Distinct gene expression by expanded clones of quiescent memory CD4$^+$ T cells harboring intact latent HIV-1 proviruses

Georg H J Weymar[1], Yotam Bar-On[2], Thiago Y. Oliveira[1], Christian Gaebler[1], Victor Ramos[1], Harald Hartweger[1], Gaëlle Breton[1], Marina Caskey[1], Lillian B Cohn[3], Mila Jankovic[1], and Michel C Nussenzweig[1, 4]

[1]Laboratory of Molecular Immunology, The Rockefeller University, New York, NY 10065, USA.

[2]Technion - Israel Institute of Technology, Haifa 32000, Israel

[3]Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA

[4]Howard Hughes Medical Institute

Address correspondence to: Michel C. Nussenzweig: nussen@rockefeller.edu

Summary

Antiretroviral therapy controls but does not cure HIV-1 infection due to a reservoir of rare $CD4^+$ T cells harboring latent proviruses. To date little is known about the transcriptional program of latent cells. Here we report on a novel strategy to enrich clones of latent cells carrying intact HIV-1 proviruses directly from blood based on their expression of unique T cell receptors. This enabled single cell transcriptomic analysis of 1,050 $CD4^+$ T cells belonging to expanded clones that harbor intact HIV-1 proviruses from 6 different individuals. The analysis revealed that most of these cells share transcriptional signatures with cytotoxic $CD4^+$ T cells, but do not express granzymes B and H, perforin 1, NKG7, and granulysin that are required for cytotoxicity. We conclude that clones of latent cells carrying intact HIV-1 proviruses persist preferentially in a distinct $CD4^+$ T cell population opening new possibilities for eradication.

Introduction

Antiretroviral therapy prevents HIV-1 viral replication but does not impact latent proviruses that are integrated into the genome of host $CD4^+$ T cells. The reservoir of latent proviruses is responsible for rapid rebound viremia in most individuals undergoing treatment interruption and is the key impediment to HIV-1 cure [1-9].

Although the precise composition of the latent compartment is not known, the relative representation of expanded clones of $CD4^+$ T cells harboring intact and defective latent proviruses increases over time such that they account for at least 50 % of the reservoir in chronically infected individuals [10-17]. Members of infected clones express the same unique T cell receptor (TCR) and have a single distinctive proviral integration site each of which can serve as a

molecular identifier for the latent clone [15,18-21]. Hypotheses about how the latent reservoir is maintained include proviral integration sites that enable cell division [22,23], homeostatic [24] and antigen driven proliferation [20,25-27].

Because cells harboring intact latent proviruses are rare [5-7,28,29], and have no well-defined markers that distinguish them from other $CD4^+$ T cells [30,31], characterizing their transcriptional program has not been possible to date. Intact proviruses are enriched among $CD4^+CD45RA^-$ $HLA\text{-}DR^+$ memory T cells [24,32-36], but other surface markers, such as CD2 [37] remain controversial. Latent cells can be identified after re-activation of HIV-1 transcription *in vitro* [18,38-42], and there are numerous cell line or tissue culture-based models of latency [37,43-47], but how well these experimental conditions and latency models reflect the physiology of latent cells in circulation is not known.

Here we present a strategy to enrich clones of quiescent latent cells from samples that were assayed directly *ex vivo* from 6 individuals living with HIV-1 based on cell surface expression of their unique TCRs. The enrichment strategy enabled analysis of the transcriptional landscape of these rare cells and identification of distinct features of this population.

Results

During development T and B lymphocytes assemble unique cell surface receptors by variable diversity and joining gene (V(D)J) recombination. This process is under feedback regulation by the receptor such that each lymphocyte expresses a single specificity, a phenomenon referred to as allelic exclusion [48]. Because antigen receptors are fixed early in development, naïve T cells

that become activated and expand produce clones of $CD4^+$ T cells that are defined by expression of a singular T cell receptor (TCR). Thus, the TCR expressed by a clone of latent cells is a unique molecular identifier for members of that clone, and because it is a cell surface protein, the TCR can be used to enrich members of the latent clone.

People living with HIV-1 frequently harbor large, expanded clones of latent $CD4^+$ T cells. We studied 6 chronically infected individuals controlled on anti-retroviral therapy that were aviremic at the time of sample collection (Table S1). Near-full length sequencing and/or viral outgrowth assays showed that the latent reservoir of each of these individuals was dominated by a single expanded intact latent proviral clone (Table S2, Figure S1) [11,18,21,26,49-51]. In all cases the members of these clones could be identified by the sequence of their HIV envelope gene (*env)* (Figure S1). The frequency of the clonally expanded latent provirus of interest in the 6 individuals ranged from 13 - $431/10^6$ total $CD4^+$ T cells based on the frequency of the specific HIV-1 *env* sequence (Table S2).

Latent cells are predominantly found in the $CD45RA^-$ memory T cell compartment [24,52]. To determine whether the $CD4^+$ T cells harboring latent proviral clones of interest are in the memory compartment we purified $CD45RA^+$ and $CD45RA^-$ cells and analysed proviral DNA by sequencing the *env* gene and producing maximum likelihood phylogenetic trees. As expected, in all cases the latent clone of interest was predominantly found in the $CD45RA^-$ population (Figure 1A and Figure S2A). Based on staining with an anti-CD45RA antibody and flowcytometric analysis of $CD4^+$ T cells from four individuals, $CD45RA^-$ cells accounted for 42 - 65 % of all

$CD4^+$ T cells and therefore purification of $CD45RA^-$ cells results in a 1.5 - 2.4-fold enrichment of the latent clone (Figure S2B).

The TCRβ locus encodes 2 different constant region genes (TRBC1 and TRBC2). Allelic exclusion ensures that all members of a $CD4^+$ T cell clone express the same TCRβ constant region. To determine which of the 2 different TCRβ constant regions is expressed by each of the latent clones of interest we performed flow cytometry experiments to purify $TRBC1^+$ and $TRBC1^-$ $CD4^+$ T cells. In each case, *env* sequencing revealed that the latent provirus was found among $CD4^+$ T cells expressing one of the two TRBC domains resulting in 1.5 - 4.5-fold enrichment (Figure 1B, and C, and Figure S3).

The TCRβ locus contains 48 functional variable domains (TRBV). To determine whether the TRBV can be used to enrich quiescent clones of latent cells, we made use of a collection of 24 different anti-TRBV monoclonal antibodies. The antibodies were divided into 8 groups of 3 each that were conjugated with either phycoerythrin (PE), or fluorescein isothiocyanate (FITC), or both (PEFITC). The clones of $CD4^+$ T cells harboring the latent provirus of interest in individuals 603, 605 and B207 express TRBV-19, TRBV11-2 and TRBV7-8 [18]. To enrich the latent clone in individual 603, $CD4^+$ T cells were magnetically sorted for $CD45RA^-$ cells and then stained with antibodies to TRBC1 and TRBV19. The latent clone was found in the $CD45RA^-TRBC1^+TRBV19^+$ population resulting in an overall 47-fold relative enrichment (Figure 2A, Table 1). The latent clone in individual 605 was enriched by 40-fold in the $CD45RA^-$ $TRBV11\text{-}2^+$ population (Figure 2B, Table 1, Figure S3B). Antibodies to TRBV7-8 expressed by the latent clone in individual B207 were not available. Therefore, B207 $CD4^+$ T cells were

stained with the 24 anti-TRBV antibodies and sorted into a CD45RA$^-$TRBC1$^+$TRBV$^-$ population resulting in only a 9-fold overall enrichment (Figure 2C, Table 1). Thus, when anti-CD45RA, -TRBC, and -TRBV enrichment is combined the latent provirus of interest is found in 20, 163, and 6 per 10^4 CD4$^+$ T cells in 603, 605 and B207 respectively (Table 2).

The TCRs expressed by CD4$^+$ T cells harboring the latent HIV-1 proviral clones in individuals 5104, 5125, and 9247 were not known. To identify the TRBV expressed by CD4$^+$ T cells that harbor the latent clone in these individuals we combined limiting dilution cell sorting with *env* sequencing. As a first screening step, CD4$^+$ T cells were stained with the 24 anti-TRBV antibodies and sorted into TRBV$^+$ and TRBV$^-$ populations. The latent clone in individual 5104 was present in the TRBV$^-$ population, yielding a 2-fold enrichment (Figure 3A). For individuals 5125 and 9247, the latent clone was found in the TRBV$^+$ population (Figure 3B and C). To identify the precise TRBV expressed by the latent clone in individual 9247, the 24 anti-TRBV antibodies were split into 2 groups of 12 antibodies and CD4$^+$ T cells were stained with either one of the 2 groups (Figure 3D). The latent clone was found among one of the 2 groups and the 12 antibodies split again into 4 groups of 3 anti-TRBV antibodies, each group containing one FITC-, one PE-, and one PEFITC-labelled antibody (Figure 3E). The latent clone was only found in one of the 4 groups, and the 3 anti-TRBV antibodies in that group were then used to stain and sort for each of the anti-TRBV individually (Figure 3F). These experiments revealed that in individual 9247, the latent clone expresses TRBV4-3 (Figure 3F). For individual 5125, the 24 TRBV antibodies were split into 8 groups of 3 and the latent clone was found in only one of the groups (Figure 3G). The 3 antibodies in that group were then used to purify the individual TRBV

expressing cells, which showed that the latent clone of interest was found among TRBV2$^+$ cells (Figure 3H).

To determine the amount of enrichment that could be achieved for 5104, 5125, and 9247 we combined anti-CD45RA, -TRBC, and -TRBV enrichment and performed limiting dilution *env* amplification and sequencing on genomic DNA (Figure 4 and Table 1). After combined enrichment the latent provirus of interest is found in 14, 15 and 18 per 10^4 CD4$^+$ T cells in 5104, 5125, and 9247 respectively (Table 2).

Although we were able to identify the TRBV expressed by expanded clones of interest in individuals 5104, and 5125, and 9247, the precise TCRαβ sequence remained unknown. Candidate TCRs were initially identified among clones of CD4$^+$ T cells in the enriched populations by single cell TCR sequencing by 10x Genomics (Figure 5A). To definitively determine the TCR expressed by the clone of CD4$^+$ T cells harboring the latent provirus we combined anti-CD45RA, -TRBC, and -TRBV staining and sorted 5 candidate cells into multi-well plates, followed by DNA and RNA extraction. *Env* amplification and sequencing from genomic DNA was used to identify wells containing latent cells. cDNA from wells containing the latent provirus or negative controls was used to amplify and sequence TCRα and β chains (Figure 5A). CD4$^+$ T cells from individuals 603 and 605 that express known TCRs were used as positive controls to validate the method. For individual 603, the TCR expressed by the latent clone was found in 63 % of Env$^+$ wells and only in 4 % of Env$^-$ wells (Figure 5B). In 605, the specific TCR was found in 90 % of the Env$^+$ wells and only in 6 % of Env$^-$ wells. In both cases there was no enrichment of irrelevant TCRs in Env$^+$ wells. All 7 Env$^+$ wells obtained from

individual 5104 contained TCRα TRAV12-3/J9 and/or TCRβ TRBV5-4/J1-1 representing a unique TCR clone in the 10x Genomics sequencing data (Figure 5C). This TCR was absent in the random selection of Env^- wells. Similarly, all 4 Env^+ wells from individual 5125 contained TRAV26-2/J32 and/or TRBV2/J1-1, that was not found among Env^- wells. Finally, in individual 9247 TRAV38-1/J33 and/or TRBV4-3/J2-3, was present in 8 out of 9 of Env^+ wells but absent in the random selection of Env^- wells. We conclude that the latent clone of interest in 5104, 5125 and 9247 express TRAV12-3/J9/TRBV5-4/J1-1, TRAV26-2/J32/TRBV2/J1-1, and TRAV38-1/J33/TRBV4-3/J2-3 respectively.

HIV-1 proviruses can integrate into $CD4^+$ T cells undergoing clonal expansion at the time they start dividing or sometime thereafter. Proviral integration in early stages of clonal expansion would yield a homogenous group of cells where the vast majority of which would harbor latent HIV-1. Integration at a later time would produce a heterogeneous $CD4^+$ T cell clone wherein only some of the cells harbor the HIV-1 provirus [20]. To estimate the fraction of infected cells within a particular clone based on its representation by TCR, we performed 10x Genomics single cell TCR sequencing on samples enriched using the antibody methods described above. We compared the TCR frequencies to the relative frequency of the specific *env* from proviral DNA in similarly enriched samples (Table 2). For individuals 5104, 5125, and 9247, the frequency of the specific provirus was similar to the frequency of corresponding TCR, but in B207, 603, and 605, the number of cells expressing the specific TCR of interest was 2.3 to 3.1 times higher than the frequency of proviral copies (Table 2). Thus, there is heterogeneity among clones of expanded $CD4^+$ T cells that harbor latent HIV-1 proviruses. In half of our samples the vast

majority of the clonally expanded $CD4^+$ T cells harbor latent HIV-1, and in the others the provirus is found in a fraction of the clone.

To determine whether $CD4^+$ T cell clones harboring latent proviruses share a transcriptional profile we combined the 10x Genomics mRNA and TCR sequencing data obtained from the enriched populations of latent cells. TCR sequencing data was used to identify $CD4^+$ T cell clones harboring the latent HIV-1 provirus but was omitted for gene expression analysis to prevent TCR-biased clustering. Uniform manifold approximation and projection (UMAP) analysis of the transcriptional profile of all the 109,217 cells from the six individuals produced 15 unique clusters of $CD4^+$ memory T cells (Figure 6A). Visual inspection revealed that $CD4^+$ T cells expressing the specific TCR associated with latent proviruses are found predominantly in gene expression cluster 7 and neighboring clusters (Figure 6A). On average 57 % of all cells expressing the TCR associated with the expanded latent clone in the 6 individuals were found in cluster 7. Except for individual 9247, the fraction of $CD4^+$ T cells belonging to the latent clone in cluster 7 was greater than all other clusters ranging from 48 - 73 % (Figure 6B, Table S3). In individual 9247, the largest fraction of latent cells was in cluster 6 (35 %), which is closely related to cluster 7, and the second largest fraction of latent cells was in cluster 7 (29 %).

To determine whether residence in cluster 7 is a general property of expanded clones of memory $CD4^+$ T cells, we determined the position of the cells in the next largest and/or smallest clones of $CD4^+$ memory T cells in the UMAP (Figure 6A). None of the neighboring $CD4^+$ T cell clones examined were predominantly found in gene expression cluster 7. In addition, some TCR clones clustered by gene expression whereas others did not. For example, in individual B207, cells in

the next largest clone to the one containing latent proviruses were found in 8 of the 15 clusters. In contrast, cells in the next smaller clone in B207 were found primarily in gene expression cluster 10 (63 %) which contains cells expressing Foxp3 (Figure 6B, Table S3). We conclude that clonal expansion per se is not sufficient for $CD4^+$ memory T cell accumulation in gene expression cluster 7.

Examination of the top 120 genes that define the UMAP clusters revealed that cluster 7 and 8 are closely related (Figure 7A). Cluster 7 is enriched in genes that encode antigen-presenting molecules or their chaperones such as HLA-DR and HLA-DP, as well as CD74, the invariant chain for major histocompatibility class II (MHCII) molecules [53] (Figure 7, and Table S4). In addition, cluster 7 is distinguished by expression of CCL5, Granzymes A and K (GZMA, GZMK), cystatin F (CST7), and the nuclear proteins LYAR [54,55] and DUSP2 (PAC-1) [56,57]. Differential expression of some of these genes has been reported in activated latent $CD4^+$ T cells [18,35,36]. However, Granzyme A, and K proteins do not accumulate specifically in latent cells because flow cytometry-based cell sorting for these markers did not enrich latent cells.

Cluster 7 shares many upregulated genes with cluster 8, namely CCL5, CST7, and GZMA (Figure 7A). The two closely related clusters differ in that cluster 8 cells also express GZMB, GZMH, PFR1 (perforin 1), NKG7, and GNLY each of which is associated with cytotoxic $CD4^+$ T cells [58,59,]. This population of cells is frequently expanded in infection and chronic inflammation [60] and is also found enriched among tumour infiltrating lymphocytes [61]. Moreover, the relative proportion of cytotoxic $CD4^+$ T cells among memory cells is expanded in HIV-1 infected individuals including those on suppressive ART [62,63].

To examine the relationship between the cells in UMAP clusters and genes associated with $CD4^+$ T cell identity we performed unsupervised clustering analysis with a collection of genes that are up or downregulated in these cells [59,60,62-65] (Figure 7B). In agreement with the UMAP cluster 8 stands out as most closely related to $CD4^+$ cytotoxic cells and cluster 7 is its closest relative. In conclusion, gene expression cluster 7 shares some of the features of cytotoxic $CD4^+$ T cells but appears to differ from this subset in the expression in key components associated with the cytotoxic machinery. Finally, gene set enrichment analysis based on genes differentially expressed by cells in cluster 7 (Table S4) shows high enrichment in pathways associated with chronic inflammation which is associated with development of cells with the cytotoxic $CD4^+$ T cell phenotype [60,65] (Figure S4).

Discussion

$CD4^+$ T cells harboring intact latent HIV-1 proviruses represent only a tiny fraction of all $CD4^+$ T cells. They are usually found among subsets of the memory cells [24,52] but there are no specific markers that facilitate the purification of these cells [30,31]. Consequently, their transcriptional program has only been studied after activation of HIV-1 expression which thereby allows their identification [18,38-41]. We have devised a novel method to enrich quiescent latent memory $CD4^+$ T cells by means of their specific antigen receptors. We find that expanded clones of memory $CD4^+$ T cells that carry intact integrated HIV-1 proviruses are enriched among cells that express a transcriptional program that is closely related to $CD4^+$ cytotoxic T cells.

$CD4^+$ cytotoxic T cells develop in response to chronic antigen burden and inflammation. They are prominently expanded in HIV-1, CMV, EBV infection, chronic inflammatory diseases and in virally induced malignancies [60-63,66]. Their polyfunctional phenotype is most closely associated with the Th1 phenotype, but they can also develop from other T cell subsets. The mechanism that regulates their development is not entirely defined but is associated with expression of T-bet, eomesodermin, Runx3, and Blimp1, and downregulation of ThPOK, Bcl6 and TCF1 [60,64,65]. In keeping with the finding that clones of $CD4^+$ T cells harboring intact latent proviruses cluster in close proximity to $CD4^+$ cytotoxic cells, these cells respond to antigens found in chronic viral infections such as HIV-1, CMV, and EBV [20,26,27,46].

Nevertheless, members of a T cell clone expressing the same TCR can adopt different fates depending on several different factors including affinity, antigen concentration and the cytokine milieu [33,67,68]. Consistent with this idea, $CD4^+$ T cells expressing the TCR associated with latent proviruses are not entirely limited to a single gene expression cluster. Nevertheless, the observation that a large fraction of the cells in the expanded clones we studied can be found in one specific transcriptional cluster among all memory $CD4^+$ T cells stands in contrast to other similar sized expanded clones obtained from the same individuals. The latter are found in several different clusters that diverge between clones and individuals, and many of the cells in these clones are widely dispersed among clusters with different transcriptional signatures. For example, among the 11 random memory $CD4^+$ T cell clones of similar size examined in the 6 individuals only 1 in individual 603 showed enrichment above 5 % in cluster 7.

In chronically infected individuals at least 50 % of the cells carrying intact proviruses belong to expanded clones [11,12], each of which can be distinguished by expression of a specific T cell receptor that is associated with a unique proviral integration site [10,15,18-20,22,23]. The intact reservoir is dynamic and while the absolute number of cells in the reservoir decreases slowly with a half-life of 4 - 5 years, clonality increases with time after infection in people on suppressive ART [4,6,10,13,16,17,33]. However, clonal expansion is not a unique feature of $CD4^+$ T cells harboring intact proviruses. Defective proviruses are also found predominantly in expanded clones [10,17,69], and clones of $CD4^+$ T cells are prominent among non-HIV-1 infected individuals [70].

Some of the genes that help define quiescent $CD4^+$ T cells in cluster 7 have also been reported to be expressed in latent cells, such as HLA-DR [18,33,35], or like CD2 or LYAR upregulated in *in vitro* models of HIV infection and latency [37,71]. In contrast, CCL5 is downregulated upon latent cell reactivation *in vitro* [18,37], but was upregulated in gene expression cluster 7. Notably, these genes had not been linked to a specific transcriptional program that differentiates latent cells from other $CD4^+$ T cells.

Cells expressing the genes that define cluster 7 are not a unique feature of HIV-1 infection and can be found in non-infected individuals [59]. Therefore, HIV-1 proviral integration and latency per se is not required for T cells to acquire the cluster 7 transcriptional program. Why latent proviruses are enriched in cells expressing this particular program is not known. One possibility among many is that the cluster 7 program favors suppression of HIV-1 gene expression during T cell activation which would permit cell division in the absence of HIV-1 virion production and

cell death. An alternative but non-exclusive possibility is that these cells are among the most likely to respond to a chronic infection, and therefore the most likely to be infected and become latent and subsequently undergo clonal expansion in response to a persistent antigen.

Our analysis is limited to large, expanded clones of latent cells in 6 individuals. Whether these observations also apply to less expanded populations of latent cells remains to be determined. Moreover, the depth of sequencing available on the 10x platform is also limiting and therefore additional elements of the cluster 7 transcriptional program remain to be defined. Despite these caveats, the observation that latent cells preferentially display a specific transcriptional program suggests that these cells could be specifically targeted for elimination.

Tables

	CD45RA	TRBC	TRBV	combined
603	1.5	2.2	14.3	47.2
605	~2 (estimate)	-	20	40
B207	1.6	2	2.9	9.3
5104	2.4	1.6	2.9	11.1
5125	~2 (estimate)	1.8	20	72
9247	1.9	5	71	674.5

Table 1: Relative enrichment per marker based on flow cytometry.

	Latent clone Env copies per 10^4 $CD4^+$ T cells	Specific TCR per 10^4 $CD4^+$ T cells	Ratio TCR/Env
603	20	45	2.3
605	163	502	3.1
B207	6	14	2.3
5104	14	16	1.1
5125	15	17	1.1
9247	18	23	1.3

Table 2: Comparison between *env* copies per 10^4 $CD4^+$ T cells (based on *env* PCR) and the frequency of the latent clone TCR (based on 10x Genomics TCR sequencing) after enrichment based on CD45RA, TRBC, and TRBV.

Figures

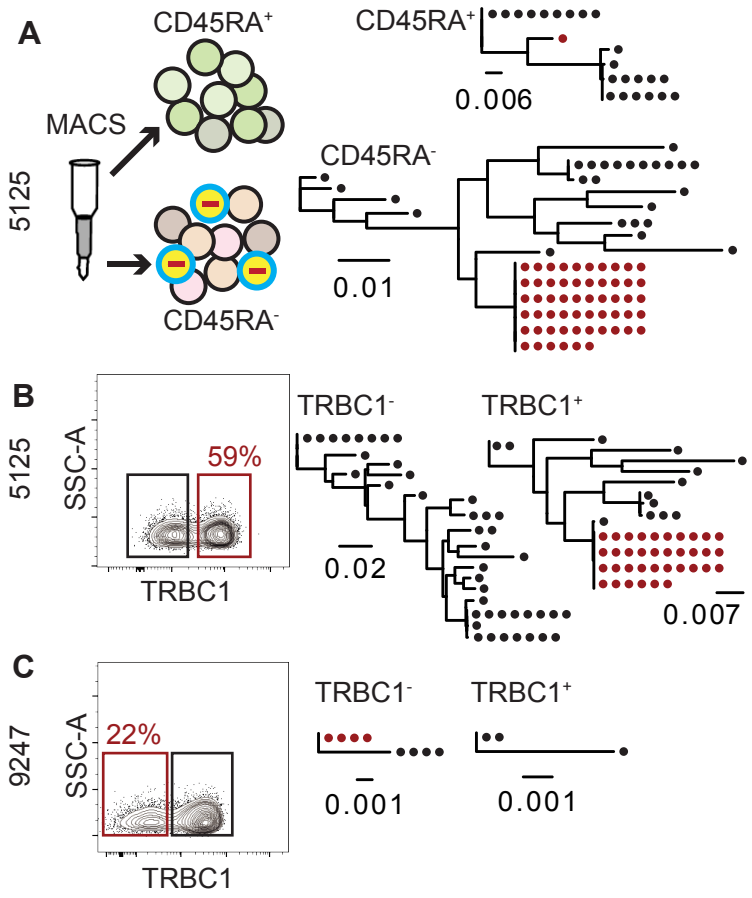


Figure 1. Screening for intact latently infected cell enrichment by sorting for CD45RA and TRBC. Red outlines indicate the population containing the clone of interest. Maximum-likelihood phylogenetic trees show *env* gene of the latent clone of interest marked in red. Each dot represents a recovered *env* sequence from the respective subpopulation of $CD4^+$ T cells. The scale bars indicate the number of substitutions per site.

(A) $CD4^+$ T cells from individual **5125** were magnetically sorted into a $CD45RA^+$ and $CD45RA^-$ population. Subsequent gDNA extraction, limiting dilution, and *env* sequencing revealed that the infected clone of interest in was enriched in the $CD45RA^-$ memory compartment.

(B) $CD4^+$ T cells from individual **5125** were stained with anti-TRBC1 antibody and sorted into $TRBC1^+$ and $TRBC1^-$ populations. The latent clone was enriched in the $TRBC1^+$ compartment. **(C)** $CD4^+$ T cells from individual **9247** were stained with anti-TRBC1 antibody and sorted into $TRBC1^+$ and $TRBC1^-$ populations. The latent clone was enriched in the $TRBC1^-$ compartment.

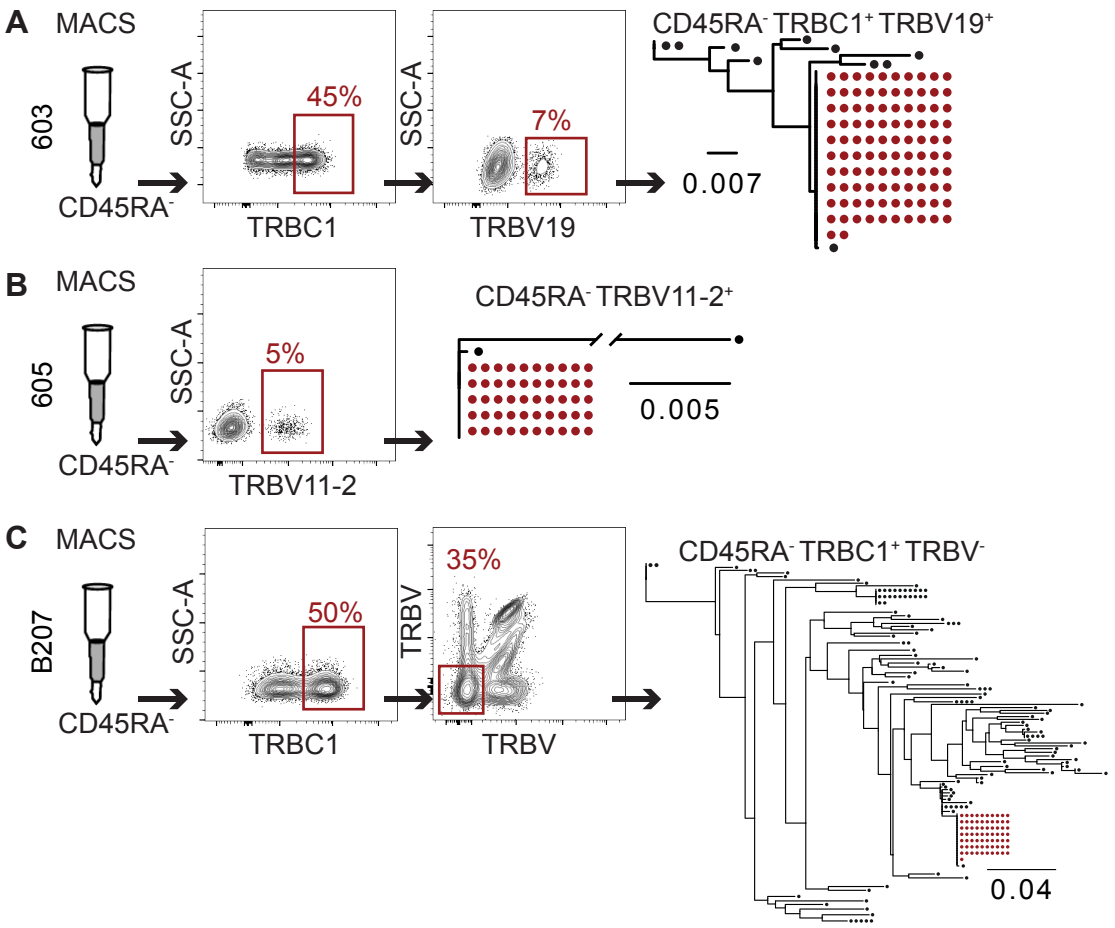
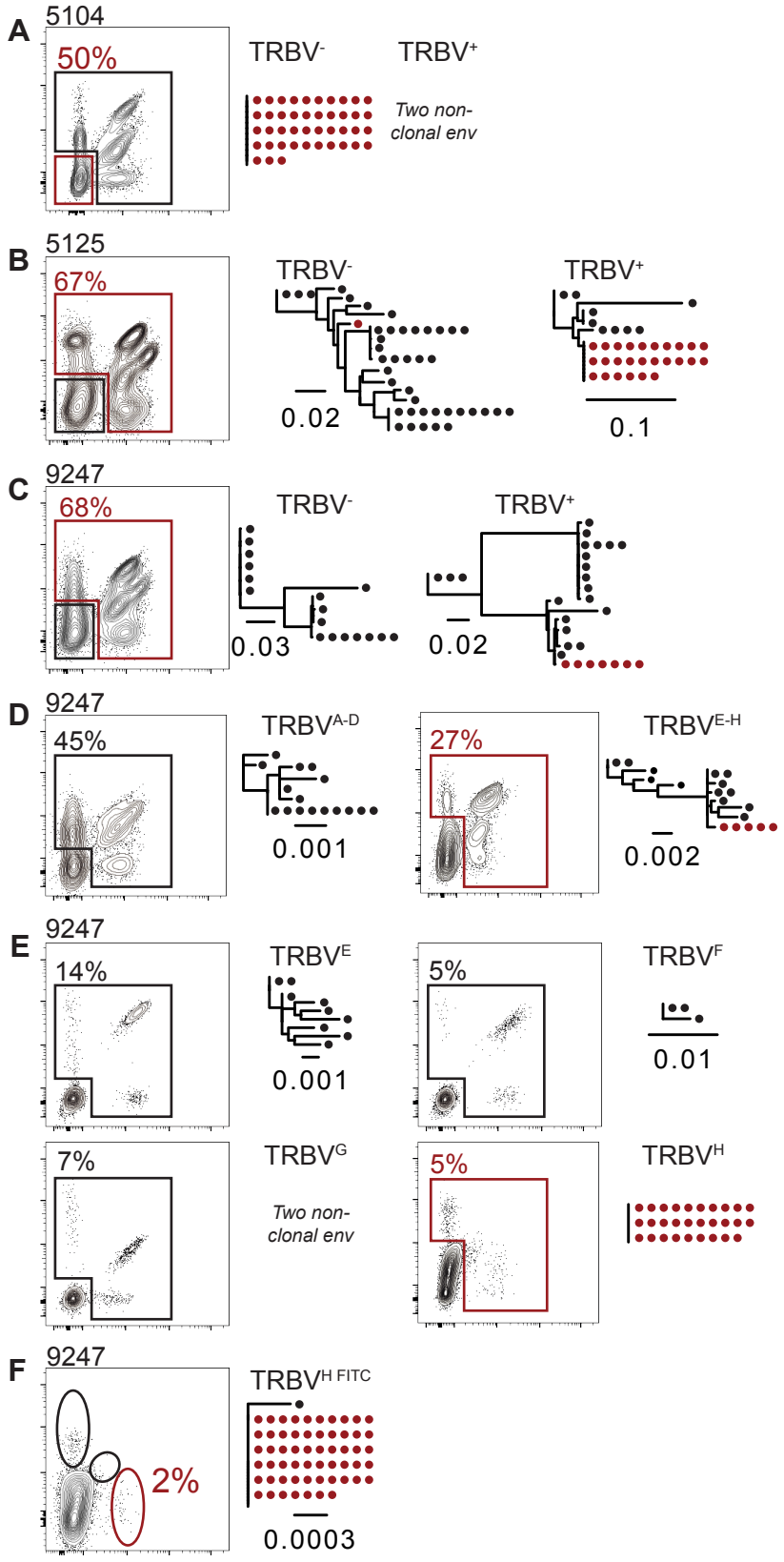


Figure 2. Combined enrichment based on CD45RA, TRBC, and TRBV expression in individuals 603, 605, and B207. Red outlines indicate the population containing the clone of interest. Maximum-likelihood phylogenetic trees of *env* sequences isolated from sorted populations, as indicated. *Env* gene of the clone of interest marked in red. The scale bars indicate the number of substitutions per site.

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

(B) Enrichment in individual **605** after purification CD45RA$^-$ memory cells followed by staining TRBV11-2$^+$ cells, which were > 90 % TRBC1$^+$ (Figure S3B). Flow cytometry plot shows TRBV11-2 staining. Percentage of CD45RA$^-$TRBV11-2$^+$ cells is indicated.

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


A
5104
50%
TRBV-
TRBV+
Two non-clonal env
B
5125
67%
TRBV-
0.02
TRBV+
0.1
C
9247
68%
TRBV-
0.03
TRBV+
0.02
D
9247
45%
TRBVA-D
0.001
27%
TRBVE-H
0.002
E
9247
14%
TRBVE
0.001
5%
TRBVF
0.01
7%
TRBVG
Two non-clonal env
5%
TRBVH
F
9247
2%
TRBVH FITC
0.0003

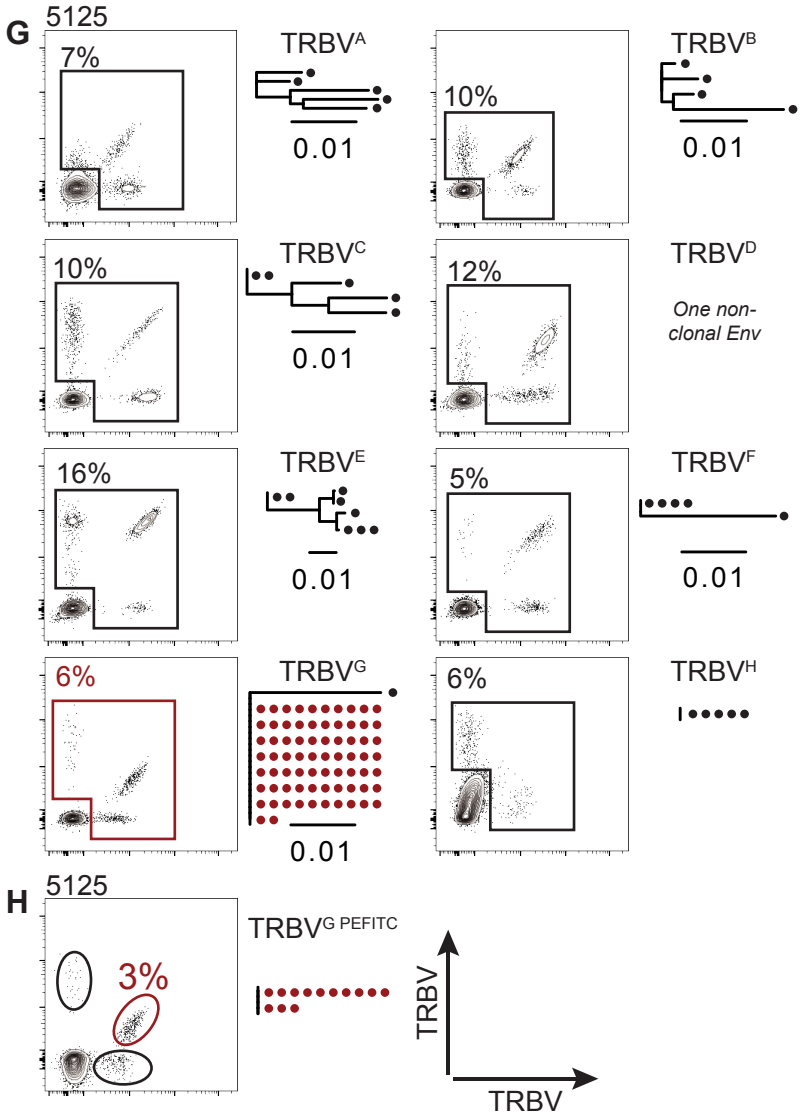


Figure 3. Screening for the TRBV expressed by the latent clone in individuals 5104, 5125 and 9247. Red outlines indicate the population containing the clone of interest. Maximum-likelihood phylogenetic trees of *env* sequences derived from sorted population, as indicated. *Env* gene of the clone of interest marked in red. The scale bars indicate the number of substitutions per site. TRBV^{A-H} indicate the grouping of monoclonal antibodies used in the cocktails (see methods).

(A – C) Flow cytometry plots show TRBV staining using a combination of 24 anti-TRBV antibodies. Percentage of $TRBV^-$ cells is indicated for individual **5104** and $TRBV^+$ cells for **9247** and **5125**.

(D) Flow cytometry plots show TRBV staining with one of 2 groups of 12 antibodies for individual **9247**. Percentages of $TRBV^+$ cells are indicated. The sorted population is indicated. The positive group was further subdivided in (E).

(E) Flow cytometry plots show TRBV staining with one of 4 groups of 3 TRBV antibodies from individual **9247**. Percentages indicate the fraction of $TRBV^+$ cells in each group. The sorted population is indicated. The clone of interest was only found in cells stained with the $TRBV^H$ cocktail indicated in red and was further subdivided in (F).

(F) Flow cytometry plots show TRBV staining with 3 different monoclonal antibodies from individual **9247**. Percentage indicates the fraction of $TRBV4\text{-}3^+$ cells.

(G) Flow cytometry plots show TRBV staining with antibody cocktails A-H from individual **5125**. The clone of interest was only found in the $TRBV^G$ cocktail which was further subdivided in (H). Percentage indicates the fraction of $TRBV^+$ cells for each cocktail.

(H) Flow cytometry plots show TRBV staining with the 3 different antibodies in the $TRBV^G$ cocktail from individual **5125**. The clone of interest was found in the $TRBV2^+$ population. Percentage indicates the fraction of $TRBV2^+$ cells.

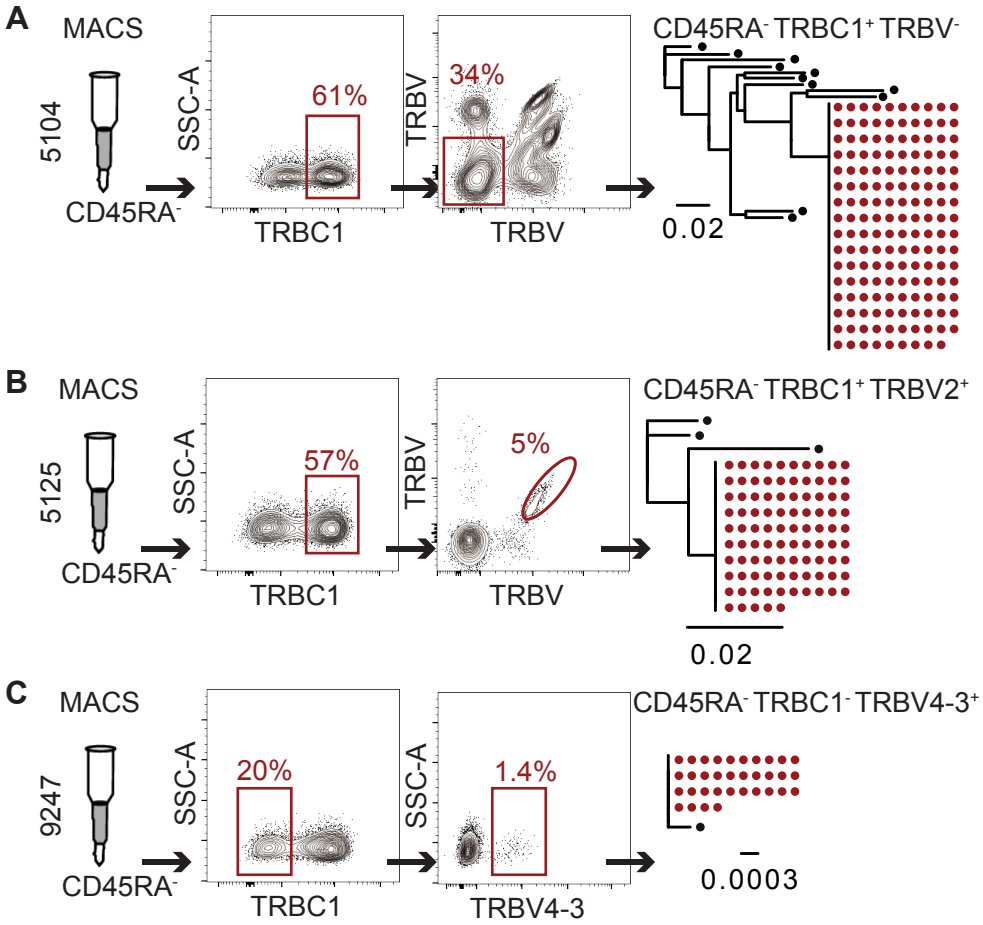


Figure 4. Combined enrichment based on CD45RA, TRBC, and TRBV expression in individuals 5104, 5125, and 9247. Red outlines indicate the population containing the clone of interest. Maximum-likelihood phylogenetic trees of *env* sequences derived from sorted population, as indicated. The scale bars indicate the number of substitutions per site. *Env* gene of the clone of interest marked in red.

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

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

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

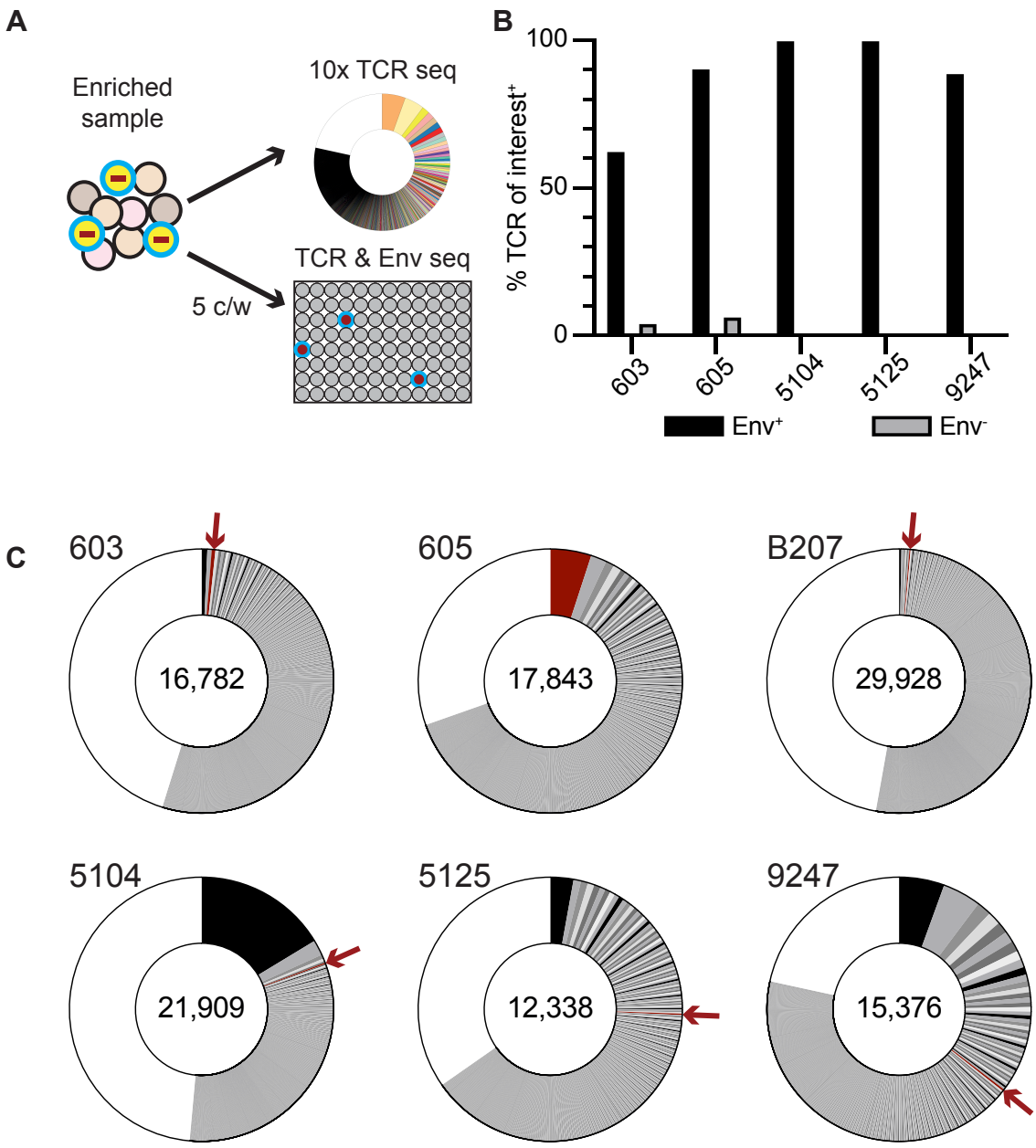


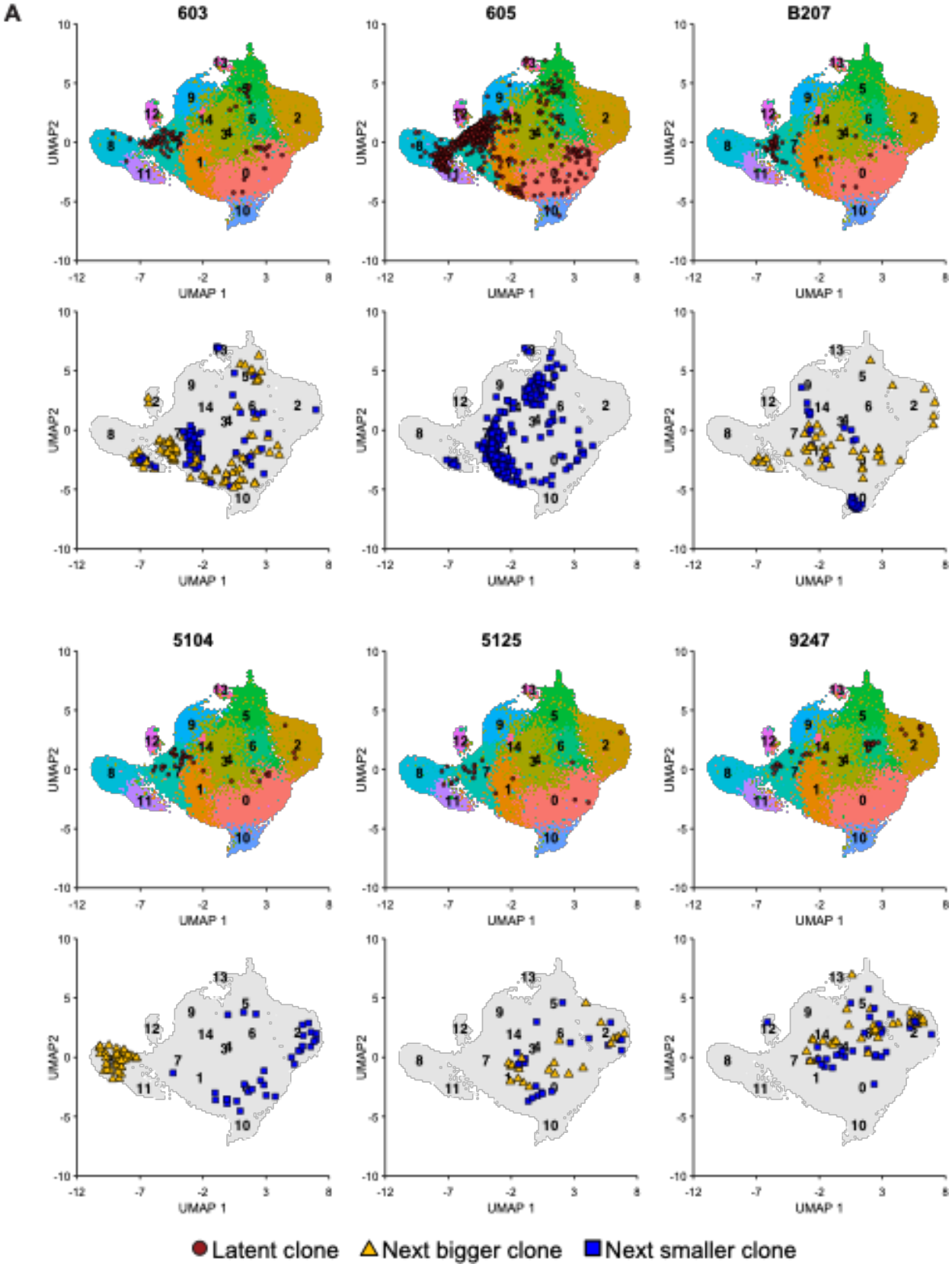
Figure 5. Identification of latent clone TCR sequences.

(A) Limiting dilution sorting strategy to identify the TCR expressed by the CD4^{+} T cells harboring the clone of interest. Each sample was enriched based on CD45RA, TRBC1, TRBV expression, and 5 cells sorted per well (c/w) into microwell plates. *Env* sequencing identified wells that contained a cell of the latent clone of interest. TCRs were sequenced from all *env*$^{+}$ and a selection of *env*$^{-}$ wells.

(B) Bar graph shows the relevant TCR clonotypes identified in *env*$^+$ and *env*$^-$ wells.

603: *env*$^+$ n=8 and *env*$^-$ n=77; **605**: *env*$^+$ n= 21 and *env*$^-$ n=62; **5104**: *env*$^+$ n=7 and *env*$^-$ n=36; **5125**: *env*$^+$ n=4 and *env*$^-$ n=31; **9247**: *env*$^+$ n=9 and *env*$^-$ n=42.

(C) Pie charts show the relative size of TCR clones as slices. The areas indicated in white represent unique TCR sequences. The number on the left above the pie chart is the donor ID for each individual. The number in the center of the pie chart represents the number of cells assayed for each individual. The clone of interest is indicated by a red arrow and pie slice.


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

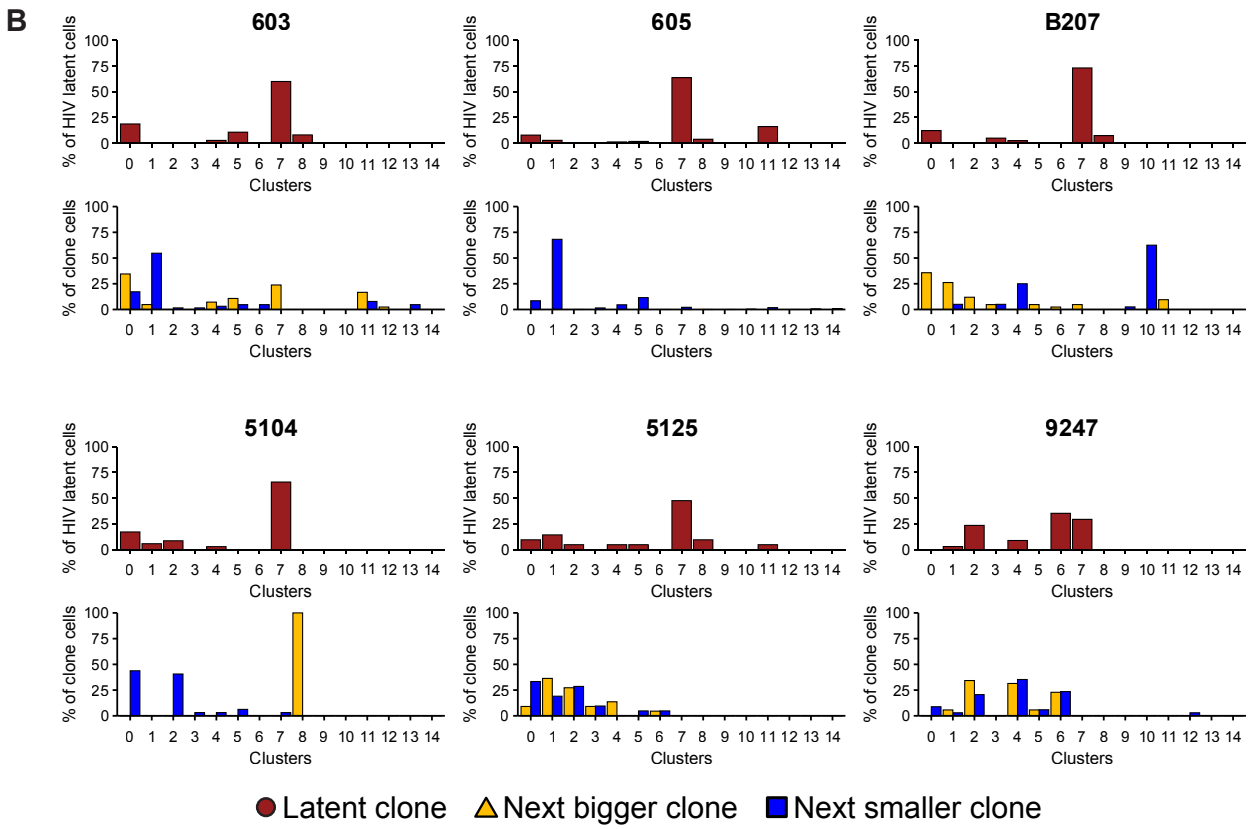


Figure 6. Uniform Manifold Approximation and Projection (UMAP) of 10x gene expression data. Data representing mRNA expression by 109,217 individual cells is shown. The latent clone of interest as well as the next bigger and next smaller clone in size were located in the UMAP by their TCR sequence.

(A) UMAPs show the position of the cells expressing the latent clone TCR for each of the 6 individuals as red dots. Underneath, UMAPs show the position of the next bigger (yellow triangles) and the next smaller (blue squares) clone in size to the clone of interest. For individual **605**, the latent clone of interest was the biggest clone, and only the next smaller clone is shown.

(B) The bar graphs show the fraction of cells in the latent clone (red bars) in each of the 15 UMAP clusters, and the fraction of cells of the next bigger (yellow bars) and the next smaller (blue bars) clone in size to the clone of interest in each of the 15 UMAP clusters (Table S3).

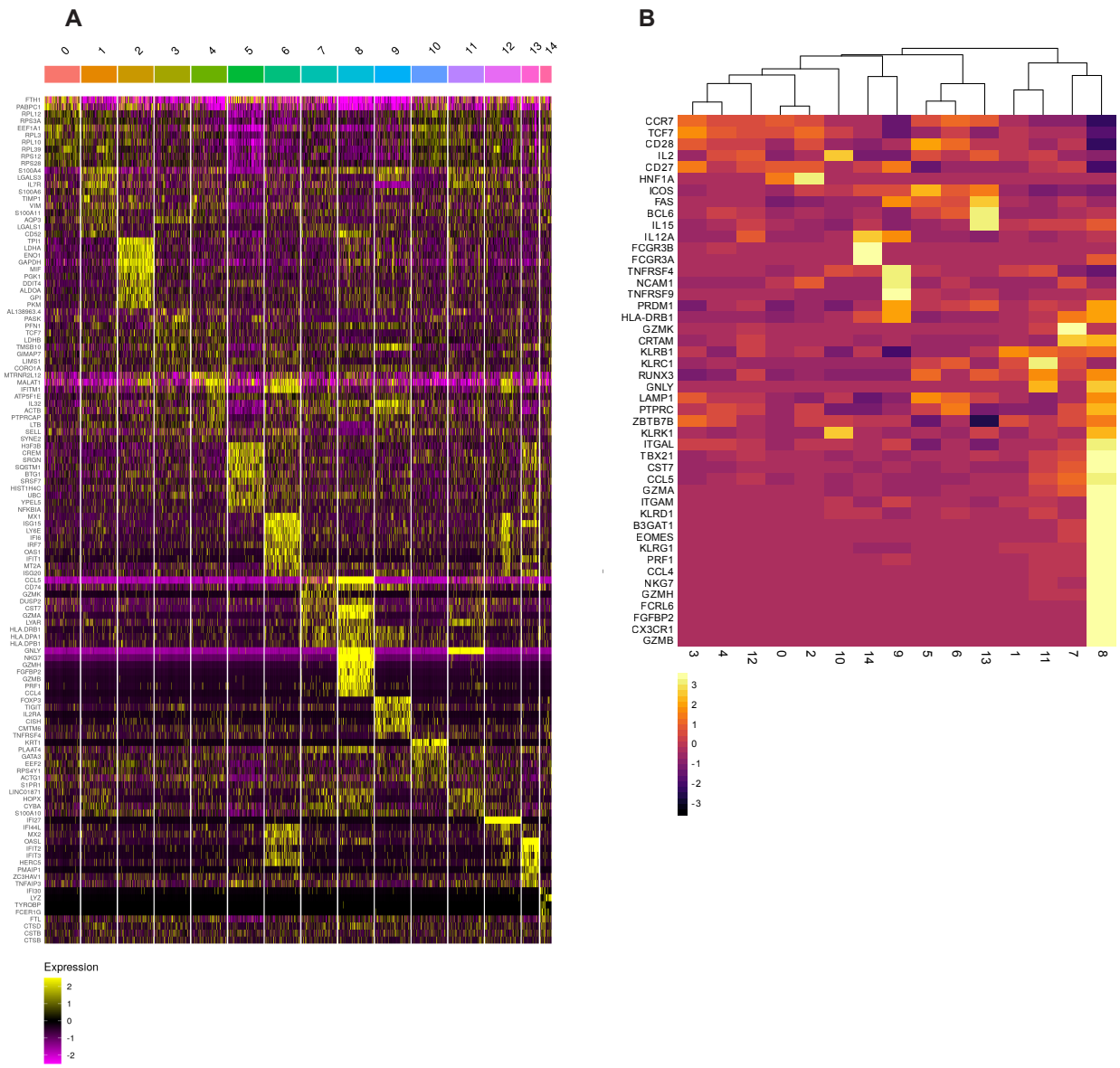


Figure 7 Heat map of cluster-defining, differentially expressed genes

(A) Heat map shows up to 10 of the most up-regulated genes per UMAP cluster compared to all other clusters. Genes are indicated on the left and clusters above. Yellow shows relatively highly expressed genes and purple relatively downregulated genes.

(B) Heat map shows unsupervised clustering based on the expression of differentially expressed genes in cytotoxic CD4+ T cells [59,60,62-65]. Genes are indicated on the left and clusters below.

Yellow shows relatively highly expressed genes and black shows relatively downregulated genes.

Supplementary Figures

Figure S1. The latent reservoir in each individual contains an expanded intact latent clone.

Maximum-likelihood phylogenetic trees show *env* gene of the clone of interest marked in red.

Each dot on the maximum-likelihood phylogenetic trees represents an *env* sequence that was recovered by *env* PCR (individuals **B207**, **603**, and **605**) or Q4PCR (individuals **5104**, **5125**, and **9247** [50,51]) from $CD4^+$ T cells before enrichment. The *env* gene that identifies the clone of interest is marked red. The scale bars indicate the number of substitutions per site. For individuals **B207**, **603**, and **605**, the *env* gene was also found in viral outgrowth assays [18]. For individuals **5104**, **5125**, and **9247**, an expanded intact clone was identified by Q4PCR [49]. All *env* sequences were extracted from near-full-length proviral sequences and aligned in trees.

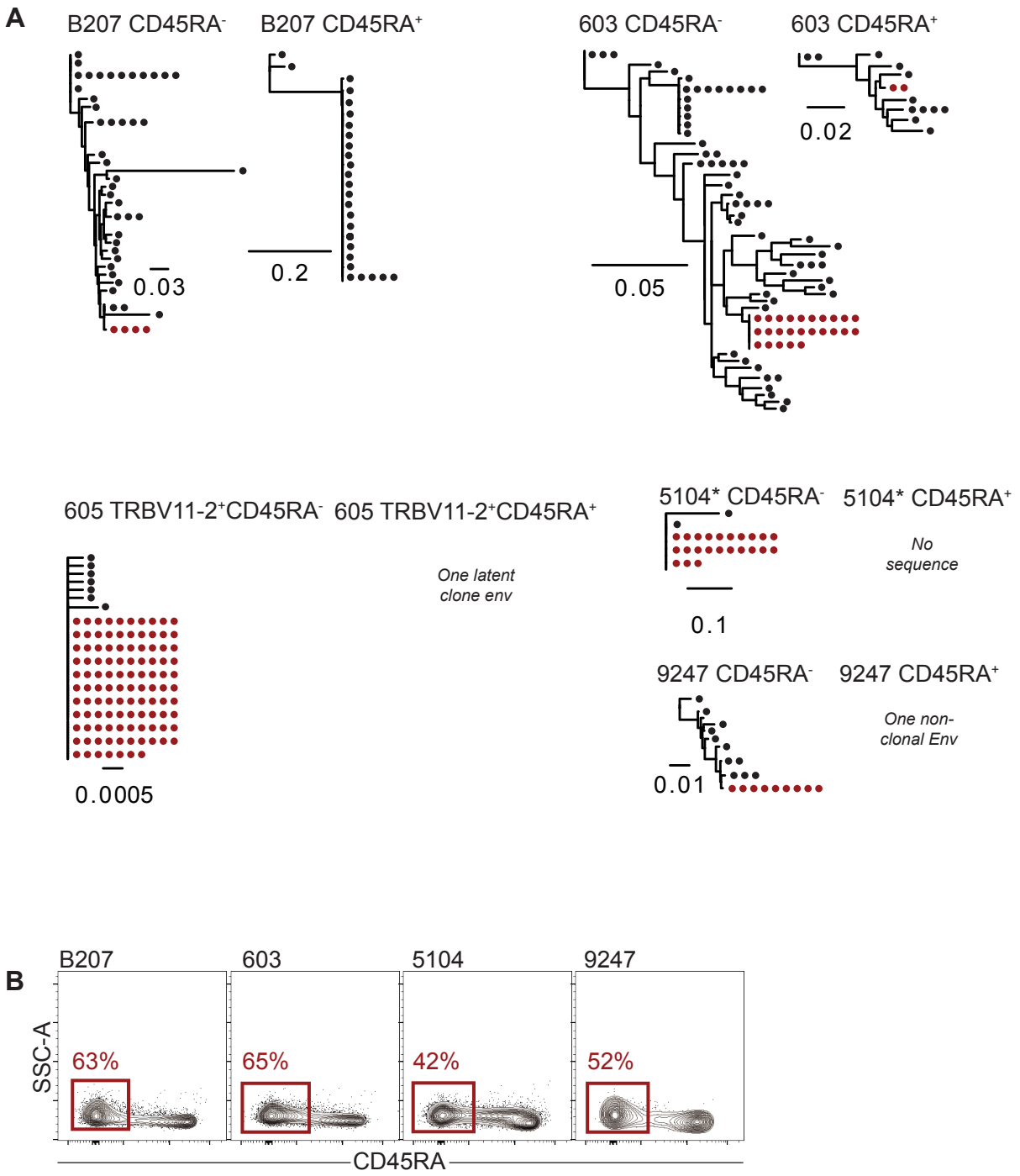


Figure S2. The latent clone of interest is enriched in the CD45RA$^-$ compartment. Maximum-likelihood phylogenetic trees show *env* gene of the clone of interest marked in red. The scale bars indicate the number of substitutions per site. Asterisks indicate trees based on *env* sequences obtained from Q4PCR [49].

(A) Magnetic negative selection of $CD4^+$ $CD45RA^-$ memory T cells enriches the latent clone of interest in all 6 individuals.

(B) Flow cytometry plots indicating the fraction of $CD45RA^-$ cells among $CD4^+$ T cells.

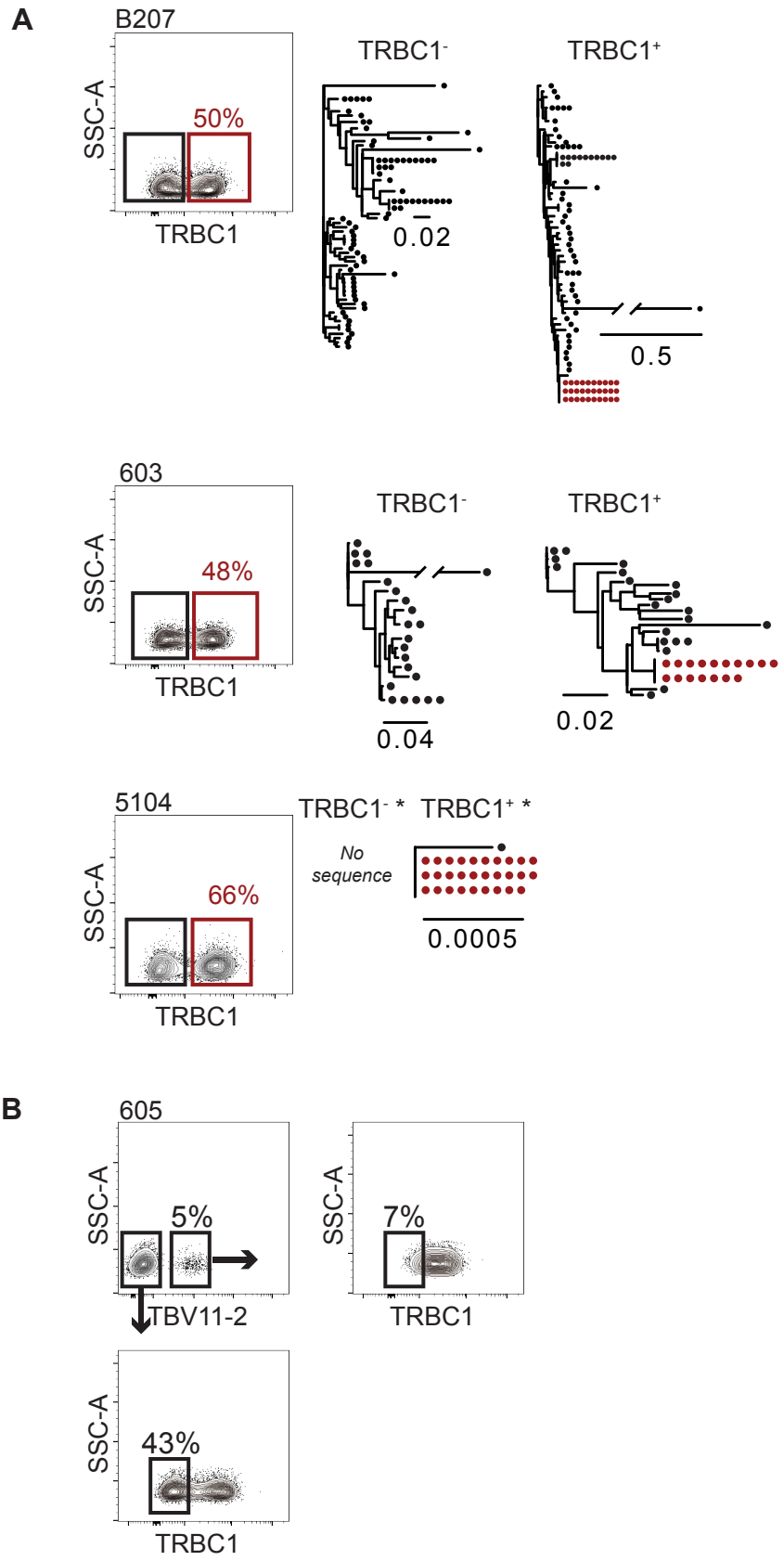

A
B207
SSC-A
50%
TRBC1
TRBC1-
TRBC1+
0.02
0.5
603
SSC-A
48%
TRBC1
TRBC1-
TRBC1+
0.04
0.02
5104
SSC-A
66%
TRBC1
TRBC1- *
TRBC1+ *
No sequence
0.0005
B
605
SSC-A
5%
TBV11-2
SSC-A
7%
TRBC1
SSC-A
43%
TRBC1

Figure S3. Identification of the TRBC domain that is expressed by the latent clone of interest. Red outlines indicate the population containing the clone of interest. Maximum-likelihood phylogenetic trees show *env* gene of the clone of interest marked in red. The scale bars indicate the number of substitutions per site. Asterisks indicate trees based on *env* sequences obtained from Q4PCR [49].

(A) Flow cytometry plots show TRBC1 staining for individuals **B207**, **603**, and **5104**. The fraction of TRBC1$^+$ CD4$^+$ T cells is indicated.

(B) Flow cytometry plots show TRBC1 staining for TRBV2$^-$ and TRBV11-2$^+$ cells in individual **605**. The fraction cells within a gate is indicated above the gate.

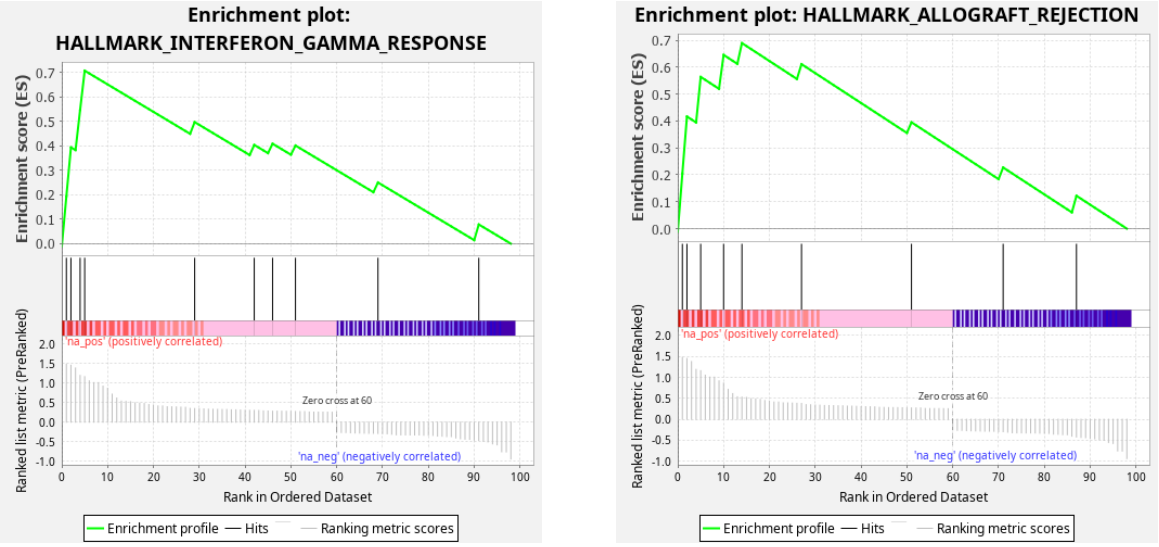


Figure S4. GSEA analysis of the 100 most differentially expressed genes (log2 fold change < - 2.5 and log2 fold change > 2.5) in gene expression cluster 7 compared to all other gene expression clusters.

Selected gene sets with significant enrichment among up-regulated genes in gene expression cluster 7 are shown (nominal p value < 1 %, FDR < 25 %).

Acknowledgments:

We thank all study participants who devoted time to our research, the Rockefeller University Hospital Research support office and nursing staff, and all members of the M.C.N. laboratory for discussions, and M. Jankovic and G. Scrivanti for laboratory support. We thank C. Zhao, H. Duan, C. Lai, and S. Huang from the Genomics Resource Center at the Rockefeller University for preparing and sequencing the 10x Genomics libraries. We also thank K. Chhosphel and K. M. Gordon for operating the cell sorters.

This work was supported by the National Institutes of Health (grants 1U01AI129825 to M.C., UM1 AI100663 and R01AI129795 to M.C.N.); REACH Delaney (grant UM1 AI164565 to M.C.), the Einstein-Rockefeller-CUNY Center for AIDS Research (grant 1P30AI124414-01A1); BEAT-HIV Delaney (grant UM1 AI126620 to M.C.) and the Robertson Fund. C.G. was supported by the Robert S. Wennett postdoctoral fellowship, the Shapiro-Silverberg Fund for the Advancement of Translational Research and by the National Center for Advancing Translational Sciences (National Institutes of Health Clinical and Translational Science Award program, grant UL1 TR001866). M.C.N. is a Howard Hughes Medical Institute (HHMI) Investigator. This article is subject to HHMI's Open Access to Publications policy. HHMI lab heads have previously granted a nonexclusive CC BY 4.0 license to the public and a sublicensable license to HHMI in their research articles. Pursuant to those licenses, the author-accepted manuscript of this article can be made freely available under a CC BY 4.0 license immediately upon publication.

Author contributions:

G.H.J.W, Y.B-O., L.B.C., M.J., and M.C.N conceived and designed experiments. C.G. and M.C. recruited participants, supervised sample collection, and collected clinical data. G.H.J.W.

performed the research. G.H.J.W., T.Y.O., V.R., L.B.C., M.J., and M.C.N analyzed data. T.Y.O. and V.R. performed bioinformatic analysis. H.H. and G.B. gave critical advice for cell sorting. G.H.J.W., L.B.C., M.J., and M.C.N. wrote the manuscript with help from all co-authors.

Competing interests: Rockefeller University has patents on anti-HIV-1 antibodies 3BNC117 and 10-1074 on which Michel Nussenzweig is an inventor that are licensed to Gilead.

Materials and methods

Study design and participants

Study participants were recruited at the Rockefeller University hospital and gave informed written consent before participation in the studies. The study protocols and procedures met the standards of Good Clinical Practice and were approved by the institutional review board of the Rockefeller University. After leukapheresis, peripheral blood mononuclear cells (PBMCs) were isolated by Ficoll separation and stored in aliquots in liquid nitrogen. At the time of sample collection, all study participants were receiving ART and were aviremic.

Magnetic-activated cell sorting of CD4 T cells and memory cells

All procedures were performed while maintaining cells at 4 °C. $CD4^+$ T cells were negatively selected from PBMCs with a $CD4^+$ T cell isolation kit (Miltenyi, cat. 130-096-533). For memory cell separation, $CD4^+$ T cells were negatively selected using magnetic CD45RA MicroBeads (Miltenyi, cat. 130-045-901).

Fluorescence-activated cell sorting

All procedures were performed while maintaining cells at 4 °C. Cells were incubated with Fc-blocking reagent (Miltenyi, cat. 130-059-901). Fixable Viability Dye eFluor 780 (Invitrogen, cat. 65-0865-14) was used for live/dead cell staining. The following antibodies were used for surface staining: PerCP/Cy5.5 anti-human CD4 (BioLegend, cat. 317428), PacificBlue anti-human CD3 (BioLegend, cat. 300431), Brilliant Violet 605 anti-human TCR Cβ1 (BD, cat. 747979), FITC anti-human TCR Vβ17 (Beckman Coulter, cat. IM1234), FITC anti-human TCR Vβ21.3 (Backman Coulter, cat. IM1483), FITC anti-human TCR Vβ7.2 (Beckman Coulter, cat.

B06666), and Beta Mark TCR Vbeta Repertoire Kit (Beckman Coulter, cat. IM3497). The Vbeta Repertoire Kit contains 24 different anti-TRBV antibodies that come in 8 vials (A – H). In each vial, there are three antibodies conjugated with PE, or FITC, or PE FITC.

For enrichment, the following population was sorted after MACS for $CD4^+CD45RA^-$ T cells: individuals B207 and 5104 $CD3^+CD4^+TRBC1^+TRBV^-$ lymphocytes; individual 603 $CD4^+TRBC1^+TRBV19^+$ lymphocytes; individual 605 $CD4^+TRBV11\text{-}2^+$ lymphocytes; individual 5125 $CD4^+TRBC1^+TRBV2^+$ lymphocytes; individual 9247 $CD4^+TRBC1^-TRBV4\text{-}3^+$ lymphocytes. Sorts were performed on BD FACS Aria IIu and BD FACS Aria III.

Flowcytometric analysis

Flowcytometry data was analysed with FACSDiva software (version 2.0.2) and FlowJo version 10.8.1 (both by BD).

Near-full length HIV provirus amplification by Q4 PCR and sequencing from bulk sorted cells

gDNA was extracted from sorted cells, quantified, and limiting dilution performed as described [26]. Near-full length genome amplification was performed as described [49]. PCR amplicons were sequenced with MiSeq 300 cycle v2 kits (Illumina, cat. MS-102-2002). Sequences were assembled with the Defective and Intact HIV Genome Assembler [26].

***Env* amplification and sequencing from bulk sorted cells**

gDNA was extracted from sorted cells, quantified, and limiting dilution performed as described [26]. All *env* amplifications were run with Platinum™ Taq DNA Polymerase High Fidelity

(ThermoFisher, cat. 11304011) by nested PCR. For individual 605, the first PCR included the primers envB3out (5′-TTGCTACTTGTGATTGCTCCATGT–3′) [11] and B3F3 (5′-TGGAAAGGTGAAGGGGCAGT-AGTAATAC-3′) [72] at 94°C for 2min; (94°C for 15s, 60.4°C for 30s, and 68°C) for 6 min for 40 cycles; and 68°C for 15 min. For individuals B207, 603, 5104, 5125, and 9247, the first PCR as well as the second PCR for all individuals was performed as previously described [11]. PCR amplicons were sequenced with MiSeq 300 cycle v2 kits (Illumina, cat. MS-102-2002). Sequences were assembled with the Defective and Intact HIV Genome Assembler [26].

Phylogenetic trees

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

10x Genomics

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

Latent clone TCR identification

Quiescent latent cells were enriched based on CD45RA, TRBC, and TRBV, and 5 cells per well were sorted into a 96-well plate containing TCL buffer (Qiagen, cat. 1031576) with 1 % beta-Mercaptethanol and snap frozen on dry ice. Plates were stored at -80 °C until further use. After thawing, magnetic bead clean-up was performed with Agencourt RNAClean XP magnetic beads (Beckman Coulter, cat. A63987). TCR mRNA was reverse transcribed into cDNA with SuperScript III Reverse Transcriptase (Invitrogen, cat. 18080044) and the following primers: AC1R (5'-ACACATCAGAATCCTTACTTTG-3'), BC1R (5'-CAGTATCTGGAGTCATTGA-3') [73]. Then, a second magnetic bead clean-up, binding TCR cDNA and gDNA, was performed, and *env* amplification from gDNA was performed for individuals 603, 5104, 5125, and 9247 as described above. Only for individual 605, a shortened nested PCR was performed to amplify a ~ 500 bp amplicon of the *env* gene in which the expanded intact latent clone differed from other proviruses that were enriched. The first PCR was performed with primers Env 1,133 F1 (5'-GAGGGGAATTTTTCTACTGTAACAC-3') and Env 1,956 R1 (5'- GTTCTGCGAATTTT-CAATTAAGGTG -3') at 94°C for 2min; (94°C for 15s, 56°C for 30s, and 68°C for 1min) for 35 cycles; and 68°C for 15 min. The second PCR was performed with primers Env 1,240 F2 (5'-ATCACACTCCGATGC-AGAATAAAAC-3') and Env 1,765 R2 (5'-TTAGGTATCT-TTCCACAGCCAGTAC-3') at 94°C for 2min; 94°C for 15s, 58°C for 30s, and 68°C for 1min ×35; and 68°C for 15 min.

TCR amplification and sequencing were performed with primers as previously described [74]. TCR α and β chain cDNA was amplified and barcoded separately and in duplicates. All PCR products were combined at equal proportions, followed by magnetic bead clean up and run on a 1.5 % agarose gel. The band between 300 and 400 bp was excised, purified with NucleoSpin Gel and

PCR Clean-up kit (Quiagen, cat. 740609.50) and sequenced on the MiSeq platform with a MiSeq 600 cycle v3 kit (Illumina, cat. MS-102-3003).

PhiX-derived reads were removed using from downstream analysis using a k-mer based approach implemented by bbduk.sh from BBTools v38.72 (https://sourceforge.net/projects/bbmap/). Samples were demultiplexed according to the combination of oligos used to uniquely identify the plate, row, column they were placed [74]. The quality control check was performed with Trim Galore package v0.6.7 (https://github.com/FelixKrueger/TrimGalore) to trim Illumina adapters and low-quality bases. Paired-overlapping reads were exported into a single read by BBMerge. Subsequently, TRUST4 [75] was used to reconstruct and annotate the T cell receptor (TCR) sequences. There is a chance of incorrect TCR assignment if the sequencing errors occur at the barcodes present in reads. A threshold was established based on the number of reads used to assembly a specific TCR contig assigned to an empty well. We obtained the putative clonotypes, defined by the 10x Genomics V(D)J single-cell sequencing, that shared either the α or β chain found in that well. The most frequent clonotype associated with Env^+ wells in each individual was selected as the clonotype associated with proviral integration.

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

Single-cell RNA-seq binary base call (BCL) files were demultiplexed and converted into FASTQ files using BCLtoFastq prior to alignment to hg38 with CellRanger (v4.0.0) and analyzed in R studio with Seurat (v4). Cells with a mitochondrial proportion greater than 5 % and/or a feature count <200 or >2,500 were discarded. Sample batches were combined, normalized and scaled with SCTransform. Uniform Manifold Approximation and Projection (UMAP) clustering was

performed selecting the first thirty principal components. Single-cell TCR-seq FASTQs were aligned with CellRanger (v4.0.0) to the default CellRanger VDJ reference. Output contig annotations were filtered and analyzed in R studio with Seurat.

1 Chun, T. W. *et al.* Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proc Natl Acad Sci U S A* **94**, 13193-13197, doi:10.1073/pnas.94.24.13193 (1997).
2 Finzi, D. *et al.* Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nat Med* **5**, 512-517, doi:10.1038/8394 (1999).
3 Wong, J. K. *et al.* Recovery of replication-competent HIV despite prolonged suppression of plasma viremia. *Science* **278**, 1291-1295, doi:10.1126/science.278.5341.1291 (1997).
4 Siliciano, J. D. *et al.* Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nat Med* **9**, 727-728, doi:10.1038/nm880 (2003).
5 Bachmann, N. *et al.* Determinants of HIV-1 reservoir size and long-term dynamics during suppressive ART. *Nat Commun* **10**, 3193, doi:10.1038/s41467-019-10884-9 (2019).
6 Crooks, A. M. *et al.* Precise Quantitation of the Latent HIV-1 Reservoir: Implications for Eradication Strategies. *J Infect Dis* **212**, 1361-1365, doi:10.1093/infdis/jiv218 (2015).
7 Peluso, M. J. *et al.* Differential decay of intact and defective proviral DNA in HIV-1-infected individuals on suppressive antiretroviral therapy. *JCI Insight* **5**, doi:10.1172/jci.insight.132997 (2020).
8 Kearney, M. F. *et al.* Origin of Rebound Plasma HIV Includes Cells with Identical Proviruses That Are Transcriptionally Active before Stopping of Antiretroviral Therapy. *J Virol* **90**, 1369-1376, doi:10.1128/JVI.02139-15 (2016).
9 Chun, T. W. *et al.* Rebound of plasma viremia following cessation of antiretroviral therapy despite profoundly low levels of HIV reservoir: implications for eradication. *AIDS* **24**, 2803-2808, doi:10.1097/QAD.0b013e328340a239 (2010).
10 Cohn, L. B. *et al.* HIV-1 integration landscape during latent and active infection. *Cell* **160**, 420-432, doi:10.1016/j.cell.2015.01.020 (2015).
11 Lorenzi, J. C. *et al.* Paired quantitative and qualitative assessment of the replication-competent HIV-1 reservoir and comparison with integrated proviral DNA. *Proc Natl Acad Sci U S A* **113**, E7908-E7916, doi:10.1073/pnas.1617789113 (2016).
12 Bui, J. K. *et al.* Proviruses with identical sequences comprise a large fraction of the replication-competent HIV reservoir. *PLoS Pathog* **13**, e1006283, doi:10.1371/journal.ppat.1006283 (2017).
13 Hosmane, N. N. *et al.* Proliferation of latently infected CD4(+) T cells carrying replication-competent HIV-1: Potential role in latent reservoir dynamics. *J Exp Med* **214**, 959-972, doi:10.1084/jem.20170193 (2017).
14 De Scheerder, M. A. *et al.* HIV Rebound Is Predominantly Fueled by Genetically Identical Viral Expansions from Diverse Reservoirs. *Cell Host Microbe* **26**, 347-358 e347, doi:10.1016/j.chom.2019.08.003 (2019).
15 Einkauf, K. B. *et al.* Parallel analysis of transcription, integration, and sequence of single HIV-1 proviruses. *Cell* **185**, 266-282 e215, doi:10.1016/j.cell.2021.12.011 (2022).
16 Antar, A. A. *et al.* Longitudinal study reveals HIV-1-infected CD4+ T cell dynamics during long-term antiretroviral therapy. *J Clin Invest* **130**, 3543-3559, doi:10.1172/JCI135953 (2020).

17 Cho, A. *et al.* Longitudinal clonal dynamics of HIV-1 latent reservoirs measured by combination quadruplex polymerase chain reaction and sequencing. *Proc Natl Acad Sci U S A* **119**, doi:10.1073/pnas.2117630119 (2022).
18 Cohn, L. B. *et al.* Clonal CD4(+) T cells in the HIV-1 latent reservoir display a distinct gene profile upon reactivation. *Nat Med* **24**, 604-609, doi:10.1038/s41591-018-0017-7 (2018).
19 Einkauf, K. B. *et al.* Intact HIV-1 proviruses accumulate at distinct chromosomal positions during prolonged antiretroviral therapy. *J Clin Invest* **129**, 988-998, doi:10.1172/JCI124291 (2019).
20 Simonetti, F. R. *et al.* Antigen-driven clonal selection shapes the persistence of HIV-1-infected CD4+ T cells in vivo. *J Clin Invest* **131**, doi:10.1172/JCI145254 (2021).
21 Huang, A. S. *et al.* Integration features of intact latent HIV-1 in CD4+ T cell clones contribute to viral persistence. *J Exp Med* **218**, doi:10.1084/jem.20211427 (2021).
22 Maldarelli, F. *et al.* HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science* **345**, 179-183, doi:10.1126/science.1254194 (2014).
23 Wagner, T. A. *et al.* HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science* **345**, 570-573, doi:10.1126/science.1256304 (2014).
24 Chomont, N. *et al.* HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nat Med* **15**, 893-900, doi:10.1038/nm.1972 (2009).
25 Douek, D. C. *et al.* HIV preferentially infects HIV-specific CD4+ T cells. *Nature* **417**, 95-98, doi:10.1038/417095a (2002).
26 Mendoza, P. *et al.* Antigen-responsive CD4+ T cell clones contribute to the HIV-1 latent reservoir. *J Exp Med* **217**, doi:10.1084/jem.20200051 (2020).
27 Gantner, P. *et al.* Single-cell TCR sequencing reveals phenotypically diverse clonally expanded cells harboring inducible HIV proviruses during ART. *Nat Commun* **11**, 4089, doi:10.1038/s41467-020-17898-8 (2020).
28 Ho, Y. C. *et al.* Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* **155**, 540-551, doi:10.1016/j.cell.2013.09.020 (2013).
29 Sengupta, S. & Siliciano, R. F. Targeting the Latent Reservoir for HIV-1. *Immunity* **48**, 872-895, doi:10.1016/j.immuni.2018.04.030 (2018).
30 Darcis, G., Berkhout, B. & Pasternak, A. O. The Quest for Cellular Markers of HIV Reservoirs: Any Color You Like. *Front Immunol* **10**, 2251, doi:10.3389/fimmu.2019.02251 (2019).
31 Cohn, L. B., Chomont, N. & Deeks, S. G. The Biology of the HIV-1 Latent Reservoir and Implications for Cure Strategies. *Cell Host Microbe* **27**, 519-530, doi:10.1016/j.chom.2020.03.014 (2020).
32 Hiener, B. *et al.* Identification of Genetically Intact HIV-1 Proviruses in Specific CD4(+) T Cells from Effectively Treated Participants. *Cell Rep* **21**, 813-822, doi:10.1016/j.celrep.2017.09.081 (2017).
33 Lee, G. Q. *et al.* Clonal expansion of genome-intact HIV-1 in functionally polarized Th1 CD4+ T cells. *J Clin Invest* **127**, 2689-2696, doi:10.1172/JCI93289 (2017).
34 Cockerham, L. R. *et al.* CD4+ and CD8+ T cell activation are associated with HIV DNA in resting CD4+ T cells. *PLoS One* **9**, e110731, doi:10.1371/journal.pone.0110731 (2014).

35 Horsburgh, B. A. *et al.* High levels of genetically intact HIV in HLA-DR+ memory T cells indicates their value for reservoir studies. *AIDS* **34**, 659-668, doi:10.1097/QAD.0000000000002465 (2020).
36 Lee, E. *et al.* Memory CD4 + T-Cells Expressing HLA-DR Contribute to HIV Persistence During Prolonged Antiretroviral Therapy. *Front Microbiol* **10**, 2214, doi:10.3389/fmicb.2019.02214 (2019).
37 Iglesias-Ussel, M., Vandergeeten, C., Marchionni, L., Chomont, N. & Romerio, F. High levels of CD2 expression identify HIV-1 latently infected resting memory CD4+ T cells in virally suppressed subjects. *J Virol* **87**, 9148-9158, doi:10.1128/JVI.01297-13 (2013).
38 Baxter, A. E. *et al.* Single-Cell Characterization of Viral Translation-Competent Reservoirs in HIV-Infected Individuals. *Cell Host Microbe* **20**, 368-380, doi:10.1016/j.chom.2016.07.015 (2016).
39 Pardons, M. *et al.* Single-cell characterization and quantification of translation-competent viral reservoirs in treated and untreated HIV infection. *PLoS Pathog* **15**, e1007619, doi:10.1371/journal.ppat.1007619 (2019).
40 Neidleman, J. *et al.* Phenotypic analysis of the unstimulated in vivo HIV CD4 T cell reservoir. *Elife* **9**, doi:10.7554/eLife.60933 (2020).
41 Liu, R. *et al.* Single-cell transcriptional landscapes reveal HIV-1-driven aberrant host gene transcription as a potential therapeutic target. *Sci Transl Med* **12**, doi:10.1126/scitranslmed.aaz0802 (2020).
42 Grau-Exposito, J. *et al.* Latency reversal agents affect differently the latent reservoir present in distinct CD4+ T subpopulations. *PLoS Pathog* **15**, e1007991, doi:10.1371/journal.ppat.1007991 (2019).
43 Marini, A., Harper, J. M. & Romerio, F. An in vitro system to model the establishment and reactivation of HIV-1 latency. *J Immunol* **181**, 7713-7720, doi:10.4049/jimmunol.181.11.7713 (2008).
44 Krishnan, V. & Zeichner, S. L. Host cell gene expression during human immunodeficiency virus type 1 latency and reactivation and effects of targeting genes that are differentially expressed in viral latency. *J Virol* **78**, 9458-9473, doi:10.1128/JVI.78.17.9458-9473.2004 (2004).
45 Telwatte, S. *et al.* Heterogeneity in HIV and cellular transcription profiles in cell line models of latent and productive infection: implications for HIV latency. *Retrovirology* **16**, 32, doi:10.1186/s12977-019-0494-x (2019).
46 Cameron, P. U. *et al.* Establishment of HIV-1 latency in resting CD4+ T cells depends on chemokine-induced changes in the actin cytoskeleton. *Proc Natl Acad Sci U S A* **107**, 16934-16939, doi:10.1073/pnas.1002894107 (2010).
47 Lassen, K. G., Hebbeler, A. M., Bhattacharyya, D., Lobritz, M. A. & Greene, W. C. A flexible model of HIV-1 latency permitting evaluation of many primary CD4 T-cell reservoirs. *PLoS One* **7**, e30176, doi:10.1371/journal.pone.0030176 (2012).
48 Nussenzweig, M. C. *et al.* Allelic exclusion in transgenic mice that express the membrane form of immunoglobulin mu. *Science* **236**, 816-819, doi:10.1126/science.3107126 (1987).
49 Gaebler, C. *et al.* Combination of quadruplex qPCR and next-generation sequencing for qualitative and quantitative analysis of the HIV-1 latent reservoir. *J Exp Med* **216**, 2253-2264, doi:10.1084/jem.20190896 (2019).

50 Gaebler, C. (submitted).
51 Gaebler, C. *et al.* Sequence Evaluation and Comparative Analysis of Novel Assays for Intact Proviral HIV-1 DNA. *J Virol* **95**, doi:10.1128/JVI.01986-20 (2021).
52 Kwon, K. J. *et al.* Different human resting memory CD4(+) T cell subsets show similar low inducibility of latent HIV-1 proviruses. *Sci Transl Med* **12**, doi:10.1126/scitranslmed.aax6795 (2020).
53 Stumptner-Cuvelette, P. & Benaroch, P. Multiple roles of the invariant chain in MHC class II function. *Biochim Biophys Acta* **1542**, 1-13, doi:10.1016/s0167-4889(01)00166-5 (2002).
54 Su, L., Hershberger, R. J. & Weissman, I. L. LYAR, a novel nucleolar protein with zinc finger DNA-binding motifs, is involved in cell growth regulation. *Genes Dev* **7**, 735-748, doi:10.1101/gad.7.5.735 (1993).
55 Yang, C. *et al.* LYAR Suppresses Beta Interferon Induction by Targeting Phosphorylated Interferon Regulatory Factor 3. *J Virol* **93**, doi:10.1128/JVI.00769-19 (2019).
56 Rohan, P. J. *et al.* PAC-1: a mitogen-induced nuclear protein tyrosine phosphatase. *Science* **259**, 1763-1766, doi:10.1126/science.7681221 (1993).
57 Jeffrey, K. L. *et al.* Positive regulation of immune cell function and inflammatory responses by phosphatase PAC-1. *Nat Immunol* **7**, 274-283, doi:10.1038/ni1310 (2006).
58 Turman, M. A., Yabe, T., McSherry, C., Bach, F. H. & Houchins, J. P. Characterization of a novel gene (NKG7) on human chromosome 19 that is expressed in natural killer cells and T cells. *Hum Immunol* **36**, 34-40, doi:10.1016/0198-8859(93)90006-m (1993).
59 Hashimoto, K. *et al.* Single-cell transcriptomics reveals expansion of cytotoxic CD4 T cells in supercentenarians. *Proc Natl Acad Sci U S A* **116**, 24242-24251, doi:10.1073/pnas.1907883116 (2019).
60 Juno, J. A. *et al.* Cytotoxic CD4 T Cells-Friend or Foe during Viral Infection? *Front Immunol* **8**, 19, doi:10.3389/fimmu.2017.00019 (2017).
61 Oh, D. Y. & Fong, L. Cytotoxic CD4(+) T cells in cancer: Expanding the immune effector toolbox. *Immunity* **54**, 2701-2711, doi:10.1016/j.immuni.2021.11.015 (2021).
62 Appay, V. *et al.* Characterization of CD4(+) CTLs ex vivo. *J Immunol* **168**, 5954-5958, doi:10.4049/jimmunol.168.11.5954 (2002).
63 Zaunders, J. J. *et al.* Identification of circulating antigen-specific CD4+ T lymphocytes with a CCR5+, cytotoxic phenotype in an HIV-1 long-term nonprogressor and in CMV infection. *Blood* **103**, 2238-2247, doi:10.1182/blood-2003-08-2765 (2004).
64 Cachot, A. *et al.* Tumor-specific cytolytic CD4 T cells mediate immunity against human cancer. *Sci Adv* **7**, doi:10.1126/sciadv.abe3348 (2021).
65 Takeuchi, A. & Saito, T. CD4 CTL, a Cytotoxic Subset of CD4(+) T Cells, Their Differentiation and Function. *Front Immunol* **8**, 194, doi:10.3389/fimmu.2017.00194 (2017).
66 Abana, C. O. *et al.* Cytomegalovirus (CMV) Epitope-Specific CD4(+) T Cells Are Inflated in HIV(+) CMV(+) Subjects. *J Immunol* **199**, 3187-3201, doi:10.4049/jimmunol.1700851 (2017).
67 Hale, J. S. & Ahmed, R. Memory T follicular helper CD4 T cells. *Front Immunol* **6**, 16, doi:10.3389/fimmu.2015.00016 (2015).
68 Chang, J. T. *et al.* Asymmetric T lymphocyte division in the initiation of adaptive immune responses. *Science* **315**, 1687-1691, doi:10.1126/science.1139393 (2007).

69 Imamichi, H. *et al.* Lifespan of effector memory CD4+ T cells determined by replication-incompetent integrated HIV-1 provirus. *AIDS* **28**, 1091-1099, doi:10.1097/QAD.0000000000000223 (2014).
70 Qi, Q. *et al.* Diversity and clonal selection in the human T-cell repertoire. *Proc Natl Acad Sci U S A* **111**, 13139-13144, doi:10.1073/pnas.1409155111 (2014).
71 DeBoer, J. *et al.* Alterations in the nuclear proteome of HIV-1 infected T-cells. *Virology* **468-470**, 409-420, doi:10.1016/j.virol.2014.08.029 (2014).
72 Salazar-Gonzalez, J. F. *et al.* Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. *J Exp Med* **206**, 1273-1289, doi:10.1084/jem.20090378 (2009).
73 Mamedov, I. Z. *et al.* Preparing unbiased T-cell receptor and antibody cDNA libraries for the deep next generation sequencing profiling. *Front Immunol* **4**, 456, doi:10.3389/fimmu.2013.00456 (2013).
74 Han, A., Glanville, J., Hansmann, L. & Davis, M. M. Linking T-cell receptor sequence to functional phenotype at the single-cell level. *Nat Biotechnol* **32**, 684-692, doi:10.1038/nbt.2938 (2014).
75 Song, L. *et al.* TRUST4: immune repertoire reconstruction from bulk and single-cell RNA-seq data. *Nat Methods* **18**, 627-630, doi:10.1038/s41592-021-01142-2 (2021).