JCI The Journal of Clinical Investigation

Divergent populations of HIV-infected naïve and memory CD4+ T-cell clones in children on antiretroviral therapy

Mary Grace Katusiime, … , Maud Mavigner, Mary F. Kearney

J Clin Invest. 2025. https://doi.org/10.1172/JCI188533.

Clinical Research and Public Health | In-Press Preview | AIDS/HIV | Virology

BACKGROUND. Naïve cells comprise 90% of the CD4+ T-cell population in neonates and exhibit distinct age-specific capacities for proliferation and activation. We hypothesized that HIV-infected naïve CD4+ T-cell populations in children on long-term antiretroviral therapy (ART) would thus be distinct from infected memory cells.

METHODS. Peripheral blood naïve and memory CD4+ T cells from 8 children with perinatal HIV on ART initiated at age 1.7-17 months were isolated by FACS. DNA was extracted from sorted cells and HIV proviruses counted, evaluated for intactness, and subjected to integration site analysis.

RESULTS. Naïve CD4+ T cells containing HIV proviruses were detected in children with 95% statistical confidence. A median of 4.7% of LTR-containing naïve CD4+ T cells also contained HIV genetic elements consistent with intactness. Full-length proviral sequencing confirmed intactness of one provirus. In the participant with the greatest level of naïve cell infection, ISA revealed infected expanded cell clones in both naïve and memory T cells with no common HIV integration sites detected between subsets. Divergent integration site profiles reflected differential gene expression patterns of naïve and memory T cells.

CONCLUSIONS. These results demonstrate that HIV persists in both naïve and memory CD4+ T cells that undergo clonal expansion and harbor intact proviruses, suggesting that infected memory T-cell clones do not frequently arise from naïve cell differentiation in children with perinatal […]

Divergent populations of HIV-infected naïve and memory CD4+ T-cell clones in children on antiretroviral therapy

Mary Grace Katusiime[1*], Victoria Neer[1*], Shuang Guo[2], Sean C. Patro[2], Wenjie Wang[2], Brian Luke[3], Adam A. Capoferri[1], Xiaolin Wu[2], Anna M. Horner[4], Jason W. Rausch[1], Ann Chahroudi[4], Maud Mavigner[4†], Mary F. Kearney[1†]

[1]HIV Dynamics & Replication Program, Center for Cancer Research, NCI, Frederick, Maryland, [2]Cancer Research Technology Program and [3]Advanced Biomedical Computational Science, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, Maryland, [4]Department of Pediatrics, Emory University School of Medicine, Atlanta, Georgia

Corresponding authors[†]:

Maud Mavigner, Ph.D.
Emory University School of Medicine
N446 HSRBII - 1750 Haygood Dr
Atlanta, GA 30322-1119
maud.mavigner@emory.edu
404-727-3618

Mary Kearney, Ph.D.
HIV Dynamics and Replication Program
Center for Cancer Research, NCI, NIH
1050 Boyles Street, Building 535, Room 109, Frederick, MD 21702-1201, USA
kearneym@nih.gov
301-305-8197

Authorship notes: [*]MGK and VN contributed equally to this work. [†]MFK and MM are co-senior authors.

The authors have declared that no conflict of interest exists.

ABSTRACT

BACKGROUND. Naïve cells comprise 90% of the CD4+ T-cell population in neonates and exhibit distinct age-specific capacities for proliferation and activation. We hypothesized that HIV-infected naïve CD4+ T-cell populations in children on long-term antiretroviral therapy (ART) would thus be distinct from infected memory cells.

METHODS. Peripheral blood naïve and memory CD4+ T cells from 8 children with perinatal HIV on ART initiated at age 1.7-17 months were isolated by FACS. DNA was extracted from sorted cells and HIV proviruses counted, evaluated for intactness, and subjected to integration site analysis.

RESULTS. Naïve CD4+ T cells containing HIV proviruses were detected in children with 95% statistical confidence. A median of 4.7% of LTR-containing naïve CD4+ T cells also contained HIV genetic elements consistent with intactness. Full-length proviral sequencing confirmed intactness of one provirus. In the participant with the greatest level of naïve cell infection, integration sites analysis (ISA) revealed infected expanded cell clones in both naïve and memory T cells with no common HIV integration sites detected between subsets. Divergent integration site profiles reflected differential gene expression patterns of naïve and memory T cells.

CONCLUSIONS. These results demonstrate that HIV persists in both naïve and memory CD4+ T cells that undergo clonal expansion and harbor intact proviruses, suggesting that infected memory T-cell clones do not frequently arise from naïve cell differentiation in children with perinatal HIV on long-term ART.

FUNDING. Center for Cancer Research, NCI and Office of AIDS Research funding to MFK, NCI FLEX funding to JWR. Children's and Emory JFF pilot to MM.

INTRODUCTION

Of the 39 million people living with HIV (PLWH) in 2022, 1.5 million were children under 15 years of age (1), most of whom acquired the virus through vertical transmission during pregnancy, labor, delivery, or breastfeeding (2). Although maternal use of antiretroviral therapy (ART) is highly effective in preventing vertical HIV transmission (3), new pediatric infections continue to occur, due largely to the high prevalence and incidence of HIV in women of childbearing age, unavailable or incomplete implementation of prevention modalities, and/or sub-optimal timing of or adherence to ART (4).

Without treatment, more than half of children who acquire HIV vertically die by 2 years of age (5), whereas the median survival time for ART-naïve adults with HIV is 11 years (6). Children living with HIV (CLWH) also tend to have higher peak viremia and take longer to reach setpoint compared to adults (7, 8). Moreover, while early ART administration clearly benefits both adults and children (9-16), anecdotal reports suggest that post-treatment control in children may be more attainable in the setting of ART initiation shortly after birth (17-19).

These distinctive features of perinatal HIV infection are attributable, at least in part, to the equally distinctive characteristics of the developing immune system. For instance, in healthy individuals, naïve cells comprise a much greater fraction of the CD4+ T-cell pool early in life relative to adulthood, declining from ~90% at birth to approximately 50% by age 40 (20). Neonatal naïve T-cell populations are also derived from different stem cells than those in adults, have lower T-cell receptor (TCR) diversity than adult naïve T-cells, are less prone to contract into memory cells following activation, and are more likely to proliferate homeostatically, particularly under lymphopenic conditions (21-24).

Although memory cells comprise the greatest fraction of CD4+ T cells with HIV proviruses in adults on ART (25, 26), HIV proviruses are also found in naïve CD4+ T cells (26-34). The primary block to HIV infection of naïve CD4+ T cells occurs at the binding stage of the virus life cycle (35, 36), as the cellular chemokine receptor CCR5, which functions as an HIV entry

co-receptor, is not well expressed in this cell type. The alternate HIV co-receptor CXCR4 is present on naïve CD4+ T cells but viruses that utilize this co-receptor are typically not associated with transmission. Given the dominance of naïve CD4+ T cells at the time of vertical HIV acquisition, we hypothesized that the resulting pool of persistently infected naïve CD4+ T cells is important to consider in studies of curative interventions for CLWH and thus sought to more deeply understand features of the naïve T-cell reservoir.

We have previously demonstrated ~1% HIV proviral intactness frequency in peripheral blood mononuclear cells (PBMCs) from children who started ART at a median of 5.4 months of life (37) and integration site analysis (ISA) showed the early emergence of clonal infected cell populations in CLWH (38). Moreover, a comprehensive two-year longitudinal analysis of provirus intactness frequencies, integration sites, and immunological consequences in treated and untreated CLWH showed that immediate ART administration in neonates can reduce reservoir size and induce a distinct immune profile (39). While each of these studies revealed features of genetic structures of provirus populations, sites of integration, and/or propensity toward clonal expansion in children on ART, none address how these properties might differ in CD4+ naïve and memory T-cell subsets.

In this work, we characterize the proviruses persisting in circulating naïve and memory CD4+ T cells sorted from 8 CLWH who received ART from early in life. Using multiple displacement amplification (MDA), an adapted intact proviral DNA assay (IPDA), and integration site analysis, we provide evidence for the presence of intact proviruses in and clonal expansion of infected naïve CD4+ T cells of CLWH on long-term ART, including in samples from one participant in which the fractional contribution of naïve T cells to the infected cell population at the time of sampling was estimated to be 40%. In-depth analysis of samples from this participant did not detect overlap between infected cell clonal populations in naïve and memory CD4+ T cells. Additionally, proviral integration sites detected in infected naïve cells were more often found in genes expressed to higher levels in that subset relative to memory. However, except for the most highly expressed genes, this pattern was not observed among

infected memory cell integration sites, perhaps indicative of persistent infected naïve cells being activated and differentiating to memory in the years prior to sample collection. Taken together, our findings demonstrate that naïve CD4+ T cells are an important and distinct reservoir in children with perinatally-acquired HIV.

RESULTS

Participants, samples, and cell sorting. Four male and 4 female children aged 5-11 years were enrolled in the study. All 8 children acquired HIV vertically, were on ART, and had levels of HIV RNA in plasma that were <20 copies/mL at the time of blood collection. The demographic and virologic characteristics of the participants are shown in Table 1. HIV subtype was known for 7/8 children, and 5 children had subtype B while 2 acquired subtype C. The children initiated ART at a median age of 8.75 months (range: 1.5 to 17 months) and had received ART for a median of 8 years (range: 5 to 10 years). Populations of naïve and memory CD4+ T cells were isolated from peripheral blood mononuclear cells (PBMCs) by FACS, with naïve CD4+ T cells defined as CD45RO-CD45RA+CD28+CD27+CCR7+CD95- and memory CD4+ T cells as CD45RO+CD95+ (Figure S1). Sorting gates were set far apart to minimize contamination, and the median purity of the sorted cells was 97% (range: 93-100%; Table 2, Table S1).

HIV DNA persists in naïve CD4+ T cells in CLWH on long-term ART. Cellular DNA, including HIV proviruses, was extracted from sorted cell populations, dispensed to microtiter plates at limiting proviral dilutions, and amplified by multiple displacement amplification (MDA) as described in Methods and previously (40). This approach enabled re-sampling of the amplified products for subsequent analyses including real-time PCR screening, integration site analysis, and proviral sequencing. To calculate the infected cell frequencies in the naïve and memory CD4+ T-cell sorts, MDA products were screened by PCR for the presence of the HIV LTR. Poisson-corrected LTR-positive well counts were divided by the total numbers of cells screened. HIV LTR DNA was detected in naïve CD4+ T-cell sorts from all participants with a median frequency of 48 (range 5 to 266) copies per million cells (Table 2, Figure 1A). HIV LTR DNA was detected in the memory T-cell sorts with a median of 977 (range 252 to 1,546) copies per million cells (Table S1, Figure 1A).

To assess the robustness of these data, we then estimated the frequency of possible contaminating memory CD4+ T cells in the sorted naïve cells. For each participant, the

expected and maximum number of possible memory T-cell contaminants (95% confidence) within the sorted naïve CD4+ T cells were calculated using the number of cells, number of HIV LTR copies, and sort purities as previously described (41) (Table 2). In turn, these values were used to calculate minimum and expected HIV LTR copies arising from naïve cells, rather than contaminating memory cells. The median number of HIV-infected naïve cells, after correcting for possible contaminating memory cells, was 33 per million (Table 2). The corresponding values were also calculated for the memory T-cell sorts resulting in a median of 975 HIV-infected memory cells per million (Table S1).

Using HIV-infected cell frequencies corrected for possible cross-contamination and frequencies of naïve and memory CD4+ T cells obtained from the cell sorting (Figure 1B), we calculated the contribution of each subset to the total HIV-infected peripheral CD4+ T-cell population (Figure 1C). Participant 5005 was excluded from this analysis because the observed Poisson-corrected measured number of infected naïve T cells was less than the number expected from contaminating memory cells in this sort. The contribution of naïve CD4+ T cells to the total level of HIV+ CD4+ T cells ranged from 4.9% to 40% across the remaining participants, with a median contribution of 13.5% (Figure 1C).

Genetically intact proviruses persist in naïve CD4+ T cells in CLWH on long-term ART. To assess whether any of the HIV LTR detected in naïve CD4+ T cells derived from intact proviruses, HIV LTR+ MDA products were screened for HIV Psi (Ψ) and Rev Response Element (RRE) by real-time PCR using a modified Intact Proviral DNA Assay (IPDA) (Figure 2). Detection of both HIV genetic elements can be used as a surrogate for full HIV DNA sequencing to infer proviral intactness, as previously reported (42). Of the more than 200 wells containing single infected naïve cells across all 8 participants, 13.1% screened positive for LTR + Psi (indicating a 3'-defective HIV), 7.1% screened positive for LTR + RRE (5'-defective), and 75.1% did not screen positive for either Psi or RRE, indicating a provirus with both 5' and 3' defects or a solo-LTR (43). The 4.7% of MDA-amplified proviruses that screened positive for both Psi and RRE are potentially intact and capable of contributing to rebound

viremia if ART is withdrawn. This intactness frequency is comparable to those previously reported for children and adults with HIV on ART, irrespective of T-cell subtype (37, 44).

Having identified 10 candidate-intact proviruses in naïve CD4+ T cells across all donors, we sought to directly confirm intactness by PCR amplification and near full-length sequencing. Despite challenges associated with segmental amplification of HIV proviruses by MDA, as well as the inherent genetic variation of viral genomes (45), we were able to assemble a contiguous full-length sequence for one provirus found in the naïve CD4+ T-cell population from participant 1001 (GenBank accession number PV031991). Sequence-intactness was confirmed using the proviral sequence annotation & intactness test (ProSeq-IT) (45). To ensure that the intact sequence was not a laboratory contaminant, we performed phylogenetic analysis by comparing the U3 and *pol* regions to laboratory strains of HIV, to sequences obtained from other donor samples in the laboratory, and to a random collection of other clinical samples of the same subtype (Table S2, Figure S2). The genetic distance between the intact proviral sequence found in naive CD4+ T cells from participant 1001 and lab-adapted HIV strains indicates that the new HIV sequence was not a laboratory contaminant. Together, our results demonstrate that intact HIV proviruses persist in naïve CD4+ T cells in children with vertically-acquired HIV on long-term ART.

Correlations among immunological, clinical, and virological study parameters. Multiple pairwise Spearman correlation analyses revealed significant correlations among immunological, clinical, and virological study parameters (Figure 3). Most notably, naïve T-cell contributions to peripheral CD4+ T-cell pools, infected naïve T-cell contributions to total infected cells, and fractions of proviruses in infected naïve T cells that are potentially intact are all negatively correlated with participant age at time of sample collection (with Spearman correlation coefficients of -0.67, -0.77, and -0.77, respectively). These findings are likely reflective of natural transition from naïve to memory CD4+ T-cell predominance during normal immune system development, and perhaps also the selective decay of infected naïve CD4+ T-cell populations harboring intact proviruses after years on ART.

Proviruses persisting in naïve CD4+ T cells in CLWH on long-term ART are predicted to be CCR5-tropic. Whereas most transmitted HIV are R5 tropic (71), naïve CD4+ T cells express CCR5 only transiently and are relatively resistent to infection with R5-tropic virus in vitro (35). To determine whether HIV+ naïve CD4+ T cells identified in samples from participants in this study were the result of infection with R5-tropic virus or the less-abundant X4-tropic virus, we sequenced and analyzed the *env* V3 loop region amplified by PCR from RRE+ MDA products from participants 0444, 0555 and 1001 (Genbank accession numbers PV031992-PV031999). Coreceptor usage was predicted in silico using both the Position Specific Score Matrices (PSSM), which generates probability scores from the identities of amino acids or nucleotides at specific postitions (46), and Geno2Pheno (47). All tested proviruses from naïve CD4+ T cells were predicted to utilize CCR5, in accordance with previously published results for naïve CD4+ T cells obtained from adults living with HIV on ART (41).

Proviruses persisting in naïve CD4+ T cells in CLWH on long-term ART can be clonally expanded and are distinct from memory CD4+ T cell proviral populations. Identifying and cataloging the sites of integration of HIV proviruses in people with HIV on ART can be informative of the molecular and subcellular determinants favorable to integration (48, 49), viral persistence determinants and selection pressures (50), and the mechanisms and degrees of infected cell clonal expansion (51). To characterize the proviral landscape in naïve and memory T-cell populations, we therefore performed integration site analyses (ISA) using HIV+ MDA products from participant 1001, selected for the high frequency of infection of naïve CD4+ T cells. Though insufficient to support a comparative analysis of naïve and memory cell integration site profiles, 26 additional integration sites identified in naïve cells from PID 0444, PID 0555, and PID 9009 have been made available in the Retrovirus Integration Sites Database (RID) (52) in connection with this study.

ISA results for naïve and memory CD4+ T cells from participant 1001 are shown graphically in Figure 4 and are also available in the RID. We found that 71.7% of the proviral integration

sites in naïve CD4+ T cells were in genes and 40.3% of proviruses were in the same orientation as the gene. Memory CD4+ T cells had similar frequencies of proviral integration sites in genes (71.9%), with 42.7% of proviruses co-oriented with the gene. In naïve cells, we identified 67 unique integration sites for 83 MDA-amplified proviruses, of which 58 were each identified exactly once (singlets; 86.5%) and 9 more than once (clones; 13.4%). In memory CD4+ T cells, 151 MDA-amplified proviruses yielded 96 unique integration sites, of which 72 were singlets (75%) and 24 were clones (25%). In this context, mapping multiple proviruses to the same integration site definitively establishes clonal expansion of the harboring T cells. However, it is important to note that the converse, i.e., detecting an integration site only once, does not demonstrate that the T cell harboring that provirus is not part of a clonally expanded population, only that we can neither confirm nor refute this assertion.

Notably, 6 of the 83 (7.2%) amplified proviruses in naïve CD4+ T cells for which we could identify an integration site mapped to a common location within the *TSEN34* gene. Likewise, 22 of 151 (14.6%) amplified proviruses in memory CD4+ T cells mapped to a common integration site in *HLA-A*. These findings are indicative of exceptionally large clonal populations relative to other HIV+ cells in this individual. Nevertheless, none of the integration sites identified, including those of proviruses in large clonal populations, were detected in both the naïve and memory CD4+ T-cell subsets.

Taken together, an important implication of these findings is that infected naïve and memory infected cell pools are not completely mixed. If they were, the binomial probability of finding even the smallest confirmed clone twice in the naïve T cell sample (2/83) and not at all among memory T cells (0/151) would be 0.0251 (or 2.51%). For larger clones, as the number of clonal naïve cell integration site detections increases (e.g., 3/81 and 4/81), the probability of finding none among memory T cells decreases (0.385% and 0.0577%), respectively. Yet none of the clones found in the naïve cells were represented among memory cells. The probability of this collective result, assuming complete mixing, equals the product of probabilities calculated for each occurrence individually, or $p=2.6086 \times 10^{-40}$, effectively zero. These results demonstrate

that naïve CD4+ T cells in children born with HIV can be infected, can clonally expand, and can persist for years on ART, and that their detection does not result from contamination of infected memory cells in the sorting process.

Distinct expression patterns of integration site genes in naïve and memory CD4+ T cells. It is now well-established that HIV preferentially integrates into expressed genes (48, 49) and that naïve and memory CD4+ T cells exhibit distinct transcriptional profiles (53). From this foundation, and to explore a potential contributing factor to the sparse overlap among integration site genes identified from sorted naïve and memory CD4+ T-cell subset collections (CHD2 and LOC339788 are the only genes in common, although the integration sites in naïve and memory cells are different even in these genes), we examined whether relative expression levels of these genes were generally greater in the CD4+ T-cell subset in which they were identified. Using a previously published RNAseq dataset from sorted peripheral naïve and memory CD4+ T cells from 8 PWH on ART as a reference (54), expression levels of integration site genes identified in sorted PID 1001 naïve and memory CD4+ T cells are presented in Figure 5A and 5B, respectively.

All but one of the 50 integration site genes identified in naïve CD4+ T cells exhibited detectable levels of expression in naïve T cells, 46 of which yielded average TPM values >2 and are presented in descending order of naïve T cell expression in Figure 5A. Remarkably, 36/46 (78%) naïve T-cell integration site genes were expressed to higher levels in naïve cells relative to memory cells by a median of 28% (range: 3%-312%). These discrete findings were supported by analyses of relative expression of naïve T cell integration site genes in aggregate (Figure S3A), i.e., naïve T cell integration site genes were expressed in naïve CD4+ T cells at significantly higher levels than in memory CD4+ T cells ($P = 0.0003$, Figure S3). The integration site gene that best exemplifies this trend is *BACH2*, with an average relative expression more than four-fold higher in naïve than in memory T cells. Collectively, these results are consistent with the general tendency of HIV to integrate into genes that are more highly expressed. However, this tendency is not absolute, as 10 naïve cell integration site

genes (22%) were expressed at a median 16% higher levels in memory T cells in the RNAseq analysis, including the most notable counterexample *IL2RB*, which is expressed 192% more in memory than in naïve T cells.

Yet the tendency for integration site genes to be more highly expressed in the subset in which they were identified is not observed among the 50 most highly expressed integration site genes identified in memory cells, as only 21 (42%) are more highly expressed in memory than in naïve T cells (Figure 5B), by a median of 14%. Instead, the majority of memory T-cell integration site genes (29/50; 58%) are actually expressed to higher levels in naïve cells by a median of 15%. These findings are likewise supported by graphical and statistical analysis of memory T cell integration site genes in aggregate (Figure S3B), which revealed no statistically significant difference in expression between the two subsets ($P = 0.3435$). One potential explanation of these results is that some naïve cells that were infected in infancy, collectively harboring proviruses in genes more highly expressed in this subset, were activated and ultimately differentiated into memory during the intervening years on ART prior to sample collection. Such cells, despite collectively having a naïve T-cell integration site profile, would sort to the memory cell collection and thus mask the trend in this subset.

DISCUSSION

In this study, HIV proviral populations in circulating naïve and memory $CD4^+$ T cells from 8 children with vertically acquired HIV on ART for a median of 8 years were quantified and characterized. HIV was consistently detected in naïve CD4+ T cell sorted collections at frequencies that exceeded the upper bounds of possible infected memory cells contaminants with 95% statistical confidence. Frequencies of naïve CD4+ T-cell infection varied among participants and were markedly lower than those determined for memory CD4+ T cells, in agreement with previously reported measurements from adult samples (26, 55-57). However, naïve cells comprised the majority of the CD4+ T-cell pool and were consequently found to substantially contribute to the HIV reservoir as previously reported in pediatric nonhuman primate models of SIV/SHIV infection (57, 58)**.** As expected, age was found to be strongly negatively correlated with the circulating levels of naïve CD4+ T cells (frequency and absolute count) as well as with their contribution to the HIV reservoir. The frequency of infection of naïve CD4+ T cells from participant 1001 was the highest of all participants, representing 40% of the total infected peripheral CD4+ T cells in this donor. It may be that variations in naïve CD4+ T-cell numbers, lineages, and fractional contribution to total peripheral T cells between the times of infection and ART administration contribute to differences in these values among participants even after years on ART. Moreover, the mode of vertical transmission might have a pronounced effect on early infection dynamics that is perpetuated longitudinally. For instance, direct HIV transmission to the fetal circulatory system in utero might introduce the virus to a T-cell population comprised almost entirely of fetal naïve cells having a distinctive lineage and differing capacities relative to older children and adults (21-24), whereas post-natal exposure through breastfeeding might instead introduce the virus to a developing gut enriched with early memory cells (59). Regardless, the relative abundance of HIV+ naïve CD4+ T cells in PID 1001 samples allowed us to deeply characterize proviruses obtained from this participant.

By screening for HIV genetic elements Psi and RRE in a modified IPDA (42), we found a mean frequency of 4.7% potentially intact proviruses in naïve CD4+ T cells. The intactness of one provirus from participant 1001 was confirmed by segmental PCR amplification and sequencing. The frequencies of intact proviral DNA we report here in naïve CD4+ T cells, ranging from <1% to 8.8% are generally consistent with frequencies reported in total CD4+ T cells for adults initiating ART during acute or chronic infection and for children initiated on ART early in life estimated using near-full length sequencing methods (37, 44). Little data exist regarding the frequency of intact proviruses within naïve CD4+ T cells. One study limited to two adults living with HIV initiated on ART during the chronic phase of HIV infection reported 3.3% and 6.6% of intact proviruses in naïve CD4+ T cells after 9 years of ART (41). Another report from the same group shows very variable levels of intact proviruses, 0.9% to 17% after 5-6 years of ART, in the naïve CD4+ T cells from 5 adult chronic progressors specifically selected for their large range of reservoir sizes (60). Generally low to undetectable levels of intact proviruses were also found in 2 groups of 5 adults treated early or late during HIV chronic infection (61).

Despite naïve T cells typically expressing CCR5 at low levels (62-64), we find that all proviruses harbored in naïve CD4+ T cells from multiple participants in this cohort were predicted R5-tropic, in agreement with previous co-receptor usage predictions for naïve CD4+ T-cell proviruses in adults (41). Moreover, higher infection frequencies, such as we observe in naïve cell samples from PID 1001, may result from transiently elevated CCR5 expression, caused by various stimuli (35, 65, 66) and perhaps coinciding with expansive virus replication early in infection. Certain conditions can enhance HIV infection of naïve CD4+ T cells, including transmission in a tonsil lymphoid tissue microenvironment ex vivo (67, 68) or in co-culture with monocyte-derived dendritic cells (MDDCs) in vitro (69), though the latter condition was only shown to enhance infection by CXCR4-tropic (X4), but not CCR5-tropic (R5), virus. Of particular note with respect to this study, pre-incubation with IL-7, a cytokine critical for naïve T-cell homeostatic proliferation and survival (70), has been shown to sensitize umbilical

cord blood naïve T cells, but not adult naïve T cells, to HIV infection (71-73). Although further study is required, it is tempting to speculate that elevated IL-7 levels in infancy or later childhood may have contributed to the relatively high frequencies of both infection and clonal expansion reported for PID 1001's naïve CD4+ T cells.

Naïve CD4+ T cells display distinct characteristics that may affect HIV reservoir dynamics. Naïve cells have long half-lives and a slow turnover rate as compared to other maturational T-cell subsets and as such exhibit the slowest HIV DNA decay over time, typically affected more by differentiation into memory cells than by cell death (74). Here, HIV integration site analysis allowed us to assess the degree to which specific clonal populations were expanded in the naïve and memory CD4+ T-cell subsets of PID 1001. Samples from this participant proved to be exceptional in this regard as well, with large clonal populations harboring proviruses integrated into *TSEN34* and *HLA-A* projected to comprise 7.2% and 14.6% of HIV+ peripheral naïve and memory CD4+ T-cell subsets, respectively, at the time of sample collection. Indeed, detection of clonal infected cell populations in either subset more than once at this level of sampling is indicative of a high degree of clonal expansion, as was the case for 13.4% of naïve and 25% of memory CD4+ T-cell integration sites from this donor. Though the frequencies of clonal expansion measured here are quite high, relative levels between the two subsets are consistent with reports indicating that the degree of naïve T-cell clonal expansion, driven by homeostatic mechanisms, is typically less than that of memory cells driven by both antigen responses and homeostatic proliferation (41, 75). However, determining absolute clone sizes or sizes relative to uninfected cell populations would require a broad TCR characterization of much larger naïve and memory CD4+ T-cell samples.

We analyzed proviral integration sites and relative expression levels of genic integration targets for evidence of differentiation of infected naïve into memory CD4+ T cells between the times of infection and sample collection. The most direct evidence of this having occurred would be the identification of an integration site common to sorted naïve and memory CD4+ T cells, but this was not observed, even for the largest expanded clones characterized by

integration into *TSEN34* and *HLA-A*. While integrations into genes *CHD2* and *LOC339788* were observed in both naïve and memory CD4+ T cells, the integrations occurred at distinct positions in these genes, and thus neither are indicative of differentiation of clonally expanded infected naïve T-cell populations. A prior analysis of HIV integration sites in naïve and memory CD4+ T-cell subsets in adults found rare instances of identical integration sites only in the memory populations, suggestive of linear differentiation from infected central memory cells into effector memory cells (76)**.** This result contrasts with a previous study in adults that reported common proviral sequences between naïve and subsets of memory CD4+ T cells (41). It is important to note that not identifying clonal population of HIV+ CD4+ T cells in both naïve and memory sorted collections does not mean that differentiation of infected naïve cells into memory does not occur. Indeed, we should perhaps expect such observations to be rare, even exceedingly so, since identification of an HIV+ clone in both subsets would only be favored if it had become highly expanded prior to antigen-mediated activation, was overrepresented relative to other memory cell clones afterwards, and both subpopulations remained elevated at the time of sample collection.

HIV preferentially integrates into transcriptionally active genes (48, 49) and ART effectively suppresses ongoing viral replication (77). One implication of these now well-established precepts is that integration sites identified in individuals on ART have the potential to serve as indicators of the transcriptional state of cells at the time of infection, as well as how drug, fitness, and immune selective pressures shape the pool of persistent infected cells over time on ART. Considering levels of gene expression and integration site frequency for the first time in segregated naïve and memory cells obtained from a CLWH, we found that nearly 4 in 5 integration site genes identified in naïve cells are more highly expressed in this subset relative to memory. The converse was true for the 5 most highly expressed integration site genes identified in HIV+ memory cells, with HLA-A being the integration site for the largest clone of memory CD4+ T cells as well as the most highly expressed gene in both memory and naïve cells. These patterns both support our conclusion that most HIV+ cells in the sorted naïve cell

collection are not memory cell contaminants and reveal relative gene expression as a useful comparative metric of subset-selective integration site preferences.

Among the specific integration site genes identified in naïve CD4+ T cells, *BACH2* and *IL2RB* are particularly worthy of note, as both have been reported to confer a survival advantage to infected T cells (50, 51). We identified three independent integration sites in *BACH2* among PID 1001's naïve CD4+ T cells – the only integration site gene meriting this distinction. Moreover, all integrations in *BACH2* were found within introns, and in the same orientation as gene transcription. *BACH2* expression levels were more than 4-fold higher in naïve CD4+ T cells than in memory, and this gene has specifically been shown to maintain T cells in a naive state by suppressing effector memory-related genes (78). Taken together, these data support the notion that the survival advantage conferred by HIV integration into BACH2 may be naïve cell selective. Unlike *BACH2*, *IL2RB* is clearly associated with T-cell memory function, encoding the beta subunit of the high-affinity form of the receptor (IL-2R) for interleukin-2 (IL-2), a cytokine that mediates activated CD4+ T-cell development into memory, among other functions (79, 80). Moreover, relative *IL2RB* expression in memory CD4+ T cells exceeded that in naïve CD4+ T cells by almost 3-fold. Surprisingly, the *IL2RB* integration site was identified twice in naïve CD4+ T cells, indicative of a clonally expanded T-cell population, but not in memory CD4+ T cells. It must be noted, however, that the site in naïve cells was scored as gene-proximal (i.e., 3176 base pairs outside of the gene), not in an intron, as the proposed mechanism for conferring a survival advantage would require. Though anecdotal, our results suggest that the capacity for integration into these genes to confer a survival advantage may be selective for naïve or memory T-cell subtypes.

It is important to note that approximately 91% of proviruses contributing to the naïve T cell integration site profile of PID 1001 were found to be defective (IPDA-) and not part of a viral reservoir capable of kindling viremic rebound when ART is withdrawn. Further, most naïve cell integration sites were identified near or within expressed genes which is consistent with this observation, as intact HIV proviruses that persist on long-term ART have been reported to be

mostly integrated in non-genic, pericentromeric, or other regions of the genome considered repressive to gene expression (81-86). Whether the distinctive integration site profiles of persistent reservoir cells can be further dissected to reveal patterns specific to naïve cells, or even naïve cells in children who acquired HIV at birth, should be a priority subject for future studies.

In this work, we provide evidence for the persistence of HIV proviruses in naïve CD4+ T cells in children with perinatally-acquired HIV on long-term ART. Frequencies of genetically intact proviruses in this subset resemble those previously reported for children and adults, providing evidence that naïve cells contribute to the HIV reservoir in children. Our findings also indicate that large HIV+ naïve CD4+ T-cell clones are not predisposed toward activation, proliferation, and/or differentiation into memory, indicating that they may be a more challenging target for latency reversal cure strategies. Collectively, these results serve as a foundation for future studies that, with greater and longitudinal sampling, may allow us to further distinguish subtype-specific and generalized characteristics of infection in people living with HIV at different ages and phases of immune development.

METHODS

Sex as a biological variable. Our study evaluated samples from male and female children with HIV, and similar findings are reported for both sexes.

Naïve and memory CD4+ T cell isolation by fluorescence-activated cell sorting (FACS). PBMCs were isolated by density gradient centrifugation and CD4+ T cells were negatively selected from freshly isolated PBMC using magnetically labelled microbeads and subsequent column purification according to the manufacturer's protocol (Miltenyi Biotec). Enriched peripheral CD4+ T cells were rested overnight in RPMI-20% FBS then stained with previously determined volumes of the following fluorescently conjugated monoclonal antibodies: CD3-AF700 (clone SP34-2), CCR7-FITC (clone 150503), CD8-APC-Cy7 (clone SK1), CD45RA-PE-Cy7 (clone 5H9), CD45RO-APC (clone UCHL1), and CD27-PE (clone M-T271) from BD Bioscience, CD28-PE-Cy5.5 (clone CD28.2) from Beckman Coulter, and CD4-BV650 (clone OKT4) and CD95-BV605 (clone DX2) from BioLegend. Circulating populations were defined as follows: naïve CD4+, CD45RO- CD45RA+ CD28+ CD27+ CCR7+ CD95- and memory CD4+, CD45RO+ CD95+. Sorting was performed on a FACSAria LSR II (BD Biosciences) equipped with FACSDiva software.

Multiple displacement amplification of single proviruses. MDA was performed as previously described (40) with minor modifications. Genomic DNA was extracted from collections of sorted naïve and memory T cells and partitioned across 96- or 384-well plates so that <30% of wells contained a single provirus. DNA in individual wells was chemically denatured and neutralized, after which the remaining reaction components were added to the final concentrations indicated: 1x Phi29 DNA pol reaction buffer (NEB), 50 mM KCl, 1 mM dNTPs, 200 mM trehalose, 50 µM random DNA pentamer (5N) or hexamer (6N), and 0.25 U/µL NEB Phi29 DNA polymerase. Final reaction volumes were scaled to 40 µL and 10 µL for 96- and 384-well plates, respectively. Reactions were incubated at 30°C (6N primer) or 28°C (5N) for 18hr, then for 10 min at 65°C to inactivate the polymerase.

HIV Long Terminal Repeat (LTR), Psi, and Rev Response Element (RRE) detection in MDA products. Aliquots of 1:3 MDA product dilutions in 5 mM Tris, pH 8 were used for sequential screening, first for HIV LTR, and then for Psi and RRE, by single-plex and multiplex real-time PCR, respectively. Dedicated sets of primers and hydrolysis probes and 2X Roche LightCycler 480 Probes Master mix were utilized in these reactions in accordance with manufacturer's instructions. Psi and RRE primers and probes (Table S3A) were adapted from the Intact Proviral DNA Assay (IPDA) (42). The purpose behind slightly modifying primer and probe sequences, together with increasing the number of PCR cycles from 50 to 55, was to maximize assay sensitivity and tolerance for to minor variation in target sequences, including across HIV subtypes. Amplifying from MDA wells rather than single-copy proviruses in genomic DNA was also advantageous in these respects.

Calculating the probability of contaminating memory CD4+ T cells in the naïve CD4+ T-cell subset. The sizes and purities of the naïve and memory populations and the observed numbers of infected cells in each participant were used to determine the number of possible memory T-cell contaminants within naïve T-cell subsets (Schema S1). Briefly, the total number of infected cells is the sum of the fraction of naïve cells that are infected (fN) times the number of naïve cells, plus the fraction of memory cells that are infected (fM) times the number of memory cells. From this equation for each subset, fN and fM are determined, where fM times the number of memory cells in the naïve population gives the expected number of infected cells arising from the memory cells. Poisson distribution is used to determine the 95% probability that the number of infected memory cells in the respective infected naïve T-cell counts is Nmax or less.

$$\sum_{i=0}^{Nmax} \frac{\lambda^i e^{-\lambda}}{i!} \geq 0.95$$

These values were in turn used to determine the minimum number infected naïve cells in the infected naïve T-cell counts.

Determining intactness of HIV proviruses by segmental PCR and long-read sequencing. Proviruses in participant 1001 naïve T-cell MDA reactions that screened positive for LTR, Psi, and RRE were reamplified by nested or double-nested PCR in segments using primers listed in Table S3B. Using HXB2 sequence as a reference, primers were designed for nested PCR of four overlapping segments (Q1-Q4) with expected outer/inner amplicon sizes of 2145/2056, 2524/2399, 2922/2841, and 1947/1876 basepairs (bp), respectively. Inner nested PCR products visible by agarose gel electrophoresis and of appropriate size were sequenced. To obtain Q1-Q4 segments that failed to amplify, MDA products or outer Q1-Q4 PCR amplicons were re-amplified in smaller sub-segments of Q1-Q4, dubbed D1-D10 (where D1&D2 comprise Q1; D3, D4 & D5, comprise Q2; D6, D7 & D8 comprise Q3; and D9 & D10 comprise Q4). Full coverage was achieved for products of one PID 1001 naïve T cell MDA reaction. Overlapping amplicons were individually subjected to Sanger sequencing and pooled for construction of an Oxford Nanopore Technologies (ONT) sequencing library using the LSK114 ligation sequencing kit. ONT sequencing was performed in a Flongle Flow Cell (R10.4.1) on a MinION Mk1C instrument and basecalled using the “superaccuracy” model (SUP) in Guppy v6.5.7 software. A near-full length provirus contig was assembled from ONT sequences and Sanger sequences were overlaid for sequence validation and to fix ambiguous base calls using Geneious Prime® 2020.2.4 software. The completed assembly was submitted for analysis using the proviral sequence annotation & intactness test (ProSeq-IT) (45) and was determined to be sequence-intact.

Phylogenetic analysis of intact provirus identified in naïve CD4+ T cells. Neighbor-joining phylogenetic analysis was used to ensure that the intact sequence detected in PID 1001 naïve T cells was not the result of laboratory or cross sample contamination. The HIV LANL Sequence Database was accessed to obtain pre-generated full-length genome alignments. Subtype B sequences that were dated between 2000-2019 were randomly sampled and

globally distributed (n=29) (Table S3). Additionally, 3 sequences were selected as they were previously amplified products in our laboratory (87) (donor C02, MT744340; donor F07, MT745572; donor R09, MT745575). To rule out potential lab strain contamination, NL4.3 (AF324493.2), JRCSF (M38429.1), 89.6 (U39362.2), YU2 (M93258.1), LAI (MN691959.1), and HXB2/LAV (K03455.1) were included in the analysis. A near full-length alignment including subtype B reference sequences, lab strains, and the predicted intact proviral sequence from PID 1001 were aligned with MAFFT v7.450 (FFT-NS-1 200PAM/k=2 algorithm (88)) in Geneious Prime® 2020.2.4. Minor manual adjustments were performed as needed. Two regions were phylogenetically analyzed: U3 and *pro-pol* for their known strong phylogenetic signal. The U3 region was from the near full-length alignment following the polypurine tract through the CATATA box (HXB2 coordinates: 9,086-9,517bp). The *pro-pol* region was from HXB2 coordinates 2253-5096 bp with part of the highly variable p6 *gag* region excluded. In both analyses, a p-distance neighboring tree with complete deletions (where sites containing alignment gaps are removed prior to analysis) were reconstructed using MEGA X with bootstrapping (n=10,000 pseudoreplicates (89)). The trees were visualized using FigTree software and were rooted on NL4.3 lab strain.

Determining genotypic co-receptor usage. To determine the co-receptor fusion specificity of proviruses found in naïve CD4+ T cells, the V3 loop of HIV-1 *env* was amplified by PCR from a subset of MDA-amplified proviruses in naïve T cells using the primer pairs described in Table S3C. PCR reactions contained the following components: 1.25X Platinum II PCR Buffer, 0.25mM dNTPs, 0.25μM primer F, 0.25μM primer R, Platinum II Taq Enzyme (Invitrogen), and molecular grade water, and the resulting amplicons were subjected to Sanger sequencing from the same primers used for PCR. Two separate coreceptor usage prediction algorithms were used for the codon-aligned V3 sequences: PSSM (SINSI and X4R5 matrices) (46) and Geno2Pheno (47). For the Geno2Pheno algorithm, the false positivity rate (FPR) was set to a value of 10% standard cutoff, meaning there is a 10% probability of classifying an R5 virus incorrectly as X4 virus. As standards when running PSSM (SINSI and X4R5 matrices) and

Geno2Pheno, reference HIV-1 isolates with the V3 sequence were used: IIIB/LAV (X4, K03455.1), 89.6 (X4, U39362.2), JR-CSF (R5, M38429.1), and YU-2 (R5, M93258.1).

Integration site analysis (ISA). ISA was performed on individual or pooled MDA reaction products from the participant designated as PID 1001 that screened positive for the HIV LTR as previously described (40, 51, 90). Briefly, bead-purified MDA products were subjected to random DNA shearing, end repair, A-tailing, and adapter ligation, which adds known sequence to fragment termini and thus enables PCR amplification of unknown host-virus junctions. Adapters also contain functional elements required for Illumina sequencing and indexes required for multiplexing, and eventual demultiplexing, of sequencing libraries. Provirus-adjacent host-virus junction sequences were selectively amplified from ligated MDA product fragments by duplex PCR using donor-specific primers complementary to the sense U3 and antisense U5 components of the HIV LTR together with an adapter specific primer. Amplicon libraries were purified, quantified, and then sequenced on an Illumina MiSeq. Donor-specific ISA primer sequences are provided in Tables S3D, and comprehensive library preparation protocols and analytical workflows for determining and characterizing sites of HIV integration are provided in previously published works (40, 90). All integration sites identified and characterized in this study have been submitted to the Retroviral Integration Sites Database (52).

Datasets used to determine subtype-specific expression of naïve and memory T-cell integration site genes. Expression levels of genes found to harbor sites of HIV-1 integration in naïve and memory CD4+ T cells in samples from participant 1001 were determined using standard RNAseq reference data from sorted cells from 8 adults with HIV on ART that has been published previously (54) and which was graciously provided by Eli Boritz (NIH/NIAID/VRC). Gating strategies and data processing methodologies are provided again here. Stained samples were sorted into CD4+ T-cell subsets using the FACSAria (BD) system by first gating for single cells that were CD3+, Aqualow and negative for CD11c, CD14, CD20, CD56 and TCR-γδ. The remaining events that were CD4+ and CD8− were then collected as

naïve (CD27+CD45RO-), central memory (CD27+CCR7+CD45RO+ or CD27+CCR7-CD45RO+), or effector memory (CD27-) CD4+ T-cell subsets. Sorted cell subsets were processed for total RNA extraction and whole transcriptome sequencing as previously described (91). TPM values calculated for naïve, central memory, and effector memory CD4+ T-cell subsets were averaged across donors and subsets, and weighted averages were calculated for memory T cells from central and effector memory averages in accordance with PBMC CD4+ T-cell subtype distribution estimates recently reported for people with HIV on ART (74).

Statistics. The Wilcoxon matched-pairs signed rank test was used to determine P-values in Figures 1, 2A, and S3. Spearman correlations (two-tailed) were calculated for values reported in Figure 3. More specialized methods for calculating the probability of contaminating memory CD4+ T-cells in the naïve CD4+ T-cell subset and determining genotypic co-receptor usage are described in dedicated sections above and Schema S1. For all pertinent statistical analyses, a P value less than 0.05 was considered significant.

Study Approval. All samples analysed in this study were obtained with the written consent of the guardians of the participants using a protocol approved by Emory University Institutional Review Board (IRB00009146) as part of the Center for AIDS Research (CFAR) specimen repository study.

Data Availability. Raw data used to produce the graphs in Figures 1-5 are provided in a Supporting Data Values file. All HIV sequences were submitted to Genbank and assigned accession numbers PV031991-PV031999. Integration sites are available in the Retrovirus Integration Database (RID) (52).

Author contributions

MGK – performed experiments and wrote paper
VN - performed experiments and wrote paper
SG – performed experiments
SP - performed experiments and bioinformatic analyses
WW - performed experiments
BL – performed statistical analyses
AC – performed bioinformatic analyses
XW – supervised experiments and bioinformatic analyses
AH - – performed experiments
JWR – performed experiments, supervised experiments, performed bioinformatic analyses, acquired funding, wrote the paper
AC – conceived of the idea, acquired samples and funding, wrote the paper
MM – conceived of the idea, acquired funding, performed experiments and analyses, wrote the paper
MFK – conceived of the idea, acquired funding, supervised experiments, wrote the paper

All authors critically reviewed and approved the submitted version of the manuscript.

Acknowledgements. This project has been funded in part with Federal funds from the Office of AIDS Research and the Center for Cancer Research, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, and under Contract No. 75N91019D00024. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. We are grateful for the support of the Emory Center for AIDS Research Clinical Core (P01 AI P30AI050409) and would like to acknowledge our collaborative interactions with the Behavior of HIV in Viral Environments Center (U54AI170855). We thank our funding sources (see funding section) and study participants.

References:

1. UNAIDS. 2022.
2. Yitayew YA, Bekele DM, Demissie BW, and Menji ZA. Mother to Child Transmission of HIV and Associated Factors Among HIV Exposed Infants at Public Health Facilities, Dessie Town, Ethiopia. *HIV AIDS (Auckl).* 2019;11:343-50.
3. Centers for Disease C, and Prevention. Achievements in public health. Reduction in perinatal transmission of HIV infection--United States, 1985-2005. *MMWR Morb Mortal Wkly Rep.* 2006;55(21):592-7.
4. Scott GB, Brogly SB, Muenz D, Stek AM, Read JS, and International Maternal Pediatric Adolescent ACTGPST. Missed Opportunities for Prevention of Mother-to-Child Transmission of Human Immunodeficiency Virus. *Obstet Gynecol.* 2017;129(4):621-8.
5. Newell ML, Coovadia H, Cortina-Borja M, Rollins N, Gaillard P, Dabis F, et al. Mortality of infected and uninfected infants born to HIV-infected mothers in Africa: a pooled analysis. *Lancet.* 2004;364(9441):1236-43.
6. Time from HIV-1 seroconversion to AIDS and death before widespread use of highly-active antiretroviral therapy: a collaborative re-analysis. Collaborative Group on AIDS Incubation and HIV Survival including the CASCADE EU Concerted Action. Concerted Action on SeroConversion to AIDS and Death in Europe. *Lancet.* 2000;355(9210):1131-7.
7. Goulder PJ, Lewin SR, and Leitman EM. Paediatric HIV infection: the potential for cure. *Nat Rev Immunol.* 2016;16(4):259-71.
8. Tobin NH, and Aldrovandi GM. Immunology of pediatric HIV infection. *Immunol Rev.* 2013;254(1):143-69.
9. Ananworanich J, Puthanakit T, Suntarattiwong P, Chokephaibulkit K, Kerr SJ, Fromentin R, et al. Reduced markers of HIV persistence and restricted HIV-specific immune responses after early antiretroviral therapy in children. *AIDS.* 2014;28(7):1015-20.
10. Bitnun A, Samson L, Chun TW, Kakkar F, Brophy J, Murray D, et al. Early initiation of combination antiretroviral therapy in HIV-1-infected newborns can achieve sustained virologic suppression with low frequency of CD4+ T cells carrying HIV in peripheral blood. *Clin Infect Dis.* 2014;59(7):1012-9.
11. Chun TW, Justement JS, Murray D, Hallahan CW, Maenza J, Collier AC, et al. Rebound of plasma viremia following cessation of antiretroviral therapy despite profoundly low levels of HIV reservoir: implications for eradication. *AIDS.* 2010;24(18):2803-8.
12. Hocqueloux L, Avettand-Fenoel V, Jacquot S, Prazuck T, Legac E, Melard 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. *J Antimicrob Chemother.* 2013;68(5):1169-78.
13. Luzuriaga K, Tabak B, Garber M, Chen YH, Ziemniak C, McManus MM, et al. HIV type 1 (HIV-1) proviral reservoirs decay continuously under sustained virologic control in HIV-1-infected children who received early treatment. *J Infect Dis.* 2014;210(10):1529-38.
14. Martinez-Bonet M, Puertas MC, Fortuny C, Ouchi D, Mellado MJ, Rojo P, et al. Establishment and Replenishment of the Viral Reservoir in Perinatally HIV-1-infected Children Initiating Very Early Antiretroviral Therapy. *Clin Infect Dis.* 2015;61(7):1169-78.
15. van Zyl GU, Bedison MA, van Rensburg AJ, Laughton B, Cotton MF, and Mellors JW. Early Antiretroviral Therapy in South African Children Reduces HIV-1-Infected Cells and Cell-Associated HIV-1 RNA in Blood Mononuclear Cells. *J Infect Dis.* 2015;212(1):39-43.
16. Zhang L, Ramratnam B, Tenner-Racz K, He Y, Vesanen M, Lewin S, et al. Quantifying residual HIV-1 replication in patients receiving combination antiretroviral therapy. *N Engl J Med.* 1999;340(21):1605-13.
17. Frange P, Faye A, Avettand-Fenoel V, Bellaton E, Descamps D, Angin M, et al. HIV-1 virological remission lasting more than 12 years after interruption of early antiretroviral

therapy in a perinatally infected teenager enrolled in the French ANRS EPF-CO10 paediatric cohort: a case report. *Lancet HIV.* 2016;3(1):e49-54.
18. Luzuriaga K, Gay H, Ziemniak C, Sanborn KB, Somasundaran M, Rainwater-Lovett K, et al. Viremic relapse after HIV-1 remission in a perinatally infected child. *N Engl J Med.* 2015;372(8):786-8.
19. Persaud D, Gay H, Ziemniak C, Chen YH, Piatak M, Jr., Chun TW, et al. Absence of detectable HIV-1 viremia after treatment cessation in an infant. *N Engl J Med.* 2013;369(19):1828-35.
20. Douek DC, McFarland RD, Keiser PH, Gage EA, Massey JM, Haynes BF, et al. Changes in thymic function with age and during the treatment of HIV infection. *Nature.* 1998;396(6712):690-5.
21. Adkins B. Peripheral CD4+ lymphocytes derived from fetal versus adult thymic precursors differ phenotypically and functionally. *J Immunol.* 2003;171(10):5157-64.
22. Mold JE, Venkatasubrahmanyam S, Burt TD, Michaelsson J, Rivera JM, Galkina SA, et al. Fetal and adult hematopoietic stem cells give rise to distinct T cell lineages in humans. *Science.* 2010;330(6011):1695-9.
23. Rudd BD. Neonatal T Cells: A Reinterpretation. *Annu Rev Immunol.* 2020;38:229-47.
24. Smith NL, Patel RK, Reynaldi A, Grenier JK, Wang J, Watson NB, et al. Developmental Origin Governs CD8(+) T Cell Fate Decisions during Infection. *Cell.* 2018;174(1):117-30 e14.
25. Chomont N, DaFonseca S, Vandergeeten C, Ancuta P, and Sekaly RP. Maintenance of CD4+ T-cell memory and HIV persistence: keeping memory, keeping HIV. *Curr Opin HIV AIDS.* 2011;6(1):30-6.
26. 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. *Nat Med.* 2009;15(8):893-900.
27. Baldanti F, Paolucci S, Gulminetti R, Maserati R, Migliorino G, Pan A, et al. Higher levels of HIV DNA in memory and naive CD4(+) T cell subsets of viremic compared to non-viremic patients after 18 and 24 months of HAART. *Antiviral Res.* 2001;50(3):197-206.
28. Brenchley JM, Hill BJ, Ambrozak DR, Price DA, Guenaga FJ, Casazza JP, et al. T-cell subsets that harbor human immunodeficiency virus (HIV) in vivo: implications for HIV pathogenesis. *J Virol.* 2004;78(3):1160-8.
29. Delobel P, Sandres-Saune K, Cazabat M, L'Faqihi FE, Aquilina C, Obadia M, et al. Persistence of distinct HIV-1 populations in blood monocytes and naive and memory CD4 T cells during prolonged suppressive HAART. *AIDS.* 2005;19(16):1739-50.
30. Fabre-Mersseman V, Dutrieux J, Louise A, Rozlan S, Lamine A, Parker R, et al. CD4(+) recent thymic emigrants are infected by HIV in vivo, implication for pathogenesis. *AIDS.* 2011;25(9):1153-62.
31. Heeregrave EJ, Geels MJ, Brenchley JM, Baan E, Ambrozak DR, van der Sluis RM, et al. Lack of in vivo compartmentalization among HIV-1 infected naive and memory CD4+ T cell subsets. *Virology.* 2009;393(1):24-32.
32. Lambotte O, Demoustier A, de Goer MG, Wallon C, Gasnault J, Goujard C, et al. Persistence of replication-competent HIV in both memory and naive CD4 T cell subsets in patients on prolonged and effective HAART. *AIDS.* 2002;16(16):2151-7.
33. Ostrowski MA, Chun TW, Justement SJ, Motola I, Spinelli MA, Adelsberger J, et al. Both memory and CD45RA+/CD62L+ naive CD4(+) T cells are infected in human immunodeficiency virus type 1-infected individuals. *J Virol.* 1999;73(8):6430-5.
34. Wightman F, Solomon A, Khoury G, Green JA, Gray L, Gorry PR, et al. Both CD31(+) and CD31(-) naive CD4(+) T cells are persistent HIV type 1-infected reservoirs in individuals receiving antiretroviral therapy. *J Infect Dis.* 2010;202(11):1738-48.
35. Dai J, Agosto LM, Baytop C, Yu JJ, Pace MJ, Liszewski MK, et al. Human immunodeficiency virus integrates directly into naive resting CD4+ T cells but enters naive cells less efficiently than memory cells. *J Virol.* 2009;83(9):4528-37.

36. Zack JA, Arrigo SJ, Weitsman SR, Go AS, Haislip A, and Chen IS. HIV-1 entry into quiescent primary lymphocytes: molecular analysis reveals a labile, latent viral structure. *Cell.* 1990;61(2):213-22.
37. Katusiime MG, Halvas EK, Wright I, Joseph K, Bale MJ, Kirby-McCullough B, et al. Intact HIV Proviruses Persist in Children Seven to Nine Years after Initiation of Antiretroviral Therapy in the First Year of Life. *J Virol.* 2020;94(4).
38. Bale MJ, Katusiime MG, Wells D, Wu X, Spindler J, Halvas EK, et al. Early Emergence and Long-Term Persistence of HIV-Infected T-Cell Clones in Children. *mBio.* 2021;12(2).
39. Garcia-Broncano P, Maddali S, Einkauf KB, Jiang C, Gao C, Chevalier J, et al. Early antiretroviral therapy in neonates with HIV-1 infection restricts viral reservoir size and induces a distinct innate immune profile. *Sci Transl Med.* 2019;11(520).
40. Patro SC, Brandt LD, Bale MJ, Halvas EK, Joseph KW, Shao W, et al. Combined HIV-1 sequence and integration site analysis informs viral dynamics and allows reconstruction of replicating viral ancestors. *Proc Natl Acad Sci U S A.* 2019;116(51):25891-9.
41. Venanzi Rullo E, Pinzone MR, Cannon L, Weissman S, Ceccarelli M, Zurakowski R, et al. Persistence of an intact HIV reservoir in phenotypically naive T cells. *JCI Insight.* 2020;5(20).
42. 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.
43. Botha JC, Demirov D, Gordijn C, Katusiime MG, Bale MJ, Wu X, et al. The largest HIV-1-infected T cell clones in children on long-term combination antiretroviral therapy contain solo LTRs. *mBio.* 2023;14(4):e0111623.
44. Bruner KM, Murray AJ, Pollack RA, Soliman MG, Laskey SB, Capoferri AA, et al. Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nat Med.* 2016;22(9):1043-9.
45. Shao W, Shan J, Hu WS, Halvas EK, Mellors JW, Coffin JM, et al. HIV Proviral Sequence Database: A New Public Database for Near Full-Length HIV Proviral Sequences and Their Meta-Analyses. *AIDS Res Hum Retroviruses.* 2020;36(1):1-3.
46. Jensen MA, Li FS, van 't Wout AB, Nickle DC, Shriner D, He HX, et al. Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences. *J Virol.* 2003;77(24):13376-88.
47. Lengauer T, Sander O, Sierra S, Thielen A, and Kaiser R. Bioinformatics prediction of HIV coreceptor usage. *Nat Biotechnol.* 2007;25(12):1407-10.
48. Mitchell RS, Beitzel BF, Schroder AR, Shinn P, Chen H, Berry CC, et al. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol.* 2004;2(8):E234.
49. Schroder AR, Shinn P, Chen H, Berry C, Ecker JR, and Bushman F. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell.* 2002;110(4):521-9.
50. Coffin JM, Bale MJ, Wells D, Guo S, Luke B, Zerbato JM, et al. Integration in oncogenes plays only a minor role in determining the in vivo distribution of HIV integration sites before or during suppressive antiretroviral therapy. *PLoS Pathog.* 2021;17(4):e1009141.
51. Maldarelli F, Wu X, Su L, Simonetti FR, Shao W, Hill S, et al. HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science.* 2014;345(6193):179-83.
52. Shao W, Shan J, Kearney MF, Wu X, Maldarelli F, Mellors JW, et al. Retrovirus Integration Database (RID): a public database for retroviral insertion sites into host genomes. *Retrovirology.* 2016;13(1):47.
53. Youngblood B, Hale JS, and Ahmed R. T-cell memory differentiation: insights from transcriptional signatures and epigenetics. *Immunology.* 2013;139(3):277-84.
54. Clark IC, Mudvari P, Thaploo S, Smith S, Abu-Laban M, Hamouda M, et al. HIV silencing and cell survival signatures in infected T cell reservoirs. *Nature.* 2023;614(7947):318-25.

55. Zerbato JM, Serrao E, Lenzi G, Kim B, Ambrose Z, Watkins SC, et al. Establishment and Reversal of HIV-1 Latency in Naive and Central Memory CD4+ T Cells In Vitro. *J Virol.* 2016;90(18):8059-73.
56. Bacchus C, Cheret A, Avettand-Fenoel V, Nembot G, Melard A, Blanc C, et al. A single HIV-1 cluster and a skewed immune homeostasis drive the early spread of HIV among resting CD4+ cell subsets within one month post-infection. *PLoS One.* 2013;8(5):e64219.
57. Saez-Cirion A, Bacchus C, Hocqueloux L, Avettand-Fenoel V, Girault I, Lecuroux C, et al. Post-treatment HIV-1 controllers with a long-term virological remission after the interruption of early initiated antiretroviral therapy ANRS VISCONTI Study. *PLoS Pathog.* 2013;9(3):e1003211.
58. Mavigner M, Habib J, Deleage C, Rosen E, Mattingly C, Bricker K, et al. Simian Immunodeficiency Virus Persistence in Cellular and Anatomic Reservoirs in Antiretroviral Therapy-Suppressed Infant Rhesus Macaques. *J Virol.* 2018;92(18).
59. Bunders MJ, van der Loos CM, Klarenbeek PL, van Hamme JL, Boer K, Wilde JC, et al. Memory CD4(+)CCR5(+) T cells are abundantly present in the gut of newborn infants to facilitate mother-to-child transmission of HIV-1. *Blood.* 2012;120(22):4383-90.
60. Pinzone MR, Weissman S, Pasternak AO, Zurakowski R, Migueles S, and O'Doherty U. Naive infection predicts reservoir diversity and is a formidable hurdle to HIV eradication. *JCI Insight.* 2021;6(16).
61. Morcilla V, Bacchus-Souffan C, Fisher K, Horsburgh BA, Hiener B, Wang XQ, et al. HIV-1 Genomes Are Enriched in Memory CD4(+) T-Cells with Short Half-Lives. *mBio.* 2021;12(5):e0244721.
62. Mitchell CJ, Getnet D, Kim MS, Manda SS, Kumar P, Huang TC, et al. A multi-omic analysis of human naive CD4+ T cells. *BMC Syst Biol.* 2015;9:75.
63. Pace MJ, Agosto L, and O'Doherty U. R5 HIV env and vesicular stomatitis virus G protein cooperate to mediate fusion to naive CD4+ T Cells. *J Virol.* 2011;85(1):644-8.
64. Wilen CB, Parrish NF, Pfaff JM, Decker JM, Henning EA, Haim H, et al. Phenotypic and immunologic comparison of clade B transmitted/founder and chronic HIV-1 envelope glycoproteins. *J Virol.* 2011;85(17):8514-27.
65. Bleul CC, Wu L, Hoxie JA, Springer TA, and Mackay CR. The HIV coreceptors CXCR4 and CCR5 are differentially expressed and regulated on human T lymphocytes. *Proc Natl Acad Sci U S A.* 1997;94(5):1925-30.
66. Tabler CO, Lucera MB, Haqqani AA, McDonald DJ, Migueles SA, Connors M, et al. CD4+ memory stem cells are infected by HIV-1 in a manner regulated in part by SAMHD1 expression. *J Virol.* 2014;88(9):4976-86.
67. Eckstein DA, Penn ML, Korin YD, Scripture-Adams DD, Zack JA, Kreisberg JF, et al. HIV-1 actively replicates in naive CD4(+) T cells residing within human lymphoid tissues. *Immunity.* 2001;15(4):671-82.
68. Kinter A, Moorthy A, Jackson R, and Fauci AS. Productive HIV infection of resting CD4+ T cells: role of lymphoid tissue microenvironment and effect of immunomodulating agents. *AIDS Res Hum Retroviruses.* 2003;19(10):847-56.
69. Groot F, van Capel TM, Schuitemaker J, Berkhout B, and de Jong EC. Differential susceptibility of naive, central memory and effector memory T cells to dendritic cell-mediated HIV-1 transmission. *Retrovirology.* 2006;3:52.
70. Tan JT, Dudl E, LeRoy E, Murray R, Sprent J, Weinberg KI, et al. IL-7 is critical for homeostatic proliferation and survival of naive T cells. *Proc Natl Acad Sci U S A.* 2001;98(15):8732-7.
71. Dardalhon V, Jaleco S, Kinet S, Herpers B, Steinberg M, Ferrand C, et al. IL-7 differentially regulates cell cycle progression and HIV-1-based vector infection in neonatal and adult CD4+ T cells. *Proc Natl Acad Sci U S A.* 2001;98(16):9277-82.

72. Ducrey-Rundquist O, Guyader M, and Trono D. Modalities of interleukin-7-induced human immunodeficiency virus permissiveness in quiescent T lymphocytes. *J Virol.* 2002;76(18):9103-11.
73. Unutmaz D, KewalRamani VN, Marmon S, and Littman DR. Cytokine signals are sufficient for HIV-1 infection of resting human T lymphocytes. *J Exp Med.* 1999;189(11):1735-46.
74. Reeves DB, Bacchus-Souffan C, Fitch M, Abdel-Mohsen M, Hoh R, Ahn H, et al. Estimating the contribution of CD4 T cell subset proliferation and differentiation to HIV persistence. *Nat Commun.* 2023;14(1):6145.
75. Bacchus-Souffan C, Fitch M, Symons J, Abdel-Mohsen M, Reeves DB, Hoh R, et al. Relationship between CD4 T cell turnover, cellular differentiation and HIV persistence during ART. *PLoS Pathog.* 2021;17(1):e1009214.
76. Cole B, Lambrechts L, Boyer Z, Noppe Y, De Scheerder MA, Eden JS, et al. Extensive characterization of HIV-1 reservoirs reveals links to plasma viremia before and during analytical treatment interruption. *Cell Rep.* 2022;39(4):110739.
77. 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 Pathog.* 2014;10(3):e1004010.
78. Tsukumo S, Unno M, Muto A, Takeuchi A, Kometani K, Kurosaki T, et al. Bach2 maintains T cells in a naive state by suppressing effector memory-related genes. *Proc Natl Acad Sci U S A.* 2013;110(26):10735-40.
79. Carrio R, Bathe OF, and Malek TR. Initial antigen encounter programs CD8+ T cells competent to develop into memory cells that are activated in an antigen-free, IL-7- and IL-15-rich environment. *J Immunol.* 2004;172(12):7315-23.
80. Dooms H, Wolslegel K, Lin P, and Abbas AK. Interleukin-2 enhances CD4+ T cell memory by promoting the generation of IL-7R alpha-expressing cells. *J Exp Med.* 2007;204(3):547-57.
81. Cole B, Lambrechts L, Gantner P, Noppe Y, Bonine N, Witkowski W, et al. In-depth single-cell analysis of translation-competent HIV-1 reservoirs identifies cellular sources of plasma viremia. *Nat Commun.* 2021;12(1):3727.
82. Dragoni F, Kwaa AK, Traut CC, Veenhuis RT, Woldemeskel BA, Camilo-Contreras A, et al. Proviral location affects cognate peptide-induced virus production and immune recognition of HIV-1-infected T cell clones. *J Clin Invest.* 2023;133(21).
83. Einkauf KB, Lee GQ, Gao C, Sharaf R, Sun X, Hua S, et al. Intact HIV-1 proviruses accumulate at distinct chromosomal positions during prolonged antiretroviral therapy. *J Clin Invest.* 2019;129(3):988-98.
84. Einkauf KB, Osborn MR, Gao C, Sun W, Sun X, Lian X, et al. Parallel analysis of transcription, integration, and sequence of single HIV-1 proviruses. *Cell.* 2022;185(2):266-82 e15.
85. Jiang C, Lian X, Gao C, Sun X, Einkauf KB, Chevalier JM, et al. Distinct viral reservoirs in individuals with spontaneous control of HIV-1. *Nature.* 2020;585(7824):261-7.
86. Lian X, Seiger KW, Parsons EM, Gao C, Sun W, Gladkov GT, et al. Progressive transformation of the HIV-1 reservoir cell profile over two decades of antiviral therapy. *Cell Host Microbe.* 2023;31(1):83-96 e5.
87. Halvas EK, Joseph KW, Brandt LD, Guo S, Sobolewski MD, Jacobs JL, et al. HIV-1 viremia not suppressible by antiretroviral therapy can originate from large T cell clones producing infectious virus. *J Clin Invest.* 2020;130(11):5847-57.
88. Katoh K, and Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30(4):772-80.
89. Kumar S, Stecher G, Li M, Knyaz C, and Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol.* 2018;35(6):1547-9.
90. Wells DW, Guo S, Shao W, Bale MJ, Coffin JM, Hughes SH, et al. An analytical pipeline for identifying and mapping the integration sites of HIV and other retroviruses. *BMC Genomics.* 2020;21(1):216.

91. Boritz EA, Darko S, Swaszek L, Wolf G, Wells D, Wu X, et al. Multiple Origins of Virus Persistence during Natural Control of HIV Infection. *Cell.* 2016;166(4):1004-15.

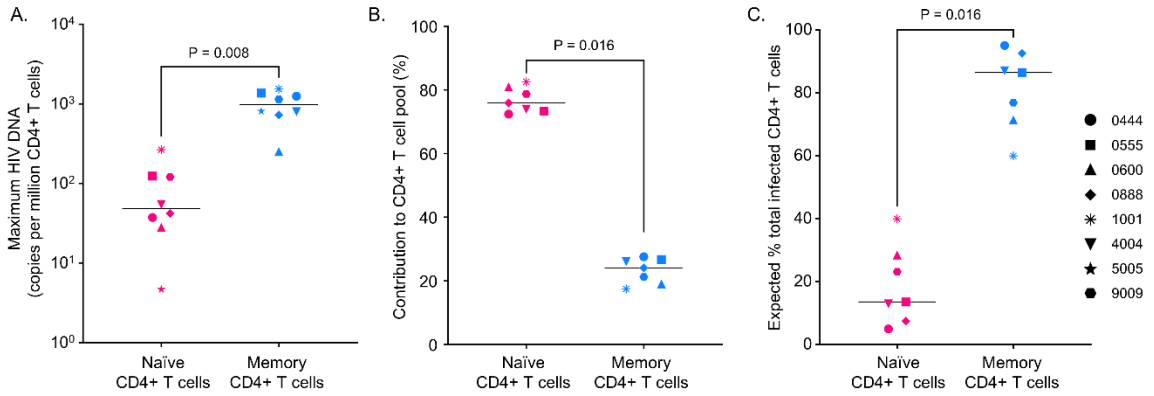

Figure 1. HIV persistence in peripheral naïve and memory CD4+ T cells from CLWH. (A) Frequency of HIV DNA in sorted populations of naïve and memory CD4+ T cells (B) Relative contribution of naïve and memory CD4+ T cells to the CD4+ T cell pool assessed by flow cytometry. (C) Relative contribution of naïve and memory CD4+ T cells to the total HIV reservoir in CD4+ T cells calculated after correction for potential contamination of the sorted cells. Horizontal bars represent the median. Wilcoxon matched-pairs signed rank test were used for statistics.

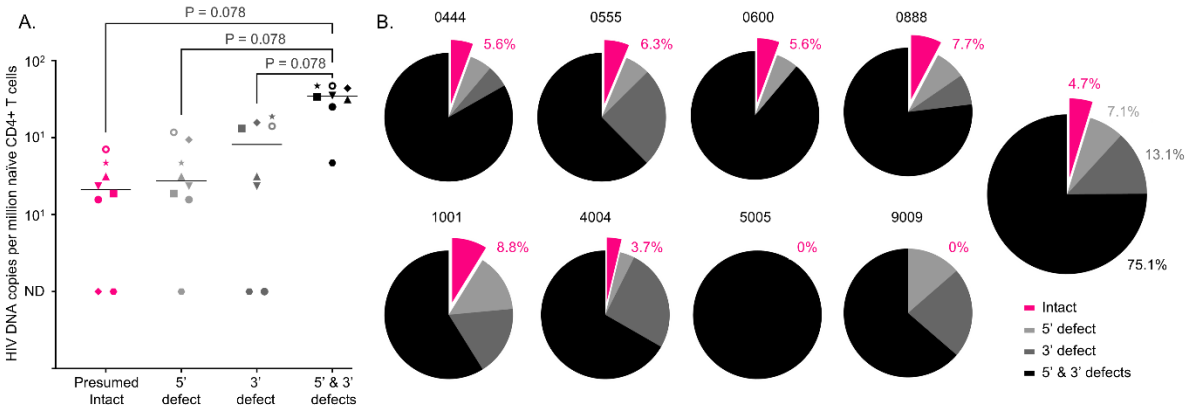

Figure 2. Persistent potentially intact and defective proviruses in naive CD4+ T cells from CLWH. (A) Levels and (B) relative frequencies of potentially intact, 5'-defective, 3'-defective and 5'&3'-defective HIV proviruses estimated by a modified Intact Proviral DNA Assay (IPDA). Symbol key provided with Figure 1. Open shapes indicate undetectable values with the limit of detection (LOD) set based on cell input. Horizontal bars represent the median. Wilcoxon matched-pairs signed rank test were used for statistics.

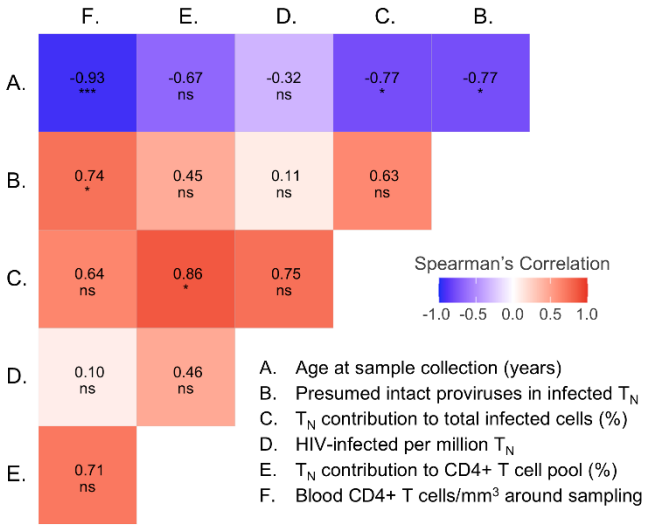

Figure 3. Correlations among immunological, clinical and virological parameters. Spearman correlations among CD4+ T cell counts around the times of sampling, several aspects of infected and uninfected naïve CD4+ T cells (T_N), and approximate participant age at time of sampling. Statistical significance: p >= 0.05 (ns); p <0.05 (*); p < 0.01 (**); p < 0.001(***).

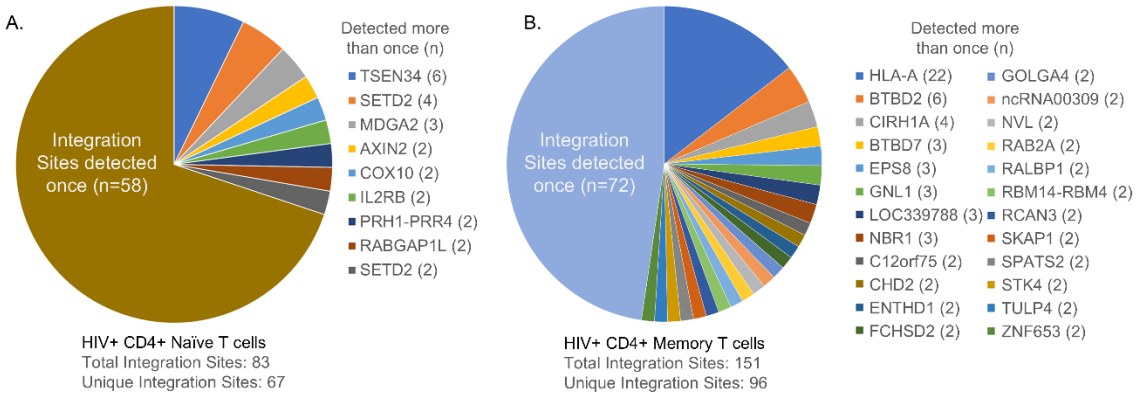

Figure 4. HIV integration site distribution in CD4+ T cells from participant 1001. Integration sites identified in (A) naïve and (B) memory CD4+ T cells. Total integration sites, numbers of unique integration sites, numbers of sites detected once, and specific genes harboring integration sites detected more than once (with number of times detected) in each sorted collection are indicated. Wedge sizes are proportional to the numbers of integration sites detected once or numbers of times specific clonal populations are detected in verified expanded clones, respectively.

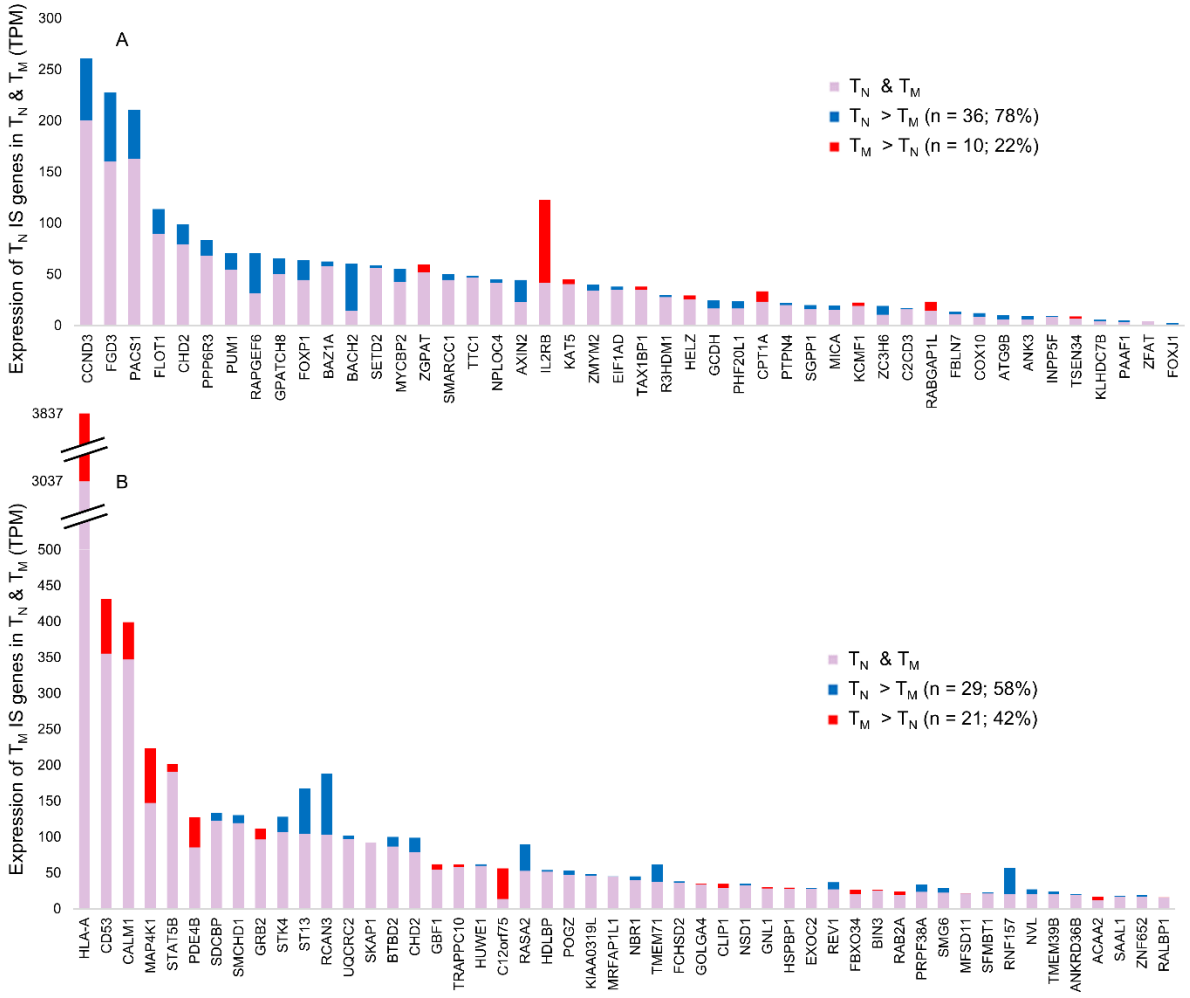

Figure 5. Genic integration sites versus expression levels in CD4+ naïve (T_N) and memory (T_M) T cells. Integration site (IS) genes identified from (A) naïve and (B) memory CD4+ T cell sorted collections are listed in descending order of expression level in the subset in which they were identified (transcripts per million, TPM, according to RNAseq data). Genes expressed to higher levels in T_N are marked by purple bars with blue caps, the heights of which indicate expression levels in T_M and T_N, respectively. Genes expressed to higher levels in T_M are marked by purple bars with red caps, the heights of which indicate expression levels in T_N and T_M, respectively.

Table 1. Participant Characteristics

PID[A]	Sex[B]	HIV subtype	Age at ART initiation (months)	ART regimen[C]	CD4 cell count (cells/mm^3)	Viral Load (copies/mL)	Age at sampling (years)
0444	M	B	13[D]	**ABC/DTG/3TC** ABC/3TC, RAL ABC/3TC, DRV/r ABC/3TC, LPV/r ABC/3TC/AZT, LPV/r ABC/3TC/AZT	917[E]	<20	10
0555	M	B	9	**BIC/FTC/TAF** ABC/3TC, LPV/r	761	<20	10
0600	F	B	1.5	**ABC/3TC, RAL** ABC/3TC, RAL, LPV/r ABC/3TC/AZT, LPV/r 3TC/AZT AZT, NVP (ppx)	1555	<20	5
0888	M	C	17	**TDF/FTC/ETR/c** ABC, RAL, AZT ABC, RAL, AZT, LPV/r 3TC/AZT, LPV/r	819[E]	<20	11
1001	F	B	8.5	**ABC, RAL, AZT** ABC, RAL, LPV/r ABC, AZT ABC/3TC, EFV 3TC/AZT, LPV/r	1506	<20	6
4004	M	C	2	**3TC/AZT, MVC, LPV/r** 3TC/TDF, AZT, LPV/r 3TC/AZT, RAL, LPV/r 3TC/AZT/NVP, LPV/r	1177	<20	7
5005	F	B	2	**FTC/TAF, DTG** ABC/3TC, RAL ABC/3TC, DRV/r FTC/TDF, ATV/r ABC/EFV/3TC ABC/3TC, LPV/r 3TC/AZT, LPV/r	1196[E]	<20	8
9009	F	ND	12[D]	**FTC/TAF, DTG** TDF, 3TC/DTG TDF, 3TC/RAL TDF, 3TC/RAL, ATV/r ABC/3TC, RAL 3TC/AZT, RAL 3TC/AZT, NFV	1018	<20	9
Median			**8.75**		**1098**	**<20**	**8.5**

[A]Participant ID. [B]Sex assigned at birth: M (male), F (female). [C]ART regimen: ABC (abacavir), ATV (atazanavir), AZT (zidovudine), BIC (bictegravir), c (cobicistat), DRV (darunavir), DTG (dolutegravir), EFV (efavirenz), ETR (etravirine), FTC (emtricitabine), LPV (lopinavir), MVC (maraviroc), NFV (nelfinavir), NVP (nevirapine), RAL (raltegravir), r (ritonavir), TAF (tenofovir alafenamide), TDF (tenofovir disoproxil fumarate), 3TC (lamivudine); ppx (prophylaxis). Current regiments are given in bold font and thereafter listed in reverse chronological order. [D]Approximate. [E]CD4 cell count was not performed at donation. Values obtained shortly prior to or after donation are reported.

Table 2. Frequencies of HIV-infected naïve CD4+ T cells.

Participant ID	Naïve sort purity (%)	HIV-infected cells/10^6 cells in naïve sort[A]	Max # HIV-infected memory cells/10^6 cells in naïve sort[B]	Expected # HIV-infected memory cells/10^6 cells in naïve sort[C]	Minimum # of HIV-infected naïve cells/10^6 cells in naïve sort[D]	Expected # of HIV-infected naïve cells/10^6 cells in naïve sort[E]
0444	99.0	38	21	13	17	25
0555	96.7	124	59	46	66	78
0600	98.3	28	9	5	19	24
0888	97.0	42	34	24	8	19
1001	96.9	266	66	49	201	217
4004	98.5	54	21	13	34	41
5005	96.7	5	42	27	<1	<1
9009	97.6[F]	121	37	28	84	93
Median	***97.3***	***48***	***36***	***26***	***27***	***33***

[A]Frequency of HIV-infected cells in the naïve sorted collection measured as LTR-positive MDA-product frequencies, Poisson-corrected. [B]Maximum (95% confidence) and [C]expected numbers of contaminating infected cells in sorted cell collections calculated as described in Materials and Methods and Supplementary Schema S1. [D]Minimum and [E]expected numbers of infected naïve cells in the sorted naïve cell collections. [F]Measured PID 9009 sort purities were unavailable, so the average of other PID sample sort purities was substituted. Naïve: naïve CD4+ T cells; memory: memory CD4+ T cells.