

## The Atlas of the *In Vivo* HIV CD4 T Cell Reservoir

Jason Neidleman<sup>1,2,†</sup>, Xiaoyu Luo<sup>1,†</sup>, Julie Frouard<sup>1,2</sup>, Guorui Xie<sup>1,2</sup>, Feng Hsiao<sup>1,2</sup>, Tongcui Ma<sup>1,2</sup>, Vincent Morcilla<sup>3</sup>, Ashley Lee<sup>3</sup>, Sushama Telwatte<sup>4</sup>, Reuben Thomas<sup>1</sup>, Whitney Tamaki<sup>5</sup>, Benjamin Wheeler<sup>5</sup>, Rebecca Hoh<sup>6</sup>, Ma Somsouk<sup>7</sup>, Poonam Vohra<sup>8</sup>, Jeffrey Milush<sup>5</sup>, Katherine Sholtis James<sup>8</sup>, Nancie M. Archin<sup>9</sup>, Peter W. Hunt<sup>10</sup>, Steven G. Deeks<sup>6</sup>, Steven A. Yukl<sup>4</sup>, Sarah Palmer<sup>3</sup>, Warner C. Greene<sup>1,5</sup>, and Nadia R. Roan<sup>1,2 \*</sup>

<sup>1</sup> Gladstone Institutes, San Francisco, CA, USA

<sup>2</sup> Department of Urology, University of California, San Francisco, CA, USA

<sup>3</sup> Centre for Virus Research, the Westmead Institute for Medical Research, The University of Sydney, Sydney, NSW 2145, Australia

<sup>4</sup> San Francisco Veterans Affairs (VA) Medical Center and University of California, San Francisco, CA, USA

<sup>5</sup> Department of Medicine, University of California, San Francisco, CA, USA

<sup>6</sup> Division of HIV, Infectious Diseases and Global Medicine, University of California San Francisco, San Francisco, CA, USA

<sup>7</sup> Department of Medicine, Division of Gastroenterology, San Francisco General Hospital and University of California, San Francisco, CA, USA

<sup>8</sup> Department of Pathology, University of California, San Francisco, CA, USA

<sup>9</sup> Division of Infectious Diseases, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

<sup>10</sup> Division of Experimental Medicine, University of California San Francisco, San Francisco, CA, USA

†Equal contribution, \*Correspondence: nadia.roan@ucsf.edu

## ABSTRACT

The latent reservoir is a main barrier for curing HIV. But because latently-infected cells cannot be phenotyped directly, the features of the *in vivo* reservoir have remained elusive. Here, we describe a method that leverages high-dimensional phenotyping using CyTOF to trace latently-infected cells reactivated *ex vivo* to their original pre-activation states. Our results suggest that contrary to common assumptions, the reservoir is not randomly distributed among cell subsets, and is remarkably conserved between individuals. However, reservoir composition differs between tissues and blood, as do cells successfully reactivated by different latency reversing agents. Most importantly, by selecting 8-10 of our 39 original CyTOF markers, we were able to isolate highly purified populations of unstimulated *in vivo* latent cells, thereby validating the PP-SLIDE approach for reservoir characterization. These purified populations were highly enriched for replication-competent and intact provirus, transcribed HIV, and displayed clonal expansion. The ability to isolate unstimulated latent cells from infected individuals enables previously impossible studies of HIV persistence.

## KEYWORDS

HIV, Latency, Atlas, Tissues, Replication-competent Reservoir, Clonal Expansion, CyTOF

## INTRODUCTION

Combination antiretroviral therapy (ART) suppresses HIV replication but does not eliminate the latent reservoir, which persists for decades in CD4+ T cells and is a major barrier to curing HIV (Chun et al., 1997; Finzi et al., 1997; Wong et al., 1997). Our understanding of the phenotypic features of latent cells *in vivo* is limited, in large part because the lack of a universal surface biomarker makes it impossible to directly identify these cells. As a result, fundamental questions remain, such as whether latent cells distribute equally among all CD4+ T cell subsets, and to what extent the reservoir is similar between individuals and between blood and tissues.

In the absence of direct ways to phenotype latent cells, characterizing the composition of the HIV reservoir has largely entailed sorting CD4+ T cells based on predefined sets of cell-surface proteins, followed by quantitating HIV through a variety of methods. The vast majority of infected cells in virally-suppressed individuals harbor defective HIV genomes (Hiener et al., 2017; Ho et al., 2013; Imamichi et al., 2016; Lee et al., 2017). Older quantitation methods based on measuring total HIV DNA/RNA levels, regardless of provirus integrity, are being replaced with methods quantitating levels of replication-competent or genome-intact provirus. These methods afford a more accurate picture of the cells leading to viral rebound during treatment interruption, which are the most relevant targets for HIV cure. The “gold standard” in virus quantitation has been the quantitative viral outgrowth assay (qVOA) (Finzi et al., 1997). qVOA has not identified significant differences in the amount of replication-competent HIV in the most commonly-studied subsets of memory CD4+ T cells – central memory (Tcm), effector memory (Tem), and transitional memory (Ttm) (Buzon et al., 2014; Kwon et al., 2020). Such observations have led many in the field to consider that the replication-competent reservoir is simply in all subsets, which if true makes targeting it more difficult.

The analysis of proviral sequences, on the other hand, has revealed differences in the frequencies of genome-intact provirus in different sorted subsets. For instance, sequencing of near full-length HIV provirus has identified Th1 and Tem, particularly those that are HLA DR+, as

subsets preferentially harboring intact viral genomes (Hiener et al., 2017; Horsburgh et al., 2020; Lee et al., 2017). Viral sequencing has also revealed many intact proviruses within clonally expanded populations, particularly within the Tem and Th1 subsets, and those expressing Ox40 or CD161 (De Scheerder et al., 2019; Hiener et al., 2017; Kuo et al., 2018; Lee et al., 2017; Li et al., 2019). These and many other studies (Cohn et al., 2020; Liu et al., 2020) have solidified the view that clonal expansion plays a major role in maintaining the long-lived HIV reservoir.

Although analyses of sorted populations have provided meaningful insights into HIV persistence, they have been limited to comparing cell populations that differ in one or two markers. Thus, they do not provide a comprehensive view of the phenotypic features of latent cells. Several methods have recently enabled the direct phenotyping of reservoir cells at single-cell resolution (Baxter et al., 2016; Cohn et al., 2018; Grau-Exposito et al., 2017; Pardons et al., 2019), providing for deeper analyses of the reservoir. These methods entail stimulating cells from ART-suppressed patients with a mitogen to induce expression of HIV RNA and/or proteins, which serve as markers for identifying or isolating cells by FACS. One important limitation with these methods, however, is that they require *ex vivo* stimulation prior to phenotyping, since viral proteins are typically not expressed at detectable levels in unstimulated latent cells (Baxter et al., 2016; Cohn et al., 2018; Pardons et al., 2019). This stimulation will induce phenotypic changes associated with T cell activation, and expression of HIV accessory genes from reactivation will result in host cell remodeling (e.g., downregulation of cell-surface CD4), which will further alter the phenotypes of the cells (Cavrois et al., 2017; Ma et al., 2020). Therefore, although current methods enable phenotyping of reactivated cells, there is at present no methodology to directly assess the phenotypes of individual latent cells in their original dormant state.

To overcome this limitation, we applied the principles behind Predicted Precursor as determined by SLIDE (PP-SLIDE), an analytic method we recently established that combines

high-dimensional single-cell phenotyping and computational approaches. We have used PP-SLIDE to predict the state of CD4+ T cells prior to phenotypic changes induced by productive HIV infection (Cavrois et al., 2017; Ma et al., 2020). Here, we applied it to determine the features of *in vivo* latent CD4+ T cells prior to their reactivation by *ex vivo* stimulation, and validated these findings with patient specimens. Through PP-SLIDE analysis, we were able to address questions fundamental to our understanding of HIV latency, and to isolate highly-enriched populations of replication-competent latent cells for mechanistic studies of reservoir maintenance.

## RESULTS

### Validation of PP-SLIDE as method to trace reactivated cells back to their latent state

PP-SLIDE was recently validated as an approach to trace HIV-remodeled cells to their original pre-infection states (Cavrois et al., 2017; Ma et al., 2020). To determine if it can trace reactivated cells to their original latent state, we tested it first in J-Lat cells, an *in vitro* model of HIV latency. J-Lat cells are Jurkat cells harboring a single integrated HIV provirus that can be reactivated upon stimulation. Multiple lines of J-Lats have been cloned (Chan et al., 2013; Jordan et al., 2003) that have diverged slightly over successive passages. We tested how well PP-SLIDE traces reactivated cells back to their original clone when given the choice of 2 parental clones (Fig. 1A). Cells from clone 6.3 were reactivated with PMA/ionomycin-stimulation and phenotyped using a CyTOF T cell panel (Ma et al., 2020), alongside unstimulated cells from clone 6.3 and another clone, 5A8. Using PP-SLIDE (see Methods), the high-dimensional CyTOF-generated information from reactivated 6.3 cells (identified by expression of HIV Gag) was matched against an atlas of concatenated unstimulated 5A8 and 6.3 cells to identify the most similar unstimulated cell for each reactivated 6.3 cell. These unstimulated cells, identified using a k-nearest neighbor (kNN) approach, are collectively termed “kNN latent cells.” The error rate, defined as the % of reactivated 6.3 cells that mapped to unstimulated 5A8 cells, was 0.8% (Fig. 1B). These results

suggest that although reactivation greatly changes the phenotype of a latent cell, some of the original “identity” of the cell is still retained, and that PP-SLIDE can be used to capture this information and thereby identify precursors of reactivated latent cells.

### **Latent cells are not randomly distributed among memory CD4+ T cells *in vivo***

We next applied PP-SLIDE to characterize latent cells from HIV-infected individuals, using a CyTOF panel ([Table S1](#)) directed at markers of the major subsets and differentiation states of CD4+ T cells, and includes receptors and intracellular proteins associated with latency (Banga et al., 2016; Buzon et al., 2014; Chomont et al., 2009; Descours et al., 2017; Fromentin et al., 2016; Hogan et al., 2018; Iglesias-Ussel et al., 2013; Khoury et al., 2016; Li et al., 2019; Serra-Peinado et al., 2019; Sun et al., 2015). Since reactivation events are rare and reducing background was key, we conjugated different anti-Gag antibodies to three different metal lanthanides, and only considered cells displaying a Gag signal in all 3 channels as true events. The full panel was validated *in vitro* ([Fig. S1](#)). We collected via leukapheresis CD4+ T cells from the blood of four ART-suppressed HIV-infected individuals ([Table S2](#)), purified memory CD4+ T cells by negative selection, and separated the cells in two pools ([Fig. 2A](#)). One pool was used to generate the atlas of unstimulated cells. This pool was phenotyped within two hours of cell isolation, and never went through any cryopreservation or *ex vivo* culture. The other pool was stimulated with PMA/ionomycin under fully suppressive ART ([Fig. S2A](#)) for 40 h, a timepoint that precedes mitogen-induced cellular proliferation ([Fig. S2B](#)). The CyTOF profiles of reactivated cells were matched against those of the unstimulated atlas cells. Assuming each reactivated cell retained some of its original pre-stimulation identity and that a phenotypically similar cell was present in the atlas of unstimulated cells, we used PP-SLIDE to identify, for each reactivated cell, the nearest-neighbor within the atlas ([Fig. 2A](#)). These identified cells, the “kNN latent cells”, harbor the predicted phenotypes of latent cells prior to their reactivation from latency.

We identified between 20 and 96 reactivated cells per specimen, corresponding to a mean frequency of 11.3 per million memory (CD45RA-CD45RO+) CD4+ T cells (range 5.6 – 17.6, [Fig. 2B](#), [Fig. S3](#)). When visualized against the atlas of unstimulated cells by t-distributed stochastic neighbor embedding (tSNE) (van der Maaten, 2009), Gag+ reactivated cells occupied a unique region, confirming their distinct phenotype relative to unstimulated cells ([Fig. 2C, left](#)). kNN latent cells mapped to several distinct regions of the atlas, ([Fig. 2C, right](#)), suggesting a non-random distribution among memory CD4+ T cells. To verify this, we implemented FlowSOM (Van Gassen et al., 2015) to subdivide the atlases from each of the four donors into 20 clusters, and visualized the distribution of kNN latent cells on the tSNE maps or on pie graphs ([Fig. 2C, center](#), and [2D](#)). We found large clusters lacking kNN latent cells as well as small clusters harboring kNN latent cells, consistent with a non-random distribution of latent cells, which was confirmed using a chi-squared test ([Fig. 2D](#)). For a detailed description of the extent of enrichment of kNN latent cells in clusters and validation of non-random distribution of the reservoir, see Methods.

### **kNN latent cells from different donors share phenotypic features**

We next examined the composition of the kNN latent cells. We focused first on the Tcm, Tem, and Ttm subsets, as they have been the most commonly studied memory CD4+ T cells in the context of HIV latency (Chomont et al., 2009; Josefsson et al., 2013; Kwon et al., 2020; Pardons et al., 2019; Soriano-Sarabia et al., 2014; von Stockenstrom et al., 2015). In all four donors, the contribution of Tem to the reservoir was greater than its contribution to total memory CD4+ T cells ([Fig. S4](#)). We did not identify any kNN latent cells with the Ttm phenotype, likely because this subset is relatively rare in blood and not efficiently captured in our samples, each of which harbored fewer than 100 kNN latent cells. When the three FlowSOM clusters of atlas cells with the highest frequencies of kNN latent cells were examined for the relative contributions of the Tcm, Tem, and Ttm subsets, we found that they were all comprised of multiple subsets ([Fig. S5](#)). Collectively, these data demonstrate that although Tem are over-represented among kNN

latent cells, the clusters most enriched for kNN latent cells – as classified by unbiased clustering based on ~40 markers – are a mixture of Tcm/Tem/Ttm subsets.

These findings could suggest that the reservoir is quite different between individuals. Alternatively, they could mean that latency markers shared between individuals are not those used to define Tcm/Tem/Ttm cells. Consistent with the latter, when datasets from the 4 donors were combined and re-run within the same tSNE, the kNN latent cells from the 4 donors resided within a similar region of the plot (Fig. 3A). These results argue that kNN latent cells between different donors are in fact similar. Consistent with this conclusion, relative to total memory CD4+ T cells, kNN latent cells consistently expressed higher levels of select markers of immune checkpoint (PD1, CTLA4), activation (CD69, CD25, HLADR), and T cell differentiation states (Tbet, CTRH2, CCR6) (Fig. 3B). We conclude that *in vivo* latent cells are not randomly distributed but instead share specific features between individuals.

Even though latent cells are similar between individuals, the inter-donor variability in latent cells could still be greater than the intra-donor variability. To test this idea, we obtained leukapheresis specimens from donor PID5003 at two timepoints spaced two months apart, and compared in an unbiased manner the kNN latent cells from those specimens to kNN latent cells from donor PID3010 (Fig. S6A, B). For each kNN latent cell in the PID3010 sample, we calculated the Euclidean distances (based on CyTOF parameters) to every kNN latent cell in the PID5003 sample. Similar calculations were conducted between kNN latent cells from the two timepoints sampled in PID5003 (Fig. S6A). The median distances were lower between specimens from the same donor than from different donors (Fig. S6C), demonstrating that despite the similarity of kNN latent cells between donors, kNN latent cells differ more between donors than within a donor.

### **Different *ex vivo* stimulations reactivate different latent cells**

We next took advantage of our ability to obtain high-resolution maps of the phenotypic features of latent cells to compare their reactivation by different *ex vivo* stimulations.

Longitudinal blood specimens were obtained from the same suppressed individual and stimulated with the broad-spectrum mitogens PMA/ionomycin or anti-CD3/CD28. Euclidean distance calculations (Fig. S6A) revealed that PMA/ionomycin-reactivatable kNN latent cells from different collection timepoints were more similar to each other than to anti-CD3/CD28-reactivatable kNN latent cells from the same timepoint (Fig. S6D, E). We next compared the latent cells reactivatable by PMA/ionomycin to those reactivatable by a combination LRA consisting of an HDAC inhibitor together with an ingenol (Romidepsin/PEP005) using 3 longitudinal blood specimens each spaced 2 months apart. tSNE visualization of the kNN latent cells revealed a region of the plot that harbored kNN latent cells from the PMA/ionomycin samples but not from the Romidepsin/PEP005 samples (Fig. 4A). Euclidean distance calculations (Fig. S6A) were again used to verify that kNN latent cells reactivatable by the same stimulation at different timepoints were more similar than those reactivatable by different stimulations within the same timepoint (Fig. 4B). These data suggest that Romidepsin/PEP005 reactivates only a subset of latent cells reactivatable by broad-spectrum mitogens such as PMA/ionomycin.

### **Phenotypic features of tissue kNN latent cells**

Because most of the HIV reservoir persists within tissues (Estes et al., 2017), we next applied the PP-SLIDE method of latent cell characterization to tissue specimens. One of the four suppressed participants who had donated a leukapheresis specimen, PID3010, also agreed to the collection of two sets of lymph node specimens via fine needle aspirates (FNAs) spaced six months apart. Upon isolation, each FNA was immediately mock-treated or stimulated for 40 h with PMA/ionomycin and then subjected to PP-SLIDE. Unlike with blood specimens, memory CD4+ T cells were not purified ahead of time since FNAs yielded limited numbers of cells. However, we pre-gated on memory CD4+ T cells for comparison to purified blood cells. The atlases from the 3 specimens (1 blood, 2 FNAs) showed that blood and FNA cells were

phenotypically distinct (Fig. 4C). The kNN latent cells from the FNAs, similar to those from blood, were not randomly distributed, and remained similar across the longitudinal specimens as demonstrated by their occupying similar regions of tSNE space (Fig. 4C). Several antigens distinguished kNN latent cells in blood vs. tissues. For instance, CD27, an antigen expressed on Tcm cells, and CD69, a marker of activation in blood but a T resident memory (Trm) marker in tissues (Cantero-Perez et al., 2019), were more highly expressed on FNA kNN latent cells (Fig. 4D, S7A). PD1 and CXCR5, markers of Tfh that are preferentially in latent cells particularly in tissues (Banga et al., 2016), were also expressed at especially high levels on kNN latent cells from FNAs (Fig. 4D, S7A). The co-stimulatory molecule ICOS was expressed at high levels on kNN latent cells exclusively in FNAs, while the exhaustion marker TIGIT was more variably expressed (Fig. 4D, S7A). These results reveal marked differences between the phenotypic features of kNN latent cells in blood vs. lymph nodes, with the notable exception of PD1, which was high in both compartments.

We also characterized the reservoir in the gut, the primary site of viral persistence during suppressive therapy (Estes et al., 2017). As longitudinal gut specimens were not available, we conducted a cross-sectional analysis comparing gut specimens from 4 donors to the blood specimens previously analyzed by PP-SLIDE. While differences in the reservoir between blood and gut were observed, some interesting similarities also emerged. kNN latent cells from the gut were mostly CD69<sup>hi</sup>CD27<sup>lo</sup>, while those from blood exhibited a much more variable expression pattern of these 2 antigens (Fig. S7B). In contrast, kNN latent cells from both gut and blood expressed high levels of PD1 but relatively low levels of CXCR5 (Fig. S7B). Another similarity was that CTLA4 and CCR6, previously reported to be preferentially expressed on latent cells from tissues (Gosselin et al., 2017; McGary et al., 2017), were both highly expressed in kNN latent cells from gut and blood. Taken together with the FNA results, these analyses identify CD69 and PD1 as antigens expressed at high levels in kNN latent cells from both gut and lymph nodes, suggesting shared features of the reservoir between different tissue compartments.

To compare the global similarities between kNN latent cells from the three compartments (blood, lymph nodes, gut), we visualized by tSNE the atlas and kNN latent cells of all donors combined (Fig. 4E). Interestingly, the FNA atlas cells appeared equidistant to the blood and gut atlas cells, an observation confirmed by Euclidean distance calculations. In contrast, kNN latent cells from FNA and gut tended to cluster more closely together, away from those from blood, which we also confirmed by Euclidean distance calculations (Fig. 4E). Together with the above analyses suggesting that kNN latent cells from both tissue compartments express high levels of CD69 and PD1, these results suggest that phenotypic features of the reservoir are shared between different tissue sites. That said, a notable difference was that kNN latent cells from FNAs included both naïve and memory cells, while those from gut only included memory cells (Fig. S7C).

### **Markers of kNN latent cells enrich for the replication-competent, genome-intact reservoir**

To determine whether surface markers differentially expressed on kNN latent cells can enrich for the replication-competent reservoir, we custom-designed sorting strategies for each of the 4 donor blood specimens. Because PD1 is arguably the most validated surface marker of latent cells (Banga et al., 2016; Chomont et al., 2009; Fromentin et al., 2016; Pardons et al., 2019), we compared the additional enrichment afforded by our sorting strategy to that afforded by PD1 selection alone. Our panels took into account the markers' expression levels in kNN latent cells, the numbers of cryopreserved cells available from the participant, and the overall abundance of the populations in each sequential gate. We arrived at panels of 8 – 10 markers individualized for each participant. For each donor specimen, we isolated a PD1+ population that we submitted to further selection with the relevant marker panel ("enriched"), and two PD1+ populations that we did not subject to further selection ("disenriched"). The disenriched

populations harbored markedly lower frequencies of kNN latent cells than the enriched one (Fig. S8). These populations were sorted (Fig. S9) and subjected to multiple assays.

We first conducted a viral outgrowth assay. Because sorting based on our 8-10 marker panels yielded few cells, we modified the qVOA accordingly. Sorted cells were stimulated in bulk for 2 days with anti-CD3/CD28, and plated into individual wells at various dilutions. The HIV-permissive MOLT-CCR5 cell line was added to propagate viral replication, and after two weeks supernatants were quantitated for viral outgrowth by p24 ELISA. Replication was scored as the proportion of wells that had outgrowth, and as the concentration of p24 in each well. Both scoring methods revealed strikingly more outgrowth in the enriched relative to both disenriched populations in all four donors (Fig. 5A). In fact, most disenriched populations did not result in any outgrowth even at the highest concentration tested (200,000 cells/well).

Because recent studies suggest that the *in vivo* reservoir is transcriptionally active (Yukl et al., 2018), we assessed HIV transcription initiation, elongation, and completion in the sorted populations. In the absence of stimulation, HIV transcriptional activity was observed in the sorted cells, with more transcription initiation in the enriched than disenriched ones in 7/8 instances (Fig. 5B). These results suggest that populations highly enriched for the replication-competent reservoir initiate HIV transcription in the absence of stimulation.

We then assessed to what extent our 8 – 10 marker panels enrich for latent cells harboring intact genomes. We sequenced by FLIPS (Hiener et al., 2017) proviruses from the enriched and from one of the disenriched populations, for each of three donors. Consistent with the PP-SLIDE and outgrowth results, enriched cells harbored higher frequencies of intact p24 sequences than disenriched ones (Fig. 5C). To assess the extent of clonal expansion, a major driver of HIV persistence (Cohn et al., 2020; Liu et al., 2020), we determined the proportions of expanded identical sequences (EIS). EIS were prominent in both populations, but the proportions of EIS with intact p24 sequences were consistently higher in the enriched ones (Fig. S10). A phylogenetic analysis of the sequences revealed that in two of the donors (PID1695 and

PID5003), EIS from the enriched and disenriched populations formed distinct lineages, while in donor PID3010 multiple EIS were shared between the enriched and disenriched populations (Fig. S11).

As the enriched population from PID5003 harbored a particularly high frequency of intact p24 sequences, we analyzed this donor in more detail. We categorized all the sequences analyzed from this individual into those that were full-length intact (i.e. no defects along the entire proviral genome), and those harboring large internal deletions, hypermutations, premature stop codons, and inversions. Remarkably, 65.2% of the enriched population was fully intact, compared to only 0.9% for the disenriched population (Fig. 5D). This corresponded to an overall frequency of 214 full-length intact sequences per million in the enriched population, as compared to only 2 per million in the disenriched population. The vast majority (95.5%) of the proviruses in the disenriched population harbored large internal deletions. EIS were prominent among both populations (Fig. 5E). Among the EIS sequences in the enriched population, 66.7% were fully intact, compared to 0% in the disenriched population. Fully-intact EIS in the enriched population comprised two groups, suggesting 2 major expansions of replication-competent reservoir cells in this population of cells highly enriched for replication-competent HIV (Fig. 5E). Collectively, these results show that surface markers identified by PP-SLIDE can be used to enrich for replication-competent latent cells, and that these cells are enriched for the transcriptionally active reservoir and exhibit clear evidence of clonal expansion of fully-intact proviruses.

### **Universal panel enriches for latent cells with replication-competent HIV**

The sort panels implemented above were tailored for each donor specimen. Given that some features of kNN latent cells are shared between individuals (Fig. 3), we set out to design a universal panel to enrich for latent cells in a donor-independent fashion. We designed a sorting strategy incorporating 8 surface antigens that enrich for kNN latent cells from all 4 blood

specimens analyzed (Fig. S12). As TIGIT was found to be expressed at high levels on kNN latent cells in some but not all donors (Fig. S12), we designed a strategy to isolate two enriched populations – one TIGIT<sup>high</sup> and one TIGIT<sup>med/low</sup> – to test the hypothesis that TIGIT is indeed not a marker of latent cells shared between donors. These populations were compared to a PD1+ population disenriched for kNN latent cells (Fig. S12). We applied this “universal” panel on four donor blood specimens (Table S2) that had not undergone PP-SLIDE analysis (Fig. S13). Viral outgrowth assays revealed robust outgrowth in the enriched but not disenriched populations (Fig. 5F). Furthermore, the TIGIT<sup>high</sup> population harbored more virus in some donors and the TIGIT<sup>med/low</sup> one in others, validating the notion that TIGIT is not a marker of latent cells shared between donors. These results suggest that shared features between kNN latent cells can be exploited to enrich for the replication-competent latent reservoir across donors.

## DISCUSSION

Although the latent reservoir has long been recognized as a main barrier to HIV cure, exactly what these cells look like *in vivo* has remained somewhat obscure. A common viewpoint, based on the limited number of studies that have conducted viral outgrowth assays on sorted cell populations, is that latent HIV persists in all types of CD4+ T cells with no preference for one subset over another. If so, targeting latently-infected cells as a curative strategy would be all the more difficult. Taking advantage of recent developments in high-parameter single-cell phenotyping and our PP-SLIDE analysis pipeline, we now show that latently-infected cells are in fact not randomly distributed among CD4+ T cells, and provide an in-depth view of the phenotypic features of the *in vivo* blood and tissue reservoir.

Multiple observations make us confident that PP-SLIDE identifies the physiologically relevant reservoir. First, our *in vitro* assay with J-Lat cells showed the method could predict the correct precursor cell 99.2% of the time. Second, cells sorted based on PP-SLIDE were enriched for genome-intact and replication-competent reservoir cells in the expected

populations. Of note, our PP-SLIDE characterization had charted the translation-competent reservoir, defined as latent cells able to produce viral proteins upon stimulation (Baxter et al., 2018), which may include some replication-incompetent HIV provirus (Baxter et al., 2016; Imamichi et al., 2020). Nevertheless, our viral outgrowth results closely aligned with the enrichment of kNN latent cells, confirming that PP-SLIDE identifies markers of the replication-competent HIV reservoir most relevant for a cure. Third, using our “universal panel,” we validated the preferential presence of the replication-competent reservoir in expected populations in four untested donors, and PP-SLIDE’s prediction that the latent reservoir is preferentially maintained in TIGIT-expressing cells in some but not all donors.

The sorting strategies used in our validation experiments were designed to enable comparison of different populations of memory CD4+ T cells expressing high levels of PD1, a marker preferentially expressed on reservoir cells (Banga et al., 2016; Chomont et al., 2009; Fromentin et al., 2016; Pardons et al., 2019) and which we found expressed at high levels on kNN latent cells. This was implemented for two reasons: First, to demonstrate that we can enrich for latent cells beyond the selection afforded by PD1 alone. Indeed both our viral outgrowth and proviral sequencing results confirmed PP-SLIDE’s prediction that some but not all PD1-expressing memory CD4+ T cells harbor high frequencies of latent HIV. Second, to enable mechanistic comparisons of highly similar populations of sorted cells harboring vastly different levels of replication-competent HIV. We found that in 7/8 cases, the enriched populations initiated more HIV transcription than their disenriched counterparts, suggesting that the transcriptionally active reservoir associates most closely with the replication-competent reservoir and therefore should be a target of HIV curative strategies. We also found that relative to disenriched populations, the enriched ones harbored higher proportions of EIS with intact p24, consistent with the importance of clonal expansion in driving maintenance of the replication-competent HIV reservoir (Bui et al., 2017; Hosmane et al., 2017; Lorenzi et al.,

2016), although it should be noted that cells with different integration sites can harbor the same proviral sequence (Patro et al., 2019).

Of particular interest was our in-depth comparison of two PD1-expressing memory CD4+ T cell populations from PID5003. A striking 65.2% of proviruses in the enriched population from this donor were fully intact, versus 0.9% in the disenriched, despite both being PD1 expressors. To put these numbers in perspective, on average only 2-3% of sequences from infected individuals treated during chronic infection are intact (Bruner et al., 2016; Hiener et al., 2017). The overall frequency of intact provirus in the enriched PID5003 population was also extremely high: at 214 per million cells, it was more than an order of magnitude higher than what is observed in Tem (range = 1–26, median, 13 per million), the subset with the highest frequencies of intact provirus as assessed by FLIPS (Hiener et al., 2017). This high frequency is not due to an unusually large overall reservoir in this donor, as the phenotypically-similar (PD1+) disenriched population sequenced from this donor exhibited a frequency of only 2 per million. Our observation that this enriched population, but not the disenriched one, exhibited robust viral outgrowth, further suggests that the intact proviruses in this population are replication-competent. Interestingly, the intact sequences in this enriched population consisted of two large clusters of EIS, suggesting two major clonal expansions in this well-defined population of memory CD4+ cells sorted on the basis of 8 surface markers. As these two clusters of viral sequences likely integrated at different genomic sites, these observations support a host- rather than virus-driven expansion of these latent cells.

That we could find such vastly different levels of intact and replication-competent reservoir cells in our enriched versus disenriched populations goes against the notion that the reservoir is “everywhere” and randomly distributed in all subsets. But how do our observations fit with prior reports that the frequencies of replication-competent provirus, as determined from outgrowth assays, were similar between sorted subsets (Buzon et al., 2014; Kwon et al., 2020)? We believe that the previously sorted subsets were not ones that differ markedly in reservoir

frequencies. Most prior subset analyses had focused on the Tem and Tcm subsets, and when focusing on just these subsets our PP-SLIDE results indeed support the notion that both Tem and Tcm were prominent among kNN latent cells. However, our findings that the proportions of Tem was greater among the kNN latent cells than among atlas cells suggest that latency is somewhat biased towards the Tem subset, a finding in line with the higher prevalence of intact provirus in Tem over Tcm (Hiener et al., 2017). Further supporting a non-random distribution of the reservoir is our identification of markers consistently expressed at higher levels on kNN latent cells relative to memory CD4+ T cells. Such markers include ones previously implicated in latency (PD1, CTLA4, Ox40) (Banga et al., 2016; Fromentin et al., 2016; Kuo et al., 2018; McGary et al., 2017), some activation markers (CD25, CD69), and markers of Th1, Th2, and Th17 cells (Tbet, CRTh2, and CCR6, respectively) suggesting preferential persistence of HIV in more differentiated cells.

Our analysis of longitudinal specimens by PP-SLIDE also provided insights into the stability of the reservoir and its capacity for reactivation. By comparing kNN latent cells from different donors versus longitudinal samples, we found intra-individual variability to be less than inter-individual variability. This observation implies that the phenotypic properties of the reservoir are stable over time, a feature that may facilitate its targeting. In fact, longitudinal stability also held for the reservoir's susceptibility to the LRAs we tested. Our observation that different stimulatory signals reactivate different latent cells further suggests that "shock and kill" strategies may require multiple LRAs, each preferentially targeting different cell subsets, as recently suggested (Baxter et al., 2016; Grau-Exposito et al., 2019; Pardons et al., 2019).

Although our study focused on the blood reservoir for logistical and technical reasons, we also conducted PP-SLIDE analyses on tissue specimens. Perhaps not surprising given the known phenotypic differences between T cells from blood versus tissues (Wong et al., 2015), we found that blood and tissue kNN latent cells differ quite dramatically. High-dimensional distance calculations revealed that kNN latent cells were more similar between lymph node and

gut than between the tissues and blood, suggesting phenotypic commonalities in the latent reservoir found in tissues. Indeed, kNN latent cells from both lymph node and gut expressed high levels of the Trm marker CD69, suggesting that Trm in these tissues, like those from the cervix (Cantero-Perez et al., 2019), preferentially harbor the reservoir. Together with our observation that kNN latent cells from all sites examined expressed high levels of PD1, these results suggest PD1-expressing Trm as possible targets to eliminate the tissue reservoir. Given recent findings that blood, lymph nodes, and gut are the major sources of HIV dissemination in ART-suppressed individuals (Chaillon et al., 2020), targeting latent cells in the compartments analyzed in this study may be sufficient to make a meaningful reduction in the reservoir.

Our findings do not support the existence of a single surface biomarker that can be used to target the entire HIV reservoir at once. However, PP-SLIDE enables in-depth charting of the latent reservoir residing in blood and, likely, any tissue we can access. For therapeutic applications, subsets of surface antigens identified by PP-SLIDE as preferentially co-expressed on latent cells can be targeted with a series of bi- or multi-specific antibodies (Klein et al., 2016). More importantly, our ability to isolate highly enriched populations of replication-competent reservoir cells enables in-depth functional analyses of the reservoir, and potentiates future implementation of discovery-based approaches such as scRNAseq and PCR-activated cell sorting (PACS) (Bradley et al., 2018; Clark and Abate, 2017). These approaches hold great promise for identifying novel markers for HIV eradication, but require a high frequency of latent cells that was not achievable previously. We envision that implementing PP-SLIDE on cells phenotyped by scRNAseq together with Antibody-seq approaches (Peterson et al., 2017; Stoeckius et al., 2017) will enable an even higher-resolution mapping, with the added advantage that whole-transcriptome approaches do not require pre-defining markers of interest, and allow for the discovery of unanticipated—and perhaps unique—markers of latent cells.

## AUTHOR CONTRIBUTIONS

J.N., J.F., G.X., F.H., V.M., A.L., S.T., W.T., and K.H.J., conducted the experiments; J.N., X.L., J.F., F.H., T.M, V.M., A.L., S.T., R.T., B.W., N.M.A., S.A.Y., S.P., and N.R.R. conducted data analysis; X.L., R.T., and B.W. wrote scripts; J.N., X.L., J.F., F.H., V.M., A.L., S.T., K.S.J., N.M.A., and N.R.R. designed the experiments; J.N., X.L., N.M.A., P.W.H., S.A.Y., S.P., W.C.G., and N.R.R. conceived ideas within the study; R.H., M.S., P.V., J.M., and S.G.D. provided specimens and advised on clinical aspects of study; N.R.R. supervised the study and wrote the manuscript. All authors read and approved the manuscript.

## ACKNOWLEDGEMENTS

This work was supported by the National Institutes of Health (R01AI127219 and R01AI147777 to N.R.R., P01AI131374 to N.R.R. and W.C.G.) and the amfAR Institute for HIV Cure Research (109301). We also acknowledge NIH for the sorter (S10-RR028962), support from CFAR (P30AI027763), and the Pendleton Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank M. McCune for insightful discussions; N. Lazarus and E. Butcher for the Act1 antibody; S. Tamaki and C. Bisbo for CyTOF assistance at the Parnassus Flow Core; N. Raman for assistance in flow cytometry at the Gladstone Flow Core; Viva Tai and Marian for assistance with the SCOPE specimens; J. Carroll for assistance in graphics; F. Chanut for editorial assistance; and R. Givens for administrative assistance.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

## FIGURE LEGENDS

**Figure 1. Validation of PP-SLIDE as method to trace reactivated J-Lat clones to their parent. A)** Schematic. J-Lat clones 5A8 or 6.3 were phenotyped by CyTOF in the absence of stimulation (a). 6.3 was additionally stimulated for 40 h with PMA/ionomycin and then phenotyped by CyTOF (b). The high-dimensional datasets of unstimulated 5A8 and 6.3 were merged to generate the “atlas” of unstimulated J-Lat cells. The high-dimensional data for each reactivated (Gag+) 6.3 cell was matched against the atlas by PP-SLIDE (c). The error rate was defined as the % of reactivated 6.3 cells that matched to unstimulated 5A8 cells (d). **B)** PP-SLIDE correctly identifies the unstimulated precursor of reactivated 6.3 cells 99.2% of the time. tSNE plots of unstimulated J-Lat 6.3 (a), unstimulated J-Lat 5A8 (b), and PMA/ionomycin-stimulated J-Lat 6.3 (c). PP-SLIDE was used to identify the kNN latent cell of each reactivated 6.3 cell among a merged atlas of unstimulated 6.3 and 5A8 cells. Note marked overlap of kNN latent cells (d) with unstimulated J-Lat 6.3 (e, 99.2%) and not J-Lat 5A8 (f, 0.8%).

**Figure 2. PP-SLIDE charting of *in vivo* latent reservoir reveals non-random distribution of latent cells. A)** Schematic of PP-SLIDE analysis of *in vivo* reservoir. Memory CD4+ T cells from infected, ART-suppressed individuals were isolated and an aliquot immediately processed for CyTOF phenotyping (a), while the remaining cells were phenotyped after treatment with PMA/ionomycin for 40 h in the presence of ART (b). PP-SLIDE identified a kNN latent cell (the most similar cell in the atlas of unstimulated cells) for each reactivated cell (c). An overlay of the kNN latent cells (red) onto the atlas (grey) reveals regions preferentially associated with latency (d). **B)** Identification of *ex vivo*-reactivated latent cells from blood of ART-suppressed individuals. Cells were phenotyped before or after stimulation for 40 h with PMA/ionomycin in the presence of ART. Reactivated cells are shown in the gates. **C)** kNN latent cells are non-randomly distributed among memory CD4+ T cells. *Left:* Reactivated cells (red) visualized by

tSNE alongside unstimulated memory CD4+ T cells (black) from the same patient. Due to phenotypic changes induced by stimulation and reactivation, the reactivated cells (stacked as tight populations) reside in distinct regions of each tSNE plot (red ovals). *Center:* Atlas of memory CD4+ T cells from each sample, clustered by FlowSOM. Each cluster is depicted as a different color. *Right:* The kNN latent cells are colored according to the cluster they belong to. **D)** Pie graphs showing relative proportions of each cluster among the atlas. “D” (“Detectable”) designates clusters harboring at least one kNN latent cell and “U” (“Undetectable”) those lacking any. The D clusters are arranged in order of the frequency of kNN latent cells they harbor, with D1 clusters harboring the highest frequencies. The existence of small D clusters and large U clusters, along with the chi-squared values, demonstrate non-random distribution of the latent reservoir.

**Figure 3. kNN latent cells from different donors are phenotypically similar. A)** The atlas (grey) and kNN latent cells (red) for 4 donors depicted within the same tSNE space. In all donors, the center region of the tSNE is devoid of kNN latent cells. **B)** Expression levels of select antigens on kNN latent cells as compared to the total population of memory CD4+ T cells. Top plots show the antigen expression levels as a heatmap within the tSNE space depicted in *panel A*. Bottom graphs show the mean signal intensity (MSI) of each antigen in the total population of memory CD4+ T cells (“Total”) compared to the kNN latent cells (“Latent”). Proteins consistently expressed at higher levels on kNN latent cells than on the total population included immune checkpoint molecules (PD1, CTLA4) and the HIV co-receptors CCR5 and CXCR4. Some activation markers (CD69, CD25, HLA-DR) but not others (CD38) were also upregulated, as were markers of Th1, Th2, and Th17 cells (Tbet, CCR6, and CCR6, respectively). Some markers of Tfh (PD1, ICOS) but not others (CXCR5) were consistently elevated. In contrast, the Tcm marker CD27 was expressed at low levels on kNN latent cells. \* p < 0.05, \*\* p < 0.01 as determined by a Student’s paired t-test. n.s.: not significant.

**Figure 4. Comparison of kNN latent cells reactivatable by different stimulations and from blood vs. tissues using longitudinal specimens.** **A)** Atlas (grey) and kNN latent cells (red) from 3 blood specimens from the same donor (PID5003) spaced 2-3 months apart, where reactivation was induced by either PMA/ionomycin or Romidepsin/PEP005. Arrows highlight regions that harbor kNN latent cells in the PMA/ionomycin but not Romidepsin/PEP005 samples. **B)** kNN latent cells reactivatable by PMA/ionomycin at 2 different timepoints are more similar to each other than to those reactivatable by Romidepsin/PEP005 at the same timepoint. The empirical cumulative distribution of median distances between kNN cells was calculated between 1) Timepoint 1 PMA/ionomycin and Timepoint 2 PMA/ionomycin samples (red), and 2) Timepoint 2 Romidepsin/PEP005 and Timepoint 3 Romidepsin/PEP005 samples (pink), and 3) Timepoint 2 PMA/ionomycin and Timepoint 2 Romidepsin/PEP005 samples (blue). **C)** Atlases of memory CD4+ T cells from fine needle aspirates (FNAs) or blood of the same donor (PID3010) with corresponding kNN latent cells (red). Red arrows point to regions of the FNA tSNEs that are concentrated in kNN latent cells. Bottom plots show each FNA atlas overlaid onto the blood atlas. **D)** Dot plots demonstrating that FNA but not blood kNN latent cells (red dots) express uniformly high levels of CD27, CD69, CXCR5, PD1, and ICOS. Atlas cells are in grey. Pie charts show proportions of kNN latent cells within each quadrant. **E)** kNN latent cells from different tissue sites are similar. *Top:* Atlas (*left*) and kNN latent cells (*right*) from 4 blood, 2 FNA, and 4 gut specimens as depicted by tSNE. *Middle/Bottom:* Empirical cumulative distribution of median distances between atlas (*left*) or kNN cells (*right*), as calculated between FNA vs. blood (*purple*) or vs. gut (*blue*).

**Figure 5. Sorting PD1+ cells based on surface markers identified by PP-SLIDE highly enriches for latent cells with replication-competent, transcriptionally active, and clonally expanded HIV.** **A)** Enriched populations of sorted cells are enriched for replication-competent

HIV. Three populations of memory CD4+ PD1+ cells – one enriched and 2 disenriched for kNN latent cells – were purified by sorting (Fig. S9) and then subjected to a viral outgrowth assay. The cells were stimulated for 2 days with anti-CD3/CD28, diluted to the indicated number of cells per well, and cultured with MOLT-CCR5 cells for 15 days. Wells were then assayed for p24 by ELISA. Data are reported as the concentration of p24 in each well (bar graphs), and the fraction of wells that exhibited viral replication as defined by a final p24 content of > 1 ng/ml (numbers above each bar). **B)** Enriched populations harbor more HIV transcription initiation. The enriched and disenriched populations from 4 donors were left unstimulated and quantitated for levels of initiated (TAR), elongated (Long LTR, Pol), and polyadenylated (PolyA) HIV RNAs. Data are normalized to the corresponding region of HIV DNA. The dotted red line shows the levels of initiated HIV transcripts for the enriched populations. **C)** Enriched populations harbored higher frequencies of total HIV and intact p24 sequences in 3 donors, as revealed by FLIPS. **D)** Pie charts showing the proportions of full-length intact and defective proviruses (including large internal deletions, hypermutations, premature stop codons, and inversions) from the enriched (*top*) and disenriched (*bottom*) populations. The enriched and disenriched populations harbored full-length intact proviruses at proportions of 66.5% and 0.9%, and at overall frequencies of 214 and 2 per million cells, respectively. **E)** Phylogenetic analysis of proviruses demonstrating enrichment of identical full-length intact proviruses in the enriched (*top*) and disenriched (*bottom*) populations. Wedges depict expanded identical sequences (EIS). Concentric circles identify proviruses with (*yellow*) or without (*black*) intact p24 sequences, and with (*green*) or without (*black*) full-length intact sequences. Blue bars depict the length of each proviral sequence. Groups 1 and 2 label the 2 expansions of full-length intact sequences in the enriched population. **F)** The universal panel enriches for replication-competent reservoir cells. Memory CD4+ PD1+ from 4 untested donors were sorted into two enriched (one TIGIT<sup>high</sup> and one TIGIT<sup>med/low</sup>) and one disenriched population (Fig. S13) and then subjected to the viral outgrowth assay described in *panel A*.

## EXTENDED METHODS

### Study Participants and Specimen Collection

Multiple leukapheresis samples were obtained from eight HIV-1 subtype-B individuals on long-term ART who initiated therapy during chronic (>1 year) infection and were stably suppressed (HIV RNA <40 copies/ml) (Table S2). Leukapheresis specimens from four of these eight individuals were subjected to full analyses by CyTOF / PP-SLIDE. Lymph node specimens were obtained from one of these four participants under local anesthesia using ultrasound-guided fine needle aspiration. All fine needle aspirate (FNA) procedures were performed without any adverse effects, and the subject reported minimal to no discomfort both during and after the procedure, and was able to perform their usual activities after the procedure. Gut specimens were obtained by sigmoid biopsies from four ART-suppressed HIV-infected participants (Table S2). All participants provided informed consent before participation. The study was approved by the University of California, San Francisco (IRB # 10-01330) and the University of North Carolina (IRB # 12-1660).

### Cells

#### *J-Lat culture and stimulation*

J-Lat clone 5A8 was previously described (Chan et al., 2013) and J-LAT clone 6.3 was obtained from the NIH AIDS Reagent Program. Both cell lines were cultured in RP10 media (RPMI 1640 medium (Corning) supplemented with 10% fetal bovine serum (FBS, VWR), 1% penicillin (Gibco), and 1% streptomycin (Gibco)). Cells were either processed for CyTOF in the absence of stimulation, or stimulated with 16 nM phorbol myristate acetate (PMA, Sigma-Aldrich) and 0.5 µM ionomycin (Sigma-Aldrich) for 24 h prior to processing for CyTOF analysis as described in the appropriate section below.

### ***Preparation of HLACs***

Uninfected and HIV-infected human lymphoid aggregate cultures (HLACs) were used as a source of cells to validate the CyTOF panel. Human tonsils from the Cooperative Human Tissue Network (CHTN) were dissected into small pieces and pressed through a 40- $\mu$ m strainer (Fisher). The resulting HLAC cells were cultured at a concentration of  $10^6$  cells/well in 96-well U-bottom polystyrene plates (VWR) in Complete Media (CM) (consisting of RPMI 1640 supplemented with 15% FBS, 100  $\mu$ g/ml gentamicin (Gibco), 200  $\mu$ g/ml ampicillin (Sigma-Aldrich), 1 mM sodium pyruvate (Sigma-Aldrich), 1% non-essential amino acids (Mediatech), 1% Glutamax (ThermoFisher), and 1% Fungizone (Invitrogen)) for 3 days in the absence or presence of 100–200 ng/ml p24<sup>Gag</sup> of F4.HSA. Both uninfected and infected cells were then treated with cisplatin and PFA as described in the appropriate section below, and analyzed by CyTOF.

### ***Memory CD4+ T cell enrichment of PBMCs***

PBMCs were isolated from leukapheresis specimens using Lymphoprep<sup>TM</sup> (StemCell Technologies). A portion of the cells were immediately cryopreserved by pelleting the cells and resuspending them at a concentration of 50-100 $\times 10^6$  cells/ml in RPMI 1640 medium (Corning) with 90% FBS and 10% DMSO (Sigma-Aldrich). CD4+ T cells were then purified by negative selection using the EasySep Human CD4+ T Cell Enrichment Kit (Stemcell Technologies), and then further enriched for memory cells by negative selection with CD45RA MicroBeads (Miltenyi Biotec) to remove naïve cells.

### ***Isolation and preparation of lymph node cells***

Lymph node specimens were obtained by FNA biopsies of 1-2 lymph nodes in the inguinal chain area. For sampling, 3cc slip tip syringes and either 23- or 22-gauge needles, 1 1/2 inch in length, were used. Generally, 4-5 passes per lymph node were performed. A portion of

the first pass was taken for fixation in 95% alcohol onto a slide, which was then stained with Toluidine Blue and then examined by microscopy. After confirmation from the cytopathologist that the sample was representative of the target lymph node, the remaining sample in the needle was transferred into RPMI medium (Corning). Due to the small numbers of cells from these specimens, the cells were neither cryopreserved nor enriched for CD4+ T cells, and were immediately prepared for PP-SLIDE analysis. A total of 2 FNAs were drawn from the same participant (PID3010) ~6 months apart.

### ***Isolation and preparation of gut cells***

Processing of sigmoid biopsies into single-cell suspensions was performed similar to a recently described (Trapecar et al., 2017). Briefly, tissue specimens were first incubated for 20 min at 37°C with rotation in intraepithelial lymphocyte (IEL) solution (consisting of 200 ml DPBS, 10 mM DTT (Sigma-Aldrich), 5 mM EDTA (Thermofisher), 10 mM HEPES (Thermofisher), and 5% FBS), after which supernatants were saved and tissues subjected to two additional rounds of treatment. Media from these incubations were combined, filtered through a 70 micron filter (Fisher), and the resulting IELs were resuspended in CM. The tissue pieces remaining after the IEL incubations were incubated for 20 min at 37°C with 10 ml Rinse Buffer (RPMI 1640 with 10 mM HEPES and 5% FBS). Tissue pieces were then placed into a gentleMACS C Tube (Miltenyi Biotec) in 1 ml of digestion solution consisting of 6 ml RPMI-1640, 10 mM HEPES, 5% FBS, 6 mg collagenase (Worthington-Biochemical Corp), and 7.5 µg/ml DNase (Sigma-Aldrich), and rotated for 30 min at 37°C. Two ml of Rinse Buffer was then added for a total of 3 ml. The sample was vortexed for 10 seconds and the sample processed through the GentleMACS Dissociator (Miltenyi Biotec) to isolate cells from the lamina propria. The tissue specimens were then mechanically dissociated for 10 passes using a 5 ml syringe and a blunt, 20G needle (Becton Dickinson). After adding 5 ml Rinse Buffer, the sample was passed through a 70 micron strainer, washed with another 10 ml Rinse Buffer, and passed through a new 70 micron strainer.

The flow-through material containing cells was then pelleted and washed once with 5 ml Rinse Buffer. The cells were then treated at 4°C for 30 min with 1 ml of DNase solution (6 ml RPMI-1640, 10 mM HEPES, 5% FBS, and 7.5 µg/ml DNase (Sigma-Aldrich)). The cells were then washed once in CM, and then combined with the IELs. Of note, these gentle processing conditions were established and optimized to minimize digestion-induced changes in surface antigen expression (Trapecar et al., 2017). Due to the small number of cells from these specimens, the cells were neither cryopreserved nor enriched for CD4+ T cells, and were immediately prepared for PP-SLIDE analysis. In one of the donors, half of the cells were used in a viral outgrowth assay.

### **Kinetics of proliferation following mitogen stimulation**

Memory CD4+ T cells from blood of two HIV-seronegative donors was purified as described above, and then resuspended at  $2 \times 10^7$  cells/ml in PBS containing 0.1% FBS. The cells were then loaded for 8 minutes with 1.5 µM of the proliferation dye CFSE (ThermoFisher). Labeling was then stopped by addition of an equal volume of pre-warmed, 100% FBS, and the labelled cells were incubated at 37°C for an additional 10 minutes. The sample was then washed 3 times in RP10. Memory CD4+ T cells not exposed to CFSE were treated identically in parallel. The time=0 specimen was immediately stained as described below. The remaining cells were cultured with or without activation with 16 nM PMA, 1 µM ionomycin, and 100 IU/ml IL2 (all three reagents from ThermoFisher). At the indicated timepoints, cells were stained with APC/Cyanine7 anti-human CD3 (Clone SK7, Biolegend), PE/Cy7 anti-human CD4 (Clone A161A1, Biolegend), Alexa Fluor 700 anti-human CD8 (Clone SK1, Biolegend), and Zombie Red (Biolegend) as a Live/Dead discriminator. Stained cells were fixed and analyzed by FACS on an LSRII (BD Biosciences). Flowjo (BD Biosciences) was used for analysis. Live, singlet

CD3+CD4+CD8- cells were assessed for proliferation by monitoring the loss of CFSE signal.

Results shown are from one of two donors which gave similar results.

### **Viral production**

293T cells were seeded in 6-well plates (Falcon) at a concentration of  $3 \times 10^5$  cells/well, and transfected the next day using FuGENE (Promega) with 0.5  $\mu$ g/well of the F4.GFP or F4.HSA proviral constructs previously described (Cavrois et al., 2017; Ma et al., 2020; Neidleman et al., 2017). Supernatants from the cultures were harvested after 2 days and p24<sup>Gag</sup> concentrations were measured with a Lenti-X p24 Rapid Titer kit (Clontech).

### ***In vitro* confirmation of the ART cocktail's suppressive activity**

To demonstrate that the cocktail of ART used in this study was fully suppressive, PBMC-derived CD4+ T cells from HIV-seronegative donors were activated for 2 days with RP10 containing 10  $\mu$ g PHA (PeproTech) and 100 IU/ml IL2 (Thermofisher), washed with RP10, and then cultured in 20 IU/ml IL2 for one additional day. The cells were then either left untreated, or pretreated for 1 h with an ART cocktail consisting of 5  $\mu$ M AZT, 5  $\mu$ M Ritonavir, 8 nM Efavirenz, 10  $\mu$ M Lamivudine, 50 nM Raltegravir, and 0.5  $\mu$ g/ml T-20 (all from NIH AIDS reagent program). Cells were then cultured for 4 days at a concentration of  $2 \times 10^6$  cells/ml with RP10 alone or in the presence of 250 ng/ml p24<sup>Gag</sup> F4.GFP in RP10 containing 20 IU/ml IL2. The cells were then stained for 30 minutes at 4°C at a concentration of  $10^6$  cells/well in FACS buffer, with anti-CD3 APC/Cyanine7, anti-CD4 PE/Cy7, anti-CD8 Alexa Fluor 700, and Zombie Aqua as a Live/Dead discriminator (all from Biolegend). The cells were then washed 3 times with FACS buffer and fixed overnight in 2% PFA (Electron Microscopy Sciences). The cells were analyzed on an LSRII (BD Biosciences) and Flowjo software was used to gate on live, singlet CD3+CD8- cells, and infection rates in these cells were monitored by GFP expression.

### Preparation of participant specimens for PP-SLIDE analysis

For atlas generation, we used freshly isolated and purified memory CD4+ T cells from blood, or freshly isolated total cells from tissues (as described above). A portion of these cells were immediately treated with cisplatin (Sigma-Aldrich) as a Live/Dead marker and fixed. Briefly,  $6 \times 10^6$  cells were resuspended at room temperature in 2 ml PBS (Rockland) with 2 mM EDTA (Corning). Next, 2 ml of PBS containing 2 mM EDTA and 25  $\mu$ M cisplatin (Sigma-Aldrich) were added to the cells. The cells were quickly mixed and incubated at room temperature for 60 seconds, after which 10 ml of CyFACS (metal contaminant-free PBS (Rockland) supplemented with 0.1% FBS and 0.1% sodium azide (Sigma-Aldrich)) was added to quench the reaction. The cells were then centrifuged and resuspended in 2% PFA in CyFACS, and incubated for 10 minutes at room temperature. The cells were then washed twice in CyFACS, after which they were resuspended in 100  $\mu$ l of CyFACS containing 10% DMSO. These fixed cells were stored at -80°C until analysis by CyTOF.

The rest of the freshly isolated cells were stimulated to allow for latent cell reactivation. These cells were diluted to  $2 \times 10^6$  cells/ml and stimulated with 16 nM PMA (Sigma-Aldrich) and 1  $\mu$ M ionomycin (Sigma-Aldrich) for 40 hours. This stimulation time was chosen as it is similar to times previously implemented (Cohn et al., 2018) and shorter stimulations under these conditions did not lead to reactivation. Where indicated, cells were instead stimulated with Dynabeads® Human T-Activator CD3/CD38 beads (Gibco), or 12 nM PEP005 (Ingenol 3-angelate, Sigma-Aldrich) with 40 nM romidepsin (Sigma-Aldrich). For gut specimens, to minimize cellular toxicity, cells were instead stimulated for 16 hours with 160 nM PMA and 1  $\mu$ M ionomycin. All stimulations were conducted in the presence of 100 IU/ml IL2 (Thermofisher). To prevent spreading infection, a cocktail of ART consisting of 5  $\mu$ M AZT, 5  $\mu$ M Ritonavir, 8 nM Efavirenz, 10  $\mu$ M Lamivudine, 50 nM Raltegravir, and 0.5  $\mu$ g/ml T-20 was added. To limit death

of reactivated cells, 10  $\mu$ M of the pan-caspase inhibitor Z-VAD-FMK (R&D Systems Inc) was added as previously described (Cohn et al., 2018; Grau-Exposito et al., 2019).

### **CyTOF staining and data acquisition**

Staining of cells for analysis by CyTOF was conducted similar to recently described methods (Cavrois et al., 2017; Ma et al., 2020; Trapecar et al., 2017). Briefly, cisplatin-treated cells were thawed and washed in Nunc 96 DeepWell<sup>TM</sup> polystyrene plates (Thermo Fisher) with CyFACS buffer at a concentration of  $6 \times 10^6$  cells/800  $\mu$ l per well. Cells were then pelleted and blocked with mouse (Thermo Fisher), rat (Thermo Fisher), and human AB (Sigma-Aldrich) sera for 15 minutes at 4°C. The samples were then washed twice in CyFACS, pelleted, and stained in a 100  $\mu$ l cocktail of surface antibodies (Table S1) for 45 minutes at 4°C. The samples were then washed 3X with CyFACS and fixed overnight at 4°C in 100  $\mu$ l of freshly prepared 2% PFA in PBS (Rockland). Samples were then washed twice with Intracellular Fixation & Permeabilization Buffer (eBioscience) and incubated in this buffer for 45 minutes at 4°C. Next, samples were washed twice in Permeabilization Buffer (eBioscience). The samples were then blocked for 15 minutes at 4°C in 100  $\mu$ l of mouse and rat sera diluted in Permeabilization Buffer, washed 1X with Permeabilization buffer, and incubated for 45 minutes at 4°C in a 100  $\mu$ l cocktail of intracellular antibodies (Table S1) diluted in Permeabilization Buffer. The cells were then washed with CyFACS and stained for 20 minutes at room temperature with 250 nM of Cell-ID<sup>TM</sup> Intercalator-IR (Fluidigm). Finally, the cells were washed 2X with CyFACS buffer, once with MaxPar® cell staining buffer (Fluidigm), and once with Cell acquisition solution (CAS, Fluidigm), and then resuspended in EQ<sup>TM</sup> Four Element Calibration Beads (Fluidigm) diluted to 1X in CAS. Sample concentration was adjusted to acquire at a rate of 200 - 350 events/sec using a wide-bore (WB) injector on a CyTOF2 instrument (Fluidigm) at the UCSF Parnassus flow core facility.

For tissue specimens where multiple specimens were combined to minimize cell loss, the Cell-ID 20-Plex Pd Barcoding Kit (Fluidigm) was used according to the manufacturer's instructions to barcode the individual cisplatin-treated samples prior to mixing. Briefly, the frozen and fixed samples were thawed and washed once with CyFACS and 2X with Maxpar® Barcode Perm Buffer (Fluidigm), and then resuspended in 800 µl of Barcode Perm Buffer. A total of 10 µl of each barcode was diluted in 100 µl of Barcode Permeabilization Buffer and added to each sample. The samples were barcoded for 30 minutes at room temperature, washed once with MaxPar® cell staining buffer, and once with CyFACS. The samples were then combined subjected to the CyTOF staining protocol as described above.

## **Cell Sorting**

Memory CD4+ T cells purified from cryopreserved PBMCs as described above were sorted using the following antibodies. Antibody clone numbers are indicated in the table further below:

### **Panel #1 (used for PID1695):**

CD27 PE-CF594; CD38 BUV737; CD45RO BUV395; CD57 FITC; CD62L BV650; CD69 BV421; CXCR5 APC-R7; OX40 PE-Cy7; PD-1 PE

### **Panel #2 (used for PID3010):**

CD45RO BUV395; CD62L BV650; CD69 APC; CXCR5 APC-R700; OX40 PE-Cy7; PD-1 PE; CD28 BUV737; CCR5 BV421; CD49d BB515; CCR7 PE-Dazzle

### **Panel #3 (used for PID5003):**

CD27 PE-CF594; CD62L BV650; CD57 PE-Cy7; CD38 BUV737; CD25 BUV395; CCR5 BV421; CD49d BB515; CD7 APC Cy7; CD127 BV 605; TIGIT APC; PD-1 PE; CXCR5 APC R7; CD28 PE-Cy5

### **Panel #4 (used for PID2053):**

CXCR5 APC R7; PD-1 PE; CD27 PE-CF594; CD62L BV650; CD45RO BUV395; CD2 BV421; CD57 FITC; TIGIT APC; CD69 APC-Cy7; OX40 PE-Cy7

**Panel #5 (universal panel used for untested donors PIDs 2402, 2781, 2161, and 2046):**

PD-1 PE; CD49d BB515; CD57 PE-Cy7; CD38 BUV737; CD27 PE-CF594; CD62L BV650; CD45RO BUV395; CD2 BV421; TIGIT APC

Dead cells were excluded using the Zombie Aqua fixable viability kit (BioLegend), while doublets were excluded based on FSC-A/FSC-H. Cells were sorted on an Aria FACS sorter (Becton Dickinson). The purity of the subsets was confirmed by analysis on the Aria immediately post-sort.

**Sorting Antibodies**

Antigen Target	Clone	Vendor (catalog number)
Live/Dead (Zombie Aqua)		BioLegend (#423102)
PE-CF594 Mouse Anti-Human CD27	M-T271	BD (#562297)
BUV737 Mouse Anti-Human CD38	HB7	BD (#564686)
BUV395 Mouse Anti-Human CD45RO	UCHL1	BD (#564291)
FITC Mouse Anti-Human CD57	NK-1	BD (#555619)
BV650 Mouse Anti-Human CD62L	DREG-56	BD (#563808)
APC Mouse Anti-Human CD69	FN50	BioLegend (#310910)
BV421 Mouse Anti-Human CD69	FN50	BD (#562884)
APC Mouse Anti-Human CD103	Ber-ACT8	BD (#563883)
APC-R700 Rat Anti-Human CXCR5 (CD185)	RF8B2	BD (#565191)
PE-Cy7 Mouse Anti-Human CD134 (OX40)	ACT35	BD (#563663)
PE Mouse anti-Human CD279 (PD-1)	EH12.1	BD (#560795)
BUV737 Mouse Anti-Human CD28	CD28.2	BD (#612815)
BB515 Mouse Anti-Human CD49d	9F10	BD (#564593)
PE/Dazzle 594 Mouse Anti-Human CD197 (CCR7)	G043H7	BioLegend (#353236)
Brilliant Violet 421 Rat Anti-Human CD195 (CCR5)	J418F1	BioLegend (#359118)
PE-Cy7 Mouse Anti-Human CD57	TB01	eBioscience (25-0577-42)
BUV395 Mouse Anti-Human CD25	2A3	BD (#564034)
APC-H7 Mouse Anti Human CD7	M-T701	BD (#564020)
BV605 Mouse Anti Human CD127	HIL-7R-M21	BD (#562662)
APC Mouse Anti Human TIGIT	A15153G	BioLegend (#372706)
PE-Cy5 Mouse Anti Human CD28	CD28.2	BioLegend (#302910)

APC-Cy7 Mouse Anti Human CD69	FN50	BD (#560737)
BV421 Mouse anti human CD2	TS1/8	BioLegend (#309218)
PE-Cy7 Mouse anti human CD57	TBO1	Invitrogen (#25-0577-42)

### **Viral outgrowth assays**

Cells sorted from blood were cultured at a density of  $10^6$  cells/mL in RPMI in the presence of 60 IU/mL recombinant human IL-2 (Thermofisher), and activated with Dynabeads® Human T-Activator CD3/CD28 (Gibco) at a ratio of 1 bead/cell. After 48 h of culture, cells were diluted to a concentration of  $10^5$  cells/mL in RPMI supplemented with 60 IU/ml recombinant human IL-2, and then plated at  $2 \times 10^5$  cells per well (dilution A),  $4 \times 10^4$  cells per well (dilution B),  $8 \times 10^3$  cells per well (dilution C), and  $1.6 \times 10^3$  cells per well (dilution D). Molt4/CCR5 cells (Laird et al., 2013) were then added at  $10^5$  cells per well. Cultures were replenished at days 3, 6, 12 and 15 by replacing half of the media with RPMI containing 5 IU/mL IL-2 without disturbing the cellular layer. On day 9, half the cultures, inclusive of both cells and media, were discarded, and the remaining cells were replenished with the same volume of RPMI containing  $10^5$  Molt4/CCR5 cells. At the end of the culture period (days 15 – 18), supernatants were quantitated for levels of p24 antigen by ELISA (Takara).

A viral outgrowth assay was conducted for one of the gut specimens (PID1223) to demonstrate the presence of replication-competent virus in the tissue. To this end, a Digital ELISA Viral Outgrowth (DEVO) assay was implemented. Pooled biopsy cells were plated in a limiting dilution format at 5 replicates at 470,000 cells per well, 12 replicates at 100,000 cells per well, and 12 replicates at 25,000 cells per well. Cells were stimulated for 24 hours with 3  $\mu$ g/mL phytohemagglutinin (PHA) (Thermofisher), 100 iU/mL IL-2, and irradiated PBMCs from an HIV-seronegative donor. The cells were washed and Molt4/CCR5 target cells were added to amplify outgrowth of HIV. The cultures were replenished with fresh media every 3-4 days, and another round of Molt4/CCR5 cells added on day 9 post-stimulation. Supernatants were harvested on days 9, 13, 16, 20, 26, and 29 post stimulation and assessed for HIV p24 using the SIMOA HD-

1 Analyzer (Quanterix, Billerica, MA, USA). Only wells exhibiting sustained and increasing p24 expression over time were scored as positive. The frequency of infection, reported as Infectious Units Per Million (IUPM) of rectal-sigmoid cells was calculated using the SLDAssay R software package (Trumble et al., 2017). The IUPM from this analysis was 0.265, with 95% confidence intervals between 0.0216 and 1.400. Of note, this IUPM corresponds to that of total gut cells.

### **HIV DNA and transcriptional profiling analysis**

Pellets of sorted cells were lysed in 1 ml TRI reagent with 2.5  $\mu$ l polyacryl carrier (Molecular Research Center). Total cellular RNA was subsequently extracted per the TRI Reagent protocol, while total cellular DNA was extracted using the ‘back extraction buffer’ (4M guanidine thiocyanate, 50mM sodium citrate, 1M Tris) as described (Yukl et al., 2018). A common reverse transcriptase (RT) reaction was used to generate cDNA for all droplet digital PCR (ddPCR) assays except TAR, where a separate two-step RT was performed as described (Yukl et al., 2018). Briefly, each 50  $\mu$ l RT reaction contained cellular RNA, 5  $\mu$ l of 10x Superscript III buffer (Invitrogen), 5  $\mu$ l of 50 mM MgCl<sub>2</sub>, 2.5  $\mu$ l of 50 ng /  $\mu$ l random hexamers (Invitrogen), 2.5  $\mu$ l of 50  $\mu$ M dT15, 2.5  $\mu$ l of 10 mM dNTPs, 1.25  $\mu$ l of 40 U /  $\mu$ l RNaseOUT (Invitrogen), and 2.5  $\mu$ l of 200 U /  $\mu$ l Superscript III RT (Invitrogen). PBMCs from uninfected donors, and water that was subjected to nucleic extraction by TRI Reagent, served as negative controls for each transcript. The thermocycling conditions were as follows: 25°C for 10 min, 50°C for 50 min, and an inactivation step at 85°C for 5 min.

For the ddPCR reactions, cDNA from each sample was assayed in duplicate wells for TAR, Long LTR, Pol, and PolyA regions using validated assays (Kaiser et al., 2017; Telwatte et al., 2018; Yukl et al., 2018). Total cell equivalents in the DNA extracted from the same samples were determined by measuring the absolute copy numbers of a nonduplicated cellular gene, Telomere Reverse Transcriptase (TERT), similar to previously described methods (Telwatte et al., 2018; Yukl et al., 2018). To account for the effects of deletions or hypermutations on the HIV

RNA measurements, all HIV RNAs were further normalized to HIV DNA measured by the same assay used for the RNA. Each 20  $\mu$ l ddPCR reaction contained 5  $\mu$ l cDNA or DNA, 10  $\mu$ l of ddPCR Supermix for Probes (no dUTP) (Bio-Rad, Hercules), 900 nM of primers, and 250 nM of probe. Following production of droplet emulsions using the QX100 Droplet Generator (Bio-Rad), samples were amplified under the following thermocycling conditions using a 7900 Thermal Cycler (Life Technologies): 10 minutes at 95°C, 45 cycles of 30 seconds at 95°C and 59°C for 60 seconds, and a final droplet cure step of 10 minutes at 98°C. Droplets were quantified using the QX100 Droplet Reader (Bio-Rad) and analyzed using the QuantaSoft software (version 1.6.6, Bio-Rad) in the “Rare Event Detection” quantification mode.

### **Full-length Individual Proviral Sequencing assay (FLIPS)**

The FLIPS assay was performed as previously reported (Hiener et al., 2017). Briefly, HIV-1 proviruses within sorted CD4+ T cell subsets were amplified and sequenced at limiting dilution to near full-length (9kb; 92% of the genome). A median of 140 individual proviruses (range 18 – 254) were sequenced per sorted population of cells from 3 of the 4 participant blood specimens analyzed by PP-SLIDE. Next-generation sequencing was conducted using the Illumina MiSeq platform with individual proviruses de novo assembled using a specifically designed workflow in CLC Genomics. Proviruses were characterized as defective if they contained INDELs, stop codons, or APOBEC3G hypermutations, or intact (full-length) if they lacked such defects. Expanded identical sequences (EIS) were identified using ElimDuplicates from the Los Alamos database

(<https://www.hiv.lanl.gov/content/sequence/elimduplicatesv2/elimduplicates.html>). All sequences that were 100% identical were considered part of an EIS cluster. Maximum likelihood phylogenetic trees using the generalized time-reversible model were estimated for each participant using Geneious. Branch support was inferred using 1000 bootstrap replicates. Annotated tree images were constructed using the iTOL software Version 5.5.1 (Letunic and Bork, 2019).

## CyTOF data export and PP-SLIDE analysis

The CyTOF data were exported as FCS files, and samples were de-barcoded according to manufacturer's instructions (Fluidigm). For the J-Lat experiments, equal numbers of the unstimulated live, singlet intact 6.3 and 5A8 were concatenated together as a source of the "atlas cells". Reactivated 6.3 cells were identified by gating on the Gag-expressing cells in the stimulated 6.3 sample. For patient samples, events corresponding to live, singlet intact CD3+CD19-CD8- T cells from unstimulated samples were exported as a source of the atlas cells. Events corresponding to reactivated cells were identified by gating upon live, singlet intact CD3+CD19-CD8- T cells expressing Gag on all three Gag channels, and exported as the population of reactivated cells. Where indicated, naïve cells (CD45RA+CD45RO-) were gated out prior to export. Data export was conducted using FlowJo (BD Biosciences) and Cytobank software. tSNE analyses were performed using the Cytobank software with default settings. All cellular markers not used in the upstream gating strategy were included in generating the tSNE plots. Non-cellular markers (e.g., live/dead stain) and the HIV gag proteins were excluded for the generation of tSNE plots. Dot plots were generated using both Cytobank and FlowJo.

PP-SLIDE analysis to identify kNN latent cells followed our previously described method (Cavrois et al., 2017; Ma et al., 2020). Using a custom script in R, each reactivated (Gag+) cell from the stimulated sample was mapped against every cell in the unstimulated atlas generated immediately after sample procurement, to identify the kNN latent cells for each patient sample.

The steps implemented for PP-SLIDE are summarized below:

### *(1) Data cleanup and standardization:*

The atlas cells and the reactivated cells were exported as described above. To prepare for PP-SLIDE, the following parameters, which do not contain useful information for identifying the original cell type, were removed from the analysis:

Non-informative markers	Live/dead staining, event length, barcodes, beads channel, DNA, time, background channel (190), and other non-cell markers
Infection marker	All 3 sets of anti-Gag antibodies
Marker highly modified by HIV and not informative for identification of kNN latent cell	CD4
Markers not expressed on CD4+ T cells	CD19, CD8, CD14

Raw expression values (signal intensity) of selected markers from each cell in the exported files were transformed by the inverse hyperbolic function (arcsinh) transformation as follows:

$$\text{arsinh}(x) = \ln(x + \sqrt{x^2 + 1})$$

This transformation is used to standardize the diverse range of raw expression level scales for the measured parameters, and minimizes the effect of outliers and extreme numbers.

*(2) Identification of kNN latent cell for each reactivated cell:*

The Euclidean distance ( $d_{F\_U}$ ) between each reactivated cell  $F$  and each atlas cell  $U$  was calculated as follows:

$$d_{F\_U} = \sqrt{\sum_{i=1}^n (F_i - U_i)^2}$$

where  $n$  is the number of parameters analyzed and  $i$  refers to the parameter being analyzed. For example, for parameter 1,  $F_i - U_i$  would correspond to the value of parameter 1 on the reactivated cell minus the value of parameter 1 on the atlas cell.

For each reactivated cell  $F$ , the  $d_{F\_U}$  of all the atlas cells  $U$  were sorted from lowest to highest to identify the shortest  $d_{F\_U}$  value. This corresponds to the  $k=1$  nearest neighbor atlas cell for that particular reactivated cell  $F$ , or the kNN latent cell. After identifying the kNN latent cell corresponding to each reactivated cell, the expression values corresponding to the original data matrix of all the kNN latent cells were combined and exported as a new FCS file for downstream analysis. These kNN latent cells correspond to a subset of the original data matrix of total atlas cells.

In the J-Lat experiments, when PP-SLIDE was implemented using all phenotyping parameters in our CyTOF panel, 6.3 cells mapped back to their parent clone with an error rate of 6.7%. Given that some parameters in our panel, such as activation markers, may not be as informative as other antigens in mapping as they markedly change upon stimulation, we undertook a systematic way to assess how removal of the most changed parameters affects error rates. We first ranked each antigen in terms of to what extent they changed upon stimulation. As expected, the top most changed parameters were mostly activation markers. We then conducted a series of PP-SLIDE analyses where we removed the top most changed parameter, then the top two most changed parameters, etc. As more parameters were removed, the error rate improved up to when the top six most changed antigens were removed, when the error rate reached a minimum of 0.6%. Thereafter, the error rate increased, likely because power was lost as fewer parameters were left to conduct the mapping. The top six changed parameters were those that exhibited a  $> 2.5$  unit difference in expression levels upon reactivation. These data demonstrate that when PP-SLIDE is implemented on all phenotyping parameters it can map reactivated J-Lat 6.3 back to its correct clone with an error rate of 6.7%, and that this can be improved to 0.8% when the mostly highly changed parameters were excluded in the analyses. Because these J-Lat PP-SLIDE results suggested that removing the most changed parameters can potentially improve the error rate, for each patient sample we conducted two sets of PP-SLIDE analyses: one using all phenotyping parameters, and one under conditions where we excluded most changed parameters (defined as a  $> 2.5$  unit difference in expression levels). The kNN latent cells identified from the two PP-SLIDE analyses exhibited similar features. The results presented correspond to those implementing all phenotyping parameters.

## **Empirical cumulative distribution analysis**

To assess the intra-individual and inter-individual variability in kNN latent cells, and to compare kNN latent cells reactivated following treatment with PMA/ionomycin, anti-CD3/CD28, and Romidepsin/PEP005, empirical cumulative distributions were calculated. The distributions of minimum distances in 32-dimensional space (corresponding to measured protein levels of antigens not used in the upstream gating strategy) corresponded to those between cells of the indicated pairs of samples. The minimum (Euclidean) distance over all cells in the first sample was computed for each cell in the second sample using their corresponding 32-dimensional protein level vectors. Then, similarly, the minimum (Euclidean) distance over all cells in the second sample was computed for each cell in the first sample using their corresponding 32-dimensional protein level vectors. The union of these two sets of minimum distances was used to determine the distribution of minimum distances between the sample pair. These resulting datasets were visualized in terms of a cumulative distribution function using the `stat_ecdf` function of the `ggplot2` package in R.

### **Distribution analysis of FlowSOM clusters**

This section describes in detail how FlowSOM was implemented to cluster the atlas and kNN latent cells, and how this information was used to demonstrate non-random distribution of kNN latent cells among memory CD4+ T cells. For simplicity, we have illustrated the process with one of the four leukapheresis donors (PID1695), but similar methods were implemented for the other three. The FlowSOM analysis for all four donor specimens revealed a non-random distribution of kNN latent cells among memory CD4+ T cells.

### ***Clustering of atlas and kNN latent cells***

The default settings of Cytobank's FlowSOM analysis pipeline (Clustering method: Hierarchical consensus; Iterations: 10; Seed: Automatic) was implemented to identify 20 clusters of memory CD4+ T cells (comprised of 20 metaclusters containing 225 sub-clusters) in

the atlas of unstimulated memory CD4+ T cells from PID1695 on the day of blood isolation. To characterize the latent cells, memory CD4+ T cells from the same leukapheresis were stimulated for 40 hours with PMA/ionomycin, and reactivated cells were identified based on intracellular expression of HIV-1 gag proteins. A total of 20 such reactivated cells were identified from this sample. PP-SLIDE was implemented to identify the corresponding 20 kNN latent cells from the atlas. The 20 kNN latent cells were then assigned to the appropriate cluster of the atlas. We found that the latent cells were distributed among 6 of the 20 clusters, while in the remaining 14 clusters no kNN latent cells were found. The distribution of all the memory CD4+ T cells is depicted as a pie graph in [Fig. 2D](#).

The following were the numbers of kNN latent cells ( $N_{\text{latent}}(\text{Cluster DX})$  and  $N_{\text{latent}}(\text{Cluster UX})$ ) in each of the clusters (Clusters beginning with “D” correspond to clusters with detectable virus, those with “U” correspond to clusters with undetectable virus):

$N_{\text{latent}}(\text{Cluster D1}) = 3$   
 $N_{\text{latent}}(\text{Cluster D2}) = 3$   
 $N_{\text{latent}}(\text{Cluster D3}) = 7$   
 $N_{\text{latent}}(\text{Cluster D4}) = 2$   
 $N_{\text{latent}}(\text{Cluster D5}) = 4$   
 $N_{\text{latent}}(\text{Cluster D6}) = 1$   
 $N_{\text{latent}}(\text{Clusters U1-U14}) = 0$  (undetectable)

The number of cells in each of the detectable clusters is represented as the numbers of red-colored circles in the top of [Fig. S5A](#).

#### ***Proportion of latent cells in each cluster calculation***

To calculate the percent of the latent reservoir in each of the detectable clusters we used the following equation:

$$P_{\text{latent}}(\text{DX}) = [( \text{number of kNN latent cells in Cluster DX} ) / ( \text{total number of kNN latent cells} )] * 100$$

This calculation gave the following percentages ( $P(\text{Cluster DX})$ ) of kNN latent cells in each detectable cluster:

$P_{latent}(Cluster D1) = 15\%$   
 $P_{latent}(Cluster D2) = 15\%$   
 $P_{latent}(Cluster D3) = 35\%$   
 $P_{latent}(Cluster D4) = 10\%$   
 $P_{latent}(Cluster D5) = 20\%$   
 $P_{latent}(Cluster D6) = 5\%$

The above percentages of the reservoir among the detectable clusters are depicted in blue in the top of [Fig. S5A](#).

### ***Demonstration of non-random distribution of latent cells***

We then set out to determine whether kNN latent cells preferentially resided within any of the clusters, or alternatively whether the kNN latent cells were stochastically distributed. To do this, we first established the null hypothesis that kNN latent cells are randomly distributed among memory CD4+ T cells. Under this null hypothesis, the percentages of kNN latent cells in each detectable and undetectable cluster ( $(P_{latent\_null}(Cluster DX)$  and  $P_{latent\_null}(Cluster UX)$ , respectively) would reflect the frequency of cells making up that cluster:

$P_{latent\_null}(Cluster D1) = 3\%$   
 $P_{latent\_null}(Cluster D2) = 3\%$   
 $P_{latent\_null}(Cluster D3) = 11\%$   
 $P_{latent\_null}(Cluster D4) = 7\%$   
 $P_{latent\_null}(Cluster D5) = 15\%$   
 $P_{latent\_null}(Cluster D6) = 12\%$   
 $P_{latent\_null}(Cluster U1) = 4\%$   
 $P_{latent\_null}(Cluster U2) = 6\%$   
 $P_{latent\_null}(Cluster U3) = 3\%$   
 $P_{latent\_null}(Cluster U4) = 5\%$   
 $P_{latent\_null}(Cluster U5) = 3\%$   
 $P_{latent\_null}(Cluster U6) = 2\%$   
 $P_{latent\_null}(Cluster U7) = 2\%$   
 $P_{latent\_null}(Cluster U8) = 4\%$   
 $P_{latent\_null}(Cluster U9) = 4\%$   
 $P_{latent\_null}(Cluster U10) = 3\%$   
 $P_{latent\_null}(Cluster U11) = 8\%$   
 $P_{latent\_null}(Cluster U12) = 1\%$   
 $P_{latent\_null}(Cluster U13) = 2\%$   
 $P_{latent\_null}(Cluster U14) = 2\%$

We then converted these percentages into the number of latent cells we would have detected in each detectable and undetectable cluster ( $N_{latent\_null}(Cluster D)$  and  $N_{latent\_null}(Cluster UX)$ ,

respectively) under the conditions whereby our total of 20 kNN cells were identified, which required running a total of 1,671,440 cells. This was accomplished by multiplying each percentage by this total number of kNN latent cells (20 cells), giving the following numbers of detectable kNN cells under the null hypothesis:

$N_{latent\_null}(Cluster\ D1) = 0.6$   
 $N_{latent\_null}(Cluster\ D2) = 0.6$   
 $N_{latent\_null}(Cluster\ D3) = 2.2$   
 $N_{latent\_null}(Cluster\ D4) = 1.4$   
 $N_{latent\_null}(Cluster\ D5) = 3$   
 $N_{latent\_null}(Cluster\ D6) = 2.4$   
 $N_{latent\_null}(Cluster\ U1) = 0.8$   
 $N_{latent\_null}(Cluster\ U2) = 1.2$   
 $N_{latent\_null}(Cluster\ U3) = 0.6$   
 $N_{latent\_null}(Cluster\ U4) = 1$   
 $N_{latent\_null}(Cluster\ U5) = 0.6$   
 $N_{latent\_null}(Cluster\ U6) = 0.4$   
 $N_{latent\_null}(Cluster\ U7) = 0.4$   
 $N_{latent\_null}(Cluster\ U8) = 0.8$   
 $N_{latent\_null}(Cluster\ U9) = 0.8$   
 $N_{latent\_null}(Cluster\ U10) = 0.6$   
 $N_{latent\_null}(Cluster\ U11) = 1.6$   
 $N_{latent\_null}(Cluster\ U12) = 0.2$   
 $N_{latent\_null}(Cluster\ U13) = 0.4$   
 $N_{latent\_null}(Cluster\ U14) = 0.4$

These numbers were then compared to the actual numbers of kNN latent cells detected in each cluster:

$N_{latent}(Cluster\ D1) = 3$   
 $N_{latent}(Cluster\ D2) = 3$   
 $N_{latent}(Cluster\ D3) = 7$   
 $N_{latent}(Cluster\ D4) = 2$   
 $N_{latent}(Cluster\ D5) = 4$   
 $N_{latent}(Cluster\ D6) = 1$   
 $N_{latent}(Clusters\ U1-U14) = 0$  (undetectable)

Comparisons of  $N_{latent}(Cluster\ X)$  to  $N_{latent\_null}(Cluster\ X)$  revealed the following (where X can correspond to either detectable clusters or undetectable clusters):

Clusters D1, D2, D3, D4, and D5 were enriched for kNN latent cells relative to the null hypothesis, as defined by  $N_{latent}(Cluster\ X) > N_{latent\_null}(Cluster)$ .

Clusters D6, U2, and U11 were disenriched for kNN cells relative to the null hypothesis, as defined by  $N_{latent\_null}(Cluster\ X) > 1$  and  $N_{latent\_null}(Cluster\ X) > N_{latent}(Cluster\ X)$ . (In other words, were latent cells equivalently distributed, kNN latent cells should have been detectable in each of these clusters at frequencies higher than that actually observed).

The remaining clusters (all undetectable ones) we could not draw conclusions on with regards to enrichment relative to the null hypothesis, as defined by the conditions  $N_{latent\_null}(Cluster\ X) \leq 1$  and  $N_{latent}(Cluster\ X)$  being undetectable. (In other words, enrichment could have occurred but too few cells were run to have detected a single kNN latent cell under the null hypothesis).

Our observation that kNN latent cells are enriched among Clusters D1-D5 and disenriched among Clusters D6, U2, and U11 refute the null hypothesis that kNN latent cells are evenly distributed among memory CD4+ T cells. These findings were further confirmed by running a chi-squared test ( $p < 0.01$ ).

### ***Frequencies of kNN latent cells among clusters***

We then set out to calculate the frequencies of kNN latent cells in each of the detectable clusters. To do this, we needed to first account for the fact that more cells were analyzed in the sample harvested 40 hours post-stimulation (referred to below as “Day 2 cells”), than in the original atlas sample. To this end, we calculated the normalization factor  $NF$  defined as the fold-increase in cells collected on Day 2 as compared to in the atlas:

$$NF = (\# \text{ of total Day 2 cells}) / (\# \text{ of total atlas cells}) = 1,671,440 / 11,974 = 139.6$$

By multiplying the numbers of cells in each detectable cluster from the atlas ( $N_{\text{day0}}(\text{Cluster DX})$ ) with  $NF$ , we obtain  $N_{\text{atlas}}(\text{Cluster DX})$ , the number of cells in each cluster we would have collected had we run by CyTOF the number of cells equivalent to that needed to detect 20 reactivated cells:

$$N_{\text{atlas}}(\text{Cluster DX}) = NF * N_{\text{day0}}(\text{Cluster DX})$$

To determine the frequencies of kNN latent cells in each cluster, we divide the number of detected kNN latent cells in each cluster ( $N_{\text{latent}}(\text{Cluster DX})$ ) by  $N_{\text{atlas}}(\text{Cluster DX})$ :

$$F_{\text{latent}}(\text{Cluster DX}) = N_{\text{latent}}(\text{Cluster DX}) / N_{\text{atlas}}(\text{Cluster DX})$$

The resulting frequencies of kNN latent cells in the six detectable clusters were as follows (reported as kNN latent cells per million cells in Cluster DX):

$$\begin{aligned} F_{\text{latent}}(\text{Cluster D1}) &= 71 \\ F_{\text{latent}}(\text{Cluster D2}) &= 52 \\ F_{\text{latent}}(\text{Cluster D3}) &= 38 \\ F_{\text{latent}}(\text{Cluster D4}) &= 17 \\ F_{\text{latent}}(\text{Cluster D5}) &= 16 \\ F_{\text{latent}}(\text{Cluster D6}) &= 5 \end{aligned}$$

These numbers are shown in the bottom of [Fig. S6A](#).

#### **Glossary of terms for stats analysis of FlowSOM clusters**

Term	Description
$N_{\text{latent}}(\text{Cluster DX})$	Number of kNN latent cells in Cluster DX
$P_{\text{latent}}(\text{Cluster DX})$	Percent of kNN latent cells belonging to Cluster DX
$P_{\text{latent\_null}}(\text{Cluster X})$	Percent of kNN latent cells belonging to Cluster X under null hypothesis of stochastic distribution
$N_{\text{latent\_null}}(\text{Cluster X})$	Number of kNN latent cells in Cluster X under null hypothesis of stochastic distribution
$N_{\text{day0}}(\text{Cluster DX})$	Number of atlas cells in Cluster X that were run
$N_{\text{atlas}}(\text{Cluster DX})$	Number of atlas cells in Cluster X that would have been run to capture 20 reactivated cells
$F_{\text{latent}}(\text{Cluster DX})$	Frequency of kNN latent cells per million cells of Cluster DX

## DATA AVAILABILITY

Raw CyTOF datasets have been made publically available through the public repository Dryad and is accessible via the following link: <https://doi.org/10.7272/Q6KK991S>

## SUPPLEMENTAL TABLES

**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
HLADR	TÜ36	Qdot (112Cd)	Life Technologies
ROR $\gamma$ t*	AFKJS-9	115 In	In-house
CD49d ( $\alpha$ 4)	9F10	141Pr	Fluidigm
CTLA4*	14D3	142Nd	In-house
NFAT*	D43B1	143Nd	Fluidigm
CCR5	NP6G4	144Nd	Fluidigm
BIRC5*	91630	145Nd	In-house
SAMHD1*	Ag3287 Immunogen	145Nd	In-house
Ki67*	B56	146Nd	In-house
CD95	BX2	146Nd	In-house
BCL6*	K112-91	146Nd	In-house
CD7	CD76B7	147Sm	Fluidigm
ICOS	C398.4A	148Nd	Fluidigm
Tbet*	4B10	149Sm	In-house
Gag (mixture)*	71-31, 91-5, 241-D, AG3.0	150Nd	In-house
CD2	TS1/8	151Eu	Fluidigm
CD103	Ber-ACT8	151Eu	Fluidigm
Gag (28B7)*	28B7	152Sm	In-house
CD62L	DREG56	153Eu	Fluidigm
TIGIT	MBSA43	154Sm	Fluidigm
CCR6	11A9	155Gd	In-house
Gag (KC57)*	FH190-1-1	156 Gd	In-house
CD8	RPA-T8	157Gd	In-house
CD19	HIB19	157Gd	In-house
CD14	M5E2	157Gd	In-house
OX40	ACT35	158Gd	Fluidigm
CCR7	G043H7	159Tb	Fluidigm
CD28	CD28.2	160Gd	Fluidigm
CD45RO	UCHL1	161Dy	In-house

CD69	FN50	162Dy	Fluidigm
CRTH2	BM16	163Dy	Fluidigm
PD-1	EH12.1	164Dy	In-house
CD127	A019D5	165Ho	Fluidigm
CXCR5	RF8B2	166Er	In-house
CD27	L128	167Er	Fluidigm
CD30	BERH8	168Er	In-house
CD45RA	HI100	169Tm	Fluidigm
CD3	UCHT1	170Er	Fluidigm
CD57	HNK-1	171Yb	In-house
CD38	HIT2	172Yb	Fluidigm
$\alpha 4\beta 7$	Act1	173Yb	In-house
CD4	SK3	174Yb	Fluidigm
CXCR4	12G5	175Lu	Fluidigm
CD25	M-A251	176Yb	In-house
Gag (KC57)*	FH190-1-1	209 Bi	In-house

\*Intracellular antibodies

**Table S2: Participant Characteristics**

Patient ID	Gender	Ethnicity	Age	Year of first HIV+ test	Viral Load (copies/ml)	CD4 count (per/mm <sup>3</sup> )	ART regimen	Specimen type
PID1695*	Female	Mixed	40	2010	<40	710	FTC/TDF, ATV, RTV	Leukapheresis
PID3010*	Male	AA	54	1990	<40	902	DRV, RTV, RPV/TAF/FTC	Leukapheresis FNA
PID5003*	Male	Hispanic	48	1993	<40	289	ATV, ABC/TCV/3TC	Leukapheresis
PID2053*	Male	Asian	60	1991	<40	333	FTC/TAF, TCV	Leukapheresis
PID01223	Female	AA	39	2000	<20	989	BIC/FTC/TAF	Gut
PID1128	Male	AA	69	1996	<40	318	BIC/FTC/TAF	Gut
PID2511	Male	White	51	2001	<40	330	TCV/RPV	Gut
PID2402	Male	AA	43	2004	<40	545	BIC/FTC/TAF	Leukapheresis Gut
PID2161	Male	White	71	1985	<40	677	3TC/DRV/RTV/TCV	Leukapheresis
PID2375	Male	White	71	1989	<40	594	EFV/TDF/FTC	Leukapheresis
PID2781	Male	White	43	2009	<40	518	ABC/TCV/3TC	Leukapheresis

\*Complete CyTOF/PP-SLIDE analyses + validations for these leukapheresis specimens

## SUPPLEMENTAL FIGURE LEGENDS

**Figure S1. CyTOF antibody validation. A)** Tonsils were used as a source of primary cells for validating the *in vivo* latency CyTOF panel, as they provide an abundant source of T and B cells which differentially express many antigens within the panel. Shown is the gating strategy to identify live, singlet cells in human lymphoid aggregate cultures (HLACs) from tonsils. **B)** Expression of antigens differentially expressed on T and B cells as assessed by CyTOF. The first 2-dimensional plot boxed in red schematizes the location of T cells (CD3+) and B cells (CD3-), the 2 main cell populations isolated from HLACs. The indicated antibodies were validated by demonstrating that the differential expression patterns of the corresponding antigens on T vs. B cells are consistent with the known expression patterns of these antigens. Cells were pre-gated on live, singlet cells. *Grey inset:* To validate the 3 sets of anti-Gag antibodies, HLACs were exposed to the HIV reporter virus F4.HSA (Cavrois et al., 2017) and the infected cultures were compared to uninfected cultures for expression of Gag as detected by these antibodies. The “Gag mix” antibody comprises mix of 4 different anti-Gag clones. **C)** Select antigens known to be expressed at higher levels on memory compared to naïve CD4+ T cells were validated by demonstrating the expected expression patterns on these 2 populations of cells. Shown on the left are the gates to identify the memory (CD45RO+CD45RA-) and naïve (CD45RO-CD45RA+) cells. Shown on the right are the expression levels of the indicated antigens in memory (blue) and naïve (red) CD4+ T cells. Cells were pre-gated on live, singlet CD4+ T cells. These validation approaches are similar to those previously implemented to validate CyTOF panels used for characterizing HIV-induced remodeling (Cavrois et al., 2017; Ma et al., 2020).

**Figure S2. Establishing conditions for *ex vivo* reactivation of latent cells. A)** A combination cocktail of ART fully suppresses HIV infection. PHA-activated PBMCs were mock-treated or

exposed to the CCR5-tropic HIV-1 reporter strain F4.GFP (Neidleman et al., 2017), either in the absence or presence of a cocktail of ART drugs targeting HIV-1 Reverse Transcriptase, Protease, and Integrase, as well as the fusion inhibitor T20. Cells were then analyzed by FACS 4 days later for infection levels. The proportion of HIV-infected (GFP+) cells is indicated within the gates. Results are pre-gated on live, singlet CD3+CD8- cells. **B)** PMA/ionomycin stimulation of T cells for 2 days does not induce proliferation. PBMC-derived T cells were stained with CFSE and then mock-treated or stimulated with PMA/ionomycin for the indicated number of days prior to FACS analysis. Cells not stained with CFSE were also analyzed at each timepoint. The results demonstrate that proliferation begins to occur 3 days post-stimulation. All stimulations in this study were conducted for less than 2 days when T cells had not yet proliferated.

**Figure S3. Gating strategy to identify ex vivo reactivated cells from 4 HIV-infected, ART-suppressed patients.** Memory CD4+ T cells were purified by negative selection from leukapheresis specimens and then either left unstimulated and fixed immediately, or stimulated for 40 h with PMA/ionomycin and then fixed. The unstimulated and stimulated specimens were then analyzed by CyTOF. Shown is the gating strategy leading to the identification live, singlet memory CD4+ T cells expressing HIV Gag proteins as detected on 2 separate channels. These cells also bound the third set of anti-Gag antibodies, and exhibited downregulation of cell-surface CD4 (data not shown).

**Figure S4. Distribution of Tcm, Tem, and Ttm among kNN latent cells and total memory CD4+ T cells.** Shown are the percent of kNN latent cells (*top*) or total memory CD4+ T cells (*bottom*) belonging to the Tcm (CCR7+CD27+), Tem (CCR7-CD27-), and Ttm (CCR7-CD27+) subsets. Tem cells were over-represented among kNN latent cells relative to total memory

CD4+ T cells. The numbers below the pie graphs correspond to the number of kNN latent cells or total memory CD4+ T cells in each sample.

**Figure S5. The distribution, frequencies, and phenotypic features of clusters most enriched for kNN latent cells.** **A)** *Top:* Shown are the distributions of kNN latent cells among all the FlowSOM “D” clusters in [Fig. 2D](#) that harbored at least one kNN latent cell, for each of the 4 participant blood specimens analyzed. Each red circle corresponds to one kNN latent cell, with the total adding up to the number of total kNN cells identified from that specimen. The percentages of kNN latent cells that belong to each of the clusters is indicated in blue. *Bottom:* Shown are the frequencies of kNN latent cells in each of the FlowSOM clusters that harbored at least one kNN latent cell. For detailed explanation of how the values in this panel were calculated, see Methods. **B)** Shown are the percent of kNN latent cells belonging to the Tcm, Tem, and Ttm subsets for D1-D3, the 3 clusters in each patient sample harboring the highest frequencies of kNN latent cells.

**Figure S6. Characterization of the inter-donor and stimulation-dependent variability in kNN latent cells.** **A)** Schematic of Euclidean distance calculations to assess relative similarities between different populations of cells. In this hypothetical scenario, one seeks to determine whether Population 2 is more similar to Population 1 or to Population 3. To do this, the median Euclidean distances, based on all the CyTOF parameters, are calculated for each pair of populations as shown. The resulting data are depicted as density plots. The pairs of samples with lower distance calculations are more similar. In this hypothetical scenario, Population 2 is most similar to Population 1. **B)** The atlas and kNN latent cells from 3 leukapheresis specimens from 2 donors represented as tSNE plots. The 2 specimens from donor PID5003 were obtained 2 months apart. **C)** The kNN latent cells of the 2 PID5003 specimens are more similar than the PID5003 specimens are to the PID3010 specimen. The empirical cumulative distribution of median distances was calculated between 1) kNN latent cells from the 2 PID5003 specimens (red), and

2) kNN latent cells between the PID3010 specimen and a PID5003 specimen (*blue*). The plot on the left shows the distribution of the data plotted as density plots. The plot on the right shows the same data plotted as empirical cumulative distribution curves, where the y-axis value corresponds to the proportion of the median distances that are less than the corresponding x-axis value. In both plots, the more similar 2 specimens are, the more the curves are shifted to the left. **D)** The kNN latent cells from 3 leukapheresis specimens from the same donor (PID5003), reactivatable with either PMA/ionomycin or anti-CD3/CD28 as indicated. **E)** The kNN latent cells reactivatable by PMA/ionomycin at the 2 different timepoints are more similar than those reactivatable by PMA/ionomycin vs. anti-CD3/CD28 at the same timepoint. The empirical cumulative distribution of median distances was calculated between 1) kNN latent cells of the Timepoint 1 PMA/ionomycin and Timepoint 2 PMA/ionomycin samples (*red*), and 2) the Timepoint 2 PMA/ionomycin and Timepoint 2 anti-CD3/CD28 samples (*blue*). The density plots and empirical cumulative distribution curves are depicted on the left and right sides, respectively.

**Figure S7. Comparison kNN latent cells from blood and tissues.** **A)** The tSNE maps presented in [Fig. 3C](#) are portrayed as heatmaps showing expression levels of CD27, CD69, PD1, CXCR5, TIGIT, and ICOS. Regions with the highest density of kNN latent cells as reflected in [Fig. 3C](#) are circled in red. Dot plot and pie graph depictions of the same datasets are presented in [Fig. 3D](#). **B)** Dot plots comparing the expression patterns of CD27, CD69, CXCR5, PD1, CTLA4, and CCR6 on kNN latent cells from gut and blood. In the dot plots, kNN latent cells are depicted in red and atlas cells in grey. The proportions of kNN latent cells that fall within each quadrant are depicted as pie charts beneath each dot plot. *Top:* Most kNN latent cells from gut reside in quadrant 3 (Q3), corresponding to cells expressing high levels of CD69 and low levels of CD27. *Middle:* Most kNN latent cells from gut reside in Q3, corresponding to cells expressing high levels of PD1 and low levels of CXCR5. *Bottom:* Most kNN latent cells from gut reside in quadrant 2 (Q2), corresponding to cells expressing high levels of CTLA4 and CCR6. **C)** kNN latent cells from

FNAs but not gut include naïve CD4+ T cells. Dot plots showing kNN latent cells (*red*) and atlas cells (*grey*) for 2 FNA and 4 gut specimens. Events were pre-gated on CD4+ T cells (CD3+CD8-CD19-). Naïve cells are visualized as those expressing high levels of CD45RA and low levels of CD45RO. These naïve cells also expressed high levels of CD27 and CCR7 (not shown).

**Figure S8. Design of tailored sort strategies to isolate 3 populations memory CD4+ PD1+ T cells harboring different frequencies of kNN latent cells.** Sort strategies were designed for each of the 4 leukapheresis donors analyzed by PP-SLIDE. Shown are the CyTOF datasets, with unstimulated atlas cells are shown as grey contours and kNN latent cells shown as red contours. Gates in black correspond to upstream gates. Each sorting strategy isolates 3 populations of memory CD4+ PD1+ T cells: 2 disenriched (*pink, purple*), and one enriched (*red*). Results were pre-gated on live, singlet memory CD4+ T cells (CD3+CD8-CD19-CD45RO+CD45RA-). The 3 functional assays applied on sorted cells are listed. The grey inset on the right shows the fold-enrichment of kNN latent cells in the final sorted enriched population relative to each of the disenriched populations. In instances where a dis-enriched population that did not harbor any kNN cells, the fold-enrichment is listed as NA (not available), because the fold-enrichment is infinity when dividing by zero.

**Figure S9. Implementation of tailored sort strategies to isolate 3 populations memory CD4+ PD1+ T cells harboring different frequencies of kNN latent cells.** The sort strategies described in [Fig. S8](#) were converted into multiparameter flow panels and used to isolate the 3 populations of memory CD4+ PD1+ cells for each of the 4 donors as shown. Shown are FACS plots from the sorts. Memory CD4+ T cells from each donor were purified by negative selection prior to sorting. The phenotypic features of each sorted population are depicted beneath the corresponding population.

**Figure S10. Relative to disenriched populations, enriched populations harbor more clonally expanded proviruses with intact p24 sequences.** Relative to disenriched populations, the enriched populations harbored more intact p24 sequences with expanded identical sequences (EIS). The 3 populations of memory CD4+ PD1+ T cells sorted in Fig. S9 were subjected to proviral sequence analysis by FLIPS. Sequences were classified as EIS or solo sequences containing intact or defective p24 sequences. EIS with intact p24 were preferentially present in the enriched populations.

**Figure S11. Phylogenetic trees of near full-length HIV sequences derived from sorted populations.** Maximum likelihood phylogenetic trees from 3 participants representing the proviral sequences from the enriched (red) and disenriched (purple/pink) populations combined within the same tree. Intact p24 sequences are depicted with circles, defective p24 sequences with squares. Clonal expansions are indicated with the shaded wedges emanating from the center. Of note, in PID5003, one proviral sequence from the disenriched population was identical to an EIS harboring 38 identical sequences from the enriched population. Because only 1/39 sequences in this EIS belonged to the disenriched population, we classified this EIS as belonging to the enriched population.

**Figure S12. Design of universal sort panel to enrich for latent cells.** A universal sort strategy was designed based on the CyTOF datasets from the 4 leukapheresis donors analyzed by PP-SLIDE. Plots are shown as contour, with the kNN populations in red and atlas cells in grey. Of note, PID1695 and PID3010 were stained with a panel that had CD103, while PID5003 and PID2053 were stained with a panel that replaced CD103 with CD2. The PID1695 and PID3010 data suggest that kNN latent cells are enriched among CD103<sup>low</sup> cells, while the PID5003 and PID2053 data suggest that kNN latent cells are enriched among CD2<sup>high</sup> cells. Because selection for CD2<sup>high</sup> cells eliminates more atlas cells lacking kNN cells and resulted in a higher enrichment

of kNN latent cells, we chose CD2 instead of CD103 to include within our panel. The final plot in the gating strategy demonstrates that kNN latent cells express high levels of TIGIT in some donors but not others (the gate most enriched for kNN latent cells is in red); both TIGIT<sup>high</sup> and TIGIT<sup>med/low</sup> cells were sorted for comparison. The PD1+ population of disenriched cells is shown in pink. Results were pre-gated on live, singlet memory CD4+ T cells (CD3+CD8-CD19-CD45RO+CD45RA-).

**Figure S13. Implementation of the universal sort strategy to isolate 3 populations memory CD4+ PD1+ T cells harboring different frequencies of kNN latent cells from 4 untested donors.** The universal sort strategy shown in Fig. S12 was converted into a multiparameter sort panel and implemented on 4 new donors. Shown are the sequential FACS plots depicting the sorting that was used to isolate, for each donor, a single population of disenriched cells, and 2 populations of enriched cells (TIGIT<sup>high</sup> vs. TIGIT<sup>med/low</sup>). Memory CD4+ T cells from each donor were purified by negative selection prior to sorting. The phenotypic features of each sorted population are depicted beneath the corresponding population.

## References

Banga, R., Procopio, F.A., Noto, A., Pollakis, G., Cavassini, M., Ohmiti, K., Corpataux, J.M., de Leval, L., Pantaleo, G., and Perreau, M. (2016). PD-1+ and follicular helper T cells are responsible for persistent HIV-1 transcription in treated aviremic individuals. *Nature medicine*.

Baxter, A.E., Niessl, J., Fromentin, R., Richard, J., Porichis, F., Charlebois, R., Massanella, M., Brassard, N., Alsahafi, N., Delgado, G.G., *et al.* (2016). Single-Cell Characterization of Viral Translation-Competent Reservoirs in HIV-Infected Individuals. *Cell host & microbe* 20, 368-380.

Baxter, A.E., O'Doherty, U., and Kaufmann, D.E. (2018). Beyond the replication-competent HIV reservoir: transcription and translation-competent reservoirs. *Retrovirology* 15, 18.

Bradley, T., Ferrari, G., Haynes, B.F., Margolis, D.M., and Browne, E.P. (2018). Single-Cell Analysis of Quiescent HIV Infection Reveals Host Transcriptional Profiles that Regulate Proviral Latency. *Cell reports* 25, 107-117 e103.

Bruner, K.M., Murray, A.J., Pollack, R.A., Soliman, M.G., Laskey, S.B., Capoferri, A.A., Lai, J., Strain, M.C., Lada, S.M., Hoh, R., *et al.* (2016). Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nature medicine* 22, 1043-1049.

Bui, J.K., Sobolewski, M.D., Keele, B.F., Spindler, J., Musick, A., Wiegand, A., Luke, B.T., Shao, W., Hughes, S.H., Coffin, J.M., *et al.* (2017). Proviruses with identical sequences comprise a large fraction of the replication-competent HIV reservoir. *PLoS pathogens* 13, e1006283.

Buzon, M.J., Sun, H., Li, C., Shaw, A., Seiss, K., Ouyang, Z., Martin-Gayo, E., Leng, J., Henrich, T.J., Li, J.Z., *et al.* (2014). HIV-1 persistence in CD4+ T cells with stem cell-like properties. *Nature medicine* 20, 139-142.

Cantero-Perez, J., Grau-Exposito, J., Serra-Peinado, C., Rosero, D.A., Luque-Ballesteros, L., Astorga-Gamaza, A., Castellvi, J., Sanhueza, T., Tapia, G., Lloveras, B., *et al.* (2019). Resident memory T cells are a cellular reservoir for HIV in the cervical mucosa. *Nat Commun* *10*, 4739.

Cavrois, M., Banerjee, T., Mukherjee, G., Raman, N., Hussien, R., Rodriguez, B.A., Vasquez, J., Spitzer, M.H., Lazarus, N.H., Jones, J.J., *et al.* (2017). Mass Cytometric Analysis of HIV Entry, Replication, and Remodeling in Tissue CD4+ T Cells. *Cell reports* *20*, 984-998.

Chaillon, A., Gianella, S., Dellicour, S., Rawlings, S.A., Schlub, T.E., De Oliveira, M.F., Ignacio, C., Porrachia, M., Vrancken, B., and Smith, D.M. (2020). HIV persists throughout deep tissues with repopulation from multiple anatomical sources. *The Journal of clinical investigation* *130*, 1699-1712.

Chan, J.K., Bhattacharyya, D., Lassen, K.G., Ruelas, D., and Greene, W.C. (2013). Calcium/calmodulin synergizes with prostratin to promote NF- $\kappa$ B dependent activation of latent HIV. *PLoS One* *8*, e77749.

Chomont, N., El-Far, M., Ancuta, P., Trautmann, L., Procopio, F.A., Yassine-Diab, B., Boucher, G., Boulassel, M.R., Ghattas, G., Brenchley, J.M., *et al.* (2009). HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nature medicine* *15*, 893-900.

Chun, T.W., Stuyver, L., Mizell, S.B., Ehler, L.A., Mican, J.A., Baseler, M., Lloyd, A.L., Nowak, M.A., and Fauci, A.S. (1997). Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proceedings of the National Academy of Sciences of the United States of America* *94*, 13193-13197.

Clark, I.C., and Abate, A.R. (2017). Finding a helix in a haystack: nucleic acid cytometry with droplet microfluidics. *Lab on a chip* *17*, 2032-2045.

Cohn, L.B., Chomont, N., and Deeks, S.G. (2020). The Biology of the HIV-1 Latent Reservoir and Implications for Cure Strategies. *Cell host & microbe* 27, 519-530.

Cohn, L.B., da Silva, I.T., Valieris, R., Huang, A.S., Lorenzi, J.C.C., Cohen, Y.Z., Pai, J.A., Butler, A.L., Caskey, M., Jankovic, M., and Nussenzweig, M.C. (2018). Clonal CD4(+) T cells in the HIV-1 latent reservoir display a distinct gene profile upon reactivation. *Nature medicine* 24, 604-609.

De Scheerder, M.A., Vrancken, B., Dellicour, S., Schlub, T., Lee, E., Shao, W., Rutsaert, S., Verhofstede, C., Kerre, T., Malfait, T., *et al.* (2019). HIV Rebound Is Predominantly Fueled by Genetically Identical Viral Expansions from Diverse Reservoirs. *Cell host & microbe* 26, 347-358 e347.

Descours, B., Petitjean, G., Lopez-Zaragoza, J.L., Bruel, T., Raffel, R., Psomas, C., Reynes, J., Lacabaratz, C., Levy, Y., Schwartz, O., *et al.* (2017). CD32a is a marker of a CD4 T-cell HIV reservoir harbouring replication-competent proviruses. *Nature* 543, 564-567.

Estes, J.D., Kityo, C., Ssali, F., Swainson, L., Makamdop, K.N., Del Prete, G.Q., Deeks, S.G., Luciw, P.A., Chipman, J.G., Beilman, G.J., *et al.* (2017). Defining total-body AIDS-virus burden with implications for curative strategies. *Nature medicine* 23, 1271-1276.

Finzi, D., Hermankova, M., Pierson, T., Carruth, L.M., Buck, C., Chaisson, R.E., Quinn, T.C., Chadwick, K., Margolick, J., Brookmeyer, R., *et al.* (1997). Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science (New York, N.Y)* 278, 1295-1300.

Fromentin, R., Bakeman, W., Lawani, M.B., Khoury, G., Hartogensis, W., DaFonseca, S., Killian, M., Epling, L., Hoh, R., Sinclair, E., *et al.* (2016). CD4+ T Cells Expressing PD-1, TIGIT and LAG-3 Contribute to HIV Persistence during ART. *PLoS pathogens* 12, e1005761.

Gosselin, A., Wiche Salinas, T.R., Planas, D., Wacleche, V.S., Zhang, Y., Fromentin, R., Chomont, N., Cohen, E.A., Shacklett, B., Mehraj, V., *et al.* (2017). HIV persists in CCR6+CD4+ T cells from colon and blood during antiretroviral therapy. *AIDS* 31, 35-48.

Grau-Exposito, J., Luque-Ballesteros, L., Navarro, J., Curran, A., Burgos, J., Ribera, E., Torrella, A., Planas, B., Badia, R., Martin-Castillo, M., *et al.* (2019). Latency reversal agents affect differently the latent reservoir present in distinct CD4+ T subpopulations. *PLoS pathogens* 15, e1007991.

Grau-Exposito, J., Serra-Peinado, C., Miguel, L., Navarro, J., Curran, A., Burgos, J., Ocana, I., Ribera, E., Torrella, A., Planas, B., *et al.* (2017). A Novel Single-Cell FISH-Flow Assay Identifies Effector Memory CD4(+) T cells as a Major Niche for HIV-1 Transcription in HIV-Infected Patients. *MBio* 8.

Hiener, B., Horsburgh, B.A., Eden, J.S., Barton, K., Schlub, T.E., Lee, E., von Stockenstrom, S., Odevall, L., Milush, J.M., Liegler, T., *et al.* (2017). Identification of Genetically Intact HIV-1 Proviruses in Specific CD4(+) T Cells from Effectively Treated Participants. *Cell reports* 21, 813-822.

Ho, Y.C., Shan, L., Hosmane, N.N., Wang, J., Laskey, S.B., Rosenbloom, D.I., Lai, J., Blankson, J.N., Siliciano, J.D., and Siliciano, R.F. (2013). Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* 155, 540-551.

Hogan, L.E., Vasquez, J., Hobbs, K.S., Hanhauser, E., Aguilar-Rodriguez, B., Hussien, R., Thanh, C., Gibson, E.A., Carvidi, A.B., Smith, L.C.B., *et al.* (2018). Increased HIV-1 transcriptional activity and infectious burden in peripheral blood and gut-associated CD4+ T cells expressing CD30. *PLoS pathogens* 14, e1006856.

Horsburgh, B.A., Lee, E., Hiener, B., Eden, J.S., Schlub, T.E., von Stockenstrom, S., Odevall, L., Milush, J.M., Liegler, T., Sinclair, E., *et al.* (2020). High levels of genetically-intact HIV in HLA-DR+ memory T-cells indicates their value for reservoir studies. *AIDS*.

Hosmane, N.N., Kwon, K.J., Bruner, K.M., Capoferri, A.A., Beg, S., Rosenbloom, D.I., Keele, B.F., Ho, Y.C., Siliciano, J.D., and Siliciano, R.F. (2017). Proliferation of latently infected CD4(+) T cells carrying replication-competent HIV-1: Potential role in latent reservoir dynamics. *The Journal of experimental medicine* **214**, 959-972.

Iglesias-Ussel, M., Vandergeeten, C., Marchionni, L., Chomont, N., and Romerio, F. (2013). High levels of CD2 expression identify HIV-1 latently infected resting memory CD4+ T cells in virally suppressed subjects. *J Virol* **87**, 9148-9158.

Imamichi, H., Dewar, R.L., Adelsberger, J.W., Rehm, C.A., O'Doherty, U., Paxinos, E.E., Fauci, A.S., and Lane, H.C. (2016). Defective HIV-1 proviruses produce novel protein-coding RNA species in HIV-infected patients on combination antiretroviral therapy. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 8783-8788.

Imamichi, H., Smith, M., Adelsberger, J.W., Izumi, T., Scrimieri, F., Sherman, B.T., Rehm, C.A., Imamichi, T., Pau, A., Catalfamo, M., *et al.* (2020). Defective HIV-1 proviruses produce viral proteins. *Proceedings of the National Academy of Sciences of the United States of America*.

Jordan, A., Bisgrove, D., and Verdin, E. (2003). HIV reproducibly establishes a latent infection after acute infection of T cells in vitro. *The EMBO journal* **22**, 1868-1877.

Josefsson, L., von Stockenstrom, S., Faria, N.R., Sinclair, E., Bacchetti, P., Killian, M., Epling, L., Tan, A., Ho, T., Lemey, P., *et al.* (2013). The HIV-1 reservoir in eight patients on long-term

suppressive antiretroviral therapy is stable with few genetic changes over time. *Proceedings of the National Academy of Sciences of the United States of America* **110**, E4987-4996.

Kaiser, P., Joshi, S.K., Kim, P., Li, P., Liu, H., Rice, A.P., Wong, J.K., and Yukl, S.A. (2017). Assays for precise quantification of total (including short) and elongated HIV-1 transcripts. *J Virol Methods* **242**, 1-8.

Khoury, G., Anderson, J.L., Fromentin, R., Hartogenesis, W., Smith, M.Z., Bacchetti, P., Hecht, F.M., Chomont, N., Cameron, P.U., Deeks, S.G., and Lewin, S.R. (2016). Persistence of integrated HIV DNA in CXCR3 + CCR6 + memory CD4+ T cells in HIV-infected individuals on antiretroviral therapy. *AIDS* **30**, 1511-1520.

Klein, C., Schaefer, W., and Regula, J.T. (2016). The use of CrossMAb technology for the generation of bi- and multispecific antibodies. *MAbs* **8**, 1010-1020.

Kuo, H.H., Ahmad, R., Lee, G.Q., Gao, C., Chen, H.R., Ouyang, Z., Szucs, M.J., Kim, D., Tsibris, A., Chun, T.W., *et al.* (2018). Anti-apoptotic Protein BIRC5 Maintains Survival of HIV-1-Infected CD4(+) T Cells. *Immunity* **48**, 1183-1194 e1185.

Kwon, K.J., Timmons, A.E., Sengupta, S., Simonetti, F.R., Zhang, H., Hoh, R., Deeks, S.G., Siliciano, J.D., and Siliciano, R.F. (2020). Different human resting memory CD4(+) T cell subsets show similar low inducibility of latent HIV-1 proviruses. *Sci Transl Med* **12**.

Laird, G.M., Eisele, E.E., Rabi, S.A., Lai, J., Chioma, S., Blankson, J.N., Siliciano, J.D., and Siliciano, R.F. (2013). Rapid quantification of the latent reservoir for HIV-1 using a viral outgrowth assay. *PLoS pathogens* **9**, e1003398.

Lee, G.Q., Orlova-Fink, N., Einkauf, K., Chowdhury, F.Z., Sun, X., Harrington, S., Kuo, H.H., Hua, S., Chen, H.R., Ouyang, Z., *et al.* (2017). Clonal expansion of genome-intact HIV-1 in functionally polarized Th1 CD4+ T cells. *The Journal of clinical investigation* **127**, 2689-2696.

Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic acids research* **47**, W256-W259.

Li, X., Liu, Z., Li, Q., Hu, R., Zhao, L., Yang, Y., Zhao, J., Huang, Z., Gao, H., Li, L., *et al.* (2019). CD161(+) CD4(+) T Cells Harbor Clonally Expanded Replication-Competent HIV-1 in Antiretroviral Therapy-Suppressed Individuals. *MBio* **10**.

Liu, R., Simonetti, F.R., and Ho, Y.C. (2020). The forces driving clonal expansion of the HIV-1 latent reservoir. *Virol J* **17**, 4.

Lorenzi, J.C., Cohen, Y.Z., Cohn, L.B., Kreider, E.F., Barton, J.P., Learn, G.H., Oliveira, T., Lavine, C.L., Horwitz, J.A., Settler, A., *et al.* (2016). Paired quantitative and qualitative assessment of the replication-competent HIV-1 reservoir and comparison with integrated proviral DNA. *Proceedings of the National Academy of Sciences of the United States of America* **113**, E7908-E7916.

Ma, T., Luo, X., George, A.F., Mukherjee, G., Sen, N., Spitzer, T.L., Giudice, L.C., Greene, W.C., and Roan, N.R. (2020). HIV efficiently infects T cells from the endometrium and remodels them to promote systemic viral spread. *Elife* **9**.

McGary, C.S., Deleage, C., Harper, J., Micci, L., Ribeiro, S.P., Paganini, S., Kuri-Cervantes, L., Benne, C., Ryan, E.S., Balderas, R., *et al.* (2017). CTLA-4(+)PD-1(-) Memory CD4(+) T Cells Critically Contribute to Viral Persistence in Antiretroviral Therapy-Suppressed, SIV-Infected Rhesus Macaques. *Immunity* **47**, 776-788 e775.

Neidleman, J.A., Chen, J.C., Kohgadai, N., Muller, J.A., Laustsen, A., Thavachelvam, K., Jang, K.S.,

Sturzel, C.M., Jones, J.J., Ochsenbauer, C., *et al.* (2017). Mucosal stromal fibroblasts markedly enhance HIV infection of CD4+ T cells. *PLoS pathogens* *13*, e1006163.

Pardons, M., Baxter, A.E., Massanella, M., Pagliuzza, A., Fromentin, R., Dufour, C., Leyre, L., Routy, J.P., Kaufmann, D.E., and Chomont, N. (2019). Single-cell characterization and quantification of translation-competent viral reservoirs in treated and untreated HIV infection. *PLoS pathogens* *15*, e1007619.

Patro, S.C., Brandt, L.D., Bale, M.J., Halvas, E.K., Joseph, K.W., Shao, W., Wu, X., Guo, S., Murrell, B., Wiegand, A., *et al.* (2019). Combined HIV-1 sequence and integration site analysis informs viral dynamics and allows reconstruction of replicating viral ancestors. *Proceedings of the National Academy of Sciences of the United States of America* *116*, 25891-25899.

Peterson, V.M., Zhang, K.X., Kumar, N., Wong, J., Li, L., Wilson, D.C., Moore, R., McClanahan, T.K., Sadekova, S., and Klappenbach, J.A. (2017). Multiplexed quantification of proteins and transcripts in single cells. *Nat Biotechnol* *35*, 936-939.

Serra-Peinado, C., Grau-Exposito, J., Luque-Ballesteros, L., Astorga-Gamaza, A., Navarro, J., Gallego-Rodriguez, J., Martin, M., Curran, A., Burgos, J., Ribera, E., *et al.* (2019). Expression of CD20 after viral reactivation renders HIV-reservoir cells susceptible to Rituximab. *Nat Commun* *10*, 3705.

Soriano-Sarabia, N., Bateson, R.E., Dahl, N.P., Crooks, A.M., Kuruc, J.D., Margolis, D.M., and Archin, N.M. (2014). Quantitation of replication-competent HIV-1 in populations of resting CD4+ T cells. *J Virol* *88*, 14070-14077.

Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P.K., Swerdlow, H., Satija, R., and Smibert, P. (2017). Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* **14**, 865-868.

Sun, H., Kim, D., Li, X., Kiselinova, M., Ouyang, Z., Vandekerckhove, L., Shang, H., Rosenberg, E.S., Yu, X.G., and Licherfeld, M. (2015). Th1/17 Polarization of CD4 T Cells Supports HIV-1 Persistence during Antiretroviral Therapy. *J Virol* **89**, 11284-11293.

Telwatte, S., Lee, S., Somsouk, M., Hatano, H., Baker, C., Kaiser, P., Kim, P., Chen, T.H., Milush, J., Hunt, P.W., *et al.* (2018). Gut and blood differ in constitutive blocks to HIV transcription, suggesting tissue-specific differences in the mechanisms that govern HIV latency. *PLoS pathogens* **14**, e1007357.

Trapecar, M., Khan, S., Roan, N.R., Chen, T.H., Telwatte, S., Deswal, M., Pao, M., Somsouk, M., Deeks, S.G., Hunt, P.W., *et al.* (2017). An Optimized and Validated Method for Isolation and Characterization of Lymphocytes from HIV+ Human Gut Biopsies. *AIDS Res Hum Retroviruses* **33**, S31-S39.

Trumble, I.M., Allmon, A.G., Archin, N.M., Rigdon, J., Francis, O., Baldoni, P.L., and Hudgens, M.G. (2017). SLDAssay: A software package and web tool for analyzing limiting dilution assays. *Journal of immunological methods* **450**, 10-16.

van der Maaten, L. (2009). Learning a parametric embedding by preserving local structure. In *JMLR Workshop and Conference Proceedings*, Vol. 5: Porceedings of the 12th International Conference on Artificial Intelligence and Statistics, D. van Dyk, and M. Welling, eds. (Clearwater Beach, FL), pp. 384-391.

Van Gassen, S., Callebaut, B., Van Helden, M.J., Lambrecht, B.N., Demeester, P., Dhaene, T., and Saeys, Y. (2015). FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. *Cytometry A* *87*, 636-645.

von Stockenstrom, S., Odevall, L., Lee, E., Sinclair, E., Bacchetti, P., Killian, M., Epling, L., Shao, W., Hoh, R., Ho, T., *et al.* (2015). Longitudinal Genetic Characterization Reveals That Cell Proliferation Maintains a Persistent HIV Type 1 DNA Pool During Effective HIV Therapy. *J Infect Dis* *212*, 596-607.

Wong, J.K., Hezareh, M., Gunthard, H.F., Havlir, D.V., Ignacio, C.C., Spina, C.A., and Richman, D.D. (1997). Recovery of replication-competent HIV despite prolonged suppression of plasma viremia. *Science (New York, N.Y)* *278*, 1291-1295.

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.

Yukl, S.A., Kaiser, P., Kim, P., Telwatte, S., Joshi, S.K., Vu, M., Lampiris, H., and Wong, J.K. (2018). HIV latency in isolated patient CD4(+) T cells may be due to blocks in HIV transcriptional elongation, completion, and splicing. *Sci Transl Med* *10*.

**Figure 1**

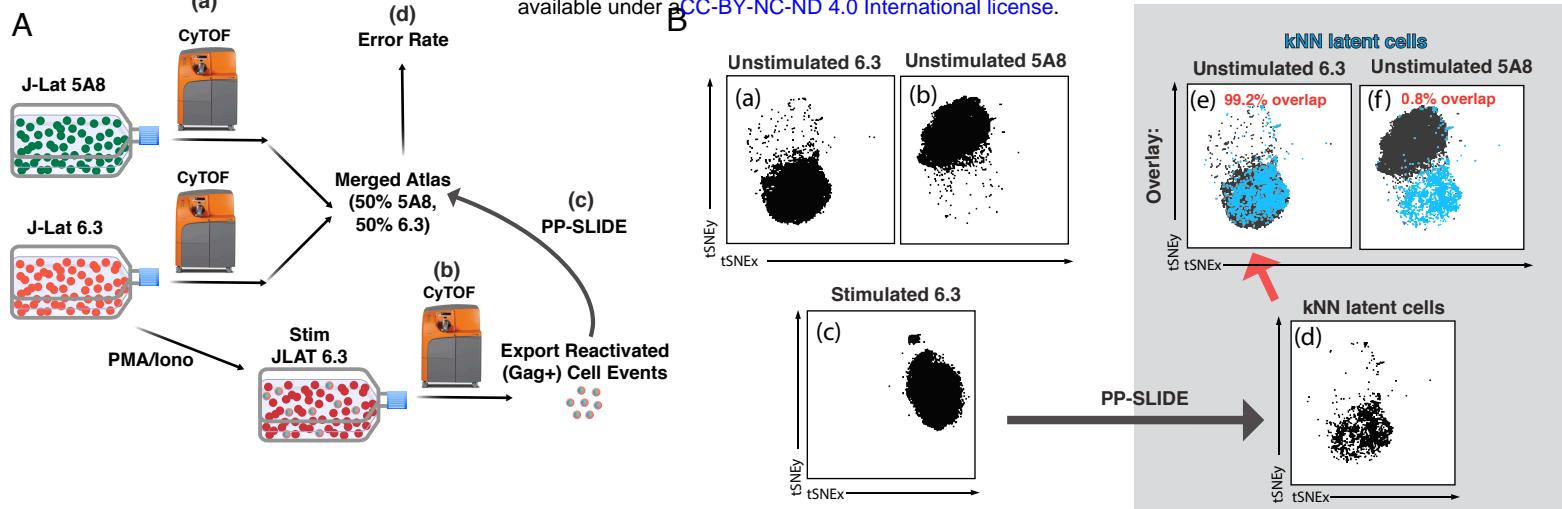


Figure 2

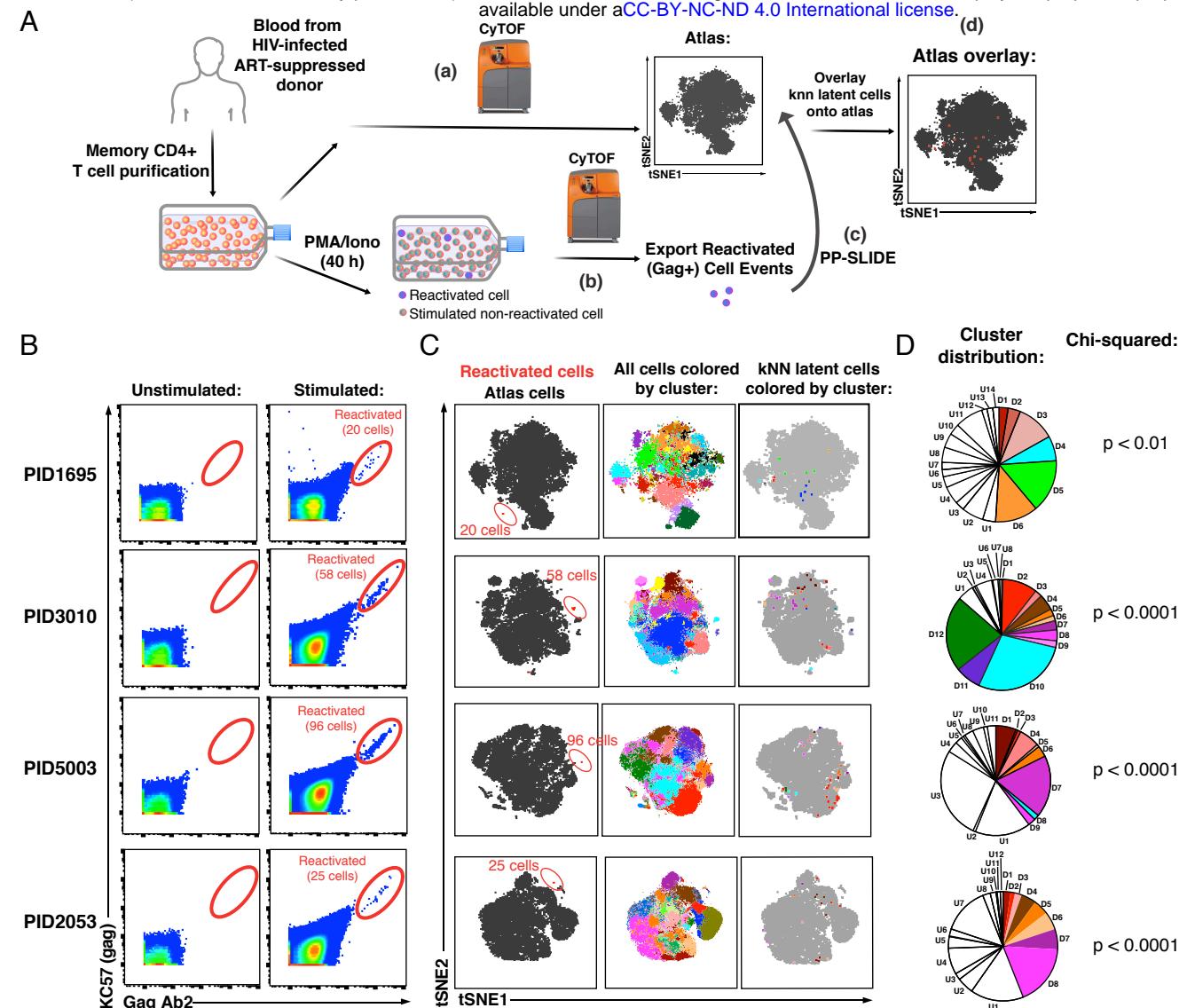
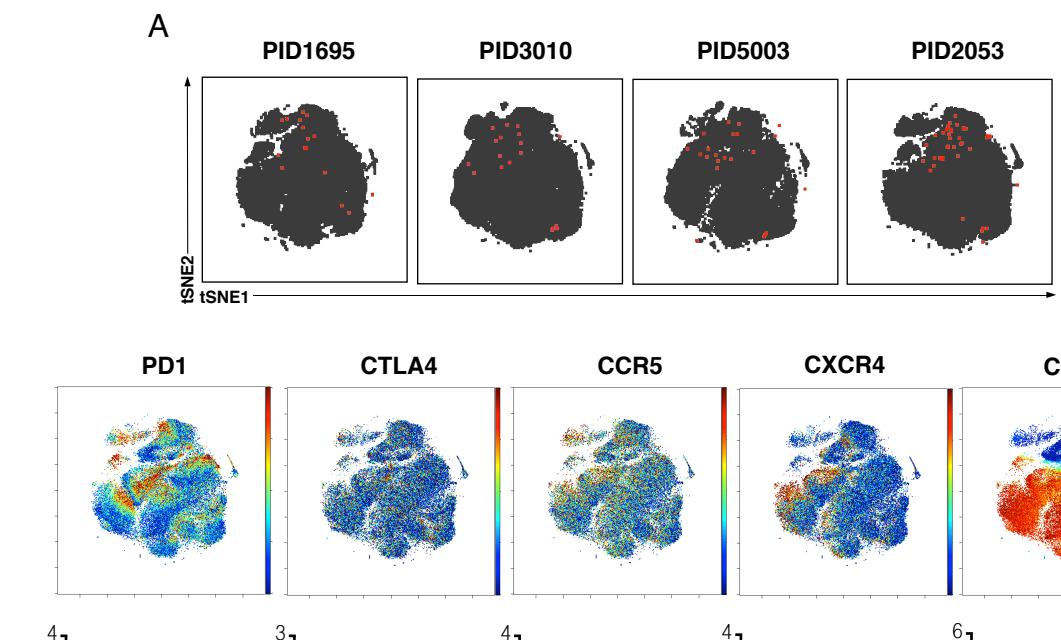
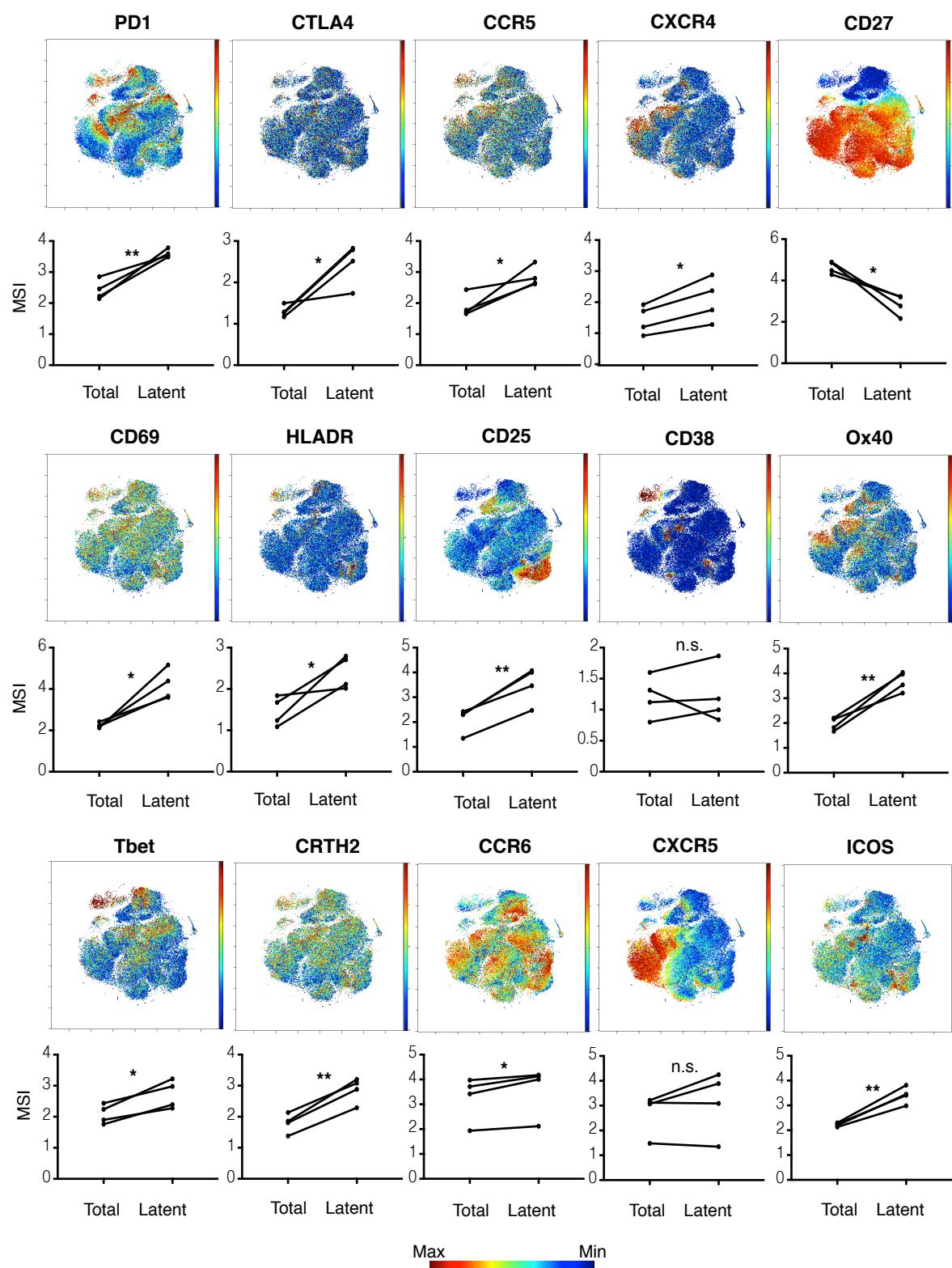
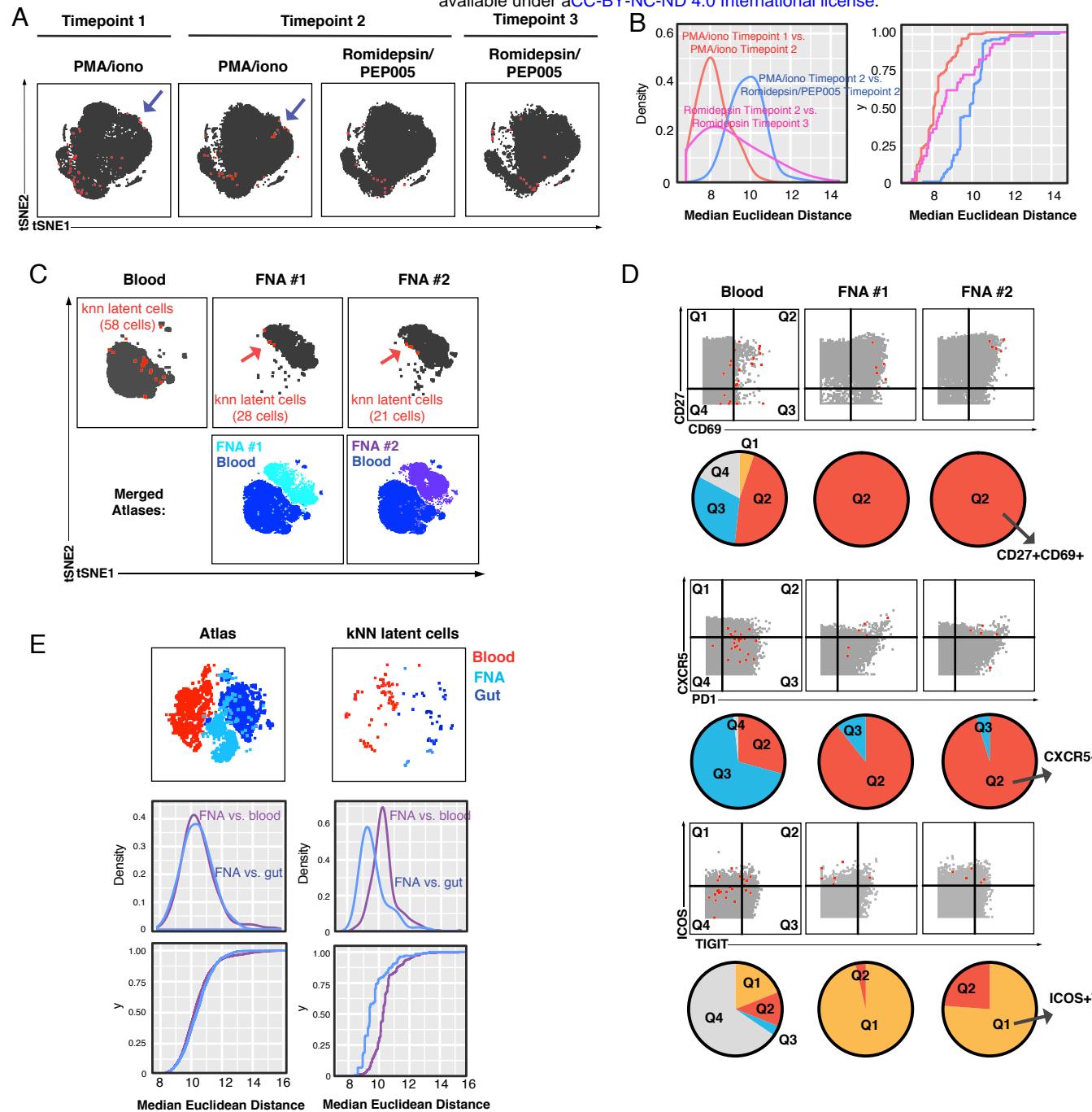


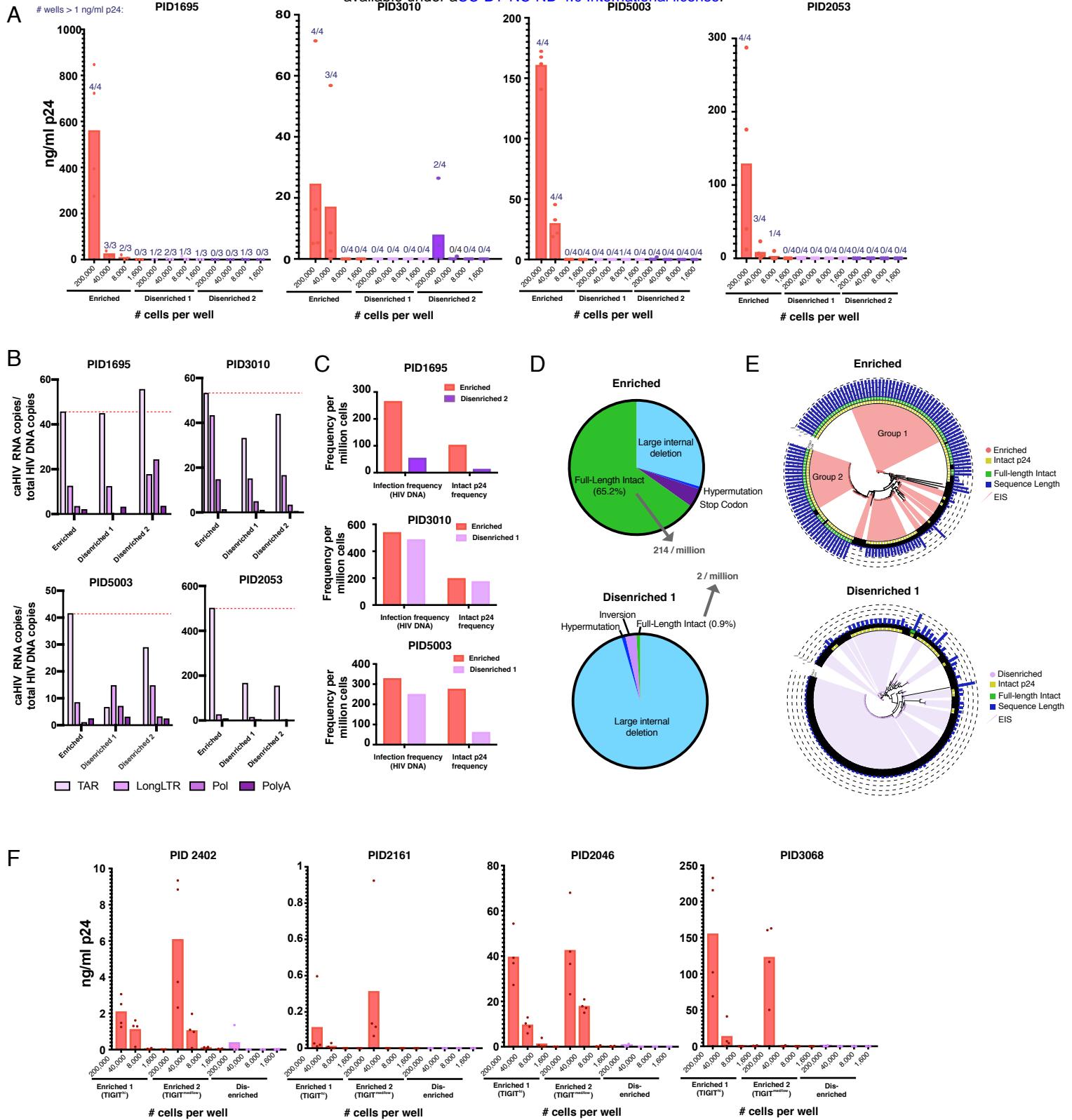
Figure 3

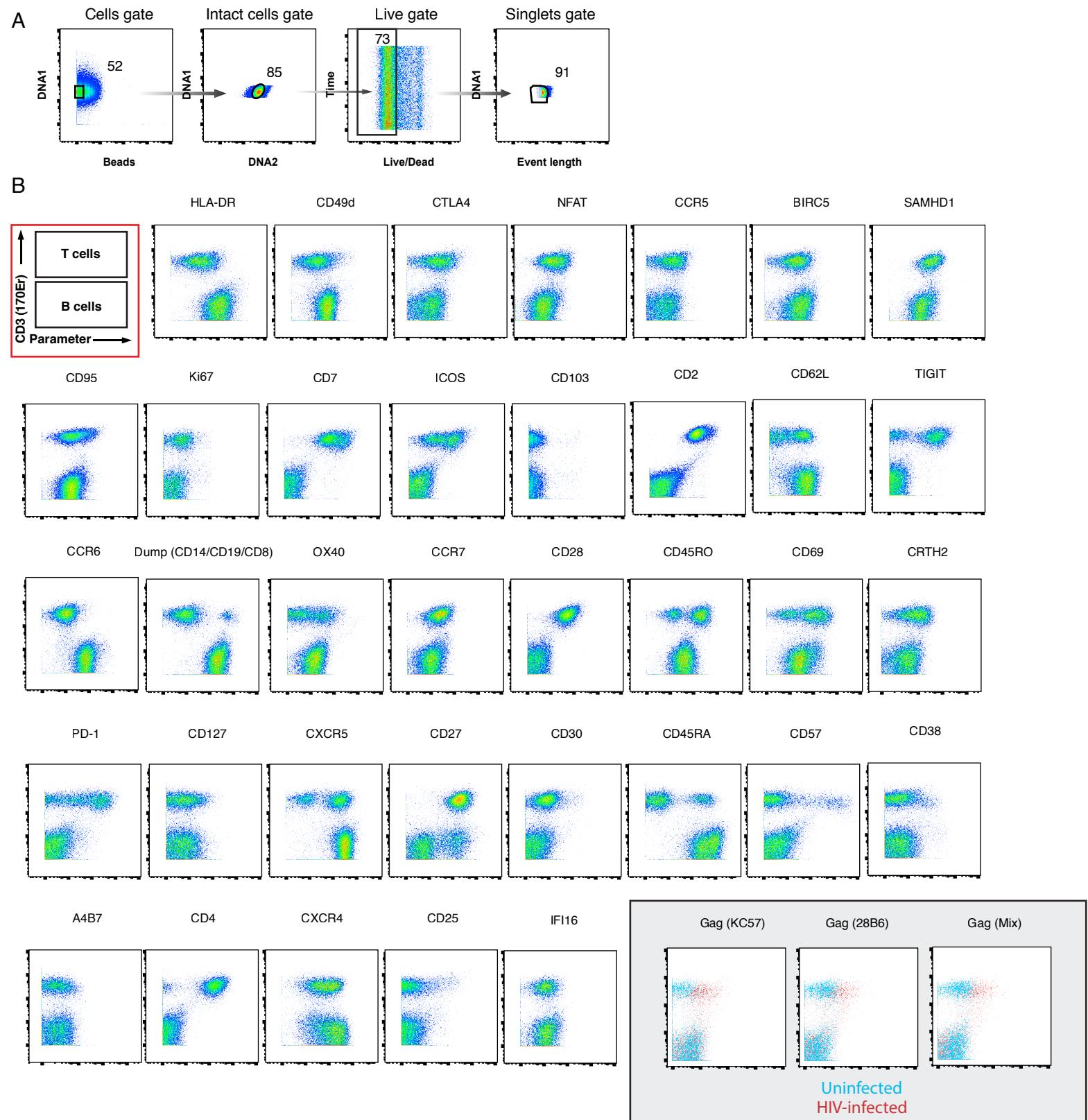


B



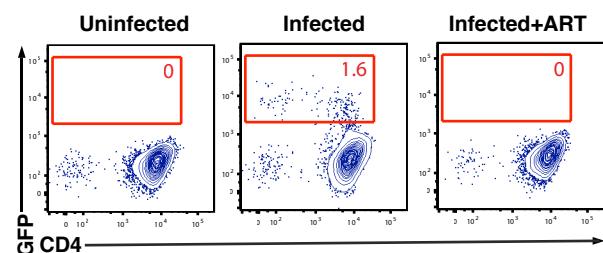




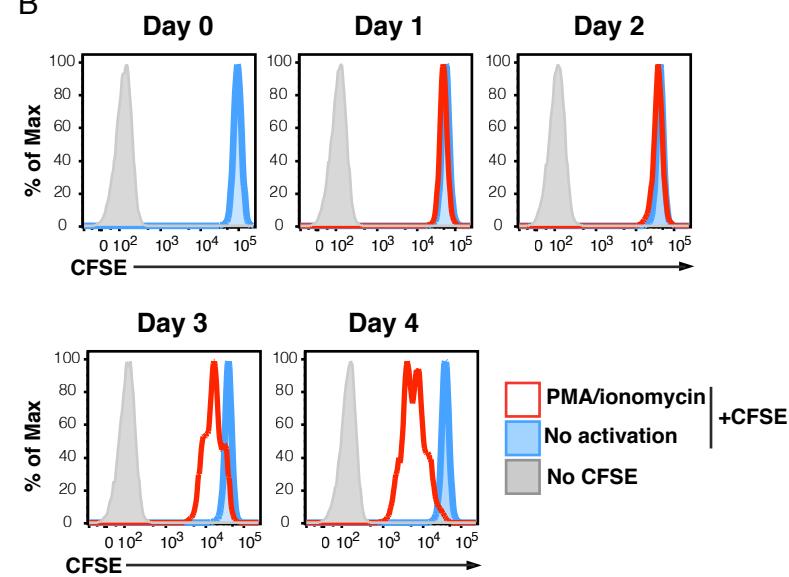


**C**

A

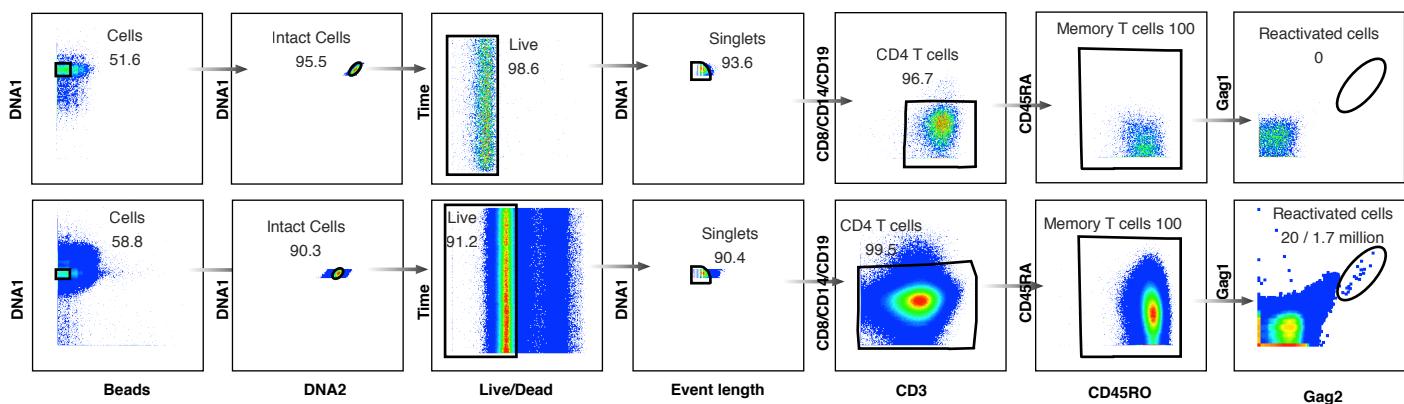


B



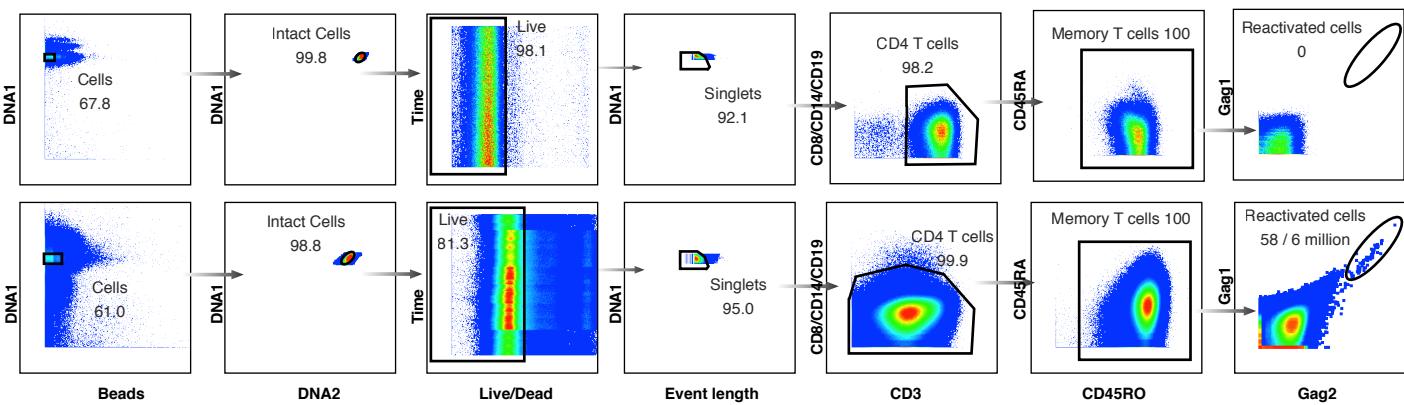
### PID1695

Unstimulated



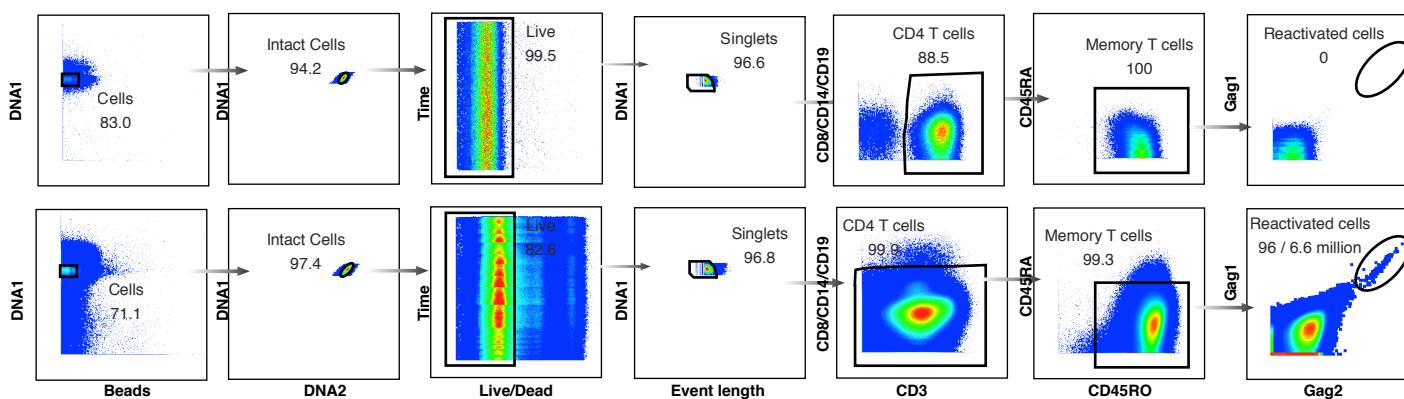
### PID3010

Unstimulated



### PID5003

Unstimulated



### PID2053

Unstimulated

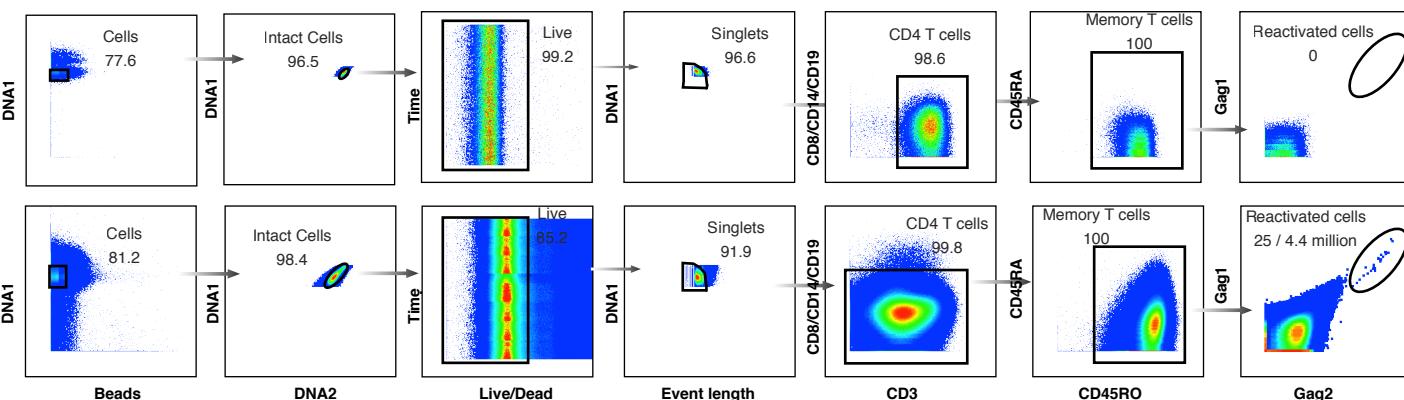
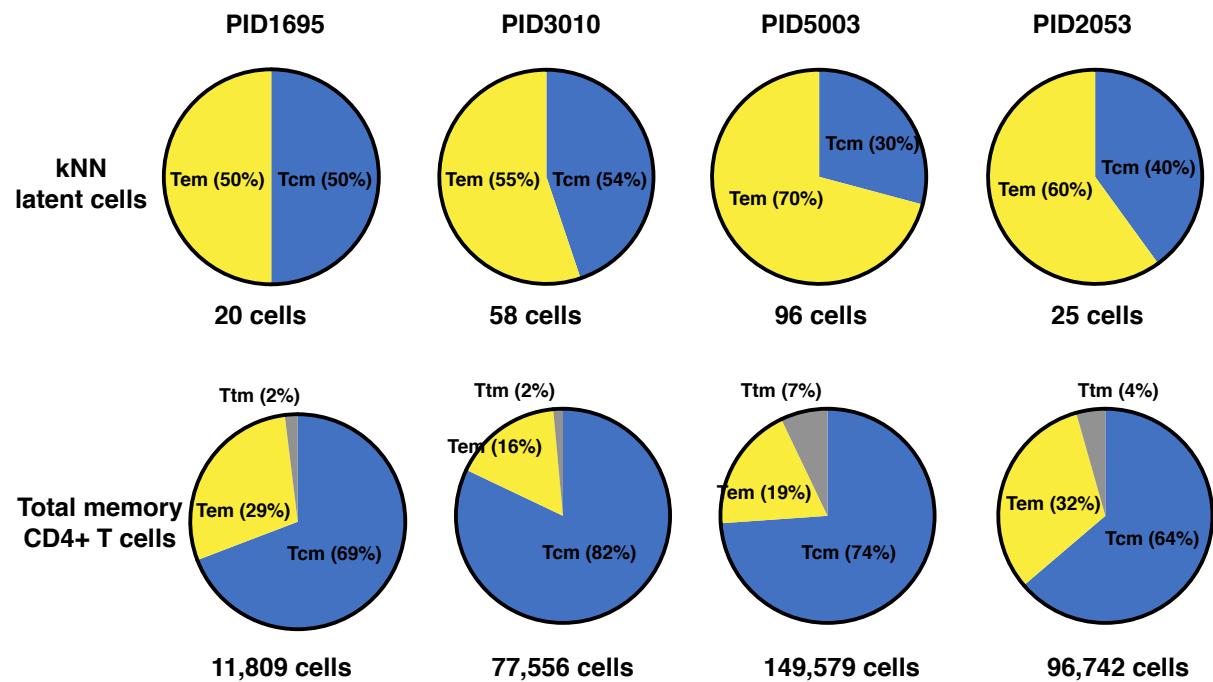
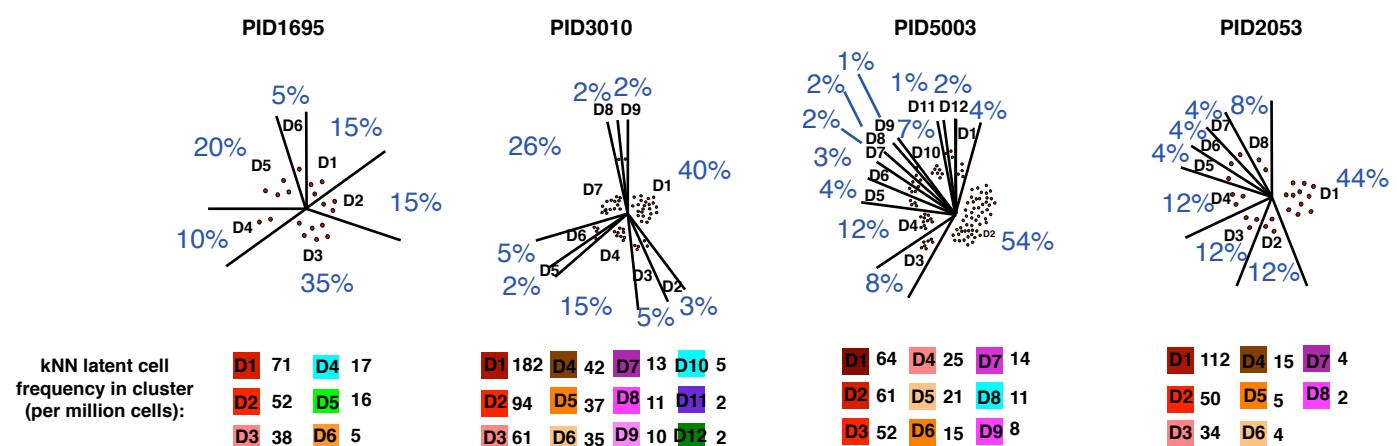


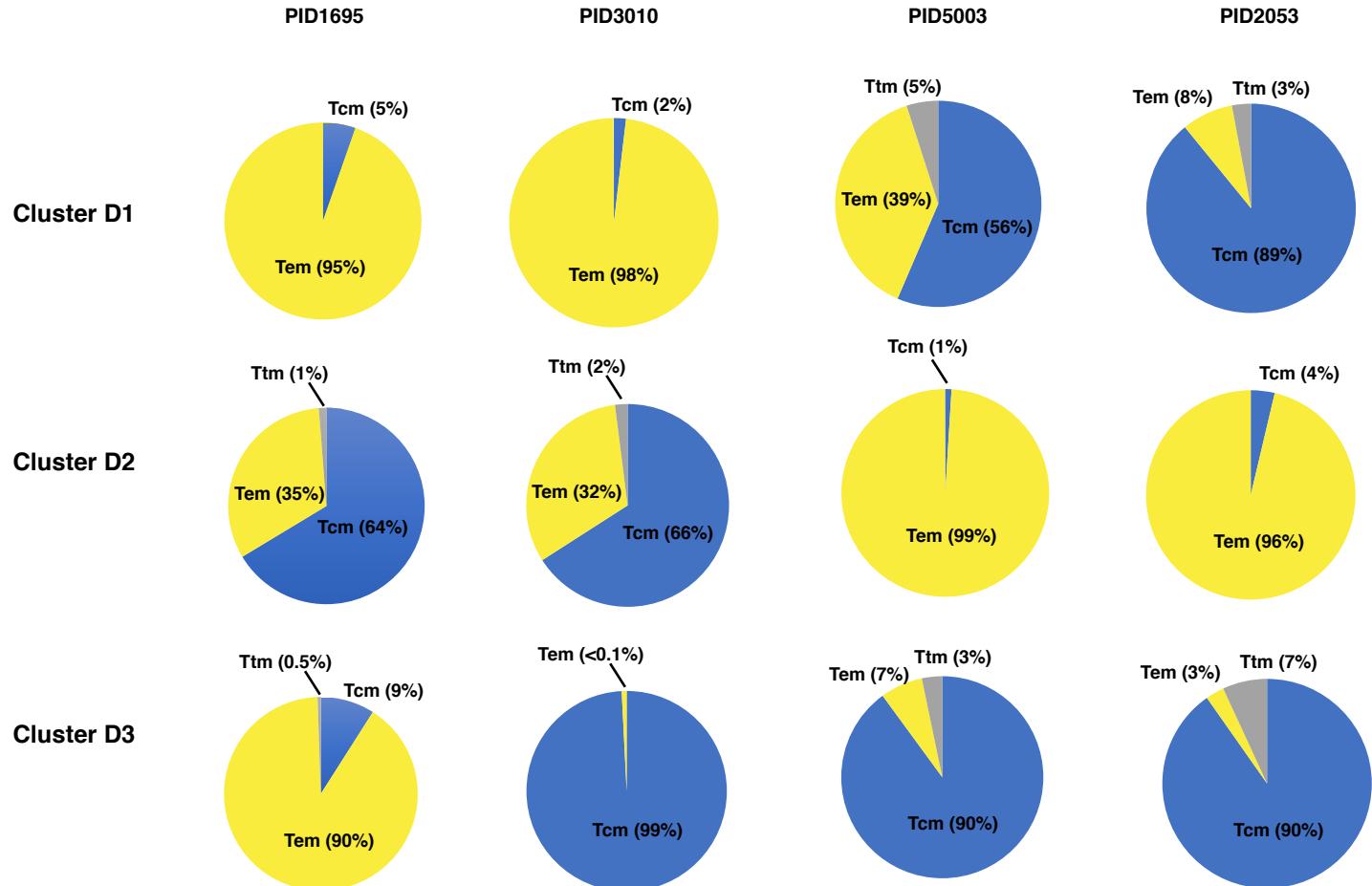
Figure S4



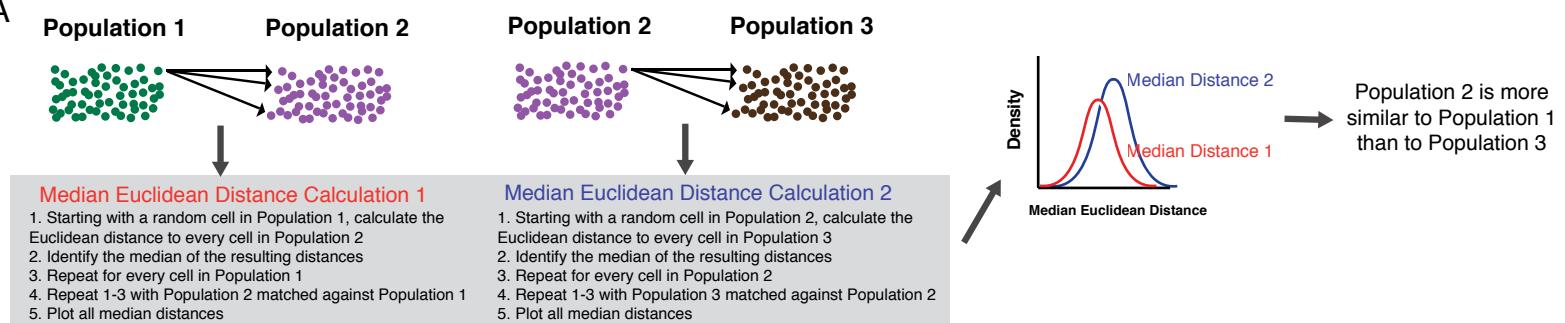
A



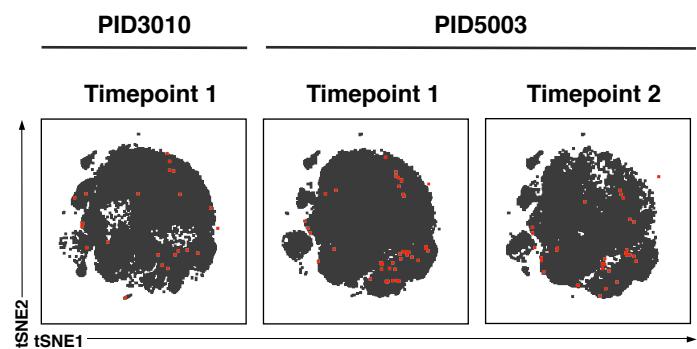
B



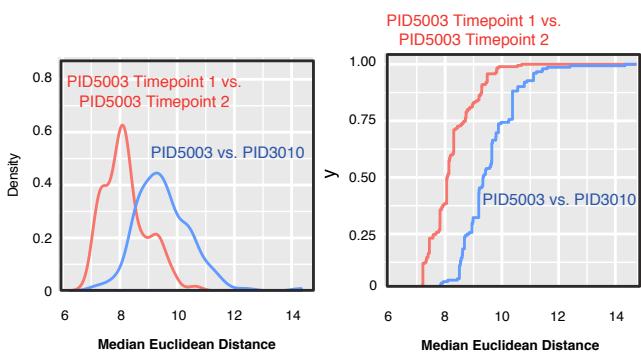
A



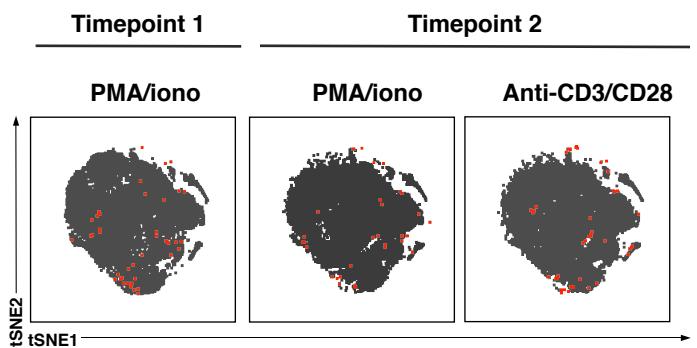
B



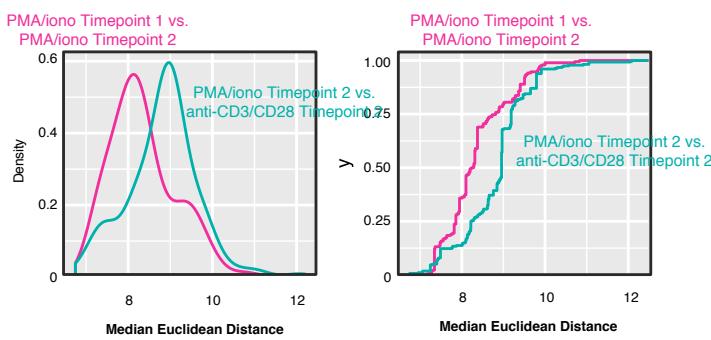
C



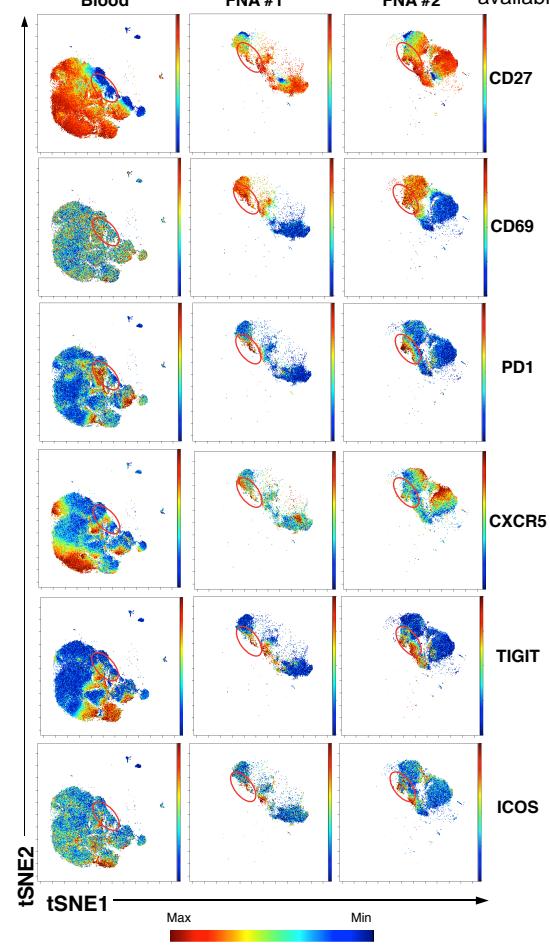
D



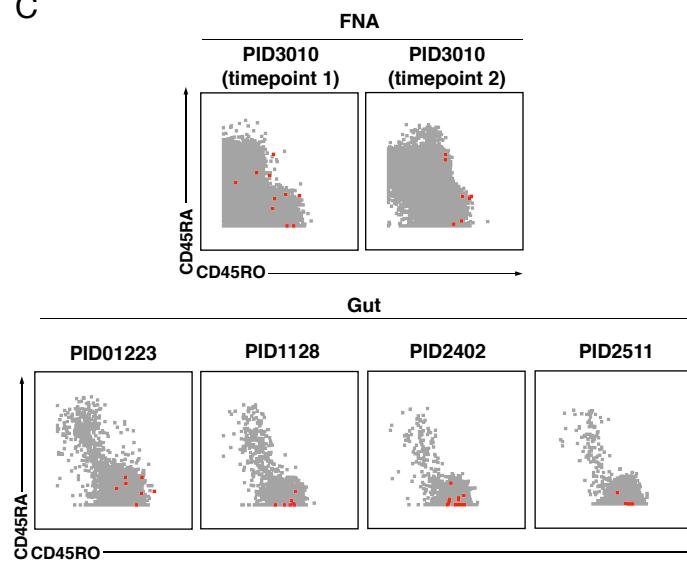
E



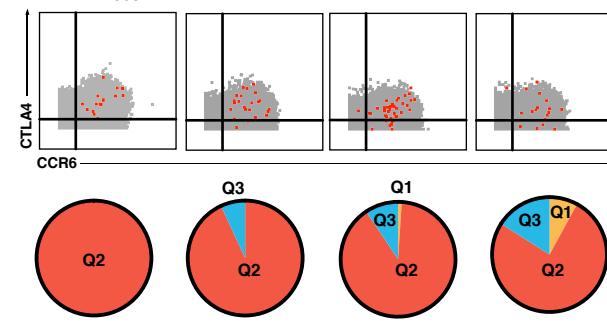
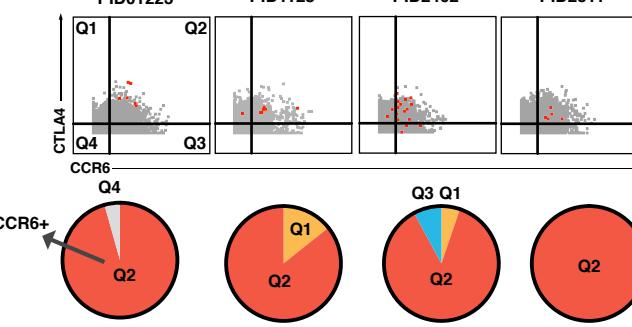
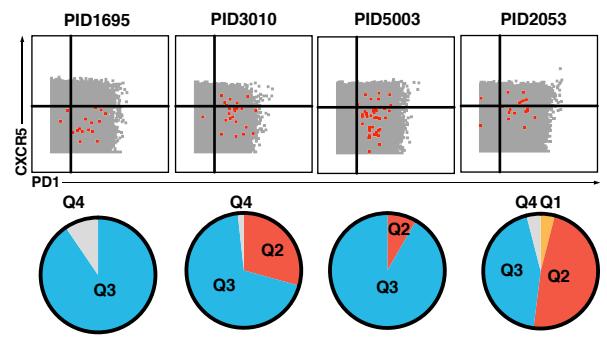
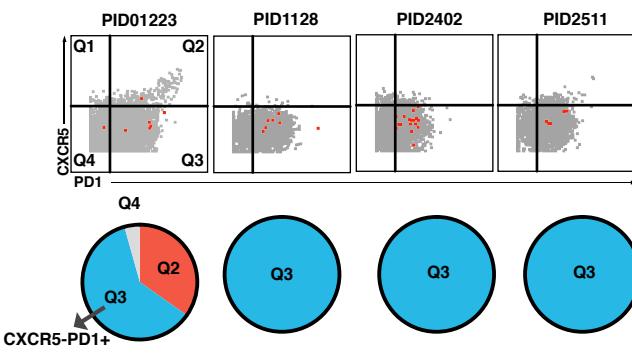
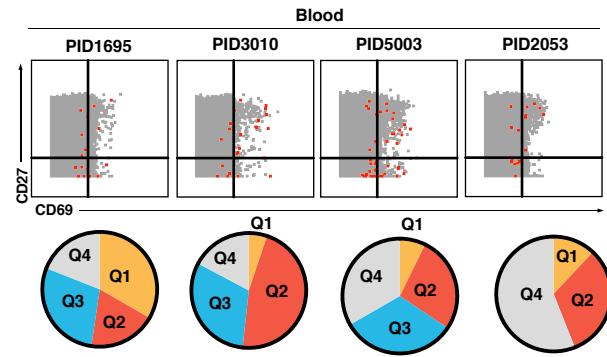
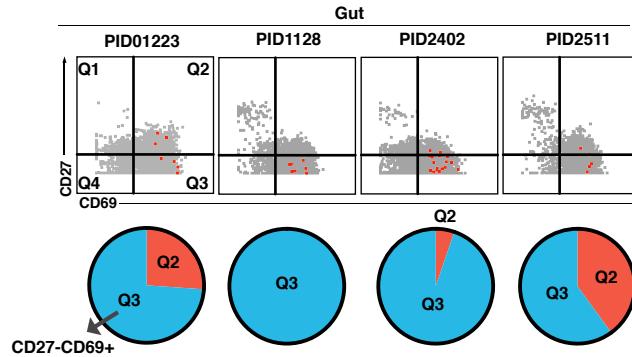
A

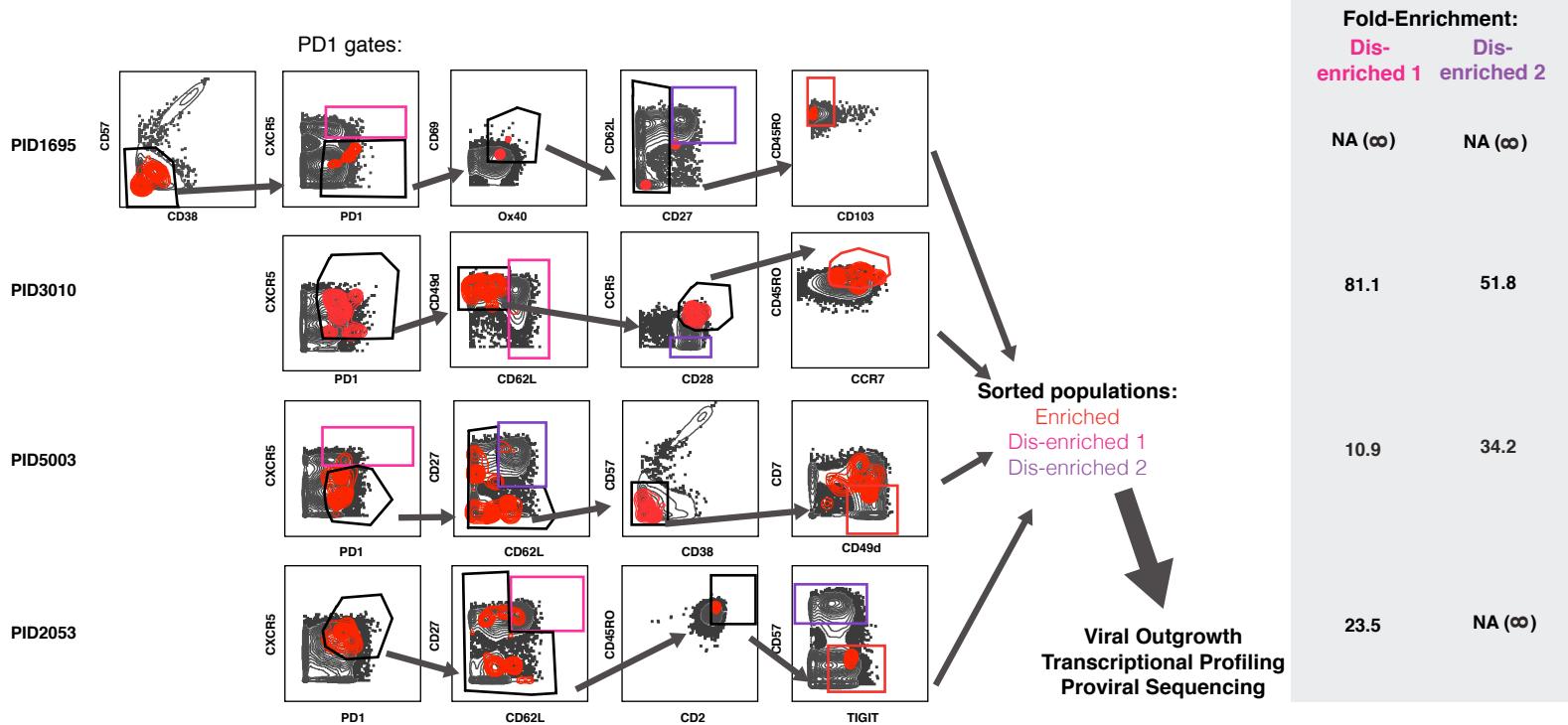


C

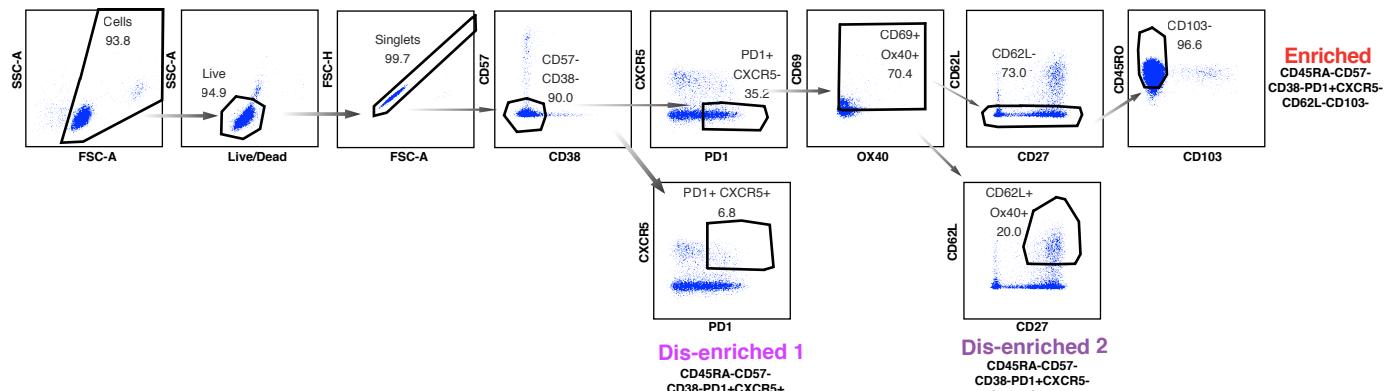


B





PID1695



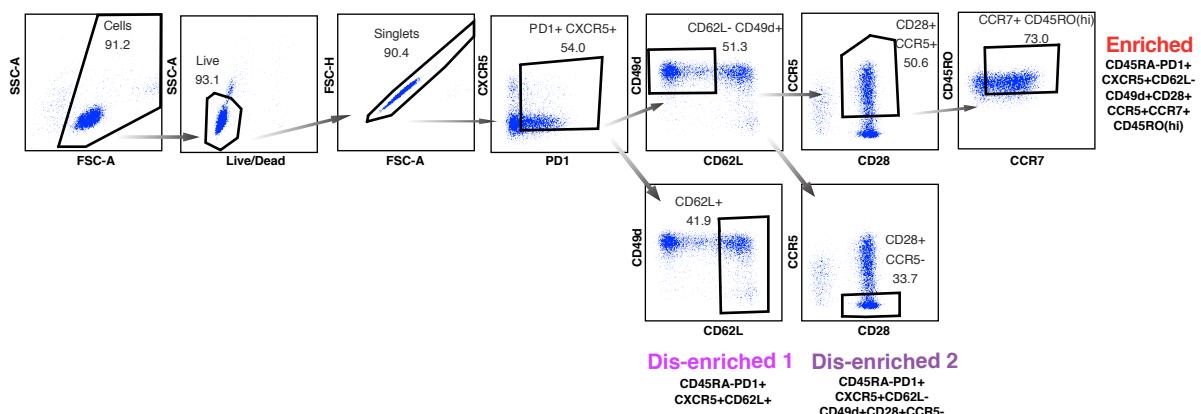
Dis-enriched 1

CD45RA-CD57-  
CD38-PD1+CXCR5+

Dis-enriched 2

CD45RA-CD57-  
CD38-PD1+CXCR5-  
CD27+CD62L+

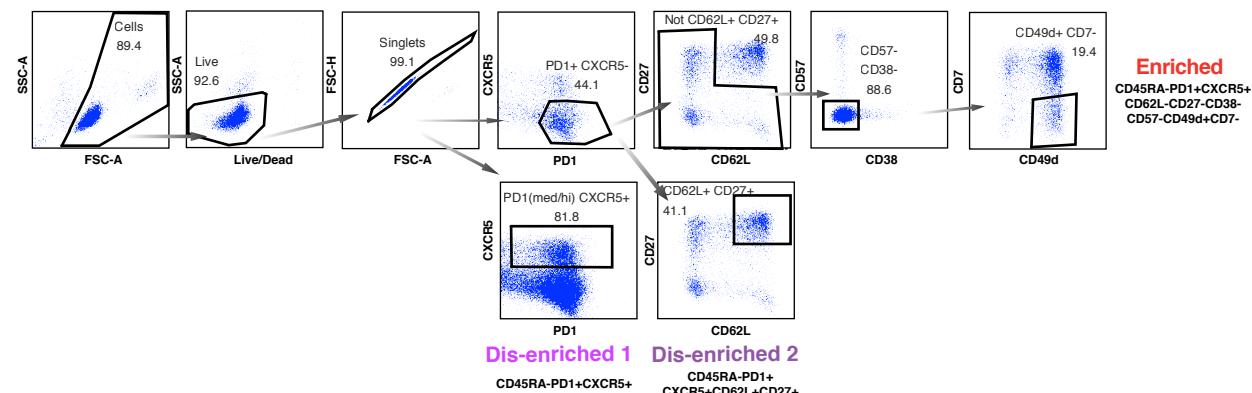
PID3010



Dis-enriched 1 Dis-enriched 2

CD45RA-PD1+  
CXCR5+CD62L-  
CD49d+CD28+CCR5-

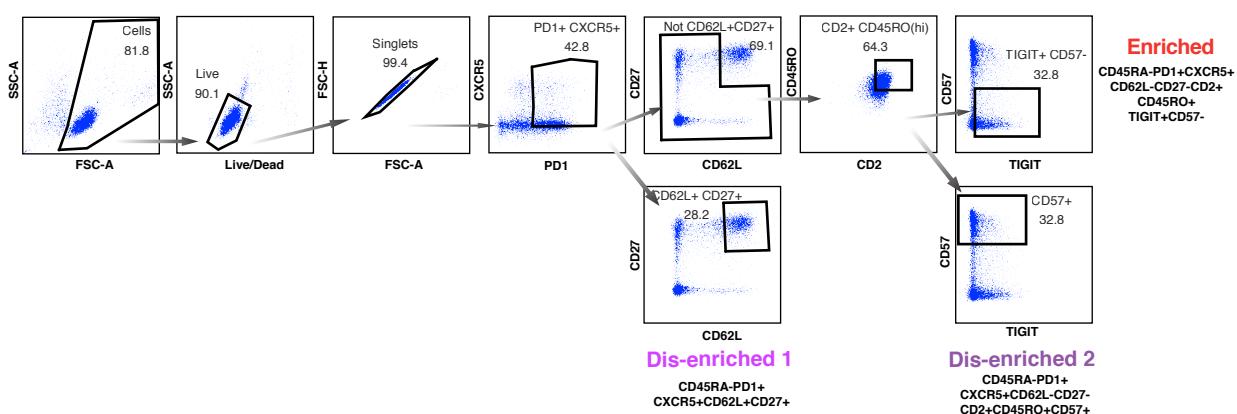
PID5003



Dis-enriched 1 Dis-enriched 2

CD45RA-PD1+CXCR5+  
CD45RA-PD1+CXCR5+CD62L+CD27+

PID2053

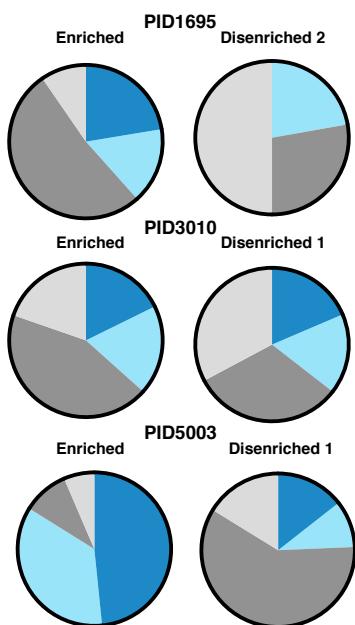


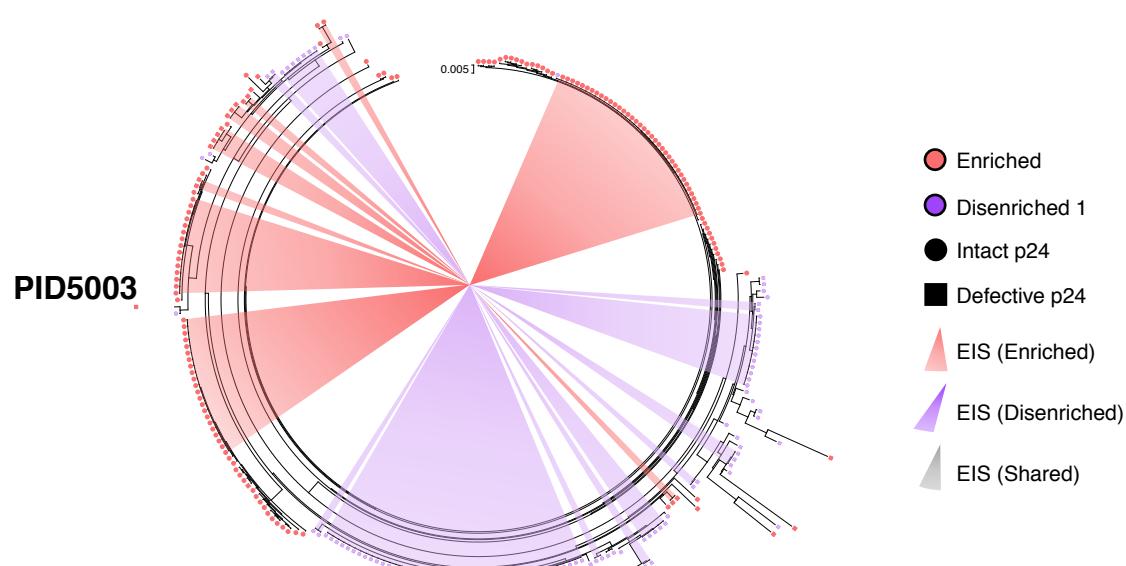
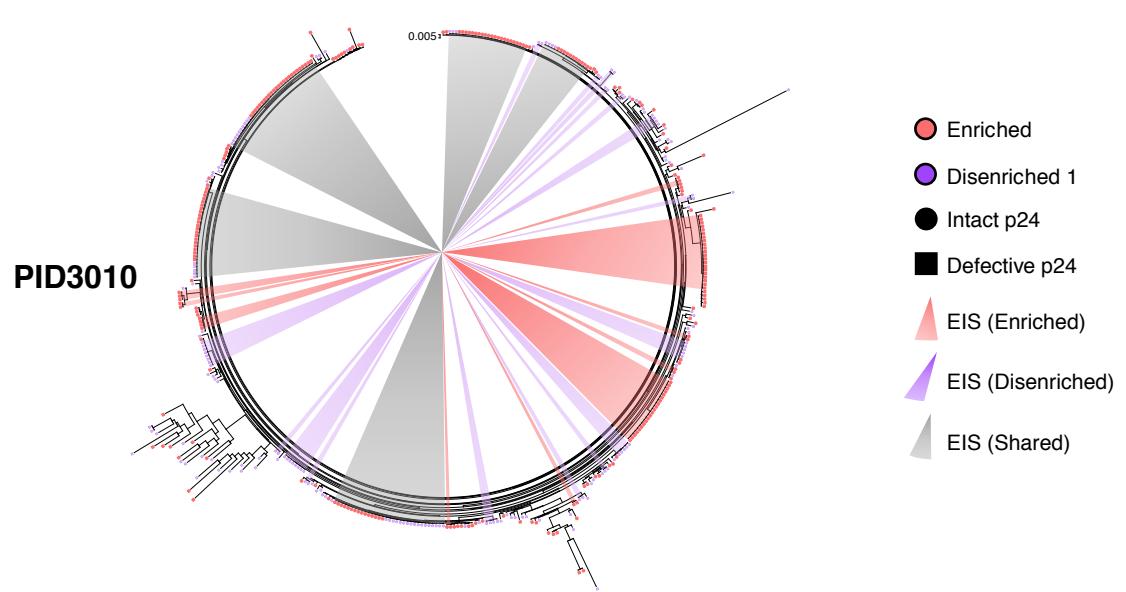
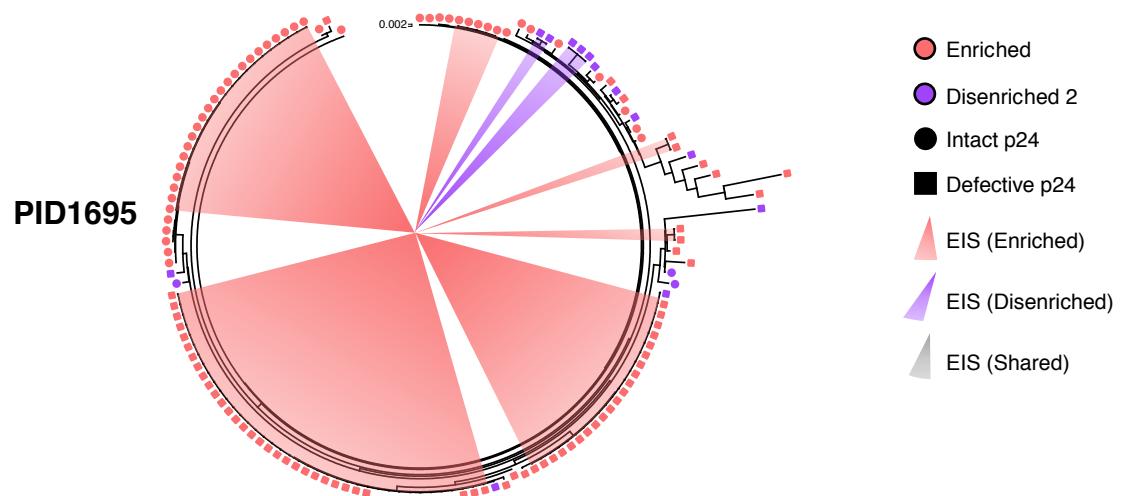
Dis-enriched 1

CD45RA-PD1+  
CXCR5+CD62L+CD27+

Dis-enriched 2

CD45RA-PD1+  
CXCR5+CD62L-CD27-  
CD2+CD45RO+CD57+





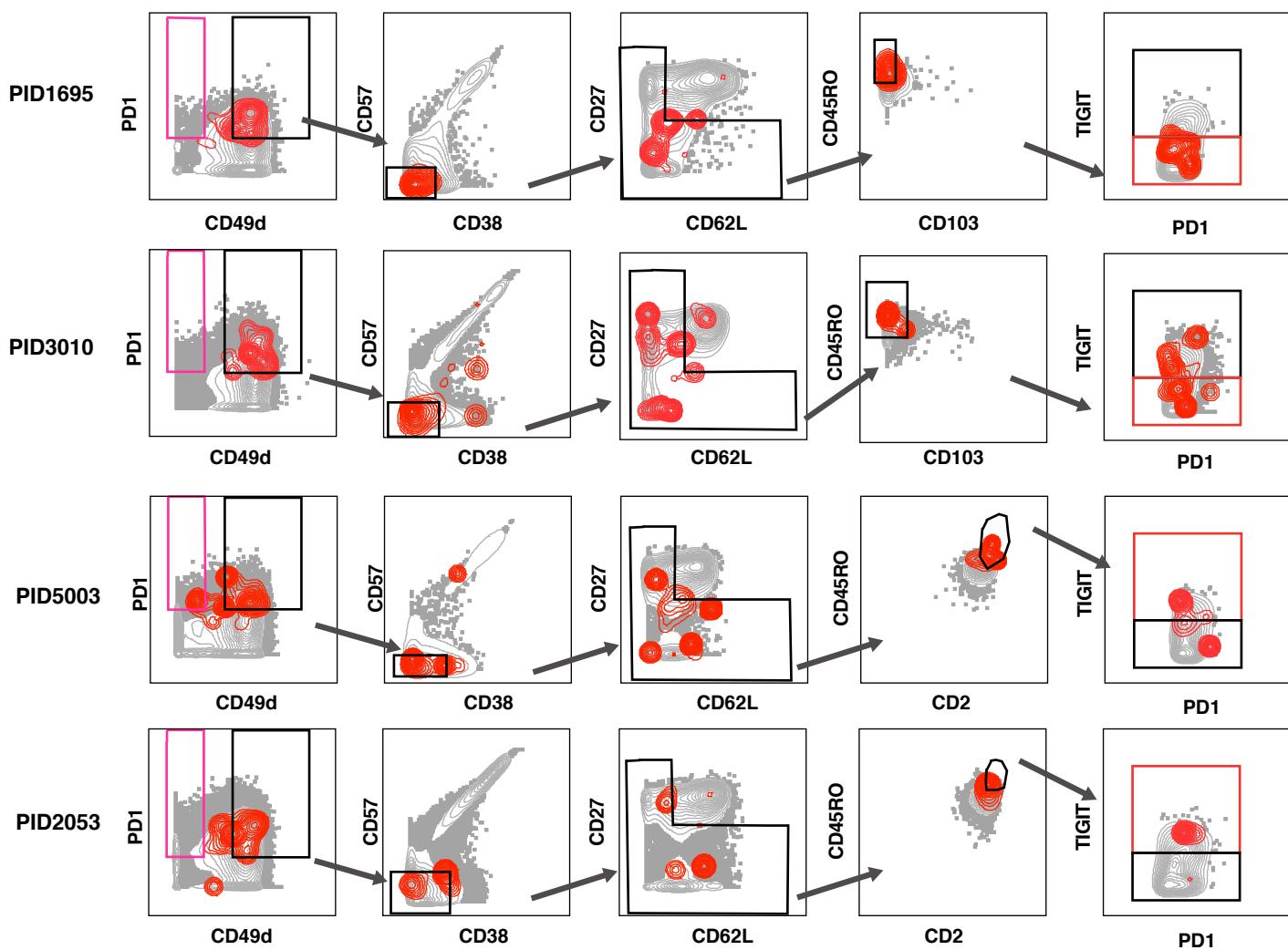


Figure S13

