

# Persistence of an intact HIV reservoir in phenotypically naive T cells

Emmanuele Venanzi Rullo,<sup>1,2</sup> Marilia Rita Pinzone,<sup>1</sup> LaMont Cannon,<sup>1,3</sup> Sam Weissman,<sup>1</sup> Manuela Ceccarelli,<sup>1,2</sup> Ryan Zurakowski,<sup>4</sup> Giuseppe Nunnari,<sup>2</sup> and Una O'Doherty<sup>1</sup>

<sup>1</sup>Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA.

<sup>2</sup>Department of Clinical and Experimental Medicine, Unit of Infectious Diseases, University of Messina, Messina, Italy.

<sup>3</sup>Center for the Study of Biological Complexity, Virginia Commonwealth University, Virginia, USA. <sup>4</sup>Department of Biomedical Engineering, University of Delaware, Newark, Delaware, USA.

Despite the efficacy of antiretroviral therapy (ART), HIV persists in a latent form and remains a hurdle to eradication. CD4<sup>+</sup> T lymphocytes harbor the majority of the HIV reservoir, but the role of individual subsets remains unclear. CD4<sup>+</sup> T cells were sorted into central, transitional, effector memory, and naive T cells. We measured HIV DNA and performed proviral sequencing of more than 1900 proviruses in 2 subjects at 2 and 9 years after ART initiation to estimate the contribution of each subset to the reservoir. Although our study was limited to 2 subjects, we obtained comparable findings with publicly available sequences. While the HIV integration levels were lower in naive compared with memory T cells, naive cells were a major contributor to the intact proviral reservoir. Notably, proviral sequences isolated from naive cells appeared to be unique, while those retrieved from effector memory cells were mainly clonal. The number of clones increased as cells differentiated from a naive to an effector memory phenotype, suggesting naive cells repopulate the effector memory reservoir as previously shown for central memory cells. Naive T cells contribute substantially to the intact HIV reservoir and represent a significant hurdle for HIV eradication.

## Introduction

With the advent of antiretroviral therapy (ART), the natural history of HIV infection has dramatically changed. Viral loads typically decline to undetectable levels. This is followed by immune recovery, which leads to an improvement in quality of life and life span (1). Nevertheless, the virus persists as a latent reservoir, which is capable of viral rebound upon ART interruption (2).

In recent years, next-generation sequencing has become a powerful tool to study the proviral landscape over time (3–8). We reasoned that studying the proviral character deeply in different cellular subsets might provide insights into their contribution to HIV persistence. Our reasoning was largely based on a prior study from our group demonstrating differential selection against proviral sequences depending on their genetic makeup (6). In that study, we observed relative enrichment of proviruses with a 5' deletion over time, consistent with refs. 9 and 10. These deletions resulted in uniform loss of the strong donor splice site D1 and preservation of the second strong donor splice site D4. We hypothesized that this enrichment may occur because 5'-deleted proviruses might be less efficient at HIV protein expression while retaining the ability to splice aberrantly to downstream genes, including genes involved in cell division/survival. Simultaneously, proviruses with a preserved D1 and an ORF were depleted over time, perhaps because encoded HIV proteins were expressed more efficiently (6).

Most efforts to study the HIV reservoir have focused on memory CD4<sup>+</sup> T cells because this heterogeneous cellular subset carries the highest levels of HIV DNA in vivo (11–19). Moreover, C-C chemokine receptor type 5-tropic (CCR5-tropic) HIV virions are responsible for transmission (20), reinforcing the central role of memory T cells, since these cells express higher levels of CCR5 compared with naive CD4<sup>+</sup> T cells (T<sub>N</sub> cells) (21). Studies of naive infection have been limited, due to the perceived dominant role of memory T cells in HIV persistence. While several groups detected HIV DNA in T<sub>N</sub> cells both in vitro and in vivo (12–19, 22–28), the levels of HIV DNA within T<sub>N</sub> cells were generally 10-fold lower than in memory cells (11–19), suggesting that the HIV DNA found in T<sub>N</sub> cells might be due to residual contamination of memory lymphocytes. Moreover, since T<sub>N</sub> cells are more resistant to direct HIV infection by CCR5-tropic

**Authorship note:** EVR, MRP, and LC contributed equally to this work.

**Conflict of interest:** The authors have declared that no conflict of interest exists.

**Copyright:** © 2020, Rullo et al. This is an open access article published under the terms of the Creative Commons Attribution 4.0 International License.

**Submitted:** September 3, 2019

**Accepted:** September 10, 2020

**Published:** October 15, 2020

**Reference information:** *JCI Insight*. 2020;5(20):e133157.

<https://doi.org/10.1172/jci.insight.133157>

<https://doi.org/10.1172/jci.insight.133157>

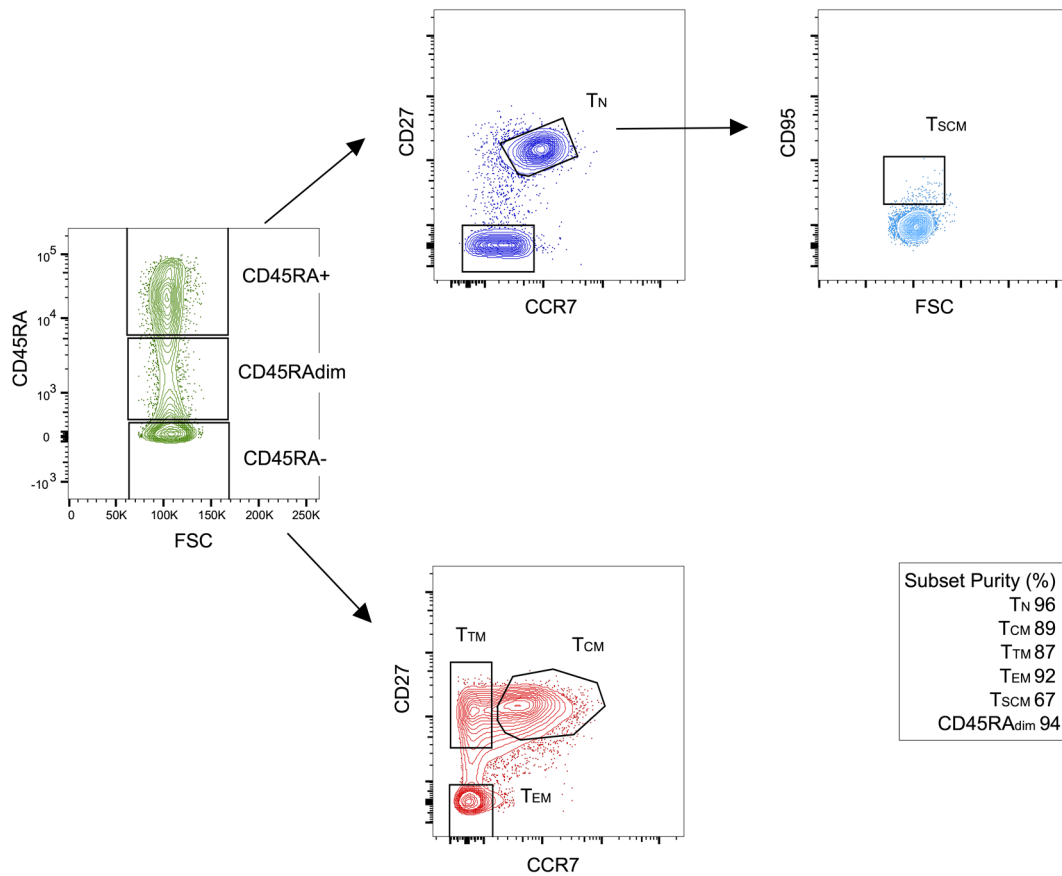
viruses in vitro (24), these cells were considered unlikely to play a central role in HIV pathogenesis. However, naive cells can be infected with C-X-C motif chemokine receptor 4-tropic (CXCR4-tropic) virus in vitro (24), as they express high levels of CXCR4. Given that HIV often evolves to express CXCR4-tropic envelope (29), there is potential for  $T_N$  cells to contribute to the reservoir. Moreover,  $T_N$  cells have unique characteristics that could lead to distinct hurdles to HIV eradication. These include longer intermitotic half-life than memory T cells (30), resistance to HIV expression after integration (24, 31–33), and the ability to give rise to all memory T cell subsets.

In the current study, we sought to study the selection pressures within and between cellular subsets using a longitudinal sequencing analysis approach. To address these questions, we sorted CD4 cellular subsets from 2 chronically infected individuals at 2 and 9 years after ART by adopting previously described sorting approaches (13, 16–18, 31, 33, 34) and obtained comparable findings in 3 other individuals through publicly available sequences (14). We performed near full-length (NFL) proviral sequencing of sorted cellular subsets and monitored dynamics of intact and defective proviruses. By tracking proviral composition over time, we uncovered evidence that infected  $T_N$  cells are a significant contributor to the reservoir with the potential to repopulate the memory reservoir.

## Results

*Sorting of CD4<sup>+</sup> T cell subsets after 2 and 9 years of ART.* To study reservoir dynamics in different cellular subsets, we chose to study 2 individuals in whom we had previously measured reservoir decay (6) (Figure 1 and Supplemental Tables 1 and 2; supplemental material available online with this article; <https://doi.org/10.1172/jci.insight.133157DS1>). Both subjects had undetectable plasma viral load at the time of the apheresis. We sorted CD3<sup>+</sup>CD8<sup>-</sup> T cells by flow cytometry into  $T_N$  cells, central memory ( $T_{CM}$ ), transitional memory ( $T_{TM}$ ), effector memory ( $T_{EM}$ ) and CD45<sub>dim</sub> cells as described in Figure 1, Table 1, Table 2, and Methods. We measured the amount of HIV DNA per million cells by quantitative PCR (qPCR). The percentage of T cells for each subset and the estimate of HIV DNA per million cells in each subset are reported in Table 1 (6). We also chose to sort CD45<sub>dim</sub> cells, as this population made up a sizeable portion of T cells but was difficult to categorize as either naive or memory. We found CD45<sub>dim</sub> cells also contained substantial levels of HIV DNA. The HIV DNA levels measured in each subset likely reflect the fraction of HIV that is integrated, since the enrolled individuals were on suppressive ART for more than 2 years at the time of apheresis (35, 36). Consistent with the literature (12–19), HIV DNA levels were lower in  $T_N$  cells compared with the combined memory population (Table 1). Given that memory T cells carry higher levels of HIV DNA than  $T_N$  cells, it was important to rule out that the few DNA sequences we identified in  $T_N$  cells might be due to contaminating memory cells. Based on postsorting purity, we estimated the fraction of proviral sequences that could be attributed to contaminating memory cells. Given that, on average, 4% of the sorted  $T_N$  cells were memory T cells, these contaminants could only account for a small fraction (less than 22%) of the HIV DNA sequences. Thus, the majority of the HIV DNA retrieved from  $T_N$  cells could not be attributed to contaminating cells with a memory phenotype. Of greater significance, sequence analysis revealed that the character and composition of proviruses in the naive subset was substantially different from those in the memory subsets.

*$T_N$  cells are a significant contributor to the intact reservoir, despite lower levels of HIV DNA.* To genetically characterize the HIV reservoir in sorted cellular subsets, we amplified and sequenced 942 full-length proviruses from Subject 1 and 975 proviruses from Subject 2 after sorting (Figure 1 and Table 2). We aligned the de novo assembled sequences to HXB2 and identified intact proviral sequences, as described in Methods and in ref. 6. To determine the number of intact proviruses contributed by each major T cell subset, we multiplied the total HIV DNA per subset (Table 1) by the fraction of intact proviruses within that subset (Table 1 and Figure 2). This provided an estimate of intact proviruses from each subset (Methods and Table 1). As an example, for Subject 1 at the first time point, we measured 655 HIV copies/million cells in  $T_N$  cells (Table 1, column A). We estimated the contribution of  $T_N$  cells to the total pool of CD4<sup>+</sup> T cells by flow cytometry (21%, column C) and the fraction of proviruses that was intact by NFL proviral sequencing (32%, column B). We used these estimates to calculate the number of intact HIV per million naive CD4 cells (column D) as  $A \times B \times C$  ( $655 \times 0.21 \times 0.32 = 43$  intact HIV/million naive CD4<sup>+</sup> T cells). In both individuals,  $T_N$  cells were a major contributor (Figure 2, A and B).  $T_N$  cell contribution to the intact reservoir was 39% and 34% in Subject 1 and 59% and 31% in Subject 2 at the first and second time points, respectively. The intact reservoir decay in PBMCs (as estimated in ref. 6) and in  $T_N$  cells was similar, with a half-life between 2 and 3 years (Figure 2 and Supplemental Table 3). Notably, when we minimized the effects of large clones by removing repeated



**Figure 1. Sorting strategy to obtain cellular subsets.** CD4<sup>+</sup> T cells were negatively selected by immunomagnetic beads from peripheral blood mononuclear cells from 2 HIV-infected donors on ART at 2 time points corresponding to 2 and 9 years after ART initiation. T<sub>N</sub> cells were enriched by sorting for CD45RA<sup>+</sup>CCR7<sup>+</sup>CD27<sup>+</sup> cells. T<sub>CM</sub> cells were CD45RA<sup>-</sup>CCR7<sup>+</sup>CD27<sup>+</sup>, T<sub>TM</sub> cells were CD45RA<sup>-</sup>CCR7<sup>-</sup>CD27<sup>+</sup>, while T<sub>EM</sub> cells were CD45RA<sup>-</sup>CCR7<sup>-</sup>CD27<sup>-</sup>. Finally, CD45RA<sup>dim</sup> cells were also collected in an effort to sample almost all CD4<sup>+</sup> T cell subsets. For selected experiments, we also sorted CD95<sup>-</sup> T<sub>N</sub> cells (CD45RA<sup>-</sup>CCR7<sup>+</sup>CD27<sup>+</sup>CD95<sup>-</sup>) and T<sub>SCM</sub> cells (CD45RA<sup>+</sup>CCR7<sup>+</sup>CD27<sup>+</sup>CD95<sup>+</sup>). Purity was determined after sorting by flow cytometry. Here, we show 1 representative experiment. For these sorting experiments, we did not sort CD45RA<sup>+</sup>CD27<sup>-</sup>CCR7<sup>-</sup> cells (T<sub>EMRA</sub> cells; ref. 66). The low purity of T<sub>SCM</sub> cells can be explained by the low levels of CD95 expressed by these cells, resulting in limited separation between the CD95<sup>+</sup> and CD95<sup>-</sup> population. T<sub>EMRA</sub><sup>+</sup> effector memory reexpressing CD45RA.

sequences, the contribution of the naive reservoir was more prominent at the later time points (Figure 2B). In contrast, total HIV DNA decayed minimally over time in both subjects (Figure 2C), which is consistent with the minimal change of defective proviral DNA between these 2 time points (Supplemental Table 3 and Figure 2, C and D). This is consistent with studies monitoring integrated HIV DNA showing a minimal proviral decline after starting ART in chronically infected individuals (36–38). In addition, the contribution of T<sub>N</sub> cells to defective proviral DNA was minor when compared with the intact reservoir. This suggests that intact proviruses may have a survival advantage within T<sub>N</sub> cells, possibly because of reduced HIV expression, which in turn might protect them from viral and/or immune cytotoxicity (39). Alternatively, intact proviruses could form more efficiently in T<sub>N</sub> cells; however, a recent study suggests greater than 90% of proviruses are intact after a single-round infection of primary CD4<sup>+</sup> T cells, making this hypothesis less likely (3).

*Proviral sequences in naive and effector memory cells have distinct characteristics that represent a continuum.* We wanted to explore the genetic pressures exerted on HIV in different cellular subsets by studying the composition of proviral DNA. We previously provided evidence that we could identify distinct categories of proviruses that are meaningful, as they undergo differential selection pressures (6). In Figure 3, we color coded proviral sequences into 6 categories based on their size and the location of deletions. These same categories could be defined by the presence or absence of D1 and D4 splice sites: intact (green, D1<sup>+</sup>D4<sup>+</sup>), nearly intact (black, D1<sup>+</sup>D4<sup>+</sup>), 3' deleted (red, D1<sup>+</sup>D4<sup>-</sup>), 5' deleted (blue, D1<sup>-</sup>D4<sup>+</sup>), massively deleted (yellow, D1<sup>-</sup>D4<sup>-</sup>), and hypermutated (purple, D1<sup>+/-</sup>D4<sup>+/-</sup>) proviruses (6). Each horizontal bar represents 1 assembled provirus aligned to HXB2. As expected, the majority of the proviruses were defective, and most had large deletions consistent with recent reports (4–6, 14).

**Table 1. Levels of total and intact HIV DNA in each subset at 2 and 9 years after ART initiation**

Subject 1	Contribution to intact reservoir by cellular subset									
	2 years of ART					9 years of ART				
	A	B	C	D	E	A	B	C	D	E
Subset	HIV DNA <sup>A</sup>	% Intact	% of total CD4	Intact HIV <sup>A</sup>	% of intact HIV	HIV DNA <sup>A</sup>	% Intact	% of total CD4	Intact HIV <sup>A</sup>	% of intact HIV
T <sub>N</sub>	655	32	21	43	39	468	6.6	23	7.1	34
T <sub>CM</sub>	4581	2	14	13	12	2641	2	20	11	51
T <sub>TM</sub>	4915	1.5	8	6	5.4	4383	0	7	0	0
T <sub>EM</sub>	3694	3.3	12	15	13	3446	0.9	11	3.2	15
CD45RA <sub>dim</sub>	2015	9.4	18	34	31	1893	0	21	0	0
Subject 2	2 years of ART					9 years of ART				
Subset	A	B	C	D	E	A	B	C	D	E
Subset	HIV DNA <sup>A</sup>	% Intact	% of total CD4	Intact HIV <sup>A</sup>	% of intact HIV	HIV DNA <sup>A</sup>	% Intact	% of total CD4	Intact HIV <sup>A</sup>	% of intact HIV
T <sub>N</sub>	1236	21	34	88	59	904	3.3	44	13	31
T <sub>CM</sub>	3719	5.8	9	19	13	2048	0	16	0	0
T <sub>TM</sub>	2601	16	4	16	11	1446	5.6	5	4	9.4
T <sub>EM</sub>	3316	14	1	4.7	3.2	740	18	2	2.6	6
CD45RA <sub>dim</sub>	1823	8.3	13	20	13	1675	6	23	23	54

<sup>A</sup>Number of copies per million CD4<sup>+</sup> T cells. The number of intact HIV/million CD4<sup>+</sup> T cells in each subset was estimated using the following formula: column D = A × B × C. The relative contribution of each subset to the total pool of intact HIV (column E) was then calculated by dividing the number of intact HIV/million cells within subset (column D) by the total number of intact HIV in all subsets. As an example, for Subject 1 at the first time point, we measured 655 HIV copies/million cells in T<sub>N</sub> cells (Table 1, column A). We estimated the contribution of T<sub>N</sub> cells to the total pool of CD4 cells by flow cytometry (21%, column C) and the fraction of proviruses that was intact by NFL proviral sequencing (32%, column B). We used these estimates to calculate the number of intact HIV/million CD4 cells (column D) as column A × B × C (655 × 0.21 × 0.316) = 43.4 (estimate) intact HIV/million CD4<sup>+</sup> T cells). The percentage of relative contribution was calculated as: 43.4/(43.4 + 12.8 + 6 + 14.6 + 34) = 39.2%. Notably, the sum of the percentage of CD4<sup>+</sup> T cells contributed by each subset is less than 100% because T<sub>EMRA</sub> cells (66), as well as cells near the boundary of sort gates, were excluded. Numbers in the table are rounded to 2 significant digits, which would cause rounding errors if used to repeat these calculations. This explains why the estimate as an explanatory example does not perfectly match the estimates provided in Table 1.

T<sub>EM</sub> cells showed the strongest evidence of selection, as they were dominated by distinct identical sequences at both time points. These identical proviral sequences detected more than once were likely to represent clones that we named “proviral clones.” This is likely a reasonable assumption, since most repeated sequences were defective proviruses and, therefore, not capable of spreading infection. For intact proviruses, it is possible that identical sequences might have different integration sites, since these intact proviruses have the capacity to release virus and cause viral spread. However, given that these individuals are on suppressive ART, this is only remotely possible.

Interestingly, there was a relative increase in blue proviral clones with 5' deletions (D1<sup>-</sup>D4<sup>+</sup> proviruses), suggesting a selective advantage for this category of proviruses consistent with refs. 6, 9, 10. In both subjects, there was a decline in the absolute level of D1<sup>+</sup> proviruses with ORFs over time (green, black and red) again consistent with (6, 9). In Subject 2, we observed a relative increase in intact proviruses, but these proviruses were all identical proviral sequences at the second time point, suggesting clonal expansion. Overall, there was evidence of large turnover of proviral DNA in the T<sub>EM</sub> subset.

The composition of the proviral DNA was strikingly different in the T<sub>N</sub> subset compared with T<sub>EM</sub> cells. Surprisingly, a large fraction of proviruses in the T<sub>N</sub> compartment were intact or nearly intact. These D1<sup>+</sup>D4<sup>+</sup> proviruses declined over time and appeared to be unique sequences, as they were only detected once. In contrast, defective proviruses within the T<sub>N</sub> compartment represented a minor fraction, and their character did not change over time, consistent with the idea that these defective proviruses were under less selective pressure in T<sub>N</sub> cells in comparison with T<sub>EM</sub> cells. To summarize, intact and nearly intact (D1<sup>+</sup>D4<sup>+</sup>) defective proviruses declined more rapidly in the T<sub>N</sub> subset relative to other defective proviral categories. In contrast, in the T<sub>EM</sub> subset, there was negative selection against some categories of defective proviruses (D1<sup>+</sup>ORF<sup>+</sup>) and relative positive selection for others (D1<sup>-</sup>D4<sup>+</sup>), resulting in the presence of distinct large clones and a high turnover rate. Thus, the genetic

**Table 2. Cells sorted and analyzed per subject at 2 and 9 years after ART initiation**

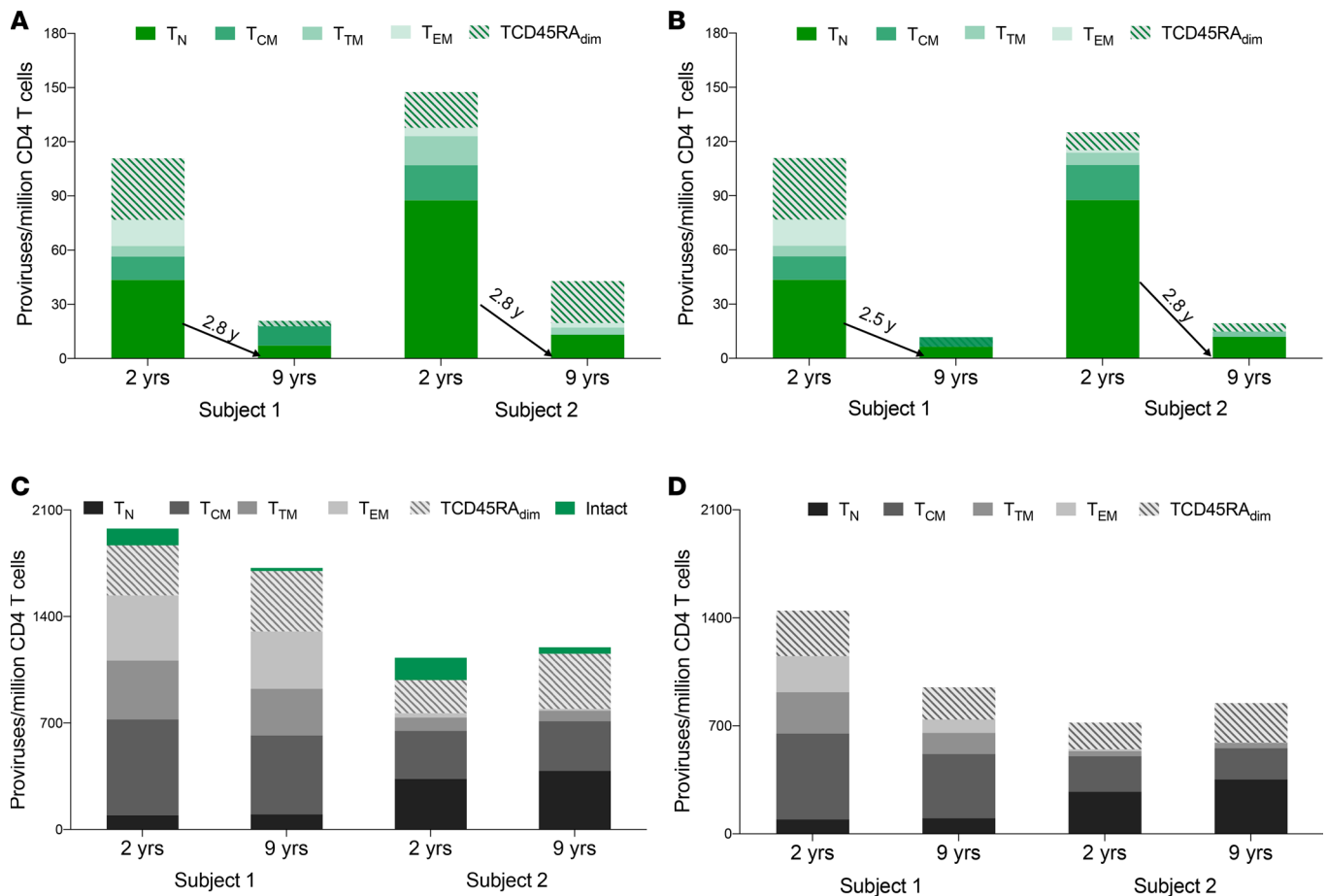
<b>Subject 1</b>										
Cellular subset	Cells processed 2 years after ART (millions)			Proviruses		Cells processed 9 years after ART (millions)			Proviruses	
	Sorted	qPCR	NFL seq	Intact (N)	Defective (N)	Sorted	qPCR	NFL seq	Intact (N)	Defective (N)
T <sub>N</sub>	2.2	0.5	1.7	23	53	7.2	2	4.5	6	77
T <sub>CM</sub>	1.4	1	0.12	2	98	6.6	2	0.2	2	98
T <sub>TM</sub>	0.8	0.7	0.1	1	65	3.3	2	0.1	0	128
T <sub>EM</sub>	1.1	0.8	0.2	3	88	3.5	2	0.2	1	117
CD45RA <sub>dim</sub>	1.6	0.9	0.7	3	29	3.1	2	0.3	0	95
CD95 <sup>-</sup> T <sub>N</sub>						8	2	1	3	50
T <sub>SCM</sub>						0.03	-	0.03	0	0
<b>Subject 2</b>										
Cellular subset	Cells processed 2 years after ART (millions)			Proviruses		Cells processed 9 years after ART (millions)			Proviruses	
	Sorted	qPCR	NFL seq	Intact (N)	Defective (N)	Sorted	qPCR	NFL seq	Intact (N)	Defective (N)
T <sub>N</sub>	2	1	0.2	10	38	8.4	2	1.7	2	87
T <sub>CM</sub>	1.1	1	0.1	3	49	14	2	1.5	0	82
T <sub>TM</sub>	0.8	0.6	0.1	7	38	6.7	2	0.3	4	68
T <sub>EM</sub>	0.2	0.1	0.1	10	60	2.9	1	1	14	66
CD45RA <sub>dim</sub>	1	0.8	0.2	4	44	2.3	2	0.2	5	78
CD95 <sup>-</sup> T <sub>N</sub>						35	2	1.2	8	207
T <sub>SCM</sub>						0.6	-	0.6	3	88

The number of cells that were sorted and used for qPCR measures and NFL proviral sequencing, as well as the number of intact and defective sequences obtained at each time point for each subject. For Subject 2 at the second time point, we were able to sort a larger number of cells in order to obtain enough proviral sequences from T<sub>SCM</sub> and T<sub>EM</sub> cells. These additional sorting experiments resulted in a large number of cells from other subsets that were processed but not used for downstream applications. yrs, years after ART initiation.

composition and longitudinal changes observed in the T<sub>N</sub> and T<sub>EM</sub> reservoir were distinct (see also Supplemental Tables 4 and 5). Deletion maps for T<sub>CM</sub> and T<sub>TM</sub> cells are provided in Supplemental Figure 1 and show intermediate phenotypes between the extremes of T<sub>N</sub> and T<sub>EM</sub> cells.

*Large proviral clones are more frequent in T<sub>EM</sub> compared with T<sub>N</sub> cells.* We aligned all the HIV sequences to compare the extent of repeated proviral sequences in T<sub>N</sub> and T<sub>EM</sub> cells. We identified 2 repeated sequences out of 212 total sequences (<1%) for T<sub>N</sub> cells and 133 repeated sequences out of 209 (64%) for T<sub>EM</sub> cells in Subject 1. Subject 2 followed the same pattern, with 10% repeated sequences within the T<sub>N</sub> compartment and 71% repeated sequences among T<sub>EM</sub> cells. The fractions of detectable proviral clones was significantly higher in T<sub>EM</sub> versus T<sub>N</sub> cells at both time points in each subject by a 2-tailed  $\chi^2$  test ( $P < 0.001$ ). Moreover, the number of clones was significantly greater at the second time point in T<sub>EM</sub> cells ( $P < 0.05$ ) but not in T<sub>N</sub> cells. Taken together, these results suggest that T<sub>N</sub> and T<sub>EM</sub> cells represent 2 ends of a continuum, with a much higher frequency of proviral clones as cells differentiate from a naive to an effector memory phenotype (Supplemental Table 4).

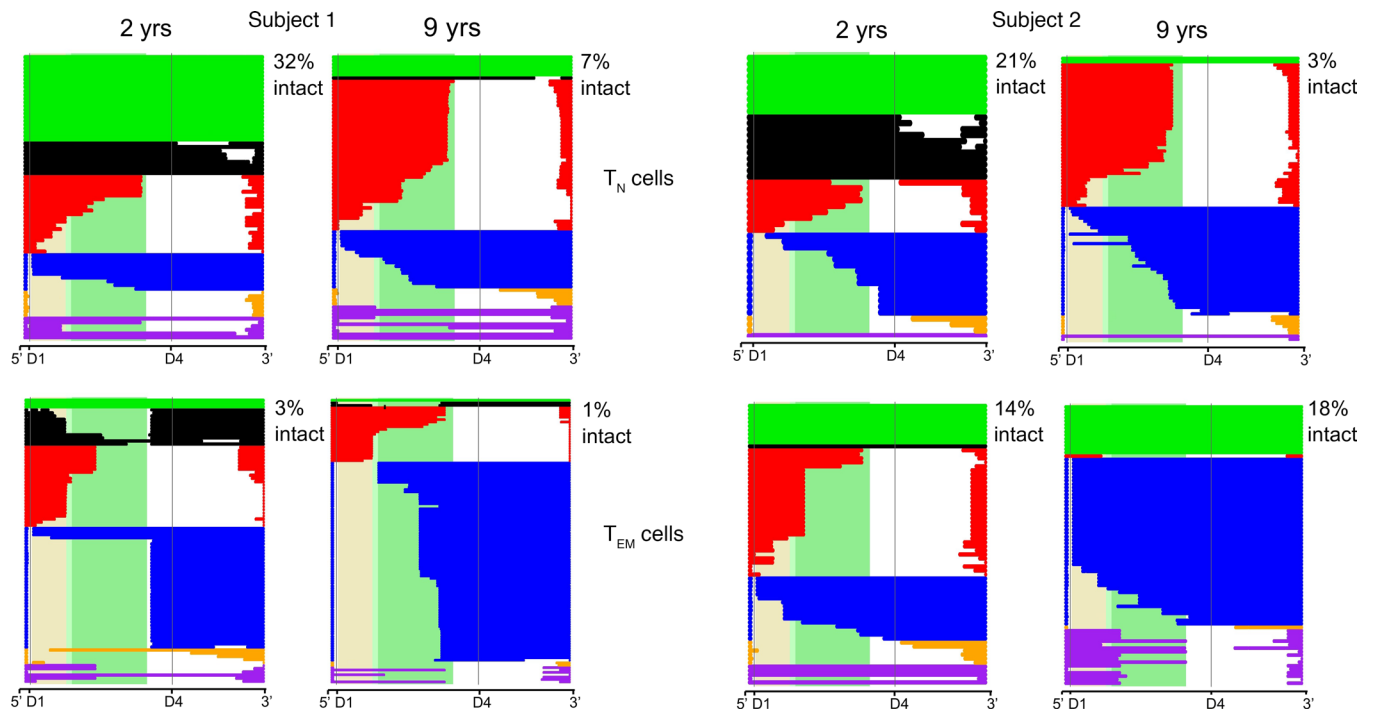
*CD95<sup>-</sup>T<sub>N</sub> and T<sub>SCM</sub> cells contain a similar fraction of intact proviruses.* The experiments presented thus far were based on a sorted CD45RA<sup>+</sup>CCR7<sup>+</sup>CD27<sup>+</sup>T<sub>N</sub> population. This population contains 0.2%–6% of stem-cell like memory (T<sub>SCM</sub>) cells, which have been reported to harbor latent HIV (15). In order to assess the contribution of T<sub>SCM</sub> cells to our results, we sorted CD45RA<sup>+</sup>CCR7<sup>+</sup>CD27<sup>+</sup>CD95<sup>-</sup> cells (CD95<sup>-</sup>T<sub>N</sub>) and CD45RA<sup>+</sup>CCR7<sup>+</sup>CD27<sup>+</sup>CD95<sup>+</sup> (T<sub>SCM</sub>) cells at the second time point, for which we had banked a sufficiently large number of cells. We measured the amount of HIV DNA in CD95<sup>-</sup>T<sub>N</sub> cells for both individuals by qPCR. For Subject 1, we estimated 411 HIV copies/million cells in CD95<sup>-</sup>T<sub>N</sub> versus 468 in T<sub>N</sub> cells. In Subject 2, we measured 798 copies/million cells in CD95<sup>-</sup>T<sub>N</sub> versus 904 copies/million cells in T<sub>N</sub> cells. This suggests that CD95<sup>-</sup>T<sub>N</sub> and T<sub>SCM</sub> cells contain similar levels of HIV DNA. For Subject 2, we amplified and sequenced proviruses from both the CD95<sup>-</sup>T<sub>N</sub> and T<sub>SCM</sub> compartment, obtaining 216 and 91 proviruses, respectively. For Subject 1, we were able to obtain proviruses only from CD95<sup>-</sup>T<sub>N</sub> cells (53 sequences), since



**Figure 2. Contribution of T cell subsets to proviral DNA at 2 and 9 years after ART initiation.** (A and B) Contribution of each T cell subset to the total number of intact proviruses at 2 time points before (A) and after removing large clonal populations (B). The half-life of intact HIV for T<sub>N</sub> cells is shown in the graph. The half-life of each subset is provided in Supplemental Table 3. T<sub>N</sub> cells represent a major contributor to the intact reservoir at both time points. Removal of repeated sequences revealed that large clones were more frequent among the memory subsets compared with the naive population. Notably, the half-life of intact proviruses in T<sub>N</sub> cells was minimally affected by removing repeated sequences in contrast to the half-life of the memory cell subsets, which was shortened. (C and D) Contribution of each T cell subset to the total number of defective proviruses at 2 time points before (C) and after removing large clonal populations (D). We calculated the level of defective HIV DNA before and after removing repeated sequences at both time points. We calculated the percentage of defective proviruses by subtracting the percentage of intact proviral sequences (depicted as green bars in C) from the total number of HIV proviruses (as estimated by qPCR). This percentage was used to estimate the absolute number of defective proviruses. The levels of defective proviruses minimally changed over time in both individuals. Moreover, T<sub>N</sub> cells contributed less to the pool of defective HIV in comparison with intact HIV.

T<sub>SCM</sub> cells were extremely rare (<0.001% of the total CD4<sup>+</sup> T cell population) (Supplemental Figure 2). In both individuals, the percentage of intact proviruses was similar to the one observed in T<sub>N</sub> cells. Finally, deletion maps did not reveal differences in the character of the proviral sequences amplified from CD95<sup>-</sup> T<sub>N</sub> and T<sub>N</sub> cells except that CD95<sup>-</sup> T<sub>N</sub> cells had fewer detectable proviral clones (Figure 4). Taken together, these results suggest that the HIV DNA detected in T<sub>N</sub> cells is not due to contaminating T<sub>SCM</sub> cells. While our in-depth sequencing analysis was limited to 2 individuals after 2 and 9 years of ART, Figure 5 shows a similar infection frequency and proviral character of T<sub>N</sub> cells from 3 additional HIV-infected individuals on ART enrolled in ref. 14. Thus, we were able to confirm our findings of an important role for T<sub>N</sub> cells in HIV persistence by applying our pipeline analysis to published data from Hiener et al. (14).

*Analysis of coreceptor tropism suggests that T<sub>N</sub> cells harbor CCR5-tropic infection.* We were next interested to determine if T<sub>N</sub> cells would only harbor CXCR4-tropic viruses, given that these cells appear to be resistant to CCR5-tropic HIV in vitro (24). We analyzed all proviruses with an intact env ORF using 2 bioinformatic tools: WebPSSM (position-specific scoring matrices) and Geno2Pheno (G2P) (40, 41). Both programs use algorithms based on critical amino acids in the V3 loops. The literature has reported limitations of both prediction tools (42, 43). Underlying this limitation is the fact that a significant fraction of HIV proviruses is dual tropic (44–49). Recognizing these limitations, we chose to analyze the proviral sequences with both

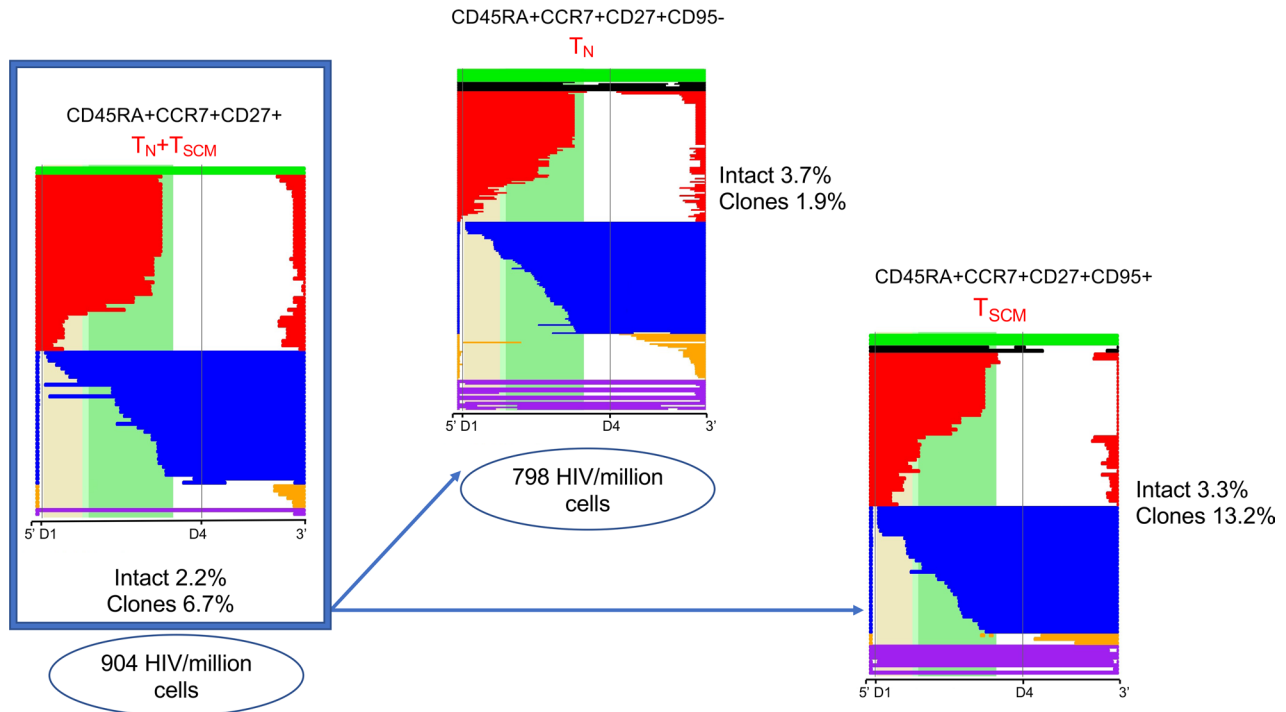


**Figure 3. Deletion maps for proviruses retrieved from  $T_N$  and  $T_{EM}$  cells at 2 and 9 years after ART initiation.** Intact and defective proviruses from  $T_N$  and  $T_{EM}$  cells were aligned to HXB2 at an early and late time point after ART initiation. Each horizontal bar represents 1 proviral sequence. Proviruses are color-coded based on size and location of deletion. Green bars are intact (D1-D4'), black bars are nearly intact (D1-D4'), red bars are 3' deleted (D1-D4'), blue bars are 5' deleted (D1-D4'), and yellow bars are massively deleted (D1-D4') proviruses. Hypermutated proviruses are represented as purple bars. The shaded beige, light green, and dark green regions correspond to the gag, gag-pol, and pol regions of HXB2, respectively. The percentage of intact proviruses is reported in the upper right corner of each deletion map.

a high and low false-positive rate (FPR). When analyzing sequences with the high FPR setting, we are likely overrepresenting CXCR4-tropic proviruses. We decided that this would be a reasonable approach to provide a lower-bound estimate of the frequency of CCR5-tropic proviruses within  $T_N$  cells. With the most restrictive criteria for classifying proviruses as CCR5-tropic (i.e., G2P 10% FPR), in both individuals, a sizeable fraction of  $T_N$  cells was predicted to be infected with CCR5-tropic HIV (35% and 14% for Subject 1 at the first and second time point, and 12% and 9% for Subject 2; Figure 6, A and B). While more investigation is needed to address these results with phenotypic studies, our findings suggest that  $T_N$  cells can harbor CCR5-tropic strains.

*UpSet plots identify relationships between cellular subsets.* We studied the relationship between cellular subsets by identifying repeated sequences (proviral clones) in different cellular subsets. These repeated sequences were used to create modified UpSet plots. The circles connected by lines represent proviral clones that were grouped based on 2 criteria: the time point and the cellular subset from which they were detected. This modified UpSet analysis revealed that different subsets often harbored the same proviral clones, suggesting that the differentiation pathway of T cells from a naive to an effector memory phenotype could play a role in HIV persistence.

The horizontal bars on the left of the UpSet capture the percentage of proviral clones in each subset at each time point (gray for defective clones and green for intact ones). The number of repeated proviral sequences detected is provided at the top of the UpSet plot, while the number of distinct sequences is shown below the UpSet plot. Proviral clones depicted in green are intact. As an example, for Subject 1, we detected 2 repeated intact proviral sequences in  $T_{EM}$  and  $T_N$  cells representing 1 distinct clonal sequence (Figure 7A, last column in green). As an additional example, for Subject 2 (Figure 7B), column 1 shows that we detected 31 repeated proviral sequences representing 1 distinct clone. This distinct proviral sequence was detected within  $T_{EM}$  and  $T_{TM}$  cells at both time points and  $T_{CM}$  cells at the second time point. Column 3 shows that we identified 7 distinct sequences in  $T_{EM}$  cells at the second time point (for a total number of 24 repeated proviral sequences) and one of them happened to be intact (as indicated by the hemi-green circle).



**Figure 4. Comparison of HIV DNA levels and reservoir composition by NFL sequencing in  $T_N$  versus  $T_{SCM}$  cells for Subject 2.** The sorting experiment described in Figure 1 was repeated, separating  $T_{SCM}$  from  $CD95^- T_N$  cells.  $CD95^- T_N$  cells were defined as  $CD45RA^+CCR7^+CD27^+CD95^-$  cells, while  $T_{SCM}$  cells were defined as  $CD45RA^+CCR7^+CD27^+CD95^+$  cells. Sorted cells were used to measure HIV DNA levels by qPCR and to obtain proviral sequences. These proviral sequences were used to generate deletion maps, as described in Figure 3.  $T_{SCM}$  and  $CD95^- T_N$  cells had similar levels of intact HIV, with a higher fraction of clones in  $T_{SCM}$  cells.

Overall, we identified 5 intact clones for Subject 2 and 1 intact proviral clone for Subject 1, as also shown in the phylogenetic tree in Supplemental Figure 3. The limited number of intact proviral clones suggests that selection against intact proviruses is stronger than against defective ones due to either direct cytotoxicity or immune pressure. Nonetheless, while selection against intact HIV is strong,  $T_N$  cells containing intact proviruses appear to repopulate the more differentiated cellular subsets.

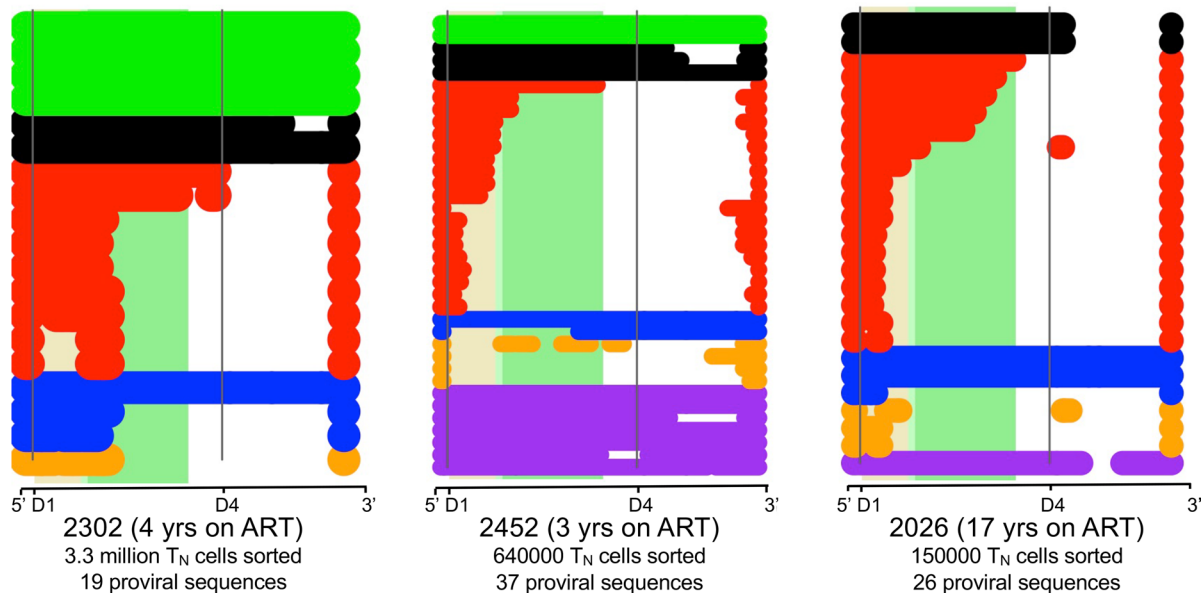
Our data are consistent with previous reports showing an increase in the proportion of proviral clones over time (6, 50). A sizeable fraction of proviral clones persisted and were detected at both time points 7 years apart from each other. The proportion of proviral clones in each subset progressed from  $T_N$  to  $T_{EM}$  cells (Supplemental Table 4, Figure 7, and Supplemental Figure 3). This follows the expected pathway of cellular differentiation with  $T_N$  cells dividing less frequently than  $T_{SCM} < T_{CM} < T_{TM} < T_{EM}$  cells (30, 51). How  $CD45RA_{dim}$  cells fit into this pathway is unclear, as these cells have not been routinely studied. In our study, we found that  $CD45RA_{dim}$  cells have a higher number of clones than  $T_N$  (and often  $T_{CM}$ ) cells but fewer than  $T_{TM}$  or  $T_{EM}$  cells, suggesting that they might represent a population in transition from a naive to a memory phenotype and vice versa.

## Discussion

The importance of  $T_N$  cell infection has generally been discounted as inconsequential, as the low level of infection found in this cellular subset have been attributed to contamination of sorted  $T_N$  cells by memory or stem-cell memory T cells and to reversion of memory cells to a naive phenotype. Our study challenges these explanations by suggesting a central role for  $T_N$  cells in HIV persistence.

We first show that naive infection cannot be due to contaminating memory cells by examining HIV DNA levels (Figure 1 and Table 1). We further support the case for naive infection with the surprising finding that these cells significantly contribute to the intact reservoir by proviral sequencing (Figure 2), as they represent one of the largest contributing subsets. We then provide genetic evidence that the character of the proviruses retrieved from infected  $T_N$  cells is distinct from any other subset (Figure 3 and Figure 4) with fewer identical sequences. The distinct nature of this subset further argues against the idea that infection could be due to contamination by or reversion from memory T cells. When we depleted  $T_{SCM}$  cells from





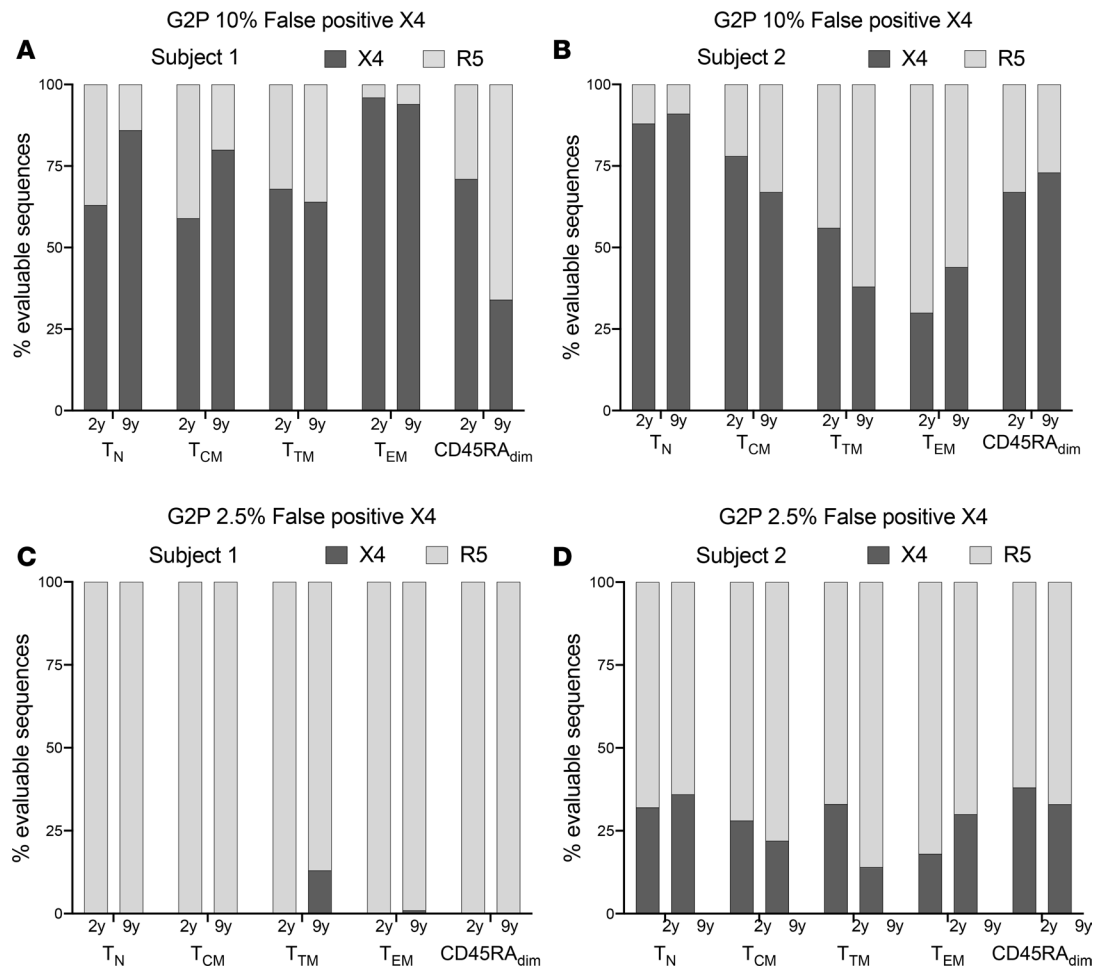
**Figure 5. Sequences from sorted naive cells support that T<sub>N</sub> cells contribute to the HIV reservoir.** Sequences from sorted T<sub>N</sub> cells were downloaded from Hiener et al. (14) and analyzed using our pipeline. The deletion maps were generated with the same bioinformatic tools used in Figure 3 and Figure 4 and Supplemental Figures 1 and 2 for our 2 subjects. No intact proviruses were detected for subject 2026, which might be explained by the low number of sorted T<sub>N</sub> cells, as well as the long time the patient has been on ART. Nonetheless, the composition of the reservoir in T<sub>N</sub> cells in these 3 individuals closely matches the proviral composition detected in T<sub>N</sub> cells in Figure 3.

the T<sub>N</sub> cells, we found the intact reservoir persisted with fewer clonal species. Thus, the infection detected in T<sub>N</sub> cells cannot be due to these rare cells (Figure 4). We used our bioinformatic pipeline to show similar findings in publicly available databases (Figure 5). Finally, we focused our analysis on the clonal populations present in CD4<sup>+</sup> T cells, showing a steady progression of clone size with T cell maturation (Figure 7).

*T<sub>N</sub> cells may repopulate the memory reservoir.* Our evidence that T<sub>N</sub> cells may give rise to other subsets includes the observation that identical sequences are shared between T<sub>N</sub> cells and other subsets. Moreover, the number of proviral clones increases as cells become more differentiated (Figure 7 and Supplemental Table 4). Our analysis shows a steady progression in the number of clonal sequences from T<sub>N</sub> to T<sub>SCM</sub> to T<sub>CM</sub> to T<sub>TM</sub> to T<sub>EM</sub> cells. Each clonal population likely arose from a single progenitor infection event. Most of the intact proviruses in T<sub>EM</sub> cells appear to be large expanded clones. Given that we sampled a very small fraction of the total proviral population (<0.00001%), detection of a proviral sequence more than twice in any sample suggests this clone made up a substantial portion of all proviruses in the blood. By contrast, the intact reservoir in the T<sub>N</sub> compartment appears to have fewer large proviral clones. Thus, a larger fraction of the intact proviruses detected in T<sub>N</sub> cells likely represents separate successful infection events. From our data, we infer that infection events leading to the formation of an intact reservoir in T<sub>N</sub> cells might be more common than previously recognized. Indeed, our evidence that clonal progeny arose from less differentiated subsets suggests that many of the expanded intact clones in the more differentiated memory subsets started as T<sub>N</sub> or T<sub>CM</sub> infections. This overarching finding of a progression from more unique to more clonal sequences as we move from naive to more differentiated memory cells is consistent with Lee et al. (8).

*T<sub>N</sub> cells provide a unique hurdle to HIV persistence.* T<sub>N</sub> cells have the longest half-life (51, 52) among cellular subsets. Evidence is accumulating that T<sub>N</sub> cells express less HIV RNA and proteins in vitro and in vivo (31–33, 53). Moreover, these cells have been suggested to be more resistant to cytotoxic T lymphocyte (CTL) killing (39). These features could result in less negative selection pressure against infected T<sub>N</sub> cells in comparison with memory T cells, which might explain, in turn, why the intact reservoir appears to be more protected in naive cells. It remains to be established whether T<sub>N</sub> cells are more resistant to latency reversal, as only a few studies have investigated T<sub>N</sub> infection (24) and fewer still have addressed latency reversal in cellular subsets (54, 55).

*HIV may play a role in proviral clonal expansion in T<sub>EM</sub> cells.* Notably, D1–D4<sup>+</sup> (5' deleted blue) proviruses in T<sub>EM</sub> cells expanded relative to all other proviral categories and were largely clonal. At first glance, this might appear counterintuitive, as we would expect enrichment for massively deleted proviruses

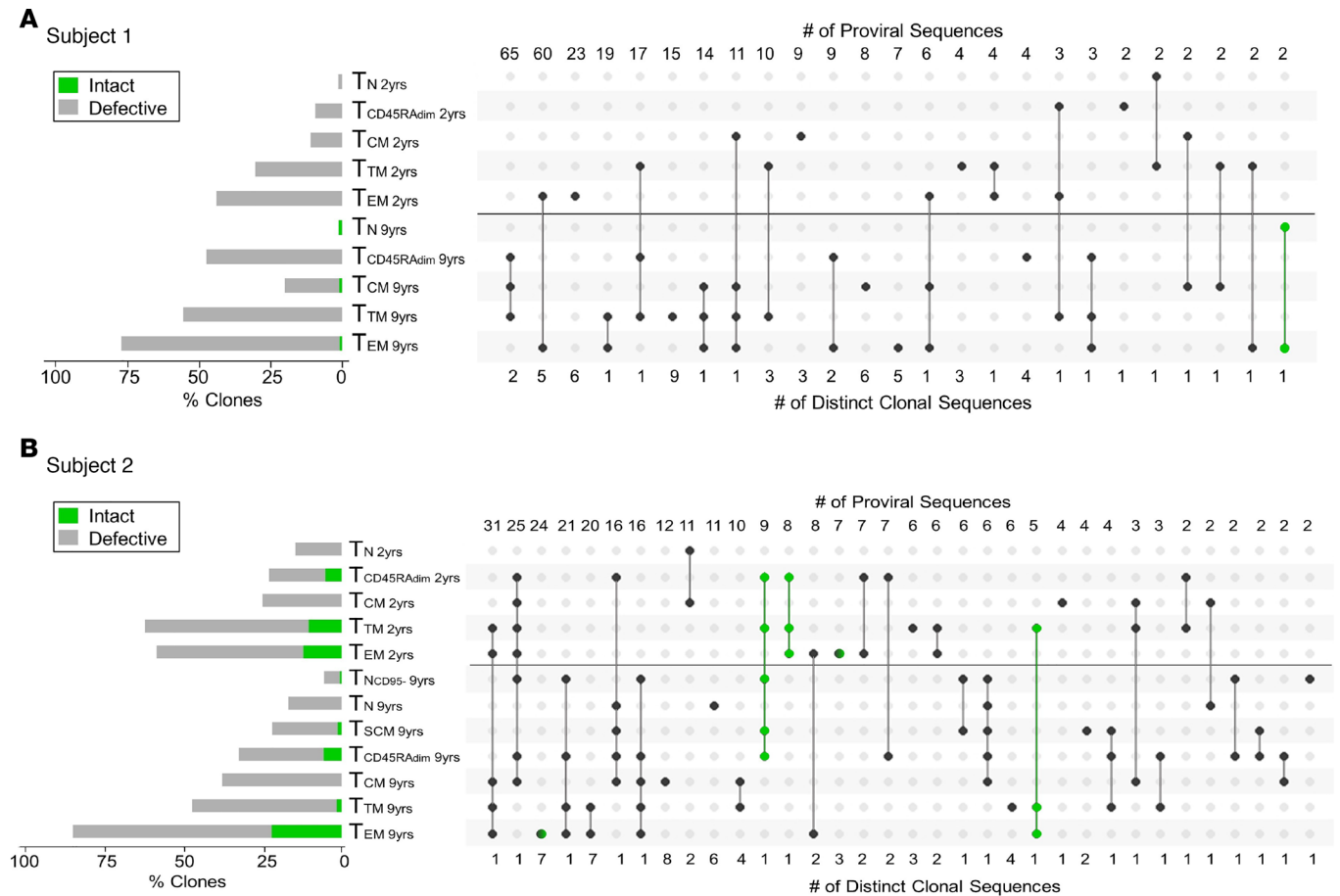


**Figure 6. Analysis of coreceptor tropism using Geno2Pheno (G2P) suggests that  $T_N$  cells might harbor CCR5-tropic HIV.** We used G2P to study coreceptor tropism of proviral sequences retrieved from different subsets. We included all the evaluable env sequences. (A and B) We used G2P with a 10% FPR, where the FPR indicates the likelihood to misclassify a provirus as CXCR4-tropic (which is based on phenotypic assays). When using G2P FPR 10%, both subjects showed a predominance of CXCR4-tropic sequences. (C and D) We used G2P with a 2.5% FPR. With G2P FPR 2.5%, most evaluable env ORFs were predicted to be CCR5-tropic for Subject 1 at both time points (C), while Subject 2 had, on average, 30% CXCR4-tropic sequences across subsets (D). Interestingly, with either 10% or 2.5% FPR,  $T_N$  cells were predicted to contain CCR5-tropic sequences in both individuals.

(yellow, D1<sup>-</sup>D4<sup>-</sup>) over other categories, since these massively deleted proviruses have a lower potential to express HIV proteins and, therefore, should be subject to limited negative selection. However, massively deleted D1<sup>-</sup>D4<sup>-</sup> proviruses were not the most persistent proviral category. This strongly suggests that the enrichment for D1<sup>-</sup>D4<sup>+</sup> proviruses we observed over time might be due to positive selection. We have hypothesized that positive selection might result from aberrant splicing of these D1<sup>-</sup>D4<sup>+</sup> proviruses to downstream oncogenes, which, in certain scenarios, could lead to increased cellular proliferation (6, 56). D1 and D4 are the strongest donor splice sequences in HIV. Given the organization of the HIV genome, with only 1 weak splice site acceptor after D4 but many splice acceptor sequences after D1, it appears reasonable that D4 would often splice aberrantly to downstream genes.

*HIV DNA levels in cellular subsets obscure important selection pressures.* Consistent with other reports (12–19), our work shows that measures of total HIV DNA provide a misleading picture of the reservoir size, since  $T_N$  cells have the lowest integration levels (12–16, 19). To estimate the real contribution by each subset, it is essential to distinguish intact and defective proviruses. Recently, Zerbato et al. showed similar levels of HIV in  $T_N$  and  $T_{CM}$  cells by viral outgrowth, even though  $T_{CM}$  cells had higher levels of integrated HIV DNA (17). This study indirectly corroborates our finding of a higher percentage of intact proviruses in  $T_N$  cells.

*Unique differences in reservoir character and selection pressures in  $T_N$  cells were revealed by NFL proviral sequencing.* It can be argued that we should further divide  $T_N$  cells into subsets based on additional markers. Historically,  $T_N$  cells were defined using the same markers we used in our study (13, 16–18, 31, 34).



**Figure 7. Modified proviral clone UpSet plots.** (A and B) Modified proviral clone UpSet plot for Subject 1 (A) and Subject 2 (B) shows that clonal expansion progressively increases with cell differentiation and is more prominent among defective proviruses. Proviral clones were identified as repeated sequences. The horizontal bars on the left side show the percentage of repeated sequences found in each subset (green for intact sequences and gray for defective ones). A black horizontal line separates the 2 and 9 years after ART initiation). Intact proviral clones are shown in green. For those proviral clones that could be detected in multiple subsets, a solid line was used to connect these subsets. The numbers at the top of the UpSet plot represent the number of proviral sequences that could be found in the same subsets within a category. The numbers below the UpSet plot represent the number of distinct clonal sequences. For example, for Subject 2 (B), we identified 308 repeated sequences. The first column shows that we detected 1 distinct clone made up by 31 proviral sequences in  $T_{EM}$  and  $T_{TM}$  at both time points, as well as  $T_{CM}$  cells at the second time point. For Subject 1, we identified 303 repeated sequences (A). We also identified 5 intact clones for Subject 2 (B) and 1 proviral clone for Subject 1 (A). yrs, years after ART initiation.

More importantly, even without using additional markers, there were clear differences in the proviral character of different sorted subsets, suggesting that our approach identified meaningful functional differences. While the  $CD45RA^+CCR7^+CD27^+CD95^-$  subset contains cells that are not truly naive from a functional perspective, the ability of this subset to recirculate (due to the presence of CCR7), their lack of recent activation (as suggested by the presence of CD45RA), and their longer half-life compared with memory T cells (51) provide a rationale for grouping these cells into one population. The limited number of proviral clones retrieved from naive cells is consistent with the properties of this population that divides rarely and recirculates continuously between the blood and lymphoid tissues.

When we analyzed the  $CD45RA^+CCR7^+CD27^+CD95^-$  naive population, we noticed similar integrated and intact HIV levels in comparison with the  $T_{SCM}$  compartment ( $CD45RA^+CCR7^+CD27^+CD95^+$ ). This would suggest that, at least in these 2 individuals,  $T_{SCM}$  cells do not differ greatly in character from  $CD95^-T_N$  cells, except for the higher frequency of repeated proviral sequences observed in  $T_{SCM}$  cells. Again, our findings are consistent with Zerbato et al., showing no significant differences in total HIV DNA and replication-competent HIV in the total  $T_N$  population (containing  $T_{SCM}$  cells) in comparison with the  $CD95^-T_N$  compartment (17).

It might seem surprising to observe CCR5-tropic sequences in  $T_N$  cells (Figure 6) since these cells are highly resistant to infection with CCR5-tropic HIV in vitro. In fact, we were unable to detect infection after in vitro inoculation of  $CD4^+$  T cells (ref. 24 and O’Doherty, unpublished observations). It is possible that

CCR5-tropic HIV can infect naive cells inefficiently due to minimal or transient CCR5 expression on  $T_N$  cells (57, 58), as CCR5 can be transiently upregulated by various stimuli (24, 59, 60). Moreover, lack of detection of CCR5 by flow cytometry would not prove its absence on the cellular surface, since flow cytometry has a detection limit of 50–100 molecules and some databases report very low levels of CCR5 RNA (61). Alternately,  $T_N$  cells may upregulate CCR5 after direct contact with antigen presenting cells in lymph nodes, since they can recirculate. Consistent with this hypothesis, it has recently been reported that  $T_N$  cells in lymph nodes harbor a high percentage of replication-competent HIV (62). In addition,  $T_N$  infection with a CCR5-tropic virus appears to occur in the presence of CCL19 and B cells (63). Memory cells harboring HIV may revert to a naive phenotype after being infected with a CCR5-tropic virus (64). It should also be noted that our analysis of coreceptor tropism is based on genotypic predictions, not phenotypic tests. As a consequence, we cannot rule out dual tropism conferred by determinants located outside of the V3 loop region, as previously described by others (44–49). While beyond the scope of this manuscript, further studies are needed to improve our understanding of the role of coreceptor tropism in  $T_N$  infection.

A major limitation of our study is the restricted number of individuals we enrolled. We made the deliberate decision to sequence a large number of proviruses per cellular subset rather than to sequence fewer proviruses in many subjects. Our study complements the results published by other groups that sorted cellular subsets in a larger number of HIV-infected individuals — but with fewer sequences per patient and fewer sequences per cellular subset (4, 5, 8, 14). To address this limitation, we applied our bioinformatic pipeline to analyze sequences obtained after sorting  $T_N$  cells from a recently published study (ref. 14 and Figure 5). While this analysis suggests our findings may apply more broadly, it also demonstrates the need for deep sequencing within each subset to provide robust conclusions.

In conclusion, our work suggests that the contribution of  $T_N$  cells to HIV persistence should be reevaluated.  $T_N$  cells may be more resistant to immune clearance and viral cytotoxicity due to their lower viral expression as reported by refs. 31–33 and 53. Thus, the  $T_N$  reservoir may be a formidable hurdle to HIV cure. Moreover, infected  $T_N$  cells can give rise to infected memory T cells through differentiation and, thereby, have the potential to continuously repopulate the memory subset, obscuring the true decay rate of infected memory cells. Further studies to explore how to perturb latency in different cellular subsets are needed.

## Methods

**Patients and samples.** We used samples from 2 HIV-1-infected individuals whose clinical history and reservoir decay curves were previously described (6). Apheresis samples were collected at 2 and 9 years after ART initiation. Before ART initiation, Subject 1 experienced a slow  $CD4^+$  T cell decline and a nadir of 295  $CD4$  cells/ $\mu$ L after 21 years of infection. Subject 2 had a more rapid progression, with a nadir of 0  $CD4$  cells/ $\mu$ L after 6 years of infection (Supplemental Tables 1 and 2). Both subjects had < 20 copies of HIV RNA/mL at the time of the apheresis.

**Sorting and purity for qPCR and NFL proviral sequencing.**  $CD3^+CD8^-$  T lymphocytes were negatively selected from peripheral blood mononuclear cells (PBMCs) by an immunomagnetic bead-based protocol (Stem Cell Technologies, catalog 19052).

Purified T lymphocytes were then sorted by fluorescence activated cell sorting (BD FACS Aria III sorter) into the following subsets:  $CD3^+CD8^- T_N$  ( $CD45RA^+CCR7^+CD27^+$ ),  $CD95^- T_N$  ( $CD45RA^+CCR7^+CD27^+CD95^-$ ),  $T_{SCM}$  ( $CD45RA^+CCR7^+CD27^+CD95^+$ ),  $T_{CM}$  ( $CD45RA^-CCR7^+CD27^+$ ),  $T_{TM}$  ( $CD45RA^-CCR7^-CD27^+$ ),  $T_{EM}$  ( $CD45RA^-CCR7^-CD27^-$ ), and  $CD45RA_{dim}$  cells (Figure 1) using PE-Cy7 anti- $CD27$  (clone 0323, Invitrogen), BV421 anti- $CD45RA$  (clone 5H9, BD Biosciences), BB700 anti- $CCR7$  (clone 3D12, BD Biosciences), and APC anti- $CD95$  (clone DX2, BD Biosciences). The following filters were used: 450/50 Violet for BV421, 780/40 Green for PeCy7, 710/50 Blue for  $CCR7$ , and 660/20 for APC (BD Biosciences). FlowJo v10.6 software was used for analysis.

**DNA quantification and NFL proviral sequencing.** Genomic DNA was purified using the Gentra Puregene kit (QIAGEN), and HIV-DNA was measured by HIV LTR qPCR (6) with several dilutions and replicates (between 9 and 15 replicates per sample). Proviruses were then amplified using NFL PCR at limiting dilution and sequenced on an Illumina MiniSeq as described in ref. 6. An in-house HIV genome analysis program developed in R studio was used to identify intact proviruses (6). The criteria to define a provirus as intact included the presence of 9 ORFs without premature stop codon and insertions/deletions, the presence of 3–4 stem loops at the psi packaging site, as well as the presence of the critical donor and acceptor splice sequences and the Rev response element (6).

*Analysis of proviral clones and classification of proviral sequences based on coreceptor usage.* Sequences were aligned to HXB2 to identify deletions using Multiple Alignment with Fast Fourier Transformation (MAFFT) (65). Proviruses with 100% sequence identity were considered clones. Identical sequences (defined as proviral clones) were grouped into clonal families. To increase the power of our analysis and detect more proviral clones, we used our complete database, including the proviral sequences retrieved from PBMCs published in ref. 6.

Proviral coreceptors were classified by analyzing the V3 region of the HIV envelope protein gp120 using WebPSSM tool (40) and G2P (41). For G2P, we performed our analysis using a 10% and 2.5% FPR. Since the WebPSSM and G2P 2.5% FPR gave comparable results, we decided to only show results from the G2P analysis in Figure 6.

*Calculation of the contribution of cellular subsets to the HIV reservoir.* Thirty-two to 304 proviral sequences were available per time point to estimate the percentage of intact proviruses. The contribution of intact proviruses to the reservoir by each subset ( $C_s$ ) was calculated with the following formula: (intact proviruses/ $1 \times 10^6$  subset cells)  $\times$  (subset cells/ $1 \times 10^6$  CD4<sup>+</sup>) =  $C_s$ .

The absolute number of intact proviruses contributed per subset was calculated as: intact proviruses = total HIV LTR DNA per subset cells  $\times$  % intact in that subset.

The percent contribution of each subset was calculated as:  $(C_s/CT_N + CCD45RA_{dim} + CT_{CM} + CT_{TM} + CT_{EM}) \times 100$ , where  $CT_N$  indicates contribution  $T_N$  cells,  $CCD45RA_{dim}$  indicates contribution  $CD45RA_{dim}$  cells,  $CT_{CM}$  indicates contribution  $T_{CM}$  cells,  $CT_{TM}$  indicates contribution  $T_{TM}$  cells, and  $CT_{EM}$  indicates contribution  $T_{EM}$  cells.

*Statistics.* A 2-tailed  $\chi^2$  test was used to compare the fraction of proviral clones in  $T_{EM}$  versus  $T_N$  cells in each individual, as well as the number of clones within the same subset at 2 time points.  $P < 0.05$  was considered significant.

For data presented in Supplemental Table 3, confidence intervals were computed as exact posteriors, using log-normal variance computed from repeated measures for qPCR measurements and binomial statistics for intact versus defective ratios. R and Excel software were used for statistical analyses. Prism Graph-Pad was used to generate the figures presented in the paper.

*Study approval.* The study was approved by the IRB at the NIH and at the University of Pennsylvania, and written informed consent was received from participants before inclusion in the study.

## Author contributions

MRP, EVR, and UO designed the study. EVR, MRP, and MC conducted the experiments. EVR, MRP, LC, SW, RZ, GN, and UO analyzed the data. UO wrote the manuscript with significant contributions from MRP, as well as additional contributions from RZ, SW, and EVR. RZ provided statistical support for data analysis. All the authors read and approved the final version of the manuscript.

## Acknowledgments

We would like to acknowledge Paul Hallberg for his technical support during sorting and flow cytometry experiments. We would like to thank David Levy for his intellectual input and Charlotte Kearns and Alfonso Ocegüera for critically reading the manuscript. This study was supported by NIH grants 1UM1AI26617, R01AI120011, R37AI150556, and R21AI143564 and by the project “Dottorati FSE XXXII ciclo Unime” CIP n. 2014.IT.05.SFOP.014/3/10.5/9.2.02/0006.

Address correspondence to: Una O’Doherty, Department of Pathology and Laboratory Medicine, University of Pennsylvania, 702 Stellar Chance, 422 Curie Boulevard, Philadelphia, Pennsylvania 19104, USA. Phone: 267.872.0750; Email: unao@penmedicine.upenn.edu.

1. Barré-Sinoussi F, Ross AL, Delfraissy JF. Past, present and future: 30 years of HIV research. *Nat Rev Microbiol.* 2013;11(12):877–883.
2. Finzi D, et al. Latent infection of CD4<sup>+</sup> T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nat Med.* 1999;5(5):512–517.
3. Pinzone MR, Bertuccio MP, VanBelzen DJ, Zurakowski R, O’Doherty U. Next-Generation Sequencing in a Direct Model of HIV Infection Reveals Important Parallels to and Differences from *In Vivo* Reservoir Dynamics. *J Virol.* 2020;94(9):e01900-19.
4. Ho YC, et al. Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell.* 2013;155(3):540–551.
5. Bruner KM, et al. Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nat Med.* 2016;22(9):1043–1049.

6. Pinzone MR, et al. Longitudinal HIV sequencing reveals reservoir expression leading to decay which is obscured by clonal expansion. *Nat Commun.* 2019;10(1):728.
7. Einkauf KB, et al. Intact HIV-1 proviruses accumulate at distinct chromosomal positions during prolonged antiretroviral therapy. *J Clin Invest.* 2019;129(3):988–998.
8. Lee GQ, et al. Clonal expansion of genome-intact HIV-1 in functionally polarized Th1 CD4+ T cells. *J Clin Invest.* 2017;127(7):2689–2696.
9. Peluso MJ, et al. Differential decay of intact and defective proviral DNA in HIV-1-infected individuals on suppressive antiretroviral therapy. *JCI Insight.* 2020;5(4):132997.
10. Bruner KM, et al. A quantitative approach for measuring the reservoir of latent HIV-1 proviruses. *Nature.* 2019;566(7742):120–125.
11. Ostrowski MA, 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):6430–6435.
12. Sáez-Cirión A, 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.
13. Chomont N, et al. HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nat Med.* 2009;15(8):893–900.
14. Hiener B, et al. Identification of Genetically Intact HIV-1 Proviruses in Specific CD4+ T Cells from Effectively Treated Participants. *Cell Rep.* 2017;21(3):813–822.
15. Buzon MJ, et al. HIV-1 persistence in CD4+ T cells with stem cell-like properties. *Nat Med.* 2014;20(2):139–142.
16. Brechley JM, et al. T-cell subsets that harbor human immunodeficiency virus (HIV) in vivo: implications for HIV pathogenesis. *J Virol.* 2004;78(3):1160–1168.
17. Zerbato JM, McMahon DK, Sobolewski MD, Mellors JW, Sluis-Cremer N. Naive CD4+ T Cells Harbor a Large Inducible Reservoir of Latent, Replication-competent Human Immunodeficiency Virus Type 1. *Clin Infect Dis.* 2019;69(11):1919–1925.
18. Fromentin R, et al. CD4+ T Cells Expressing PD-1, TIGIT and LAG-3 Contribute to HIV Persistence during ART. *PLoS Pathog.* 2016;12(7):e1005761.
19. Bacchus 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.
20. Margolis L, Shattock R. Selective transmission of CCR5-utilizing HIV-1: the ‘gatekeeper’ problem resolved? *Nat Rev Microbiol.* 2006;4(4):312–317.
21. Nicholson JK, et al. CCR5 and CXCR4 expression on memory and naive T cells in HIV-1 infection and response to highly active antiretroviral therapy. *J Acquir Immune Defic Syndr.* 2001;27(2):105–115.
22. Wightman F, 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):1738–1748.
23. Fabre-Mersseman V, et al. CD4+ recent thymic emigrants are infected by HIV in vivo, implication for pathogenesis. *AIDS.* 2011;25(9):1153–1162.
24. Dai J, 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):4528–4537.
25. Baldanti F, 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.
26. Ganesan A, et al. Immunologic and virologic events in early HIV infection predict subsequent rate of progression. *J Infect Dis.* 2010;201(2):272–284.
27. Douek DC, et al. HIV preferentially infects HIV-specific CD4+ T cells. *Nature.* 2002;417(6884):95–98.
28. Heeregrave EJ, et al. Lack of in vivo compartmentalization among HIV-1 infected naive and memory CD4+ T cell subsets. *Virology.* 2009;393(1):24–32.
29. Brumme ZL, et al. Molecular and clinical epidemiology of CXCR4-using HIV-1 in a large population of antiretroviral-naive individuals. *J Infect Dis.* 2005;192(3):466–474.
30. Vrisekoop N, et al. Sparse production but preferential incorporation of recently produced naive T cells in the human peripheral pool. *Proc Natl Acad Sci USA.* 2008;105(16):6115–6120.
31. Grau-Expósito J, et al. A Novel Single-Cell FISH-Flow Assay Identifies Effector Memory CD4+ T cells as a Major Niche for HIV-1 Transcription in HIV-Infected Patients. *mBio.* 2017;8(4):e00876-17.
32. Pardons M, et al. Single-cell characterization and quantification of translation-competent viral reservoirs in treated and untreated HIV infection. *PLoS Pathog.* 2019;15(2):e1007619.
33. Baxter AE, et al. Single-Cell Characterization of Viral Translation-Competent Reservoirs in HIV-Infected Individuals. *Cell Host Microbe.* 2016;20(3):368–380.
34. Soriano-Sarabia N, et al. Quantitation of replication-competent HIV-1 in populations of resting CD4+ T cells. *J Virol.* 2014;88(24):14070–14077.
35. Besson GJ, et al. HIV-1 DNA decay dynamics in blood during more than a decade of suppressive antiretroviral therapy. *Clin Infect Dis.* 2014;59(9):1312–1321.
36. Mexas AM, et al. Concurrent measures of total and integrated HIV DNA monitor reservoirs and ongoing replication in eradication trials. *AIDS.* 2012;26(18):2295–2306.
37. Pinzone MR, et al. Monitoring Integration over Time Supports a Role for Cytotoxic T Lymphocytes and Ongoing Replication as Determinants of Reservoir Size. *J Virol.* 2016;90(23):10436–10445.
38. Murray JM, Kelleher AD, Cooper DA. Timing of the components of the HIV life cycle in productively infected CD4+ T cells in a population of HIV-infected individuals. *J Virol.* 2011;85(20):10798–10805.
39. Liu J, Roederer M. Differential susceptibility of leukocyte subsets to cytotoxic T cell killing: implications for HIV immunopathogenesis. *Cytometry A.* 2007;71(2):94–104.
40. Jensen MA, 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):13376–13388.
41. Lengauer T, Sander O, Sierra S, Thielen A, Kaiser R. Bioinformatics prediction of HIV coreceptor usage. *Nat Biotechnol.*

- 2007;25(12):1407–1410.
42. Delgado E, et al. Evaluation of genotypic tropism prediction tests compared with in vitro co-receptor usage in HIV-1 primary isolates of diverse subtypes. *J Antimicrob Chemother.* 2012;67(1):25–31.
  43. Mulinge M, et al. HIV-1 tropism determination using a phenotypic Env recombinant viral assay highlights overestimation of CXCR4-usage by genotypic prediction algorithms for CRF01\_AE and CRF02\_AG. *PLoS One.* 2013;8(5):e60566.
  44. Toma J, Whitcomb JM, Petropoulos CJ, Huang W. Dual-tropic HIV type 1 isolates vary dramatically in their utilization of CCR5 and CXCR4 coreceptors. *AIDS.* 2010;24(14):2181–2186.
  45. Loftin LM, et al. Constrained use of CCR5 on CD4+ lymphocytes by R5X4 HIV-1: efficiency of Env-CCR5 interactions and low CCR5 expression determine a range of restricted CCR5-mediated entry. *Virology.* 2010;402(1):135–148.
  46. Svicher V, et al. HIV-1 dual/mixed tropic isolates show different genetic and phenotypic characteristics and response to maraviroc in vitro. *Antiviral Res.* 2011;90(1):42–53.
  47. Morris SR, et al. Dual-mixed HIV-1 coreceptor tropism and HIV-associated neurocognitive deficits. *J Neurovirol.* 2013;19(5):488–494.
  48. Huang W, et al. Coreceptor tropism in human immunodeficiency virus type 1 subtype D: high prevalence of CXCR4 tropism and heterogeneous composition of viral populations. *J Virol.* 2007;81(15):7885–7893.
  49. Pastore C, et al. Conserved changes in envelope function during human immunodeficiency virus type 1 coreceptor switching. *J Virol.* 2007;81(15):8165–8179.
  50. Wagner TA, McKernan JL, Tobin NH, Tapia KA, Mullins JI, Frenkel LM. An increasing proportion of monotypic HIV-1 DNA sequences during antiretroviral treatment suggests proliferation of HIV-infected cells. *J Virol.* 2013;87(3):1770–1778.
  51. De Boer RJ, Perelson AS. Quantifying T lymphocyte turnover. *J Theor Biol.* 2013;327:45–87.
  52. Vrsekoop N, et al. Quantification of naive and memory T-cell turnover during HIV-1 infection. *AIDS.* 2015;29(16):2071–2080.
  53. Pace MJ, et al. Directly infected resting CD4+T cells can produce HIV Gag without spreading infection in a model of HIV latency. *PLoS Pathog.* 2012;8(7):e1002818.
  54. Zerbato JM, et al. Establishment and Reversal of HIV-1 Latency in Naive and Central Memory CD4+ T Cells In Vitro. *J Virol.* 2016;90(18):8059–8073.
  55. Grau-Expósito J, et al. Latency reversal agents affect differently the latent reservoir present in distinct CD4+ T subpopulations. *PLoS Pathog.* 2019;15(8):e1007991.
  56. Cesana D, et al. HIV-1-mediated insertional activation of STAT5B and BACH2 trigger viral reservoir in T regulatory cells. *Nat Commun.* 2017;8(1):498.
  57. Pace MJ, Agosto L, 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):644–648.
  58. Wilen CB, et al. Phenotypic and immunologic comparison of clade B transmitted/founder and chronic HIV-1 envelope glycoproteins. *J Virol.* 2011;85(17):8514–8527.
  59. Bleul CC, Wu L, Hoxie JA, Springer TA, Mackay CR. The HIV coreceptors CXCR4 and CCR5 are differentially expressed and regulated on human T lymphocytes. *Proc Natl Acad Sci USA.* 1997;94(5):1925–1930.
  60. Tabler CO, 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):4976–4986.
  61. Mitchell CJ, et al. A multi-omic analysis of human naive CD4+ T cells. *BMC Syst Biol.* 2015;9:75.
  62. Fisher K, et al. Intact proviruses from naive and effector memory T cells match persistent viremia. Conference on Retroviruses and Opportunistic Infections. <https://www.croiconference.org/abstract/intact-proviruses-from-naive-and-effector-memory-t-cells-match-persistent-viremia/>. Accessed September 16, 2020.
  63. Gerberick A, Sluis-Cremer N, Piazza P, Delucia D, Rappocciolo G. B cells mediate R5-Tropic HIV infection of CCR5neg naive CD4 T cells. *J Virus Erad.* 2019;5(Suppl\_3):1–58.
  64. Mclean AR, Michie CA. In vivo estimates of division and death rates of human T lymphocytes. *Proc Natl Acad Sci USA.* 1995;92(9):3707–3711.
  65. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002;30(14):3059–3066.
  66. Larbi A, Fulop T. From “truly naïve” to “exhausted senescent” T cells: when markers predict functionality. *Cytometry A.* 2014;85(1):25–35.