacMillan D, Reddy T, et al. Pol-Driven Replicative Capacity Impacts Disease Progression in HIV-1 Subtype C Infection. *Journal of virology* 2018; **92**(19): e00811-18.

67. Claiborne DT, Prince JL, Scully E, et al. Replicative fitness of transmitted HIV-1 drives acute immune activation, proviral load in memory CD4+ T cells, and disease progression.

Proceedings of the National Academy of Sciences of the United States of America 2015; **112**(12): E1480-9.

68. Janes H, Herbeck JT, Tovanabutra S, et al. HIV-1 infections with multiple founders are associated with higher viral loads than infections with single founders. *Nature medicine* 2015; **21**: 1139.

69. Gottlieb GS, Nickle DC, Jensen MA, et al. Dual HIV-1 infection associated with rapid disease progression. *The Lancet* 2004; **363**(9409): 619-22.

70. Stacey AR, Norris PJ, Qin L, et al. Induction of a Striking Systemic Cytokine Cascade prior to Peak Viremia in Acute Human Immunodeficiency Virus Type 1 Infection, in Contrast to More Modest and Delayed Responses in Acute Hepatitis B and C Virus Infections. *Journal of virology* 2009; **83**(8): 3719.

71. Day CL, Kaufmann DE, Kiepiela P, et al. PD-1 expression on HIV-specific T cells is associated with T-cell exhaustion and disease progression. *Nature* 2006; **443**(7109): 350-4.

72. Giorgi JV, Hultin LE, McKeating JA, et al. Shorter Survival in Advanced Human Immunodeficiency Virus Type 1 Infection Is More Closely Associated with T Lymphocyte Activation than with Plasma Virus Burden or Virus Chemokine Coreceptor Usage. *The Journal of infectious diseases* 1999; **179**(4): 859-70.

73. Borrow P, Lewicki H, Wei X, et al. Antiviral pressure exerted by HIV-l-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. *Nature medicine* 1997; **3**(2): 205-11.

74. Liu Y, McNevin JP, Holte S, McElrath MJ, Mullins JI. Dynamics of Viral Evolution and CTL Responses in HIV-1 Infection. *PLOS ONE* 2011; **6**(1): e15639.

75. Veazey RS, DeMaria M, Chalifoux LV, et al. Gastrointestinal Tract as a Major Site of CD4+T Cell Depletion and Viral Replication in SIV Infection. *Science (New York, NY)* 1998; **280**(5362): 427-31.

76. Brenchley JM, Schacker TW, Ruff LE, et al. CD4+ T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract. *The Journal of experimental medicine* 2004; **200**(6): 749-59.

77. Mellors JW, Munoz A, Giorgi JV, et al. Plasma Viral Load and CD4+ Lymphocytes as Prognostic Markers of HIV-1 Infection. *Annals of Internal Medicine* 1997; **126**(12): 946-54.

78. Okoye AA, Picker LJ. CD4+ T-cell depletion in HIV infection: mechanisms of immunological failure. *Immunological Reviews* 2013; **254**(1): 54-64.

79. Brenchley JM, Price DA, Schacker TW, et al. Microbial translocation is a cause of systemic immune activation in chronic HIV infection. *Nature medicine* 2006; **12**(12): 1365-71.

80. Schacker TW, Nguyen PL, Beilman GJ, et al. Collagen deposition in HIV-1 infected lymphatic tissues and T cell homeostasis. *The Journal of Clinical Investigation* 2002; **110**(8): 1133-9.

81. Nasi M, De Biasi S, Gibellini L, et al. Ageing and inflammation in patients with HIV infection. *Clinical & Experimental Immunology* 2017; **187**(1): 44-52.

82. French MA, King MS, Tschampa JM, Silva BAd, Landay AL. Serum Immune Activation Markers Are Persistently Increased in Patients with HIV Infection after 6 Years of Antiretroviral Therapy despite Suppression of Viral Replication and Reconstitution of CD4+ T Cells. *The Journal of infectious diseases* 2009; **200**(8): 1212-5.

83. Ji JP, Loeb LA. Fidelity of HIV-1 reverse transcriptase copying RNA in vitro. *Biochemistry* 1992; **31**(4): 954-8.

84. Phillips RE, Rowland-Jones S, Nixon DF, et al. Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. *Nature* 1991; **354**(6353): 453-9.

85. Martinez-Picado J, Prado JG, Fry EE, et al. Fitness Cost of Escape Mutations in p24 Gag in Association with Control of Human Immunodeficiency Virus Type 1. *Journal of virology* 2006; **80**(7): 3617.

86. Li B, Gladden AD, Altfeld M, et al. Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution. *Journal of virology* 2007; **81**(1): 193-201.

87. Cleland A, Watson HG, Robertson P, Ludlam CA, Brown AJ. Evolution of zidovudine resistance-associated genotypes in human immunodeficiency virus type 1-infected patients. *J Acquir Immune Defic Syndr Hum Retrovirol* 1996; **12**(1): 6-18.

88. Ruff CT, Ray SC, Kwon P, et al. Persistence of Wild-Type Virus and Lack of Temporal Structure in the Latent Reservoir for Human Immunodeficiency Virus Type 1 in Pediatric Patients with Extensive Antiretroviral Exposure. *Journal of virology* 2002; **76**(18): 9481.

89. Tomaras GD, Yates NL, Liu P, et al. Initial B-Cell Responses to Transmitted Human Immunodeficiency Virus Type 1: Virion-Binding Immunoglobulin M (IgM) and IgG Antibodies Followed by Plasma Anti-gp41 Antibodies with Ineffective Control of Initial Viremia. *Journal of virology* 2008; **82**(24): 12449.

90. Korber B, Gaschen B, Yusim K, Thakallapally R, Kesmir C, Detours V. Evolutionary and immunological implications of contemporary HIV-1 variation. *British Medical Bulletin* 2001; **58**(1): 19-42.

91. Bonsignori M, Liao H-X, Gao F, et al. Antibody-virus co-evolution in HIV infection: paths for HIV vaccine development. *Immunological Reviews* 2017; **275**(1): 145-60.

92. Ho YC, Shan L, Hosmane NN, et al. Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* 2013; **155**(3): 540-51.

93. Bruner KM, Murray AJ, Pollack RA, et al. Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nature medicine* 2016; **22**: 1043.

94. Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* 2003; (1476-4687 (Electronic)).

95. Siliciano JD, Siliciano RF. Enhanced Culture Assay for Detection and Quantitation of Latently Infected, Resting CD4+ T-Cells Carrying Replication-Competent Virus in HIV-1-Infected Individuals. In: Zhu T, ed. Human Retrovirus Protocols: Virology and Molecular Biology. Totowa, NJ: Humana Press; 2005: 3-15.

96. Bruner KM, Wang Z, Simonetti FR, et al. A quantitative approach for measuring the reservoir of latent HIV-1 proviruses. *Nature* 2019; **566**(7742): 120-5.

97. Hiener B, Horsburgh BA, Eden JS, et al. Identification of Genetically Intact HIV-1 Proviruses in Specific CD4(+) T Cells from Effectively Treated Participants. 2017; (2211-1247 (Electronic)).

98. Martin AR, Siliciano RF. Progress Toward HIV Eradication: Case Reports, Current Efforts, and the Challenges Associated with Cure. *Annual Review of Medicine* 2016; **67**(1): 215-28.

99. Kim Y, Anderson JL, Lewin SR. Getting the "Kill" into "Shock and Kill": Strategies to Eliminate Latent HIV. *Cell Host & Microbe* 2018; **23**(1): 14-26.

100. Archin NM, Liberty AL, Kashuba AD, et al. Administration of vorinostat disrupts HIV-1 latency in patients on antiretroviral therapy. *Nature* 2012; **487**(7408): 482-5.

101. Dash PK, Kaminski R, Bella R, et al. Sequential LASER ART and CRISPR Treatments Eliminate HIV-1 in a Subset of Infected Humanized Mice. *Nature Communications* 2019; **10**(1): 2753.

102. Rothenberger MK, Keele BF, Wietgrefe SW, et al. Large number of rebounding/founder HIV variants emerge from multifocal infection in lymphatic tissues after treatment interruption. *Proceedings of the National Academy of Sciences* 2015; **112**(10): E1126.

103. Vibholm LK, Lorenzi JCC, Pai JA, et al. Characterization of Intact Proviruses in Blood and Lymph Node from HIV-Infected Individuals Undergoing Analytical Treatment Interruption. *Journal of virology* 2019; **93**(8): e01920-18.

104. Lu C-L, Pai JA, Nogueira L, et al. Relationship between intact HIV-1 proviruses in circulating CD4+ T cells and rebound viruses emerging during treatment interruption. *Proceedings of the National Academy of Sciences* 2018; **115**(48): E11341.

105. Kearney MF, Wiegand A, Shao W, et al. Origin of Rebound Plasma HIV Includes Cells with Identical Proviruses That Are Transcriptionally Active before Stopping of Antiretroviral Therapy. *Journal of virology* 2016; **90**(3): 1369.

106. Perreau M, Savoye A-L, De Crignis E, et al. Follicular helper T cells serve as the major CD4 T cell compartment for HIV-1 infection, replication, and production. *The Journal of experimental medicine* 2013; **210**(1): 143.

107. Gosselin A, Monteiro P, Chomont N, et al. Peripheral blood CCR4+CCR6+ and CXCR3+CCR6+CD4+ T cells are highly permissive to HIV-1 infection. *J Immunol* 2010; **184**(3): 1604-16.

108. Chomont N, El-Far M, Ancuta P, et al. HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nature medicine* 2009; **15**(8): 893-900.

109. Buzon MJ, Sun H, Li C, et al. HIV-1 persistence in CD4+ T cells with stem cell–like properties. *Nature medicine* 2014; **20**: 139.

110. Joos B, Fischer M, Kuster H, et al. HIV rebounds from latently infected cells, rather than from continuing low-level replication. *Proceedings of the National Academy of Sciences* 2008; **105**(43): 16725.

111. Lorenzo-Redondo R, Fryer HR, Bedford T, et al. Persistent HIV-1 replication maintains the tissue reservoir during therapy. *Nature* 2016; **530**: 51.

112. Josefsson L, von Stockenstrom S, Faria NR, et al. The HIV-1 reservoir in eight patients on long-term suppressive antiretroviral therapy is stable with few genetic changes over time. *Proceedings of the National Academy of Sciences* 2013; **110**(51): E4987.

113. Kearney MF, Spindler J, Shao W, et al. Lack of Detectable HIV-1 Molecular Evolution during Suppressive Antiretroviral Therapy. *PLoS pathogens* 2014; **10**(3): e1004010.

114. von Stockenstrom S, Odevall L, Lee E, et al. Longitudinal Genetic Characterization Reveals That Cell Proliferation Maintains a Persistent HIV Type 1 DNA Pool During Effective HIV Therapy. *The Journal of infectious diseases* 2015; **212**(4): 596-607.

115. Brodin J, Zanini F, Thebo L, et al. Establishment and stability of the latent HIV-1 DNA reservoir. 2016; (2050-084X (Electronic)).

116. Abrahams M-R, Joseph SB, Garrett N, et al. The replication-competent HIV-1 latent reservoir is primarily established near the time of therapy initiation. *Science Translational Medicine* 2019; **11**(513): eaaw5589.

117. Blankson JN, Finzi D, Pierson TC, et al. Biphasic decay of latently infected CD4+ T cells in acute human immunodeficiency virus type 1 infection. *The Journal of infectious diseases* 2000; **182**(6): 1636-42.