Supplemental Experimental Procedures

Generation of HLACs from tonsil tissues

Human tonsils from the Cooperative Human Tissue Network (CHTN) were dissected into 4 mm^3 pieces, passed through a 40-μm strainer, and cultured in 96-well U-bottomed polystyrene plates (10^6 cells/well) in 200 μl/well of Tonsil Media, consisting of RPMI supplemented with 15% heat-inactivated fetal bovine serum (FBS), 100 μg/ml gentamicin, 200 μg/ml ampicillin, 1 mM sodium pyruvate, 1% non-essential amino acids (Mediatech), 1% Glutamax (Thermo Fisher), and 1% Fungizone (Invitrogen). Following isolation, HLACs were immediately used for viral fusion or infection assays. Of note, HLAC cells were never stimulated with a mitogen at any point during *in vitro* culture.

Construction and validation of reporter viruses

We previously described the molecular approach for generating *Renilla* luciferase (LucR)-expressing, replication competent HIV-1 proviral infectious molecular clones (IMCs), in which the ectodomain of gp160/Env is encoded by heterologous HIV-1 strain *env* sequences, and Nef expression is driven by a modified IRES element (referred to as 6ATRi) as part of the bicistronic LucR.IRES-*nef* cassette (Alberti et al., 2015). The construct expressing the BaL *env* sequence was named pNL-LucR.6ATRi-BaL.ecto, and its Nef-deficient derivative was named pNL-LucR.6ATRi-Nef^{stop}-BaL.ecto. For the present study, we modified this approach by (i) inserting the heterologous Env ectodomain-coding region from HIV-1 subtype C transmitted/founder strain C.ZM109F.PB4 (corresponding to nt 123-2020 of GenBank accession number AY424138.2) in place of the BaL *env* sequence (Fig. S1A), and (ii) by replacing the LucR reporter gene with the mouse Heat Stable Antigen (HSA) gene [corresponding to CDS in GenBank entries NM_009846.2 (nt 86-316) and BC075622.1 (nt 66-296)]. The HSA ORF was amplified from pNL4-3.HSA.R+ (He et al., 1995), obtained from the NIH AIDS Reagent program (catalog # 3419). HSA encodes a protein that is anchored to the cell surface via glycosylphosphatidylinositol (GPI) and is sensitively detected by flow cytometry when exogenously expressed. HSA is particularly useful as a marker for proviral gene expression from HIV reporter viruses because of the relatively small size of its ORF (231 nucleotides encoding 77 amino acids). The two proviral constructs generated for this study, both of which

harbor the C.ZM109F.PB4 *env* and the gene for HSA, are pNL-HSA.6ATRi-C.ZM109F.PB4.ecto (referred to as HIV-F4.HSA) and its Nef-deficient counterpart pNL-HSA.6ATRi-Nef^{stop}-C.ZM109F.PB4.ecto (referred to as HIV-F4.HSA.dNef).

Testing Nef functionality by MHC class I and CD4 down modulation

We assessed for Nef-mediated MHC class I down-regulation (Fig. S1B) as a measure of Nef functionality, essentially as previously described (Alberti et al., 2015). In brief, J2574-R5 cells, which are stably transduced with a construct where HIV-1 LTR drives expression of GFP (Shishido et al., 2012), were infected for 2 d with 13 x 10^6 TZM-bl infectious units (I.U.) of HIV-F4.HSA or 6.8 x 10^6 TZM-bl I.U. of HIV-F4.HSA.dNef. MHC class I expression was then measured by flow cytometry following staining of cells with APC-conjugated anti-HLA-ABC (Clone G46-2.6, from BD Biosciences). CD4 downregulation was tested in HLACs infected for 3 d with either HIV-F4.HSA or HIV-F4.HSA.dNef (400 ng/ml p24^{Gag}) by flow cytometry (Fig. S1C) after staining cells with V450-conjugated anti-CD3 (Clone UCHT1, from BD Biosciences), APC-H7-conjugated anti-CD8 (Clone SK1, from BD Biosciences), PE-Cy5.5-conjugated anti-CD4 (Clone SK-3, from eBioscience), and FITC-conjugated anti-murine HSA (Clone M1/69, from BD Biosciences). Zombie Aqua^{TM} (Biolegend) was used as a live/dead marker. After staining, cells were washed 3x with FACS buffer (PBS + 2% FBS + 2 mM EDTA) and run on an LSR-II flow cytometer (BD Biosciences). Data were gated on live, singlet CD3+CD8- cells, and CD4 downregulation was assessed on cells expressing high levels of HSA.

Metal conjugation to antibodies

Antibodies requiring in-house conjugation to metal isotopes were conjugated using X8 antibody labeling kits (Fluidigm). All metals were purchased from Fluidigm, except for 139La ($LaCl_3$ from Sigma-Aldrich), 155Gd, and 157Gd ($GdCl_3$ from Trace Sciences). Conjugation was conducted according to manufacturer's instructions. After the final wash, samples were eluted in a total of 100 µl, quantitated for protein content by Nanodrop 2000c, and diluted 1:1 with PBS-based Antibody Stabilizer (Boca Scientific) supplemented with 0.05% sodium azide, and stored at 4ºC.

HIV fusion assays for CyTOF analysis

To generate BlaM-Vpr-containing virions for the fusion assay, 60 µg of HIV-F4.HSA or HIV-F4.HSAdNef proviral DNA, 20 µg of pCMV-BlaM-Vpr, and 10 µg of pAdVAntage vectors were transfected per 175 mm flask of 80% confluent 293T cells, similar to approaches previously described (Cavrois et al., 2014). Prior to use, virions were concentrated by ultracentrifugation of transfection supernatants and quantified by the FlaQ assay (Gesner et al., 2014). For deep phenotyping of virion fusion-positive cells by CyTOF (Fig. 1), HLACs were first depleted of B cells using CD19 microbeads (Miltenyi), resulting in yields of 100-750 x 10^6 cells for each donor. Incubations with BlaM-Vpr-containing virions were performed in 15 ml falcon tubes using 1 µg of $p24^{Gag}$ in 240 µl for 50 x 10^6 cells aliquoted per tube, similar to approaches described (Cavrois et al., 2014). Briefly, after a 2 h incubation at 37˚C, cells were washed and loaded with CCF2 dye for 1 h at room temperature. Cells were then washed and cleavage of CCF2 by BlaM was allowed to proceed for 6 h, resulting in color shift from green to blue in cells that supported HIV fusion. Aliquoted cells from the same donor were then combined and stained for 15 min at room temperature with the LIVE/DEAD fixable Red Dead Cell Stain kit (Molecular Probes) at a 1:1000 dilution in FACS buffer. After two washes, cells that supported fusion were sorted at 4˚C with a FACS AriaII (BD biosciences) under BSL3 conditions, following the gating strategy presented Fig. S2. For each donor, the total number of sorted cells that supported fusion ranged from 53,000 to 212,000 cells. From an uninfected parallel culture, cells harboring uncleaved CCF2 were sorted (Fig. S2) as controls. Sorted cells were then stained with the CyTOF antibody panel (Table S1) using approaches described in detail below.

HIV infection assays for CyTOF analysis

To generate viral stocks, 293T cells seeded in 6-well plates (3 x 10^5 cells/well) were transfected with HIV-F4.HSA or HIV-F4.HSA.dNef proviral DNA (0.5 µg/well) by Fugene (Promega). Supernatants containing viruses were harvested 2 d later and assessed for $p24^{Gag}$ concentrations using the Lenti-X p24 Rapid Titer kit (Clontech). HLACs were inoculated with the viral stocks (400 – 800 ng $p24^{Gag}$ /ml) or mock-infected for 2 h, after which cultures were replaced with fresh Tonsil Media. After 3 d of incubation at 37˚C, a small aliquot of cells was analyzed by flow cytometry for HSA surface expression to confirm infection. The rest of the cells were cultured for an additional 20 h and then processed for CyTOF analysis.

Staining of cells for CyTOF analysis

For each sample, a total of 6 x 10^6 cells was stained. For sorted HIV-fused and uninfected cells, which averaged 150,000 cells per donor, 6 x 10^6 RAMOS B cells (ATCC) were added as carriers to alleviate the loss of cells during the subsequent staining and wash steps. Cells were processed in Nunc™ 96 DeepWell™ polystyrene plates (Thermo Fisher). Cells were first blocked for 15 min at 4ºC with sera from mouse (Thermo Fisher), rat (Thermo Fisher), and human (AB serum, Sigma-Aldrich). Cells were then washed 2x with 800 μl CyFACS buffer, consisting of metal contaminant-free PBS (Rockland) supplemented with 0.1% bovine serum albumin and 0.1% sodium azide. Cells were then stained for 45 min at 4ºC with the cocktail of primary antibodies in a total volume of 100 μl/well. Given that the Act1 antibody recognizing the α4β7 complex could not be successfully conjugated to a metal isotope (N. Lazarus and E. Butcher, unpublished observations), we used APC-conjugated Act1 in the primary cocktail mix, followed by 3 washes with CyFACS, followed by a secondary stain for 30 min at 4ºC with an anti-APC antibody (Clone APC003 from Biolegend) conjugated to 173Yb. After 3 additional washes with CyFACS, cells were treated for 30 min at 4ºC with 1.7 μg/ml 139In-DOTA maleimide (Macrocyclics) as a live/dead marker. Cells were then washed 2x with CyFACS and fixed overnight with 2% PFA (diluted from 16% stock purchased from Electron Microscopy Sciences) in contaminant-free PBS (Rockland). The next day, cells were washed 2x with Permeabilization Buffer (eBioscience), and permeabilized for an additional 45 min at 4ºC. Cells were then washed 1x with CyFACS, and treated for 20 min at 25ºC with 250 nM Cell-ID™ Intercalator-Ir (Fluidigm). Samples were then washed 2x with CyFACS, 1x with PBS (Rockland), and 3x with pure water. Immediately prior to sample acquisition, samples were diluted to 6.2 x 10^5/ml in a 1:10 dilution of EQ™ calibration beads (Fluidigm). Samples were acquired at a rate of 400-800 events/sec on a CyTOF2 instrument (Fluidigm) at the Stanford Shared FACS Facility (SSFF).

CyTOF gating and data export

Data were normalized to EQ™ calibration beads (which causes their presence to no longer be detectable in sequential gating plots) and imported into both Flowjo (Treestar) and Cytobank for population gating and data analysis. Events corresponding to T cells were extracted following sequential gating to identify these

cells based on DNA content, viability, cell length, and presence in the CD3+CD19- gate (Fig. S4). Because RAMOS cells express CD19, this gate removed these carrier cells from the fusion data analysis (Fig. S4A). Productively-infected cells were identified as live, CD3+CD19-HSA^{high} cells which had downregulated CD4 (Fig. S4B). Where indicated, data for memory CD4+ T cells (Tm, defined as CD3+CD8-CD45RA-CD45RO+ cells) expressing CD57+ or CD127+ were exported as new FCS files for further analysis. For analysis by Principal Component Analysis, agglomerative hierarchical clustering, and SLIDE/PP-SLIDE, gated events were exported as arcsinh transformed expression values using Cytobank. For comparing proportions of gated populations in multiple donors, pairwise statistical comparisons between the different cellular subsets were conducted using a paired *t*-test where the alternative hypothesis was that the difference in the means between the two groups is less than 0.

ImagestreamX MarkII (Amnis) imaging of Tdp^{high} and Tdp^{intermediate}

To image the two distinct Tdp (CD45RO+CD45RA+) T cell populations we observed in HLACs (Fig. S5A) on the ImagestreamX MarkII from Amnis (Fig. S5B), 2 x 10^6 HLAC cells were stained at room temperature for 30 min with FITC-conjugated anti-CD3 (Clone UCHT1, from BD Biosciences), PE-CF594-conjugated anti-CD45RO (Clone UCHL1, from BD Biosciences), and APC-conjugated anti-CD45RA (Clone HI100, from BD Biosciences) supplemented with a 1:500 dilution of the LIVE/DEAD fixable Violet Death Stain Kit (Molecular Probes). After washing 2x in FACS buffer, cells were fixed in PBS containing 2% FBS and 1% paraformaldehyde, filtered through a 40 μm mesh, and analyzed on the MARK II ImageCytometer equipped with the 40x objective. Compensation controls were acquired without bright field and SSC. Images were analyzed using IDEAS software. Cells corresponding to live, CD3+ cells within the focus plane were analyzed for CD45RA and CD45RO expression. The indicated gated populations were then analyzed for cell morphology.

Time course analysis of HIV infection of Tm subsets

For the time course experiments (Fig. 6A), a total of 25 x10^6 HLAC cells were infected for 2 h at 37˚C with 5 μg p24^{Gag} HIV-F4.HSA containing BlaM-Vpr, after which cells were washed in Tonsil Media. One set of samples comprising 20% of the culture was kept for immediate fusion analysis, while the remaining sets

were returned to 37˚C for incubation for 2 - 5 d. After loading the cells with CCF2 as described above and in (Cavrois et al., 2014), the cells were stained with a LIVE/DEAD fixable Green Dead Cell Stain kit (Molecular Probes) at a 1:100 dilution in FACS buffer. Cells were then washed, and stained at 4ºC for 30 min with a cocktail of antibodies diluted in a 1:1 mixture of FACS buffer and the Brilliant Stain Buffer (BD Biosciences). The antibody cocktail consisted of APC-H7-conjugated anti-CD3 (Clone SK7, from BD Biosciences), PerCP5.5-conjugated anti-CD19 (Clone HIB19, from BD Biosciences), BUV737-conjugated anti-CD4 (Clone SK3, from BD Biosciences), BUV395-conjugated anti-CD45RA (Clone HI100, from BD Biosciences), ECD-conjugated anti-CD45RO (Clone UCHL1, from Beckman Coulter), PE-H7-conjugated anti-CD57 (Clone TB01, from eBioscience), and BV650-conjugated anti-CD127 (Clone HIL-7R-M21, from BD Biosciences). Fluorescence minus one (FMO) controls for CD127 staining were also included. Compensation controls using Comp Beads (BD Biosciences) were acquired at the same time. The cultures that were allowed to complete infection at 37˚C were sampled daily. On each day, cells were stained with a LIVE/DEAD fixable Green Dead Cell Stain kit (Molecular Probes) at a 1:500 dilution in FACS buffer, and then processed as described above for the set of fusion assay samples, except that BV421-conjugated anti-CD127 (Clone HIL-7R-M21, from BD Biosciences) was used instead of BV650-conjugated CD127, and PE-conjugated anti-HSA (Clone M1/69, from BD Biosciences) was added to detect productive infection.

Analysis of HIV infection of CD127-enriched Tm

For the CD127 enrichment experiments (Fig. S10), HLACs were simultaneously depleted of both B cells and naïve T cells by negative selection using CD45RA beads (Miltenyi), followed by positive selection using the CD127 MicroBead Kit (Miltenyi) to enrich for CD127+ Tm. Total Tm or CD127-enriched Tm (4 x10^6/sample) were each infected for 2 h at 37˚C with 2.5 µg $p24^{Gag}$ HIV-F4.HSA containing BlaM-Vpr. After 2 washes, 25% of each culture was analyzed for viral fusion, while the remaining 75% was returned to 37˚C for 3 d to assess productive HIV infection. Cells analyzed for fusion were stained with a LIVE/DEAD fixable Green Dead Cell Stain kit (Molecular Probes) at a 1:100 dilution in FACS buffer, washed, and then stained at 4ºC for 30 min with an antibody cocktail consisting of APC-H7-conjugated anti-CD3 (Clone SK7, from BD Biosciences), PerCP5.5-conjugated anti-CD19 (Clone HIB19, from BD Biosciences), BUV737-conjugated anti-CD4 (Clone SK3, from BD Biosciences), BUV395-conjugated

anti-CD45RA (Clone HI100, from BD Biosciences), ECD-conjugated anti-CD45RO (Clone UCHL1, from Beckman Coulter), PE-H7-conjugated anti-CD57 (Clone TB01, from eBioscience), and Alexa Fluor 647-conjugated CD127 (Clone HIL-7R-M21, from BD-Biosciences). To assess productive infection, cells were first stained with a LIVE/DEAD fixable Green Dead Cell Stain kit (Molecular Probes) at a 1:500 dilution in FACS buffer, and then with the same antibody cocktail used for analyzing HIV-fused cells, except that BV421-conjugated anti-CD127 (Clone HIL-7R-M21, from BD Biosciences) was used in place of Alexa Fluor 647-conjugated CD127, and PE-conjugated anti-HSA (Clone M1/69, from BD Biosciences) was added to detect productive infection.

Immunofluorescence (IF) analysis of lymph node sections

Use of lymph node samples was approved by the UCSF Institutional Review Board (IRB # 14-13840). For analysis of lymph node sections (Fig. 6B), 2 μm cross-sections from formalin-fixed and paraffin-embedded (FFPE) lymph node blocks were used. An uninfected lymph node from an HIV- patient was used as a negative control for *in situ* hybridization (ISH). The HIV+ lymph node was from an HIV-infected donor being evaluated for lymphadenopathy who was viremic and not on suppressive anti-retroviral therapy at the time of biopsy. The viral load was 114,606 copies/ml, with no evidence of any X4-tropic virus by the Trophile assay. The final pathology report from the studied lymph node indicated reactive lymphadenopathy. HIV-RNA was detected by ISH (Deleage et al., 2016) using the RNASCOPE® 2.5 HD brown kit (Advanced Cell Diagnostics) according to manufacturer's protocol. HIV RNA probes (cat #416111) were used for ISH, and signals were developed using the TSA® Plus Cyanine 3 (Cy3) detection kit (PerkinElmer). For IF and ISH multiplexing, sections were incubated in 10% fetal bovine serum (FBS) (Thermo Fisher Scientific) for 30 min, and then incubated overnight at 4ºC with 1:500 rabbit anti-CD127 (IgG polyclonal, from LifeSpan Biosciences) and mouse 1:500 anti-CD57 (Clone TB01, from eBioscience). Sections were washed 2x with TTBS (20 mM Tris base, 150 mM NaCl. 0.1% Tween 20, pH 7.6) for 2 min, and then incubated with Alexa Fluor 650-conjugated anti-rabbit and Alexa Fluor 488-conjugated anti-mouse secondary antibodies (both from Thermo Fisher Scientific) for 1 h at room temperature. The sections were washed 2x with TTBS for 2 min and mounted using ProLong Diamond

anti-fade mountant with DAPI (Thermo Fisher Scientific). Slides were scanned using a Zeiss Axio Scan.Z1 slide scanner at 20x (Carl Zeiss).

t-Distributed stochastic neighbor embedding (t-SNE) analysis

FlowJo was used to generate all t-SNE plots of uninfected, HIV-fused, and productively-infected cells presented in the main manuscript. In t-SNE plots, the position of every dot corresponding to a single cell is based on a non-linear projection of the combinatorial expression of all proteins, where each protein corresponds to a single dimension. Cells classified as being more similar to one another are displayed closer together on the t-SNE plot. For our studies, t-SNE plots were generated based on all parameters not used in the gating strategy, or with select markers additionally removed as indicated. Populations of interest were mapped onto t-SNE plots by first identifying the desired populations through sequential gating on standard two-dimensional dot plots, and then mapping that population on the t-SNE plot through a feature offered in FlowJo version 10. Different populations are depicted with different colors on the t-SNE plots. To generate t-SNE plots colored by raw intensity values of individual antigens, the vi-SNE feature available in Cytobank was used with default settings.

Agglomerative hierarchical clustering and Principal Component analysis (PCA)

The raw data corresponding to uninfected, HIV-fused, or HSA^{high} T cells were exported from Cytobank as arcsinh-transformed values. For each donor, the population of uninfected, HIV-fused, or HSA^{high} cells were coded by adding a categorical parameter (uninfected, fused or infected) and then combined. Agglomerative hierarchical clustering and PCA were performed in the Partek Genomics Suite. Agglomerative hierarchical clustering was performed using the Euclidiean dissimilarity score and the Ward's method (Sen et al., 2014).

SPADE analysis

Following population gating, the FCS files of the indicated populations were subjected to SPADE analysis on Cytobank (Qiu et al., 2011). SPADE trees were constructed using default settings with a target node of 25 clusters. All cell surface antigens were used as clustering channels, except for markers that were used in the gating strategy (CD3, CD19, HSA, CD45RA, CD45RO, CD57, and CD127 as indicated). The nodes

were annotated into subpopulations based on the expression of CXCR5, PD1, CCR6, CCR4, CD57, or CD127.

Scaffold

Scaffold maps were generated as described (Spitzer et al., 2015). To generate landmark nodes, uninfected, live, singlet CD3+CD19- events corresponding to naïve CD8+ T cells (CD45RA+CD45RO-CD8+), naïve CD4+ T cells (CD45RA+CD45RO-CD4+), Tdp^{high} (CD45RA+CD45RO+CD4+), memory CD8+ T cells (CD45RA-CD45RO+CD8+), memory CD4+CD57+ T cells (CD45RA-CD45RO+CD4+CD57+CD127-), memory CD4+CD57-CD127- T cells (CD45RA-CD45RO+CD4+CD57-CD127-), and memory CD4+CD57-CD127+ T cells (CD45RA-CD45RO+CD4+CD57-CD127+) were each combined into one file. Landmarks were created separately for the fusion (Fig. A8A) or infection (Fig. A8C) datasets. Landmark nodes (black) are placed in locations reflecting their dissimilarity.

To generate cluster nodes (blue in Fig. A8A and A8C), the datasets corresponding to live, CD3+CD19- cells for three HIV fusion experiments and four HIV infection experiments were each clustered independently into 50 clusters using the *clara* function in R. These cluster nodes were then connected to their most similar landmark nodes by lines termed "edges". The lengths of the edges between landmark nodes and clusters are defined as the cosine similarity between vectors defining the median marker value of the cells within the landmarks and nodes. Graphs were laid out using an in-house R implementation of the ForceAtlas2 algorithm from the graph visualization software Gephi as previously described (Spitzer et al., 2015).

Analysis of viral remodeling using SLIDE

Data Standardization

Within Cytobank, the raw expression data corresponding to uninfected, HIV-fused, and HIV-infected cells were exported as arcsinh transformed values as follows:

$$X = log_e\left(\tilde{X} + \sqrt{1 + \tilde{X}^2}\right) \quad (1)$$

where X denotes the arcsinh transformed value of a given expression marker and $\tilde{X}$ denotes the value of the expression marker in the original scale. This standardization is an inbuilt function of Cytobank, and reduces

the effects of outliers and extreme records that may be randomly generated in gargantuan datasets. Global standardization with equation 1 was applied for all subsequent analyses.

For the sub-analysis PP-SLIDE, where the levels of specific markers on different populations of cells are compared, another standardization procedure is necessary. This is because due to the nature of CyTOF data, the distribution of the random variable X will typically have a spike at 0 that corresponds to those instances when the marker was not expressed at levels detectable by CyTOF, and thus the probability distribution of X must appropriately weigh these undetectable values. To that effect, we model the distribution of X as a mixture with probability density function given by

$$f(x) = p\delta_0(x) + (1-p)g(x) \qquad (2)$$

where p is the probability of non-expression, δ_0 is the point mass at 0, and g is a non-parametric probability density function of the non-zero expression values. This models the expression values in the datasets as a mixture, and allows marker expression data to be standardized between the different samples. Equation (2) was used to test for significant differences in expression levels of specific receptors between infected cells and predicted precursor cells, as described below in equation (6).

Identification of uninfected nearest neighbor for each HIV-fused and infected cell

For subsequent analysis by SLIDE we first identified the cells in the uninfected group (termed "UI" in the equations below) that are the closest to cells in the HIV-fused or productively-infected groups (both termed "I" in the equations below). For clarity, the uninfected cells identified here were termed predicted precursor cells. Identification of the closest cells is based on a nearest neighbor approach (Cover and Hart, 1967; Devroye et al., 2013). Given that group I is smaller in size than group UI, we find the nearest neighbor of each I cell (v) among all cells (u) in the UI group. In other words,

$$for\ each\ v\ \in I, find: u_v = \arg min_{u\in UI} \sum_{i=1}^{M} |v_i - u_i|\ and\ let\ d_1(v) = \sum_{i=1}^{M} |v_i - u_{iv}|$$

where the cells u, v and u_v are M tuples holding the expression values of the M markers.

These identified nearest-neighbor cells $\{u_v\}$ are used for SLIDE analysis. Note that in this nearest neighbor identification step, we use empirical distances based on the agglomerative L1 norm across proteins, and not probability models as described in Equation (2) because such probability models implement protein expression levels that may be highly variable between the infected and uninfected populations.

For all identified uninfected nearest-neighbor cells u_v identified as described above, we then identify their uninfected nearest neighbors. That is, for each u_v we find its nearest neighbor in population $UI - u_v$, as follows:

$$for\ each\ v\ \in I, u_v \in UI, find: u_{u_v}$$

$$= \arg min_{u \in UI - u_v} \sum_{i=1}^{M} |u_{iv} - u_i| \ and\ let\ d_2(v) = \sum_{i=1}^{M} |u_{iv} - u_{iu_v}|$$

If N_I, N_{UI} denote the number of cells in groups I and UI respectively, then the computational complexity of this procedure is $O(N_I N_{uI})$.

Under the null hypothesis of no remodeling, the distribution of $d_1(v)$ would be statistically similar to the distribution of $d_2(v)$, and thus the remodeling score given by the ratio

$$r(v) = \frac{d_1(v)}{d_2(v)}, v \in I \qquad (3)$$

should be symmetrically distributed around a value close to 1. Departures from the null hypothesis would be characterized by r having a mean higher than 1, because in the presence of remodeling, the average distance between cells $\{v\}$ and $\{u_v\}$ will be greater than the average distance between the cells $\{u_v\}$ and $\{u_{u_v}\}$. To test the null hypothesis of no remodeling versus remodeling, we test $H_0: E(r) = 1$ against the alternative $H_1: E(r) > 1$, and employ a one-sided non-parametric Wicoxon test. Rejection of the null hypothesis (H_0) provides statistical evidence of remodeling.

To compare the extent of remodeling between different populations of cells that originated from the same donor, as was the case with the infection datasets comparing cells productively infected with HIV-F4.HSA vs. HIV-F4.HSA.dNef, we test whether the average remodeling score of one population is significantly different from that of the other. The cellular populations to be compared come from the same donor i, where $i = 1,2,..,N$. We test $H_0: E(r_{Wt}^i) = E(r_{dNef}^i)$ against $H_1: E(r_{Wt}^i) > E(r_{dNef}^i)$ using a one-sided two sample Wilcoxon test (Fay and Proschan, 2010). If q_i is the p−value obtained from this test, we estimate the overall p−value of this test for all the N donors as

$$q = 1 - \prod_{i=1}^{N}(1 - q_i) \qquad (4)$$

q is the probability that the average remodeling score of at least 1 donor is significantly different between cells productively infected with HIV-F4.HSA vs. HIV-F4.HSA.dNef. For q to be small (i.e., to reject the null hypothesis and conclude that SLIDE scores are different), all individual q_i values must be small. Therefore, q is a conservative estimate of the overall p-value.

When the cellular populations originate from different donors, however, one cannot use the above approach of estimating the overall p−value. Such cases include comparisons between (i) Infection and Fusion datasets, and (ii) cells that had fused with either HIV-F4.HSA versus HIV-F4.HSA.dNef, where different donors were used since typically not enough cells were available from a single HLAC donor to assess fusion to both HIV-F4.HSA and HIV-F4.HSA.dNef virus. In these cases, we combine the data on remodeling scores of all $N1$ donors in population 1 into a single vector so that $r_1 = \{r_1^1, r_1^2, \ldots., r_1^{N_1}\}$, and combine the data on remodeling scores of all $N2$ donors in population 2 into a single vector $r_2 = \{r_2^1, r_2^2, \ldots., r_2^{N_2}\}$. We then test

$$H_0: E(r_1) = E(r_2)\ against\ \ H_1: E(r_1) > E(r_2) \qquad (5)$$

using a one-sided two sample Wilcoxon test. The p-value associated with this test is then used to assess the significance of the null hypothesis of no difference in average remodeling scores between the two populations. The above approach of testing by combining the remodeling scores of all donors in a population is based on the assumption that the donors, although different, are essentially homogenous as far as the remodeling behavior of their cells is concerned.

To test for significant differences in expression levels of specific receptors between the predicted precursor and infected cells (PP-SLIDE sub-analysis), for a given a marker j, we test for evidence of statistical difference in the expression level of marker j in populations $\{u_v\}$ and $\{v\}$. Let the expression of the j^{th} marker be given by X_{ij}^k where $i = 1,2, \ldots n_k,\ j = 1,2, .. M,\ k = u_v, v$ and n_k denotes the number of cells in the k^{th} group. In our model given by equation 2, X_{ij}^k has a mixture density,

$$f_{k,j}\left(x_{ij}^k\right) = p_{k,j}\delta_0\left(x_{ij}^k\right) + (1 - p_{k,j})g_{k,j}\left(x_{ij}^k\right) \quad (6)$$

where, as in equation 2, $p_{k,j}$ is the probability of non-expression of marker j in population k, δ_0 is the point mass at 0, and $g_{k,j}$ is a non-parametric probability density function of the non-zero expression values of marker j in population k.

A difference in the expression level of marker j across the two populations could be driven by (i) the differences in the probability of non-expression and/or (ii) the differences in the distribution of non-zero expression values. Therefore, to test for statistical difference in the expression level of marker j in the two populations, the null hypothesis H_0 of there being no statistical difference expression levels of marker j between the two populations must involve both the probability of non-expression $p_{k,j}$ and the distribution of the non-zero expression values $g_{k,j}$ and is hence given by

$$H_0 : p_{u_v,j} = p_{v,j} \text{ and } g_{u_v,j} = g_{v,j} \quad (7)$$

Our alternative hypothesis H_1, that there is a statistical difference in expression levels of marker j between the two populations, is driven by the PP-SLIDE box plots. When the PP-SLIDE box plots suggest that the medians of the two populations do not exhibit any notable difference, we take $H_1 : p_{u_v,j} \neq p_{v,j}\ and\ g_{u_v,j} \neq g_{v,j}$. The failure to reject the null hypothesis in this case will establish evidence in favor of no statistical difference between the two populations for marker j. On the other hand, when the PP-SLIDE box plots demonstrate that the median of population $\{u_v\}$ is higher than that of $\{v\}$ (or the median of $\{v\}$ is higher than that of $\{u_v\}$) then we take $H_1 : p_{u_v,j} \neq p_{v,j}\ and\ g_{u_v,j} > g_{v,j}$ or vice versa. A rejection of the null hypothesis in this case establishes evidence in favor of statistical difference between the two populations for marker j.

One can view the null hypothesis H_0 as $H_0 = H_0^1 \cap H_0^2$ where $H_0^1: p_{u_v,j} = p_{v,j}$ and $H_0^2: g_{u_v,j} = g_{v,j}$. For testing H_0^1, we use a two-sided Binomial test of equality of proportions, and for testing H_0^2 we use a two sample Wilcoxon test. The $p-$value of the intersection test, q_j, is taken to be the minimum of the $p-$values obtained from testing H_0^1 and H_0^2. When testing for all the M markers simultaneously, the q_j's are corrected for multiplicity using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995).

Table S1: List of CyTOF antibodies used in study. Antibodies were either purchased from the indicated vendor or prepared in-house using commercially available MaxPAR conjugation kits per manufacturer's instructions (Fluidigm).

Antigen Target	Clone	Elemental Isotope	Vendor
HLA-DR	TÜ36	Cd112	Life Technologies
Live/Dead (In-DOTA maleimide)*		In115	Sigma-Aldrich
CD161	HP-3G10	139La	In-house
CD49d	9F10	141Pr	Fluidigm
CD11a	HI111	142Nd	Fluidigm
CD57	HCD57	143Nd	In-house
CCR5§	NP-6G4	144Nd	Fluidigm
CD31	WM59	145Nd	Fluidigm
CD8	RPA-T8	146Nd	Fluidigm
CD7	CD7-6B7	147Sm	Fluidigm
CD19*	HIB19	148Nd	In-house
CCR4	205410	149Sm	Fluidigm
HSA (mouse)⌘	M1/69	150Nd	Fluidigm
ICOS	DX29	151Eu	Fluidigm
CCR10	1B5	152Sm	In-house
CD62L§	DREG-56	153Eu	Fluidigm
CD3*	UCHT1	154Sm	Fluidigm
CCR6	11A9	155Gd	In-house
CD29	TS2/16	156Gd	Fluidigm
CCR2	K036C2	157Gd	In-house
OX40	ACT35	158Gd	Fluidigm
CD90	5E10	159Tb	Fluidigm
CD28§	CD28.2	160Gd	Fluidigm
CD45RO	UCHL1	161Dy	In-house
CD69	FN50	162Dy	Fluidigm
CXCR3	G025H7	163Dy	Fluidigm
PD-1	EH12.1	164Dy	In-house
CD127	A019D5	165Ho	Fluidigm
CCR7	150503	166Er	In-house
CD27	L128	167Er	Fluidigm
CD18	TS1/18	168Er	In-house
CD45RA	HI100	169Tm	Fluidigm
CD54	HA58	170Er	Fluidigm
CXCR5	51505	171Yb	Fluidigm
CD38	HIT2	172Yb	Fluidigm
a4b7	Act1	173Yb	In-house
CD4§	SK3	174Yb	Fluidigm
CXCR4	12G5	175Lu	Fluidigm
CD25	M-A251	176Yb	In-house
DNA*		191Ir & 193Ir	Fluidigm

*Not used in t-SNE and clustering because used in gating scheme
⌘Not used in t-SNE and clustering because infection marker
§The four markers excluded during t-SNE analysis as known proteins modulated by HIV infection

Table S2: List of flow cytometry antibodies used in study

Antigen Target	Clone	Vendor (catalog number)
v450 mouse anti-human CD3	UCHT1	BD (#560365)
FITC mouse anti-human CD3	UCHT1	BD (#555332)
APC-H7 mouse anti-human CD3	SK7	BD (#641397)
PerCP-Cy5.5 Mousse Anti-Human CD19	HIB19	BD (#561295)
APC mouse anti-human HLA-ABC	G46-2.6	BD (#555555)
APC-H7 mouse anti-human CD8	SK1	BD (#561423)
PE-Cyanine5.5 mouse anti-human CD4	SK3	eBioscience (35-0047-41)
BUV mouse anti-human CD4	SK3	BD (#564305)
ECD mouse anti-human CD45RO	UCHL1	Beckman Coulter (#IM2712U)
PE-CF594 mouse anti-human CD45RO	UCHL1	BD (#562299)
BUV395 mouse anti-human CD45RA	HI100	BD (#740298)
APC mouse anti-human CD45RA	HI100	BD (#561884)
PE-conjugated mouse anti-human OX40	L106	BD (#340420)
PE-cyanine7 mouse anti-human CD57	TB01	eBioscience (25-0577-42)
BV-421 mouse anti-human CD127	HIL-7R-M21	BD (#562436)
BV650 mouse anti-human CD127	HIL-7R-M21	BD (#563225)
Alexa Fluor 647 mouse anti-human CD127	HIL-7R-M21	BD (#558598)
PE rat antimouse CD24 (HSA)	M1/69	BD (#553262)
FITC rat anti-mouse CD24 (HSA)	M1/69	BD (#553261)

References

Alberti, M.O., Jones, J.J., Miglietta, R., Ding, H., Bakshi, R.K., Edmonds, T.G., Kappes, J.C., and Ochsenbauer, C. (2015). Optimized Replicating Renilla Luciferase Reporter HIV-1 Utilizing Novel Internal Ribosome Entry Site Elements for Native Nef Expression and Function. AIDS Res Hum Retroviruses.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met *57*, 289-300.

Cavrois, M., Neidleman, J., and Greene, W.C. (2014). HIV-1 Fusion Assay. Bio Protoc *4*.

Cover, T.M., and Hart, P.E. (1967). Nearest Neighbor Pattern Classification. Ieee T Inform Theory *13*, 21-+.

Deleage, C., Wietgrefe, S.W., Del Prete, G., Morcock, D.R., Hao, X.P., Piatak, M., Jr., Bess, J., Anderson, J.L., Perkey, K.E., Reilly, C.*, et al.* (2016). Defining HIV and SIV Reservoirs in Lymphoid Tissues. Pathog Immun *1*, 68-106.

Devroye, L., Gyorfi, L., and G., L. (2013). A probabilistic theory of pattern recognition. Springer Science & Business Media *31*.

Fay, M.P., and Proschan, M.A. (2010). Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. Stat Surv *4*, 1-39.

Gesner, M., Maiti, M., Grant, R., and Cavrois, M. (2014). Fluorescence-linked Antigen Quantification (FLAQ) Assay for Fast Quantification of HIV-1 p24Gag. Bio Protoc *4*.

He, J., Choe, S., Walker, R., Di Marzio, P., Morgan, D.O., and Landau, N.R. (1995). Human immunodeficiency virus type 1 viral protein R (Vpr) arrests cells in the G2 phase of the cell cycle by inhibiting p34cdc2 activity. J Virol *69*, 6705-6711.

Qiu, P., Simonds, E.F., Bendall, S.C., Gibbs, K.D., Jr., Bruggner, R.V., Linderman, M.D., Sachs, K., Nolan, G.P., and Plevritis, S.K. (2011). Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. Nat Biotechnol *29*, 886-891.

Sen, N., Mukherjee, G., Sen, A., Bendall, S.C., Sung, P., Nolan, G.P., and Arvin, A.M. (2014). Single-cell mass cytometry analysis of human tonsil T cell remodeling by varicella zoster virus. Cell reports *8*, 633-645.

Shishido, T., Wolschendorf, F., Duverger, A., Wagner, F., Kappes, J., Jones, J., and Kutsch, O. (2012). Selected drugs with reported secondary cell-differentiating capacity prime latent HIV-1 infection for reactivation. J Virol *86*, 9055-9069.

Spitzer, M.H., Gherardini, P.F., Fragiadakis, G.K., Bhattacharya, N., Yuan, R.T., Hotson, A.N., Finck, R., Carmi, Y., Zunder, E.R., Fantl, W.J.*, et al.* (2015). IMMUNOLOGY. An interactive reference framework for modeling a dynamic immune system. Science (New York, NY *349*, 1259425.

Supplemental Figures & Figure Legends

Figure S1

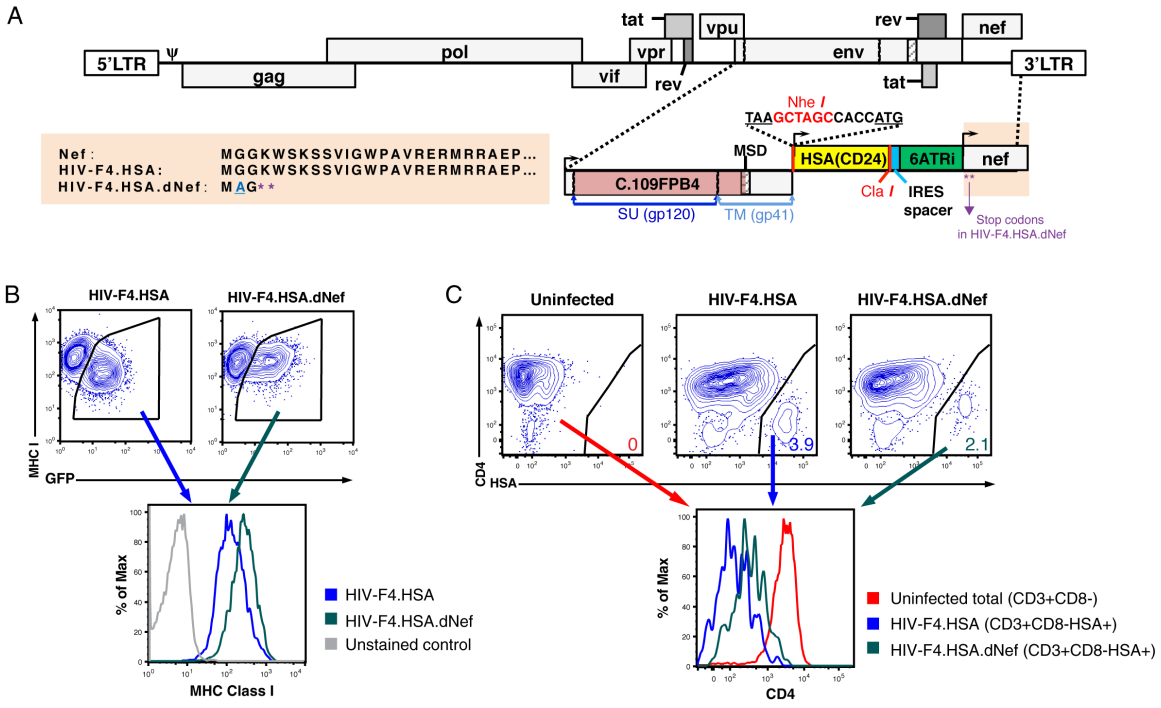


Figure S1 | Construction and characterization of reporter HIV-1 with CCR5-tropic envelope 109FPB4, related to Experimental Procedures.

(A) Schematic of HIV-F4.HSA provirus and its Nef-deficient counterpart HIV-F4.HSA.dNef. The pNL-LucR.6ATRi-Env.ecto proviral plasmid backbone (Alberti et al., 2015) contains an IRES spacer region (blue) preceding a modified EMCV IRES element, 6ATRi (green), that drives the expression of wildtype Nef. This construct was modified to encode the HSA reporter and the ectodomain of a primary CCR5-tropic Env (109FPB4) (Derdeyn et al., 2004). The HSA open reading frame (encoding murine CD24) was inserted into the *Nhe I* and *Cla I* cloning site (yellow). The coding region for the 109FPB4 Env ectodomain (pink) consisting of SU/gp120 and the extracellular portion of TM/gp41 up to the membrane-spanning domain (MSD, hatched box) of 109FPB4 was inserted as previously described (Alberti et al., 2015). Both the signal peptide and the cytoplasmic tail of Env are encoded by the backbone HIV-1 strain, NL4-3. HIV-F4.HSA.dNef was generated by combining a Gly_2→Ala mutation with 2 premature stop codons in place of

nef codons 4 and 5 (**), resulting in loss of Nef expression, as outlined in the amino acid alignment on the left (tan box). Arrows indicate translation start sites.

(B) Confirmation of Nef activity as assessed by MHC class I downregulation. J2574-R5 cells, a T-cell line that expresses GFP under the control of the HIV-1 LTR (Shishido et al., 2012), were infected for 2 d with HIV-F4.HSA or HIV-F4.HSA.dNef resulting in 39% and 30% of GFP+ cells, respectively. MHC class I expression was examined in GFP+ cells by immunostaining. The overlaid histogram plots show lower levels of MHC class I expression in cells infected with HIV-F4.HSA (blue) than in those infected with HIV-F4.HSA.dNef (green).

(C) Confirmation of Nef activity by CD4 downregulation. HLACs were infected for 3 d with HIV-F4.HSA or HIV-F4.HSA.dNef. The cells were then immunostained with antibodies against CD3, CD4, CD8, and HSA. Contour plots illustrate the proportion of HSA+ cells in cultures infected with HIV-F4.HSA and HIV-F4.HSA.dNef. The histogram plot shows the overlay of uninfected cells (red), cells infected with HIV-F4.HSA (blue), and cells infected with HIV-F4.HSA.dNef (green), analyzed for CD4 expression.

Figure S2

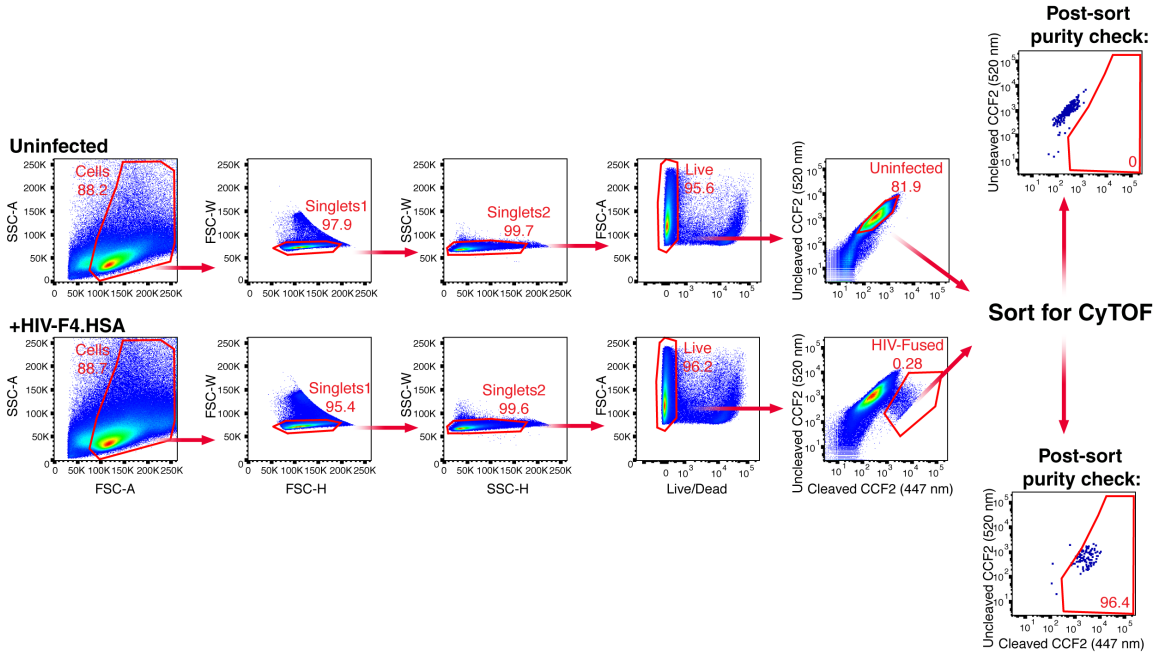


Figure S2 | Gating strategy for sorting cells that support HIV entry, related to Fig. 1.

Gating strategy employed to sort HIV-fused HLACs. HLACs were mock-treated or incubated with BlaM-Vpr-containing HIV-F4.HSA virions for 2 h, loaded with CCF2, and assessed for viral fusion by the CCF2 green-to-blue color conversion. Cells were sorted according to the gating strategy illustrated. The gates corresponding to HIV-fused cells were always set on the uninfected samples, where <0.01% of the cells exhibited a shift in fluorescence. Post-sort purity analyses were performed systematically on a small fraction of sorted cells.

Figure S3

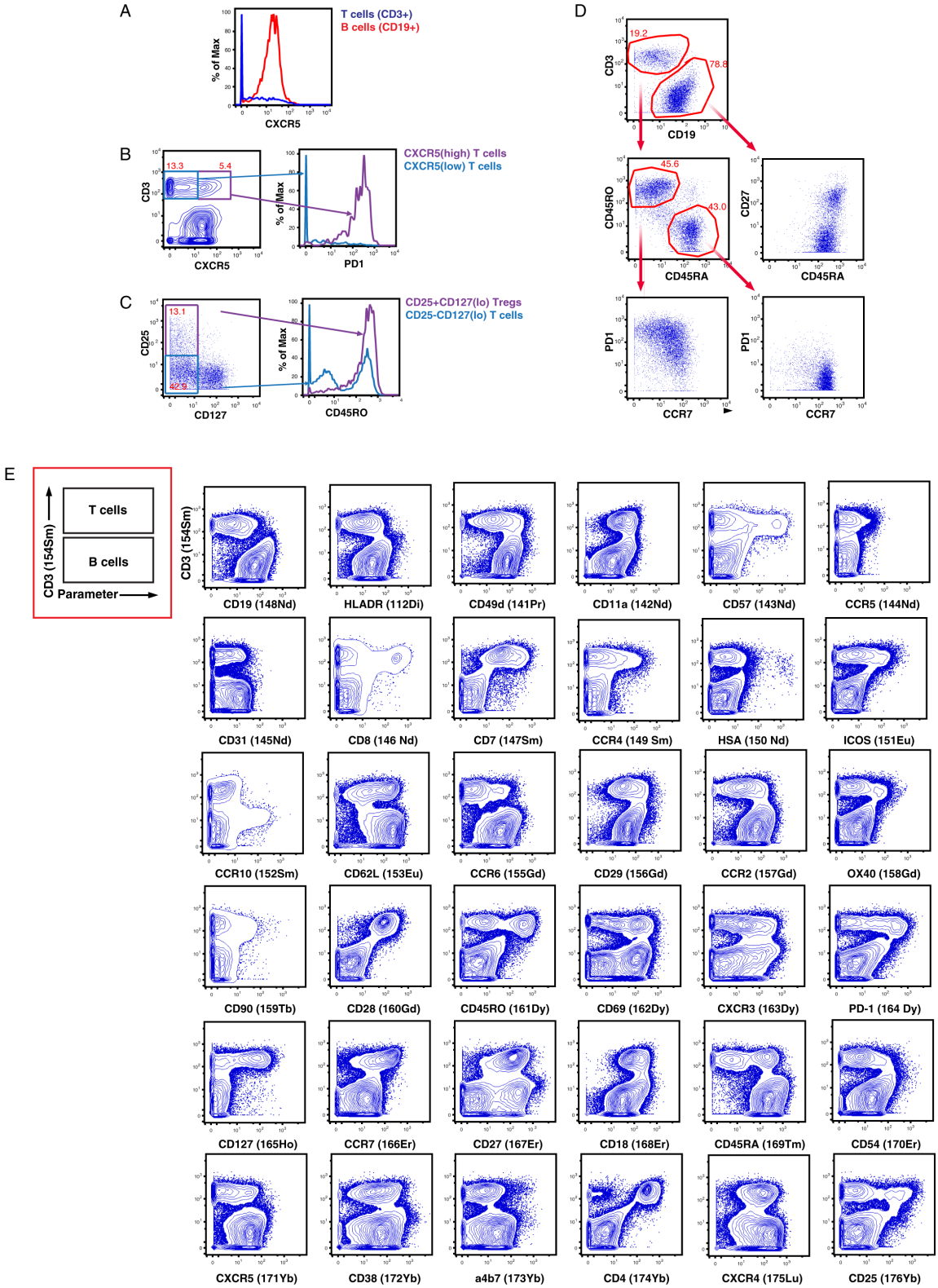

A
T cells (CD3+)
B cells (CD19+)
% of Max
CXCR5
B
CXCR5(high) T cells
CXCR5(low) T cells
13.3
5.4
CD3
CXCR5
% of Max
PD1
C
CD25+CD127(lo) Tregs
CD25-CD127(lo) T cells
13.1
42.9
CD25
CD127
% of Max
CD45RO
D
19.2
78.8
CD3
CD19
45.6
43.0
CD45RO
CD45RA
CD27
CD45RA
PD1
CCR7
PD1
CCR7
E
T cells
B cells
CD3 (154Sm)
Parameter
CD3 (154Sm)
CD19 (148Nd)
HLADR (112Di)
CD49d (141Pr)
CD11a (142Nd)
CD57 (143Nd)
CCR5 (144Nd)
CD31 (145Nd)
CD8 (146 Nd)
CD7 (147Sm)
CCR4 (149 Sm)
HSA (150 Nd)
ICOS (151Eu)
CCR10 (152Sm)
CD62L (153Eu)
CCR6 (155Gd)
CD29 (156Gd)
CCR2 (157Gd)
OX40 (158Gd)
CD90 (159Tb)
CD28 (160Gd)
CD45RO (161Dy)
CD69 (162Dy)
CXCR3 (163Dy)
PD-1 (164 Dy)
CD127 (165Ho)
CCR7 (166Er)
CD27 (167Er)
CD18 (168Er)
CD45RA (169Tm)
CD54 (170Er)
CXCR5 (171Yb)
CD38 (172Yb)
a4b7 (173Yb)
CD4 (174Yb)
CXCR4 (175Lu)
CD25 (176Yb)

Figure S3 | Examples of CyTOF antibody validation, related to Experimental Procedures.

(A) Validation of the CXCR5 antibody by confirming differential expression on B and T cells. CXCR5, a chemokine receptor that directs lymphocytes to B-cell zones of the lymph node, is expressed at uniformly high levels on tonsillar B cells (red, as identified by CD19 expression), and in a more widely distributed pattern in T cells (blue, as identified by CD3 expression).

(B) Validation of the CXCR5 antibody by confirming co-expression with PD1. T cells expressing high (purple) or low (light blue) levels of CXCR5 were examined for expression levels of PD1. As expected, cells expressing high levels of PD1 were enriched among cells expressing high levels of CXCR5, consistent with CXCR5 and PD1 as markers of T follicular helper cells (Tfh).

(C) Validation of CD25 and CD127 antibody staining of T cells. CD45RO expression was examined within Tregs (CD25+CD127(lo)) and CD25-CD127- T cells. CD25+CD127(lo) cells, which exhibit regulatory T-cell activity (Seddiki et al., 2006), are primarily CD45RO+, consistent with prior reports (Booth et al., 2010).

(D) Validation of CD45RO, CD45RA, CD27, PD1, and CCR7 antibodies on T- and B-cell populations. Tonsillar T and B cells express the expected distribution of CD45RO, CD45RA, CD27, PD1, and CCR7. B cells (CD19+) are CD45RA+ and exhibit bimodal CD27 expression. T cells (CD3+) include memory cells (CD45RO+CD45RA-) which exhibit both high and low levels of PD1 and CCR7 expression, and naïve cells (CD45RO-CD45RA+) which are mostly $PD1^{low}CCR7^{high}$. These data are consistent with published literature (Campbell et al., 2001; Morikawa et al., 1991; Steiniger et al., 2005; Wong et al., 2015; Yu et al., 2015).

(E) Differential expression of antigens on T cells and B cells as assessed by CyTOF. The first two-dimensional plot of the series distinguishes T cells (CD3+CD19-) from B cells (CD3-CD19+). Given that CD19 and CD3 expression is mutually exclusive, using CD3 alone on the y-axis allows differentiation of T cells (top half) from B cells (bottom half) (see schematic in the red box). Each antibody of our panel was therefore validated by standard two-dimensional plots showing expression of CD3 (y-axis) versus each indicated antibody (x-axis). The observed expression patterns are consistent with known expression patterns on T cells and B cells. For example, lymphoid B cells are known to express high levels of CD62L, CCR6, and CD45RA, while subsets of T cells are known to express high levels of CD57, CD28, and CD25.

Figure S4

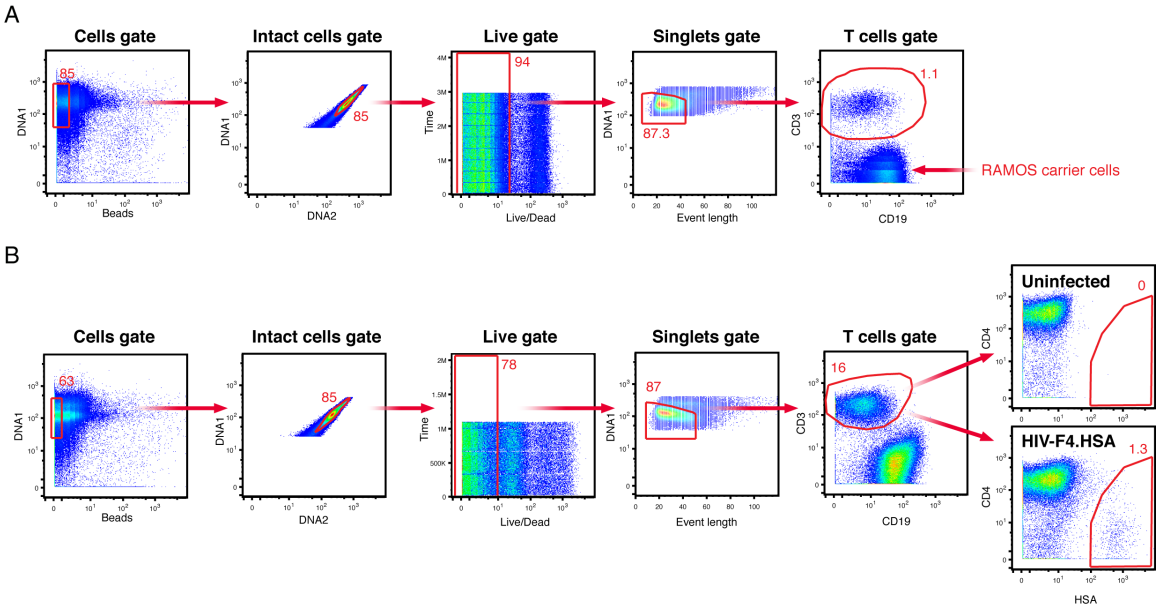


Figure S4 | Gating strategy for CyTOF data analysis, related to Fig. 1 and 2.

(A) Gating strategy for CyTOF analysis of cells supporting fusion to HIV-F4.HSA. Sorted cells combined with carrier RAMOS B cells were stained with a cocktail of metal-conjugated antibodies, and analyzed on a CyTOF2 instrument. Shown is the gating strategy used to exclude RAMOS cells and to identify live, intact CD3+ T cells.

(B) Gating strategy used for CyTOF analysis of cells productively infected with HIV-F4.HSA. Infected cells were defined as those that expressed high levels of HSA, most of which had downregulated CD4.

Figure S5

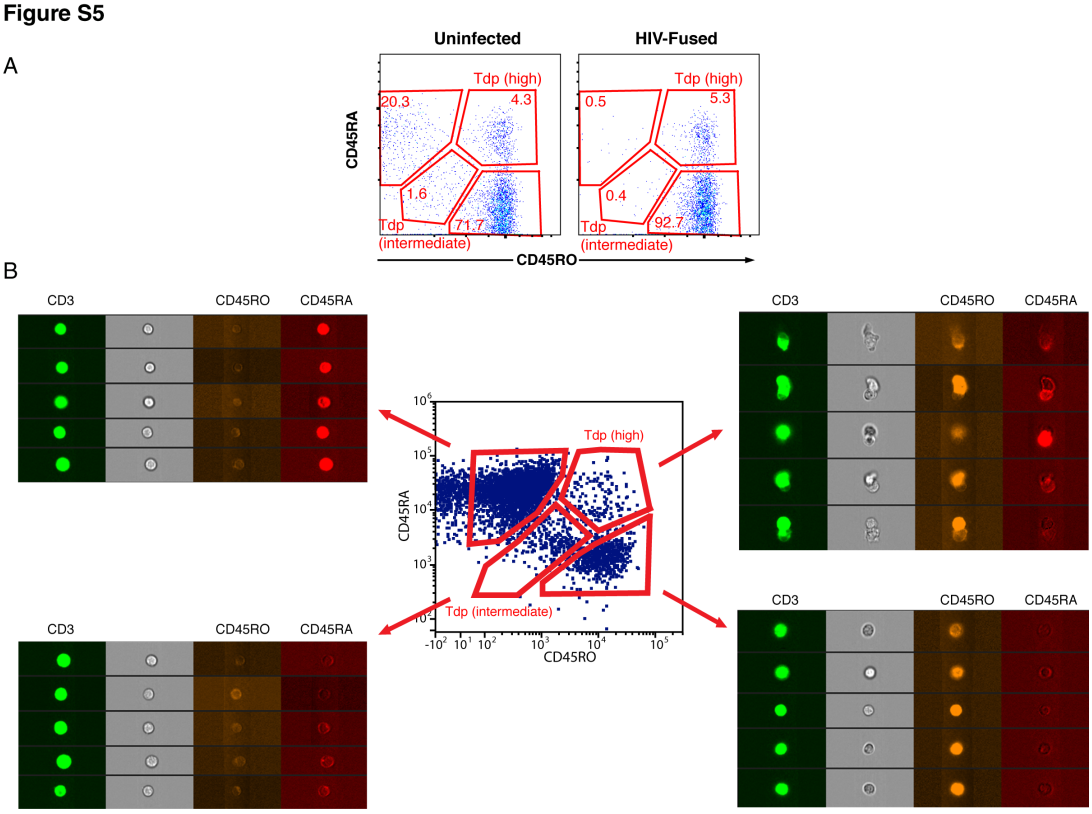


Figure S5 | Characterization of Tdp cells observed in HLACs, related to Fig. 1.

(A) Identification of two types of Tdp dually expressing CD45RA and CD45RO in HLAC cultures. Displayed are plots from the CyTOF fusion datasets showing expression levels of CD45RA and CD45RO among T cells. Tdp^{high} were relatively abundant (5.3%) among HIV-fused cells, while Tdp^{intermediate} were present at low frequency (0.4%).

(B) Characterization of Tdp^{high} and Tdp^{intermediate} using the ImagestreamX MarkII from Amnis. HLACs were stained for CD3, CD45RA, and CD45RO. The center dot plot depicts expression of CD45RA and CD45RO among live T cells. Representative images of the events depicted in the dot plot are shown. Note that Tdp^{high} were comprised primarily of doublets while Tdp^{intermediate}, CD45RA+CD45RO-, and CD45RA-CD45RO+ cells were comprised of singlets.

Figure S6

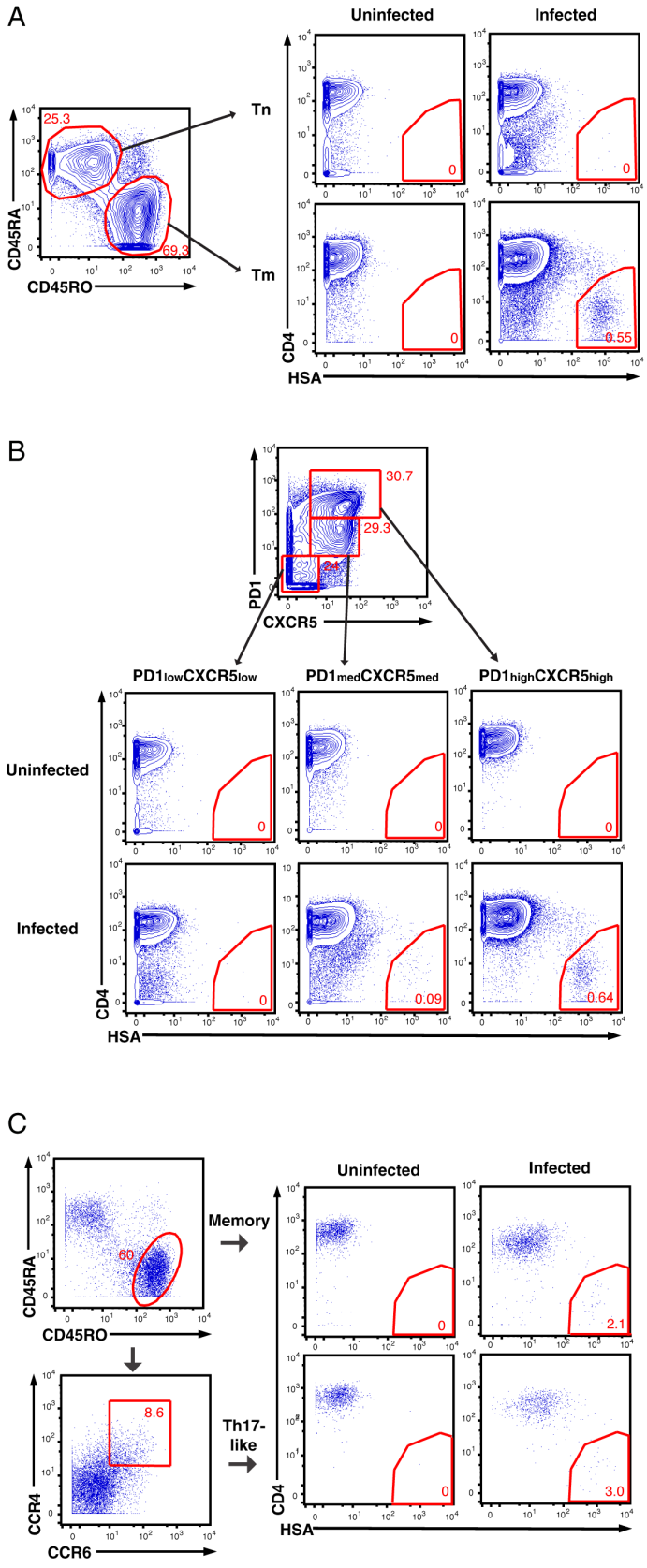

A
Uninfected
Infected
25.3
69.3
Tn
Tm
CD45RA
CD45RO
CD4
HSA
0
0
0
0.55
B
30.7
29.3
PD1
CXCR5
PD1lowCXCR5low
PD1medCXCR5med
PD1highCXCR5high
Uninfected
Infected
0
0
0
0
0.09
0.64
CD4
HSA
C
60
CD45RA
CD45RO
Memory
8.6
CCR4
CCR6
Th17-like
Uninfected
Infected
0
2.1
0
3.0
CD4
HSA

Figure S6 | HIV-F4.HSA preferentially infects Tm subsets, related to Fig. 2.

(A) HIV preferentially infects memory T cells. Live, singlet CD3+CD8- T cells were gated into Tn (CD45RA+CD45RO-) and Tm (CD45RA-CD45RO+) populations and analyzed for percentage of HIV-infected cells in mock-infected versus HIV-infected cultures. Gates correspond to infected (HSA+) cells that have down-regulated CD4.

(B) HIV infects germinal center (GC) Tfh more efficiently than non-GC Tfh and non-Tfh. Live, singlet CD3+CD8- T cells were gated into populations expressing low, medium, or high levels of PD1 and CXCR5 as indicated, and examined for percentage of HIV-infected cells in mock-infected versus HIV-infected cultures. Gates correspond to infected (HSA+) cells that have down-regulated CD4. The $PD1^{high}CXCR5^{high}$ cells correspond to GC Tfh, the $PD1^{med}CXCR5^{med}$ cells correspond to non-GC Tfh, and the $PD1^{low}CXCR5^{low}$ cells correspond to non-Tfh.

(C) Th17-like memory T cells tend to be infected at slightly higher levels compared to bulk memory CD4+ T cells. Live, singlet CD3+CD8- T cells were gated into Tm (CD45RA-CD45RO+) and Th17-like Tm (CD45RA-CD45RO+CCR4+CCR6+) populations and analyzed for percentage of HIV-infected (HSA+) cells in mock-infected versus HIV-infected cultures.

Figure S7

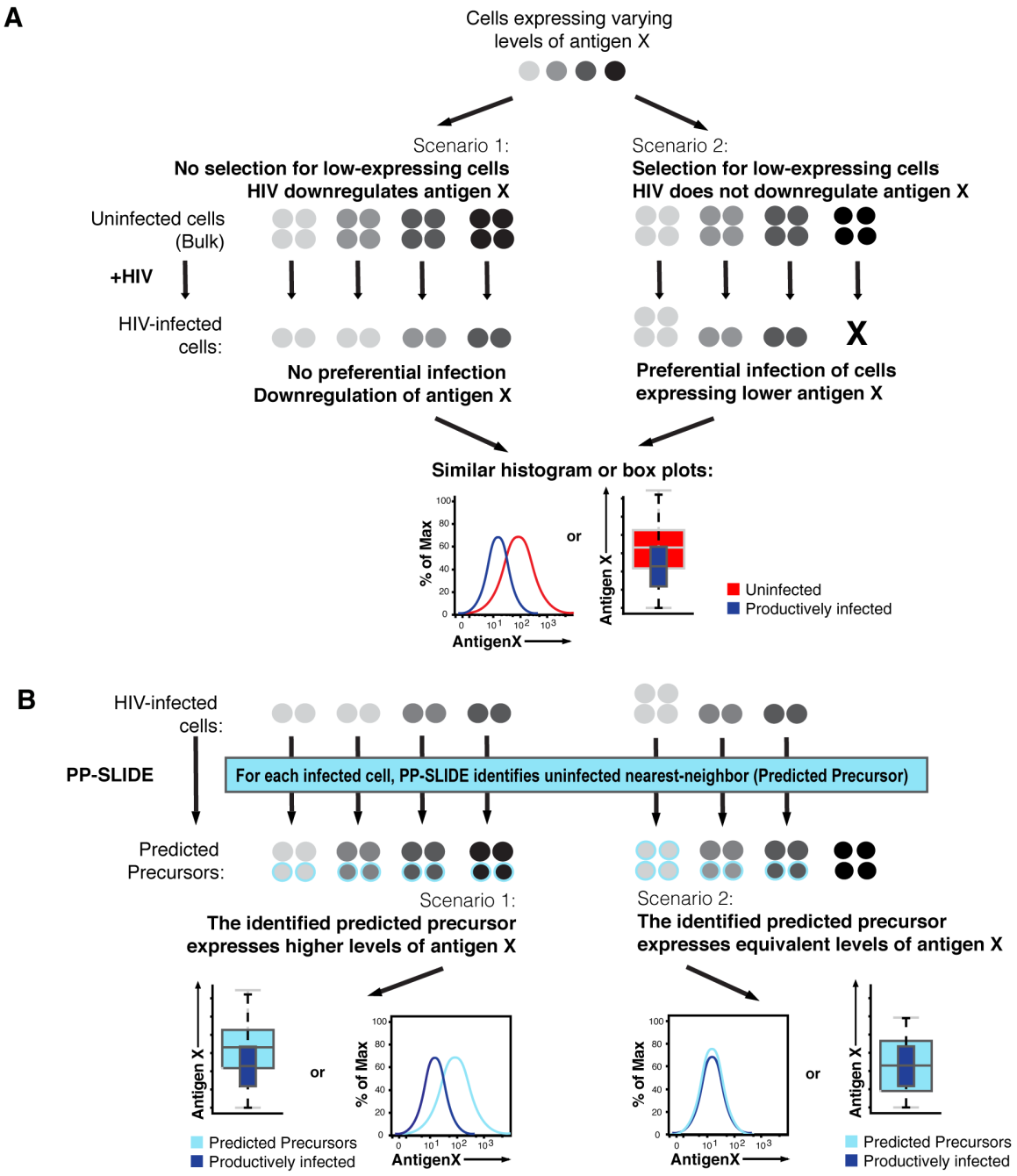


Figure S7 | PP-SLIDE to distinguish HIV-induced downregulation of an antigen from preferential infection of cells expressing low levels of the antigen, related to Fig. 4.

(A) A hypothetical example of cells expressing an antigen X at varying levels (with lowest levels represented as light grey and highest levels represented as black). In the "remodeling" scenario 1 (*left*),

HIV infects equally cells expressing different levels of antigen X, with infection leading to downregulation of antigen X by one increment, such that the infected cells now express lower levels of antigen X compared to their precursor state. In the "selection" scenario 2 (*right*), HIV preferentially infects cells with lower levels of antigen X. In this second scenario, HIV is unable to infect cells with the highest levels of antigen X and infects most frequently cells with the lowest levels of antigen X. Infection in this second scenario does not induce any downregulation of antigen X. In both scenarios, the overlay of uninfected and infected cells leads to similar-looking histograms and box plots where the distribution of infected cells (dark blue) exhibits lower antigen X expression compared to the distribution of uninfected cells (red).

(B) PP-SLIDE allows one to distinguish the two scenarios. For each infected cell, PP-SLIDE identifies its nearest neighbor in the uninfected population (cells highlighted by the light blue circles). Given that infected cells will likely retain most original features when transitioning from an uninfected to an infected state, the nearest neighbor uninfected cell identified from a high-dimensional dataset will likely be phenotypically similar to the precursor of the infected cell, and is therefore referred to as the predicted precursor cell. Comparing the infected cell population (dark blue) to the predicted precursor cell population (light blue), instead of to the bulk uninfected cell population (red), filters out the cells that HIV did not infect (e.g., the high expressers of antigen X in the "selection" scenario 2). The resulting histogram and box plots now differ between the two scenarios.

Figure S8

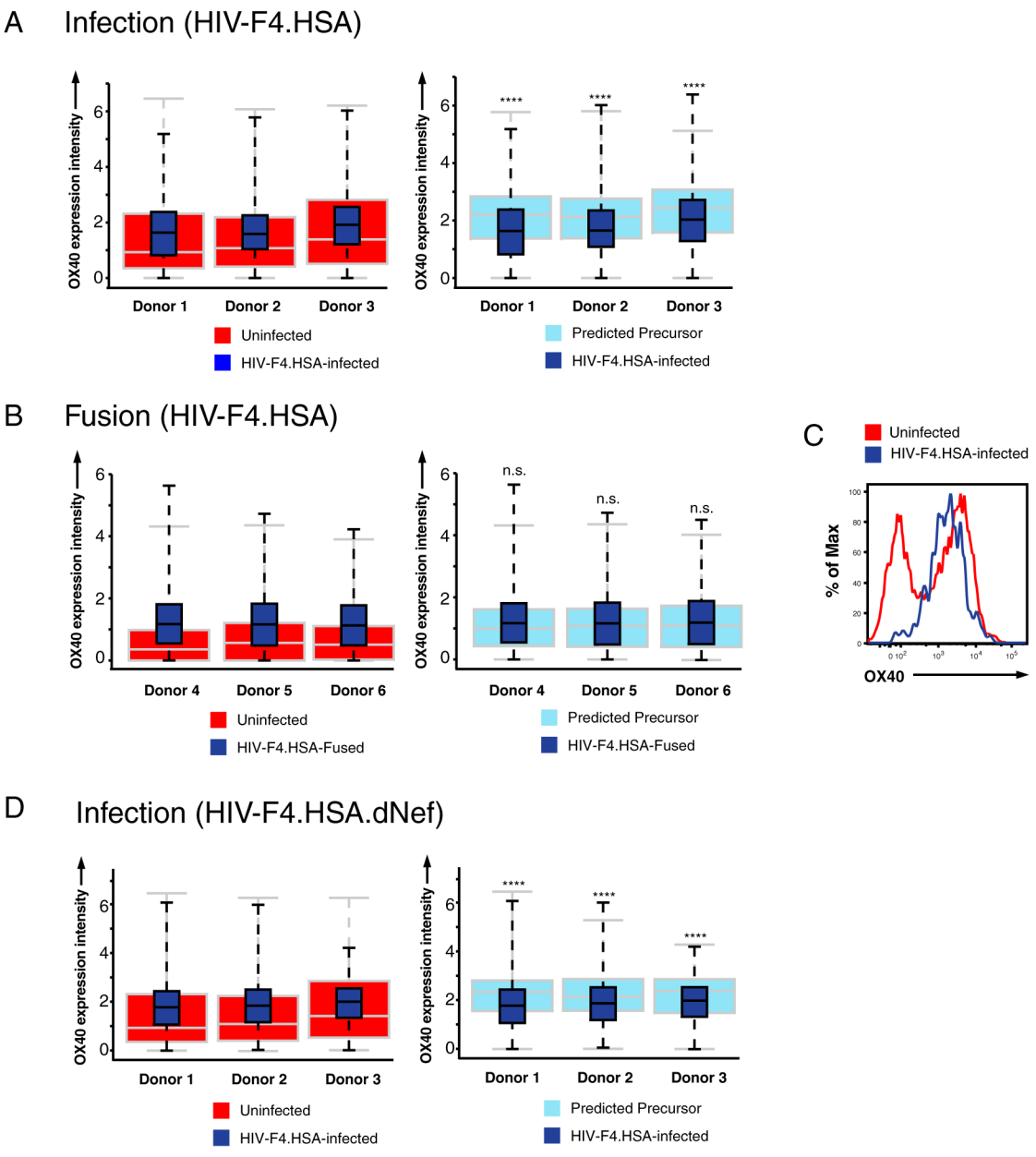


Figure S8 | PP-SLIDE analysis identifies OX40 as a receptor whose cell surface expression levels are altered by HIV infection, related to Fig. 4.

(A) PP-SLIDE analysis of OX40 levels on cells infected with HIV-F4.HSA. The plot on the left shows the standard analysis, where the expression levels of OX40 on infected cells (dark blue) is compared to that of bulk uninfected cells (red). The plot on the right shows the data subsequent to PP-SLIDE analysis, where the expression levels of OX40 on infected cells (dark blue) is instead compared to that of the predicted

precursors (light blue). Note that the MFI of OX40 is higher on infected cells relative to bulk uninfected cells (*left*) but significantly lower on infected cells relative to predicted precursors (*right*), suggesting that HIV-F4.HSA preferentially infects $OX40^{high}$ cells and then downregulates OX40 upon productive infection. ****$p<0.0001$ (see Statistics section in Supplemental Experimental Procedures for details).

(B) PP-SLIDE analysis of OX40 expression levels on cells that supported fusion to HIV-F4.HSA. The plot on the left shows the standard analysis, where the expression levels of OX40 on fused cells (dark blue) is compared to that of bulk uninfected cells (red). The plot on the right shows the data subsequent to PP-SLIDE analysis, where the expression levels of OX40 on HIV-fused cells (dark blue) is instead compared to that of the predicted precursors (light blue). Note that there are no significant differences in OX40 expression after PP-SLIDE analysis, demonstrating a lack of OX40 downregulation in fused cells 2 h post-infection. n.s.: non-significant.

(C) Standard flow cytometric analysis confirms the differences in OX40 expression on infected and uninfected cells. HLACs were mock-treated or infected for 4 d with HIV-F4.HSA, and analyzed by flow cytometry for cell surface levels of OX40. Uninfected cells (red) correspond to live, singlet CD3+CD8- cells, while HIV-F4.HSA-infected cells (blue) correspond to live, singlet CD3+CD8-HSA+ cells. Taken together with the data in panel *A*, the results suggest that the HIV-F4.HSA-infected cells originated from the $OX40^{high}$ subset of cells.

(D) PP-SLIDE analysis of OX40 expression levels on cells infected with HIV-F4.HSA.dNef. Plots are presented in the same format as panel *A*. ****$p<0.0001$.

All statistical analyses in this figure refer to box plots on the right-hand side comparing predicted precursors to either HIV-infected or HIV-fused cells.

Figure S9

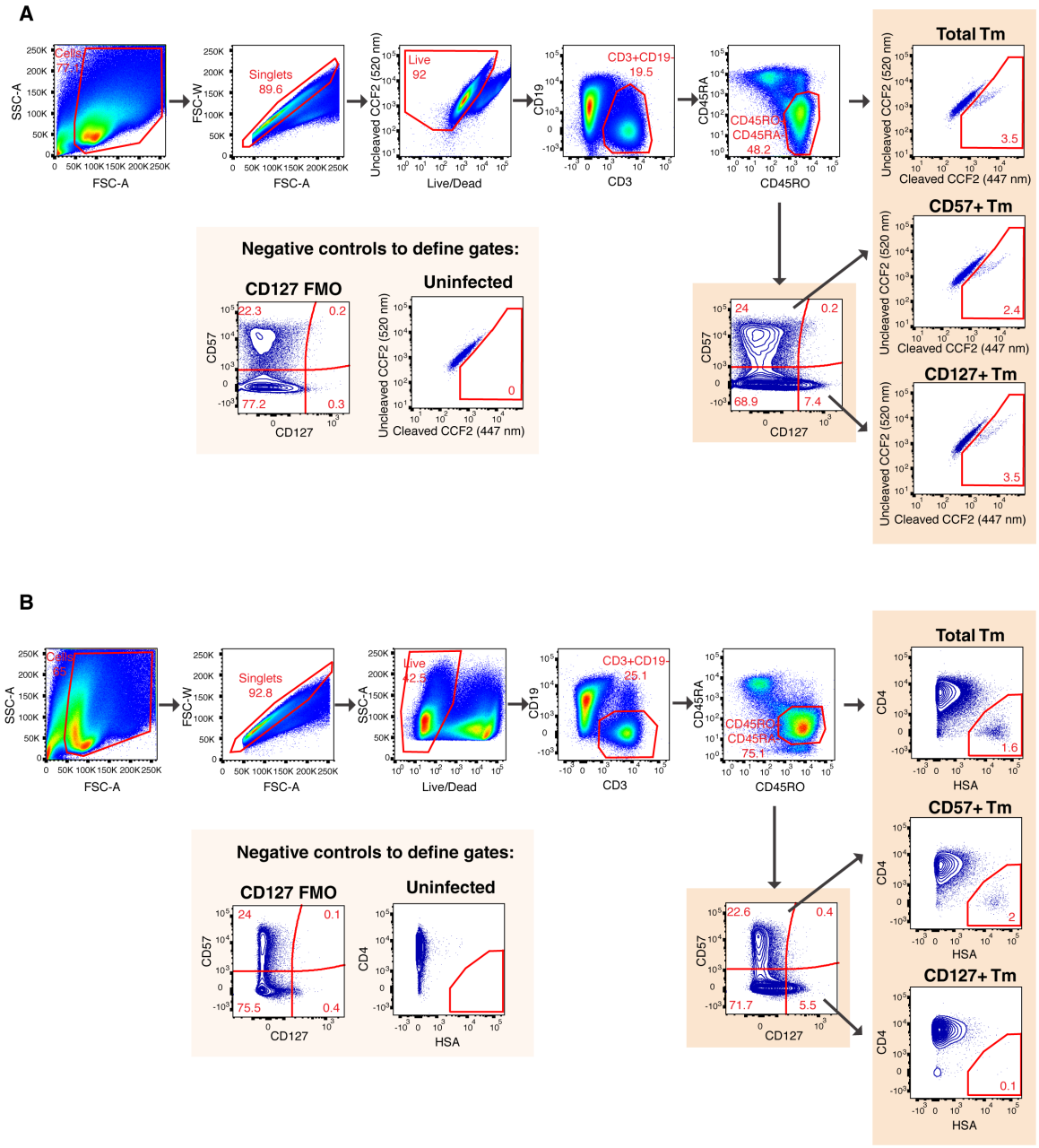


Figure S9 | Gating strategy for kinetic analysis of HIV infection of CD57+ and CD127+ Tm, related to Fig. 6.

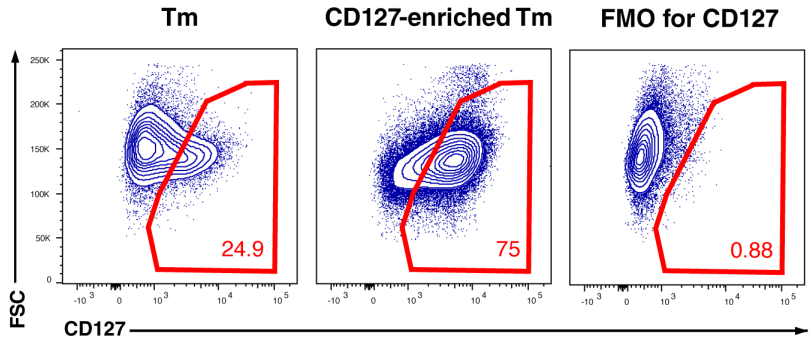
(A) Flow cytometric gating strategy to identify cells supporting HIV fusion among the bulk Tm population, and in Tm subsets expressing either CD57 or CD127. *Inset on lower left*: CD127 FMO was used to establish the gates defining CD127+ cells, while an uninfected sample was used to establish the gate

defining HIV-fused cells. Percentage of cells supporting fusion was measured among total Tm, CD57+ Tm, and CD127+ Tm.

(B) Flow cytometric gating strategy to identify productively-infected cells among the bulk Tm population, and in Tm subsets expressing either CD57 or CD127. *Inset on lower left:* CD127 FMO was used to establish the gates defining CD127+ cells, while an uninfected sample was used to establish the gate defining productively-infected cells. Percentage of infected cells was measured among total Tm, CD57+ Tm, and CD127+ Tm.

Figure S10

A



B

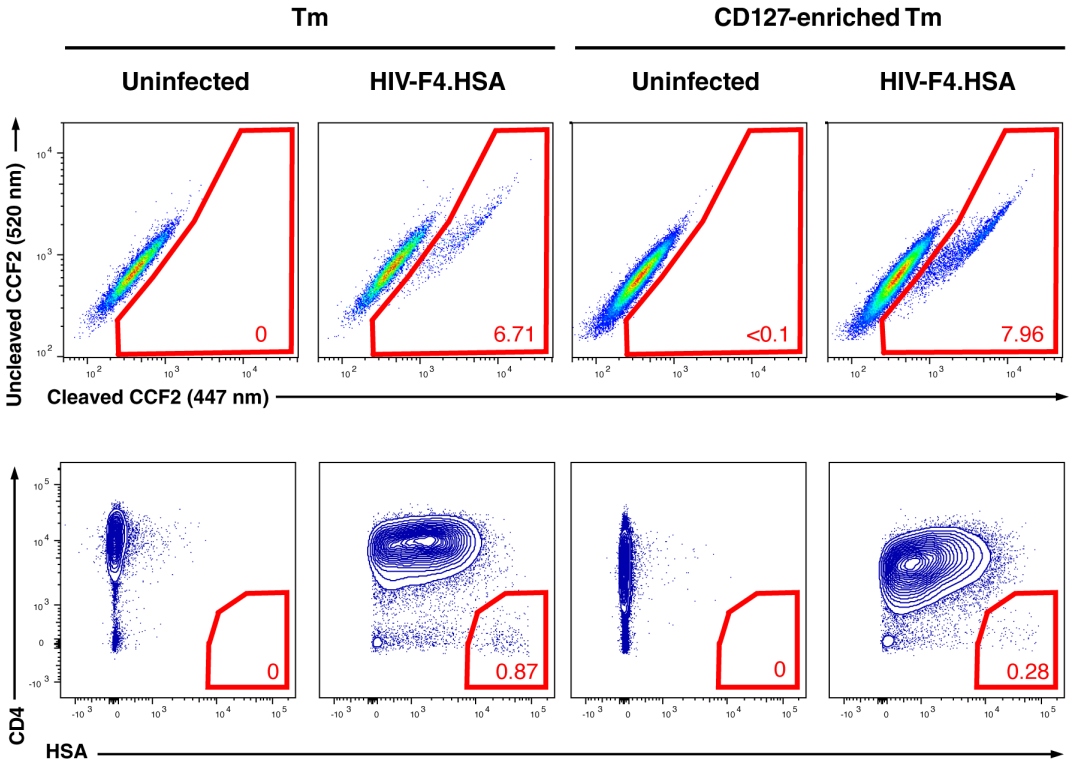


Figure S10 | Low levels of productive infection in Tm enriched for CD127+ cells, despite efficient viral entry, related to Fig. 6.

HLACs were depleted of B cells and naïve T cells by negative selection using CD45RA magnetic beads to isolate total Tm, and then positively selected for CD127+ cells to enrich for CD127+ Tm. Total Tm and CD127+ Tm were infected with BlaM-Vpr-containing HIV-F4.HSA. After 2 h, cells were washed and 25%

of the culture was analyzed for viral fusion, while the remaining 75% was cultured for 3 d after which levels of productive infection were assessed. Shown is one of two representative experiments (using cells from different donors) that gave similar results.

(A) Contour plots comparing expression of CD127 before and after CD127 enrichment as assessed by flow cytometry. Bulk or CD127-enriched HLACs were gated on live, singlet Tm. Note an enrichment of CD127+ Tm by ~3-fold (from ~25% to 75%) with the bead-based purification. The plot on the right shows the CD127 FMO, which was used to establish the gate defining CD127+ cells.

(B) HIV-F4.HSA efficiently fuses to Tm and CD127-enriched Tm but poorly replicates in the latter. The top row of plots depicts the levels of HIV fusion within the Tm or CD127-enriched Tm populations. The bottom row of plots depicts the levels of productive infection in these populations 3 d later. Comparison of the fusion and infection data suggests that CD127-enriched Tm supported high levels of HIV-F4.HSA entry but low levels of productive infection.

References

Alberti, M.O., Jones, J.J., Miglietta, R., Ding, H., Bakshi, R.K., Edmonds, T.G., Kappes, J.C., and Ochsenbauer, C. (2015). Optimized Replicating Renilla Luciferase Reporter HIV-1 Utilizing Novel Internal Ribosome Entry Site Elements for Native Nef Expression and Function. AIDS Res Hum Retroviruses.

Booth, N.J., McQuaid, A.J., Sobande, T., Kissane, S., Agius, E., Jackson, S.E., Salmon, M., Falciani, F., Yong, K., Rustin, M.H.*, et al.* (2010). Different proliferative potential and migratory characteristics of human CD4+ regulatory T cells that express either CD45RA or CD45RO. J Immunol *184*, 4317-4326.

Campbell, J.J., Murphy, K.E., Kunkel, E.J., Brightling, C.E., Soler, D., Shen, Z., Boisvert, J., Greenberg, H.B., Vierra, M.A., Goodman, S.B.*, et al.* (2001). CCR7 expression and memory T cell diversity in humans. J Immunol *166*, 877-884.

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

Morikawa, K., Oseko, F., and Morikawa, S. (1991). The role of CD45RA on human B-cell function: anti-CD45RA antibody (anti-2H4) inhibits the activation of resting B cells and antibody production of activated B cells independently in humans. Scandinavian journal of immunology *34*, 273-283.

Seddiki, N., Santner-Nanan, B., Martinson, J., Zaunders, J., Sasson, S., Landay, A., Solomon, M., Selby, W., Alexander, S.I., Nanan, R.*, et al.* (2006). Expression of interleukin (IL)-2 and IL-7 receptors discriminates between human regulatory and activated T cells. The Journal of experimental medicine *203*, 1693-1700.

Shishido, T., Wolschendorf, F., Duverger, A., Wagner, F., Kappes, J., Jones, J., and Kutsch, O. (2012). Selected drugs with reported secondary cell-differentiating capacity prime latent HIV-1 infection for reactivation. J Virol *86*, 9055-9069.

Steiniger, B., Timphus, E.M., Jacob, R., and Barth, P.J. (2005). CD27+ B cells in human lymphatic organs: re-evaluating the splenic marginal zone. Immunology *116*, 429-442.

Wong, M.T., Chen, J., Narayanan, S., Lin, W., Anicete, R., Kiaang, H.T., De Lafaille, M.A., Poidinger, M., and Newell, E.W. (2015). Mapping the Diversity of Follicular Helper T Cells in Human Blood and Tonsils Using High-Dimensional Mass Cytometry Analysis. Cell reports *11*, 1822-1833.

Yu, H., Khalid, M., Heigele, A., Schmokel, J., Usmani, S.M., van der Merwe, J., Munch, J., Silvestri, G., and Kirchhoff, F. (2015). Lentiviral Nef proteins manipulate T cells in a subset-specific manner. J Virol *89*, 1986-2001.

ADDENDUM OF CONFIRMATORY HIGH DIMENSIONAL ANALYSES

Introduction

Because CyTOF datsets generate a high volume of information, high-dimensional visualization tools and clustering approaches are essential for effective analysis and interpretation the datasets. This Addendum contains a set of four confirmatory analyses (grouped into four sections) using alternative high-dimensional data analysis approaches that were not detailed in the main manuscript. Although the information presented here is not necessary to interpret the findings presented in the main manuscript, it provides independent confirmation using alternative high-dimensional analysis approaches, and thereby strengths the conclusions drawn in our study.

Section 1 – Complementary high dimensional analyses to confirm findings related to HIV fusion.

In the main manuscript, we demonstrated HIV-fused cells localized to a region of the t-SNE plot corresponding to Tm and Tdp, but not Tn (Fig. 1B and C). To further confirm the selective entry of HIV into Tm and Tdp, we employed two other types of analyses: Principal Component Analysis (PCA) and agglomerative hierarchical clustering. PCA is a dimensionality reduction approach but unlike t-SNE, it uses a linear transformation to convert multidimensional data into a plot that can be visualized in three orthogonal directions that account for maximum variability of the data. As in a t-SNE plot, each dot in a PCA plot corresponds to a single cell, and cells judged as phenotypically similar by the algorithm are placed closer together on the PCA plot. PCA plots were generated using the same 35 parameters used for t-SNE analysis in Fig. 1, and colored to illustrate the locations of uninfected (red) and HIV-fused (blue) cells (Fig. A1A, *left*), or by expression levels of CD4, CD8, CD45RA, and CD45RO (Fig. A1A, *right*). Consistent with the t-SNE analysis, PCA segregated the main subsets of CD4+ T cells [the Tn (CD4+CD45RA-CD45RO+), Tdp (CD4+CD45RA+CD45RO+), and Tm (CD4+CD45RA+CD45RO+) subsets] and revealed that Tm and Tdp were the major populations of cells supporting HIV fusion.

A deeper characterization of subpopulations can be obtained by agglomerative hierarchical clustering to create a heatmap that groups the cells based on surface marker expression (Sen et al., 2014) (Fig. A1B). Events corresponding to uninfected and HIV-fused cells were clustered based on the expression of the 35 parameters used for the t-SNE analysis in Fig. 1 by the Euclidean dissimilarity test and

Figure A1

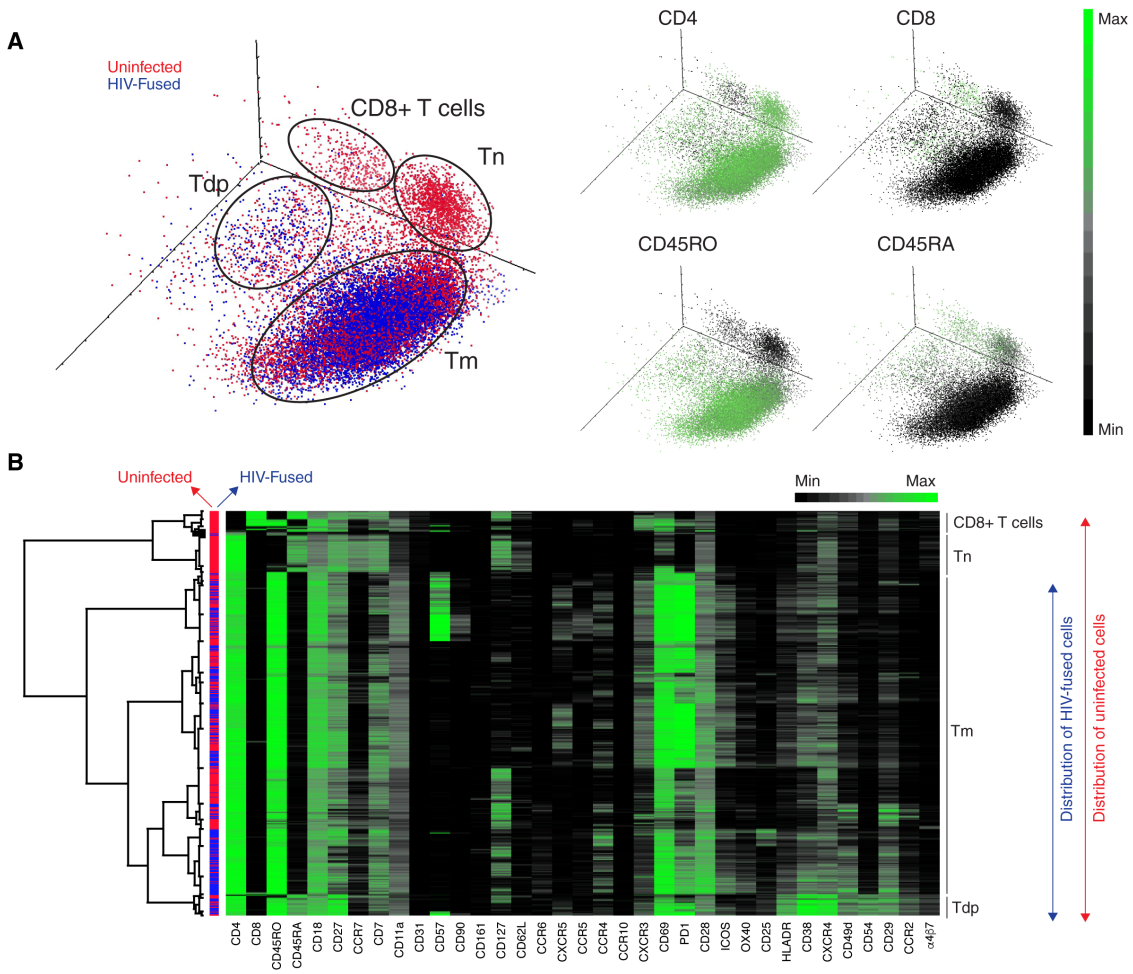


Figure A1 | Principal component analysis (PCA) and agglomerative clustering of uninfected and HIV-fused cells.
(A) PCA was used to convert the multidimensional data cloud into three orthogonal directions that account for the maximum variability of the data. *Left*: 3D representation by PCA, showing distribution of uninfected (red) and HIV-fused (blue) cells. *Right*: PCA plots colored for expression of CD4, CD8, CD45RA, and CD45RO [ranging from black (minimum) to green (maximum)], which were used to define the populations circled on the left plot. The coordinates of each cell are defined by the first three principal components based on the expression of 35 surface proteins (Table SI). Areas of the plot corresponding to CD8+ T cells (CD4-CD8+), naïve CD4+ T cells (Tn: CD4+CD8-CD45RA+CD45RO-), CD4+ T cells expressing both isoforms of CD45 (Tdp: CD4+CD8-CD45RA+CD45RO+), and memory CD4+ T cells (Tm: CD4+CD8-CD45RA-CD45RO+) are indicated. Note that HIV-fused cells overlap with regions occupied by uninfected Tdp and Tm. Results are representative of one of two donors analyzed in a similar fashion.
(B) Heatmap displaying agglomerative hierarchical clustering of uninfected (red) and HIV-fused (blue) cells. Each row represents a single cell, and each column the indicated surface marker. Dendrograms on the left correspond to the Euclidean distance between the clustered cells. Areas of the heatmap corresponding to CD8+ T cells, Tn, Tdp, and Tm are highlighted on the right based on expression patterns of CD4, CD8, CD45RA, and CD45RO. Note that the CD8+ and Tn populations are largely absent among the fused cells. Results are representative of one of two donors analyzed in a similar fashion.

Ward's minimum variance method. Expression levels of each antigen were represented by heatmap coloring, with rows corresponding to individual cells and columns corresponding to the listed antigen (Fig.

A1B). The heatmap confirms t-SNE and PCA analyses by showing that HIV-fused cells were distributed throughout the population of Tm and Tdp, and rarely observed among Tn or CD8+ T cells.

To demonstrate reproducibility of these findings, we returned to analysis by t-SNE. The t-SNE plots presented in Fig. 1 were generated with the FlowJo software that displays user-defined gated populations of cells onto the plots. This approach simplifies visualization as multiple populations can be readily displayed on a single t-SNE plot using different colors for different cell populations. Cytobank software offers a different method of visualizing t-SNE data. Instead of overlaying pre-gated populations of cells on a t-SNE plot, Cytobank-generated t-SNE plots can be colored in heatmap fashion according to expression levels of each marker, allowing a more objective assessment of the data. Given that each Cytobank-generated t-SNE plot can only display expression levels of one antigen per plot, multiple t-SNE plots need to be presented to identify a given cell population. We chose to present the FlowJo-generated t-SNE plots in the main manuscript as they illustrate our data with a smaller number of t-SNE plots, while in this Addendum we let the reader assess better the raw data by presenting the Cytobank-generated t-SNE plots. Fig. A2A shows four donors' worth of t-SNE plots displaying raw expression levels of CD4, CD8, CD45RA, and CD45RO for both uninfected and HIV-fused cells. In all four donors, HIV fused to Tdp and Tm and not Tn or CD8+ T cells, consistent with the findings reported in Fig. 1B and C. To confirm the presence of Tcm, Ttm, Th1-like, Th2-like, Th-17-like, Treg, and Tfh subsets among HIV-fused cells (Fig. 1D), we show that HIV-fused cells from all four donors include these subsets (Fig. A2B). All together, these secondary analyses strengthen the conclusions drawn from Fig. 1 of the main manuscript by reaching the same conclusion with linear dimensionality-reduction algorithms (PCA and agglomerative hierarchical clustering) and by demonstrating reproducibility among different donors.

Section 2 – Complementary high dimensional analyses to confirm findings related to HIV replication.

In the main manuscript, we demonstrated that unlike HIV-fused cells, infected cells segregated in a unique region of the t-SNE plot (Fig. 2B). Likewise, by PCA and agglomerative hierarchical clustering analyses, infected cells segregated apart from the uninfected cells (Fig. A3), demonstrating that linear dimensionality reduction methods also identify HIV-infected cells as phenotypically distinct from uninfected cells. In the main text, we further showed that when four receptors (CD4, CCR5, CD28, and

Figure A2

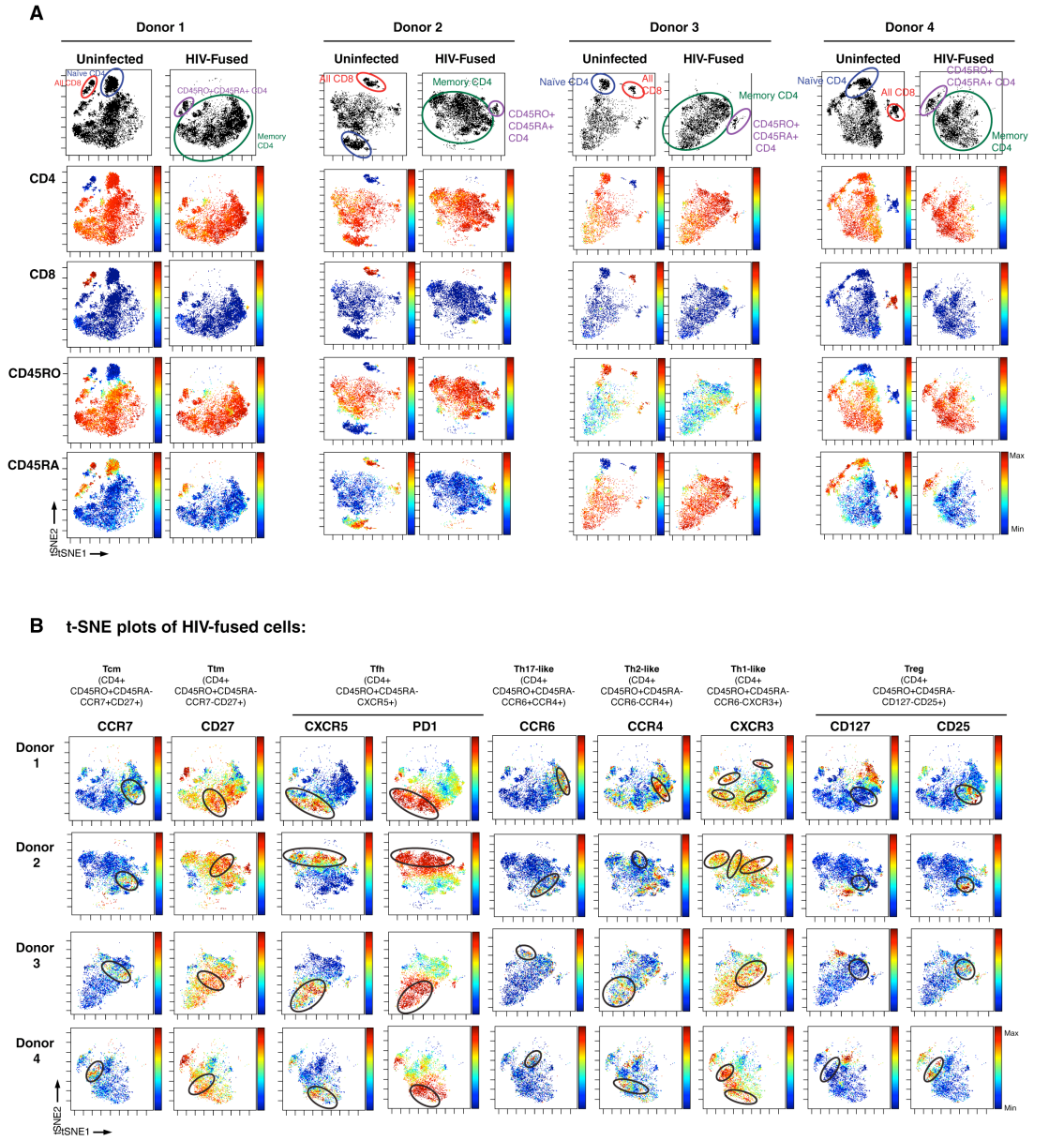


Figure A2 | HIV-F4.HSA fuses to a diverse array of memory CD4+ T cells as revealed by t-SNE maps displaying raw antigen expression values.

(A) t-SNE plots of CD4+ and CD8+ T-cell subsets in uninfected and HIV-fused cells from 4 donors based on expression of 35 surface markers. The top four pairs of t-SNE plots correspond to uninfected (*left* plots of the pairs) or HIV-fused (*right* plots of the pairs) tonsillar T cells. Each dot in a t-SNE plot corresponds to a T cell (CD3+CD19-), with similar cells being placed closer together on the plot relative to cells that are more phenotypically distinct, such that groups of similar cells form "islands". These black and white t-SNE plots were then color-coded for expression levels of CD4, CD8, CD45RA, or CD45RO, as indicated in the lower plots, ranging from blue (minimum) to red (maximum). Based on expression patterns of CD4, CD8, CD45RA, and CD45RO, ovals corresponding to CD8+ T cells (CD4-CD8+, red oval), Tn (CD4+CD8-CD45RA+CD45RO-, blue oval), Tdp (CD4+CD8-CD45RA+CD45RO+, purple oval), and Tm (CD4+CD8-CD45RA-CD45RO+, green oval) were then drawn. Note the absence of CD8+ T cells and Tn among the HIV-fused cells. Because the values on the axes of t-SNE plots have no intrinsic meaning, t-SNE visualizations can be arbitrarily rotated and therefore the various subpopulations are located in different orientations in the different donors.

(B) Identification of Tm subsets among the HIV-fused T cells in 4 donors. The t-SNE plots of HIV-fused cells were

analyzed for expression levels of CCR7, CD27, CXCR5, PD1, CCR6, CCR4, CXCR3, CD127, and CD25. Ovals correspond to populations of Tcm, Ttm, Tfh, Th17-like, Th2-like, Th1-like, and Treg subsets, and were drawn based on the expression patterns of the markers listed below each indicated subset.

CD62L) previously reported to be down-regulated by Nef (Garcia and Miller, 1991; Michel et al., 2005; Swigut et al., 2001; Vassena et al., 2015) are excluded from the t-SNE analysis, infected cells were dispersed primarily among the Tm region of the t-SNE plot (Fig. 2C). This was confirmed using a Cytobank-generated t-SNE analysis where raw expression values of CD4, CD8, CD45RA, and CD45RO were displayed on t-SNE plots generated with or without the four HIV-modulated receptors (Fig. A4A and B). The lack of CD4 expression among infected cells in both sets of t-SNE plots is consistent with the known down-modulation of this receptor during productive infection.

In the main manuscript, we further demonstrated that Th17-like cells and Tfh, known to be particularly susceptible to productive infection by HIV (Alvarez et al., 2013; Kohler et al., 2016), are present among tonsillar T cells infected with HIV-F4.HSA and include cells expressing a variety of activation markers (Fig. 2E-I). These findings were also observed in the Cytobank-generated t-SNE plots, where the raw expression values of the 8 activation markers of our panel are displayed, and which showed that infected cells express a diverse pattern of activation markers (Fig. A4C and D). Finally, in the main manuscript, we demonstrated that HIV-F4.HSA preferentially infects CD57+ Tm but not CD127+ Tm (Fig. 5). These findings were confirmed by t-SNE in a total of four donors (Fig. A5), where Cytobank-generated t-SNE plots displaying the raw expression values of CD57 and CD127 (Fig. A5, grey box), clearly show expression of CD57 but not CD127 by productively-infected cells.

Section 3 – Scaffold shows that HIV-infected cells do not exhibit features of CD127+ Tm, confirming PP-SLIDE's prediction that CD127 is not down-modulated upon infection.

The absence of CD127-expressing cells in the productively-infected population (Fig. 5 and 6) was surprising given that CD127+ Tm supported HIV fusion as efficiently as other Tm (Fig. 5). This apparent discordance can be explained in one of 2 ways: (1) HIV infection triggers CD127 downmodulation, leading to the absence of CD127-expressing HSA+ cells in the infected population, or (2) the CD127+ Tm cells

Figure A3

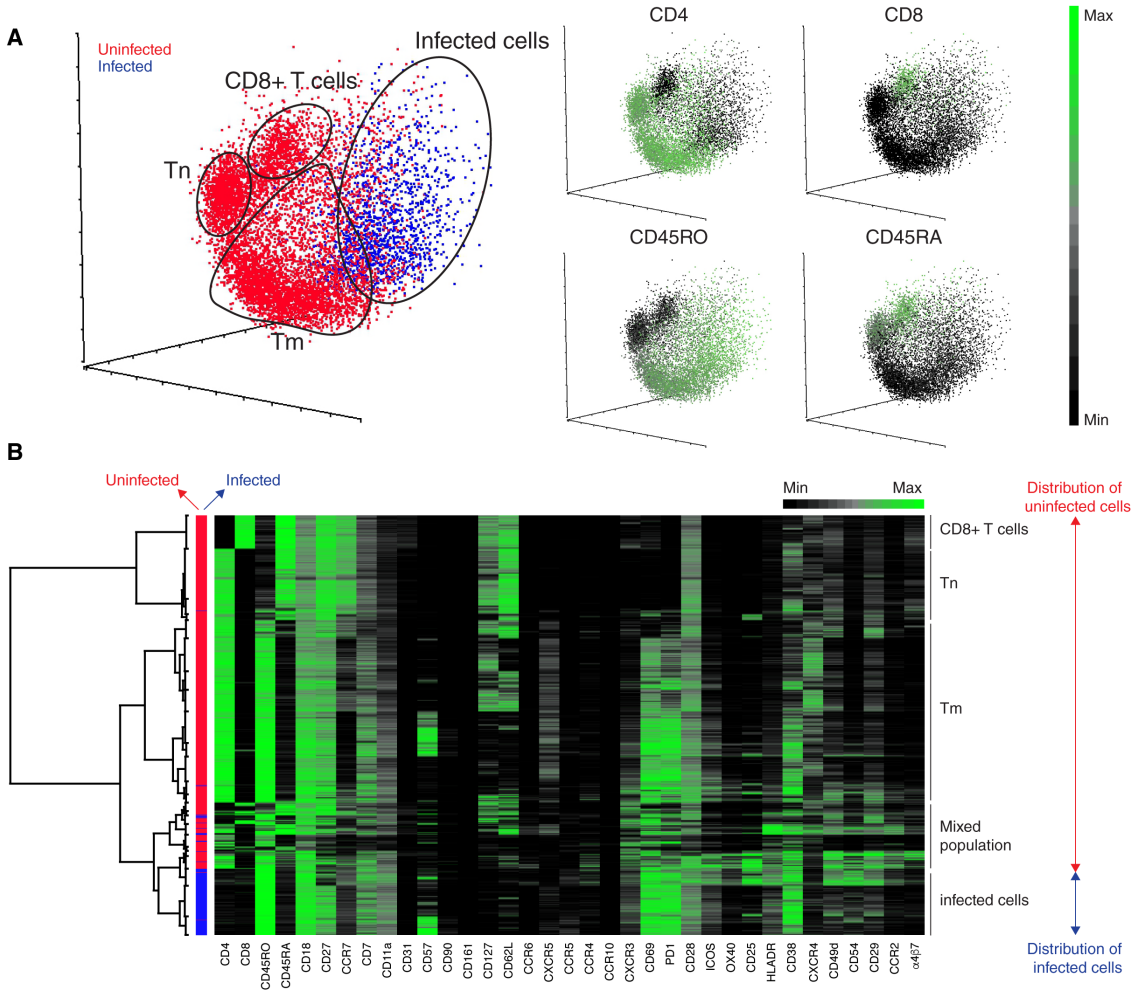


Figure A3 | PCA and agglomerative clustering of uninfected and HIV-infected cells.
(A) *Left*: 3D representation by PCA, showing distribution of uninfected (red) and productively-infected (blue) cells. *Right*: PCA plots colored for expression of CD4, CD8, CD45RA, and CD45RO [ranging from black (minimum) to green (maximum)], which were used to define the populations circled on the left plot. The coordinates of each cell are defined by the first three principal components based on the expression of 35 surface proteins (Table SI). As in Fig. A1A, areas of the plot corresponding to CD8+ T cells, Tn, Tdp, and Tm are indicated. Unlike fused cells, infected cells occupy a unique region not occupied by cells of the corresponding uninfected sample. Results are representative of one of two donors analyzed in a similar fashion.
(B) Heatmap displaying agglomerative hierarchical clustering of uninfected (red) and productively-infected (blue) cells. Each row represents a single cell and each column the indicated protein. Dendrograms on the left correspond to the Euclidean distance between the clustered cells. Unlike the HIV-fused cells (Fig. A1B), infected cells occupy a unique region on the dendrogram. Results are representative of one of two donors analyzed in a similar fashion.

simply did not allow HIV to progress in its viral life cycle and therefore remained HSA-. By showing that the infected cells expressed CD127 at similar levels as their "predicted precursors" (i.e., the most phenotypically similar cells found in the uninfected samples), PP-SLIDE predicted that CD127 expression was not down-modulated upon HIV infection and that, among Tm cells, only Tm lacking expression of CD127 became productively infected (Fig. 4C).

Figure A4

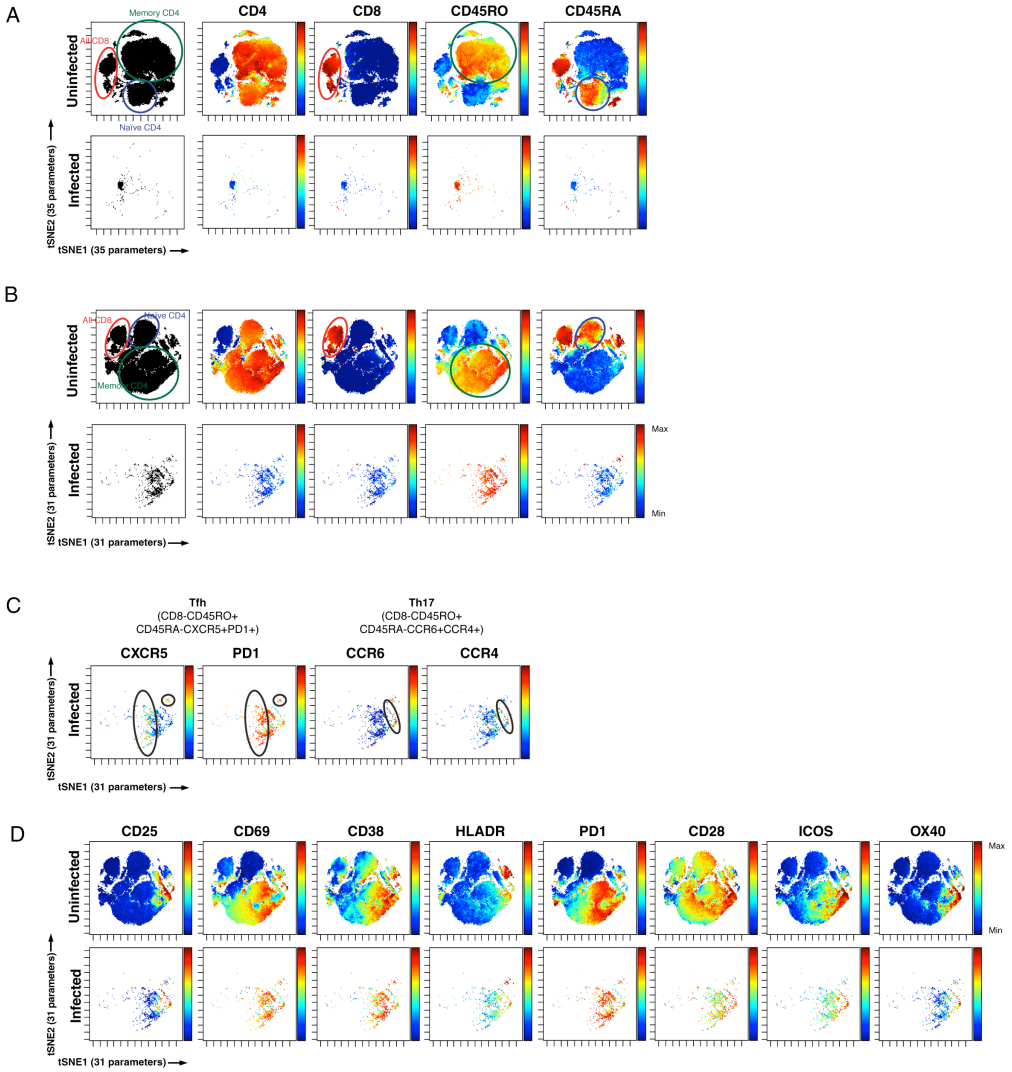


Figure A4 | HIV-infected tonsillar T cells are phenotypically distinct and include activated cells and the Th17 and Tfh subsets, as revealed by t-SNE maps displaying raw expression values.
(A) t-SNE plots of uninfected and infected cells defined by expression of 35 cell surface markers. The first column of t-SNE plots correspond to uninfected (*top*) and infected (*bottom*) tonsillar T cells. These black and white t-SNE plots were then color-coded by expression of CD4, CD8, CD45RA, or CD45RO, as indicated in the subsequent columns. Ovals designating the CD8+, Tn, Tdp, and Tm subsets were drawn using the strategy described in Fig. A2. Note that infected cells occupy a distinct region that does not overlap with occupied regions in the t-SNE plot of uninfected cells.
(B) t-SNE plots of uninfected and infected cells defined by 31 cell surface markers (after excluding CD4, CCR5, CD28, and CD62L, receptors previously reported to be modulated in HIV-infected cells). Ovals designating the CD8+, Tn, Tdp, and Tm subsets were drawn using the strategy described in Fig. A2. Note that infected cells now predominantly localize in a region overlapping with the Tm region of the t-SNE plot of uninfected cells.
(C) t-SNE plots of uninfected and infected cells defined by 31 cell surface markers (as in panel *B*) indicate that cells productively infected with HIV-F4.HSA include both Tfh (CD3+CD19-CD8-CD45RA-CD45RO+CXCR5+PD1+) and Th17 cells (CD3+CD19-CD8-CD45RA-CD45RO+CCR6+CCR4+). Ovals correspond to the Tfh and Th17 subsets, and were drawn based on the expression patterns of the markers listed below each designated subset.
(D) Expression of activation markers by uninfected and infected cells. t-SNE plots of uninfected (*top row*) and infected (*second row*) cells defined by 31 cell surface markers (as in panels *B* and *C*) were colored by expression levels of 8 cell surface receptors associated with T-cell activation. Most productively-infected cells (plots in 2nd row) reside in a region of the t-SNE plot associated with expression of multiple activation markers, as indicated by presence of red or yellow dots in these plots.

Figure A5

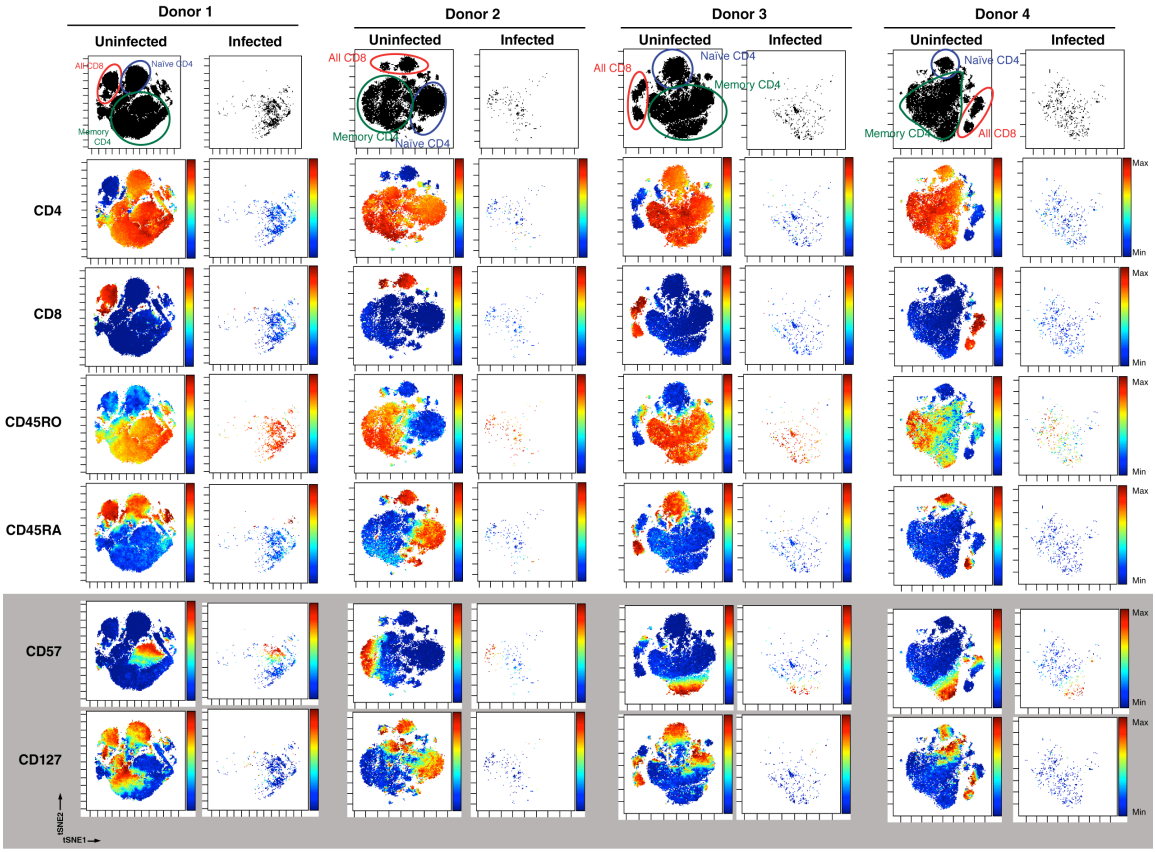


Figure A5 | CD57+ and CD127+ Tm exhibit differential susceptibility to productive HIV infection in 4 donors. CD57+ but not CD127+ Tm are well-represented among HIV-infected cells. The four pairs of plots in the top row display the locations of the main subsets of cells, defined by expression levels of CD4, CD8, CD45RA and CD45RO, as indicated in the next four rows of plots. t-SNE plots were generated based on expression of 31 surface markers (by excluding CD4, CCR5, CD28, and CD62L, previously reported to be modulated in HIV-infected cells, as in Fig. A4B) to better enable visualization of the infected cells. The grey box at the bottom highlights the same t-SNE plots colored by CD57 or CD127 expression levels as indicated. In all four donors, CD57+ but not CD127+ Tm are found among HIV-infected cells. Expression intensity ranges from blue (minimum) to red (maximum).

To confirm these predictions with an alternative bioinformatics approach, we applied a recently developed clustering and data visualization approach termed Single-cell analysis by fixed force- and landmark-directed (Scaffold) maps (Spitzer et al., 2015). Scaffold combines unsupervised clustering based on PAM (partitioning around mediods) (Kaufman and Rousseeuw, 1990) with designation of landmarks from a reference sample, which enables deviations from a reference dataset to be readily visualized (Spitzer et al., 2015). We defined cells from uninfected cultures as the reference sample, and HIV-fused/infected cells as the deviated population of cells. In our Scaffold map of uninfected T cells, we defined a total of seven landmarks (black nodes) corresponding to various subsets of CD4+ and CD8+ T cells (Fig. A6A and

C). Three landmarks were dedicated to Tm: CD57+CD127- Tm (denoted Memory CD4+CD57+), CD57-CD127+ Tm (denoted Memory CD4+CD127+), and CD57-CD127- Tm (denoted Memory CD4+CD57-CD127-) (grey regions of Fig. A6A and C). Cell clusters (indicated by the blue nodes in Fig. A6A and C) with similar phenotypes were linked to the most phenotypically similar landmark (black nodes) by lines, whose length reflects dissimilarity between a particular cluster and its associated landmark.

Focusing in on the Tm region (grey), the Scaffold maps of both uninfected (Fig. A6B, *left*) and HIV-fused cells (Fig. A6B, *right*) showed clusters associated with all three Tm landmarks. By coloring these clusters according to antigen expression levels (heatmap with dark red corresponding to highest expression), we found that all clusters expressed high levels of CD4 (Fig. A6B, top maps). As expected, CD127 was expressed at high levels only on clusters associated with the memory CD4+CD127+ landmark (Fig. A6B, middle maps), and CD57 was expressed at high levels only on clusters associated with the memory CD4+CD57+ landmark (Fig. A6B, middle and bottom maps). These maps confirm efficient HIV entry into both CD57+ and CD127+ Tm.

Although the same seven landmarks were identified in the infection datasets (Fig. A6C), a very different Scaffold map emerged as compared to the fusion datasets. Focusing on the Tm subsets of the infected cells (Fig. A6D, top maps), Scaffold effectively clustered the infected cells lacking CD4 expression around CD4+ landmarks (as highlighted by the two areas circled in blue). Therefore, even though infected cells down-modulated CD4 upon infection, these cells still shared phenotypic features of CD4+ cells, allowing Scaffold to cluster them around CD4+ landmarks. In stark contrast, none of the productively-infected cells clustered around the memory CD4+ CD127+ landmark (Fig. A6D, area circled in green), indicating that none of the infected cells harbored phenotypic features of memory CD4+CD127+ cells and thereby suggesting that none of the HIV-fused CD127+ Tm transitioned onto becoming productively-infected. Therefore, our Scaffold analysis confirms PP-SLIDE's prediction that CD127 is not downregulated during productive infection and, instead, favors a model in which CD127+ Tm cells, despite supporting HIV fusion, did not progress through the HIV life cycle to express HSA.

Figure A6

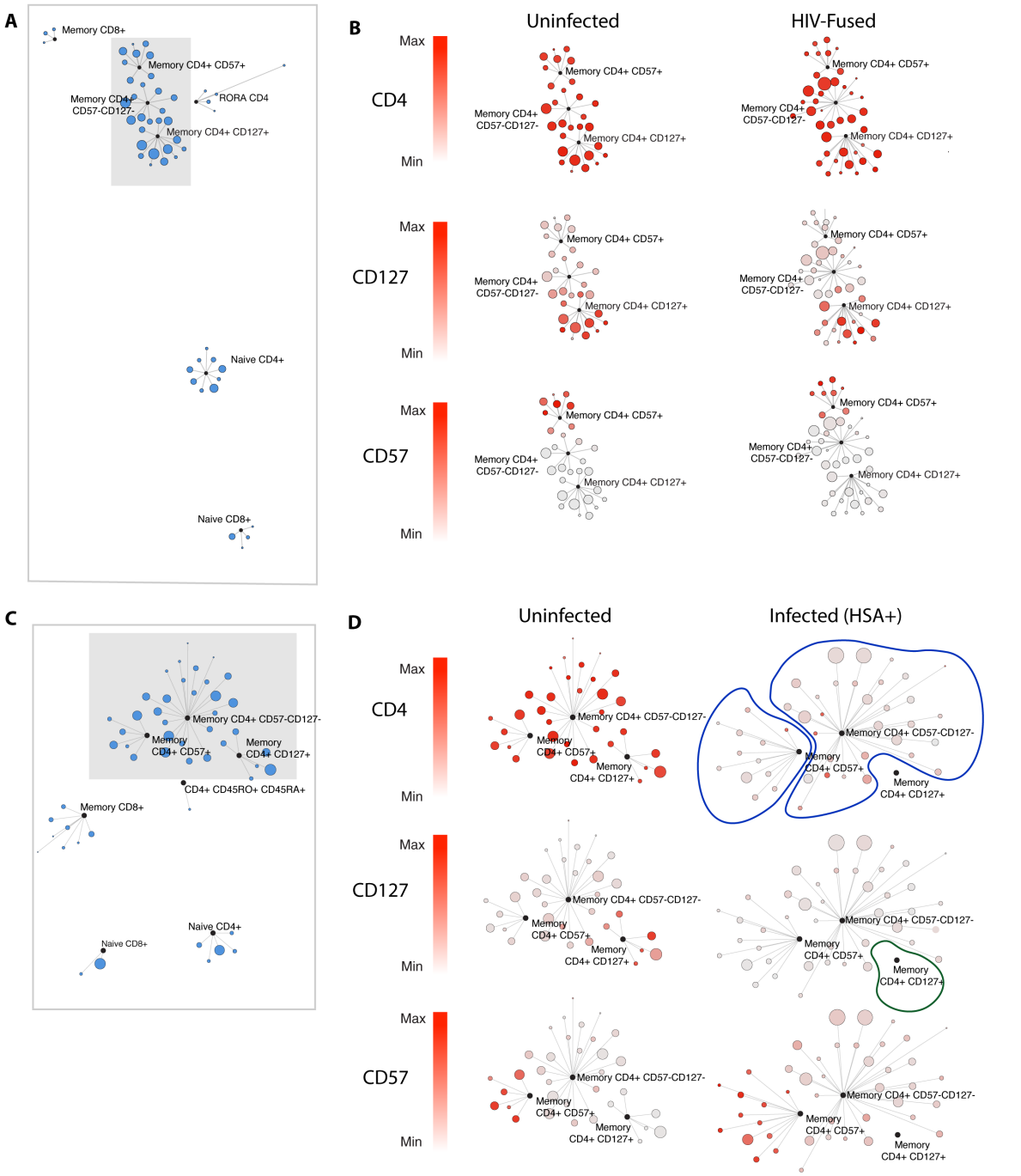


Figure A6 | Scaffold maps of uninfected, HIV-fused, and HIV-infected cells predict lack of CD127 downregulation on infected cells.

(A) Scaffold map showing landscape of uninfected T cells from the fusion dataset. Black nodes denote landmarks based on manually gated populations from four donors; blue nodes correspond to unsupervised cell clusters from one donor. The lengths of the lines connecting the blue nodes to the black nodes is proportional to the dissimilarity of the cells in these clusters to the landmark populations. The size of the blue nodes is proportional to the relative number of cells within that cluster. Only the highest scoring (most similar) relationships between landmarks and clusters are shown. The three Tm landmarks and their associated clusters, highlighted in grey, are examined in further detail in panel *B*.

(B) Scaffold map showing landscape of uninfected (*left*) and HIV-fused Tm (*right*), colored by relative expression of CD4 (*top*), CD127 (*middle*), or CD57 (*bottom*), with dark red corresponding to highest intensity. Clusters of HIV-fused cells associate with all three Tm landmarks. All clusters express CD4, while CD57 is expressed only in clusters associated with

the Memory CD57+CD127- landmark and CD127 is expressed only in clusters associated with the Memory CD57-CD127+ landmark.
(C) Scaffold map showing landscape of uninfected T cells from the infection dataset. Black nodes denote landmarks based on manually gated populations from four donors; blue nodes correspond to unsupervised cell clusters from one donor. The three Tm landmarks and their associated clusters, highlighted in grey, are examined in further detail in panel *D*.
(D) Scaffold map showing landscape of uninfected (*left*) and productively-infected (*right*) Tm, colored by relative expression of CD4 (*top*), CD127 (*middle*), or CD57 (*bottom*), with dark red corresponding to highest intensity. The map of infected cells colored by CD4 expression levels (*top right*) reveals the presence of "CD4-like" clusters (circled in blue) associated with the memory CD4+ T cells even though these cells no longer express the CD4 receptor (as evident by the grey color of these clusters). The circled green region highlights the absence of clusters associated with the Memory CD4+CD127+ landmark, showing the absence of "CD127-like" cells in the infected population, thereby suggesting against the notion that HIV had productively-infected CD127+ memory T cells and then subsequently downregulated the CD127 receptor.

Section 4 – t-SNE and SPADE analyses reveal profound phenotypic differences between CD57+ and CD127+ Tm.

In the main manuscript, we showed that CD57+ and CD127+ Tm are located in different regions of the lymph node (Lim and Kim, 2007)(Fig. 6B), suggesting that these Tm subsets serve different functions and harbor broader phenotypic differences than just differential expression of CD57 or CD127. One way to characterize these phenotypic differences is to locate the CD57+ and CD127+ Tm on a t-SNE plot. Fig. A7A shows that CD57+ and CD127+ Tm polarized to opposite regions of the t-SNE plots, suggesting that these subsets are indeed highly dissimilar. To assess the contribution of the CD57 and CD127 expression in this polarization, we investigated whether CD57+ and CD127+ Tm would still segregate separately if these two markers were excluded from the t-SNE analysis. To this end, we conducted t-SNE under conditions where we excluded both CD57 and CD127 from the analysis (Fig. A7B). After generating this t-SNE plot we colored the plots according to CD57 and CD127 expression. The resulting plots suggested that these two cellular subsets still segregated away from each other, indicating that markers other than CD57 and CD127 are differentially expressed by CD57 and CD127 Tm.

Another way to assess whether CD57+ and CD127+ Tm harbor more profound differences than just differential CD57 and CD127 expression is to cluster the cells with SPADE using either all of the clustering parameters, or by excluding CD57 and CD127 from the clustering part of the analysis. Then, by coloring the SPADE tree according to either CD57 or CD127, we can retrospectively see where the CD57 Tm and the CD127 Tm ended up clustering. As shown in Fig. A7C and D, we found that clusters corresponding to CD57+ Tm segregated away from CD127+ Tm whether or not CD57 and CD127 were including in the clustering.

Therefore, both t-SNE and SPADE analyses suggested that CD57+ and CD127+ Tm are highly phenotypically distinct and can even be distinguished one from another without using CD57 and CD127 as identifiers. This phenomenon, known as "immune redundancy", refers to the fact that, in multidimensional analyses, cellular subsets can often be distinguished without using canonical markers given that these subsets will differentially express multiple other antigens within a large panel.

Figure A7

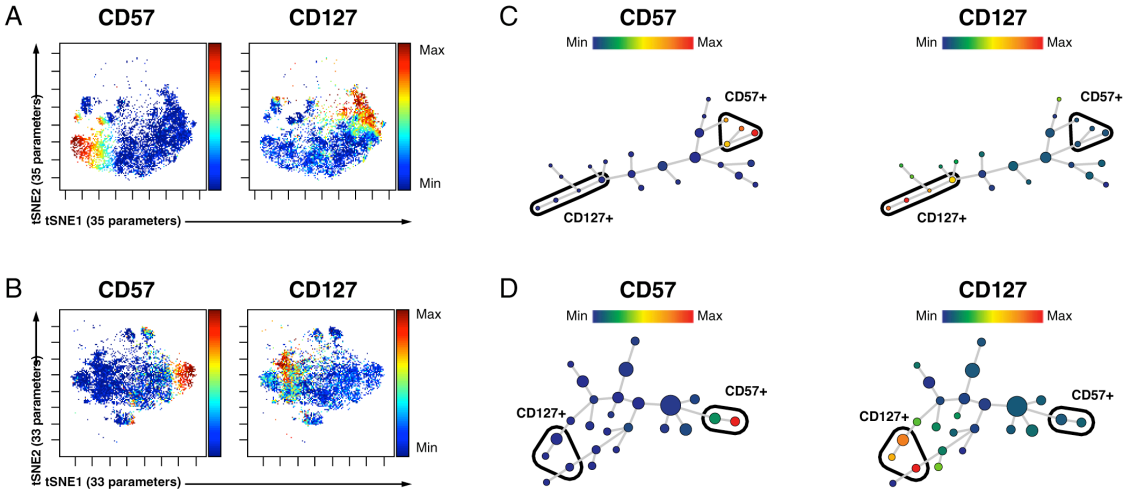


Figure A7 | CD57+ and CD127+ Tm are phenotypically distinct subsets independent of CD57 and CD127 expression levels

(A) t-SNE plots of HIV-fused cells generated based on expression of 35 surface markers and colored by expression levels of CD57 (*left*) or CD127 (*right*). This analysis is taken from Donor 1 in Fig. A2A, and direct comparisons of these plots with the map identifying the various T-cell subsets in this donor reveals that the HIV-fused populations expressing high levels of CD57 or CD127 fall within the Tm population.

(B) t-SNE plots of the same HIV fusion dataset analyzed in panel *A*, but generated based on expression of 33 surface markers by excluding CD57 and CD127 as analysis parameters. The plots were then colored by expression levels of CD57 (*left*) or CD127 (*right*) to visualize where populations expressing these receptors reside. Note that although CD57 and CD127 did not influence positioning of the cells on these t-SNE plots, cells expressing high levels of CD57 clustered away from cells expressing high levels of CD127, demonstrating that in addition to differences in expression of CD57 and CD127, these two subsets of cells express a distinct profile of the other 33 antigens within the CyTOF panel.

(C) HIV-fused Tm were distributed by SPADE into 25 nodes based on 35 cell-surface markers including CD57 and CD127. This analysis was also shown in Fig. 5F. Clusters are designated by circles, the size of which is proportional to number of cells and whose similarity to other clusters is reflected by the connections between the clusters. The color of the nodes indicates expression of CD57 (*left*) or CD127 (*right*), ranging from blue (minimum) to red (maximum). Clusters were annotated into subpopulations and circled in black based on expression of CD57 and CD127.

(D) Same as in *C*, except that the cells were clustered based on expression of 33 surface markers by excluding CD57 and CD127 as clustering parameters. Note that clusters of HIV-fused cells expressing CD57 versus those expressing CD127 segregate separately even when the contribution of CD57 and CD127 were not included during clustering.

References

Alvarez, Y., Tuen, M., Shen, G., Nawaz, F., Arthos, J., Wolff, M.J., Poles, M.A., and Hioe, C.E. (2013). Preferential HIV infection of CCR6+ Th17 cells is associated with higher levels of virus receptor expression and lack of CCR5 ligands. Journal of virology *87*, 10843-10854.

Garcia, J.V., and Miller, A.D. (1991). Serine phosphorylation-independent downregulation of cell-surface CD4 by nef. Nature *350*, 508-511.

Kaufman, L., and Rousseeuw, P.J. (1990). Finding Groups in Data. Wiley, New York.

Kohler, S.L., Pham, M.N., Folkvord, J.M., Arends, T., Miller, S.M., Miles, B., Meditz, A.L., McCarter, M., Levy, D.N., and Connick, E. (2016). Germinal Center T Follicular Helper Cells Are Highly Permissive to HIV-1 and Alter Their Phenotype during Virus Replication. J Immunol *196*, 2711-2722.

Lim, H.W., and Kim, C.H. (2007). Loss of IL-7 receptor alpha on CD4+ T cells defines terminally differentiated B cell-helping effector T cells in a B cell-rich lymphoid tissue. J Immunol *179*, 7448-7456.

Michel, N., Allespach, I., Venzke, S., Fackler, O.T., and Keppler, O.T. (2005). The Nef protein of human immunodeficiency virus establishes superinfection immunity by a dual strategy to downregulate cell-surface CCR5 and CD4. Curr Biol *15*, 714-723.

Sen, N., Mukherjee, G., Sen, A., Bendall, S.C., Sung, P., Nolan, G.P., and Arvin, A.M. (2014). Single-cell mass cytometry analysis of human tonsil T cell remodeling by varicella zoster virus. Cell reports *8*, 633-645.

Spitzer, M.H., Gherardini, P.F., Fragiadakis, G.K., Bhattacharya, N., Yuan, R.T., Hotson, A.N., Finck, R., Carmi, Y., Zunder, E.R., Fantl, W.J*., et al.* (2015). IMMUNOLOGY. An interactive reference framework for modeling a dynamic immune system. Science (New York, NY *349*, 1259425.

Swigut, T., Shohdy, N., and Skowronski, J. (2001). Mechanism for down-regulation of CD28 by Nef. The EMBO journal *20*, 1593-1604.

Vassena, L., Giuliani, E., Koppensteiner, H., Bolduan, S., Schindler, M., and Doria, M. (2015). HIV-1 Nef and Vpu Interfere with L-Selectin (CD62L) Cell Surface Expression To Inhibit Adhesion and Signaling in Infected CD4+ T Lymphocytes. J Virol *89*, 5687-5700.