Extensive heterogeneity of the HIV-1 infected CD4+ T-cell reservoir revealed by single-cell viral ASAPseq

Vincent H. Wu[1,3], Son Nguyen[1,#], Jayme M.L. Nordin[1,3], Jaimy Joy[,2,3], Felicity Mampe[2,3], Perla M. del Rio Estrada[4], Fernanda Torres-Ruiz[4], Mauricio González-Navarro[4], Yara Andrea Luna-Villalobos[4], Santiago Ávila-Ríos[4], Gustavo Reyes-Terán[5], Pablo Tebas[2,3], Luis J. Montaner[3,6], Katharine J. Bar[2,3], Laura A. Vella[3,7*], Michael R. Betts[1,3*]

Affiliations

[1] Department of Microbiology, Perelman School of Medicine, University of Pennsylvania (Philadelphia, PA, USA)
[2] Department of Medicine, Perelman School of Medicine, University of Pennsylvania (Philadelphia, PA, USA)
[3] Center for AIDS Research, University of Pennsylvania (Philadelphia, PA, USA)
[4] Centro de Investigación en Enfermedades Infecciosas, Instituto Nacional de Enfermedades Respiratorias (Mexico City, Mexico)
[5] Institutos Nacionales de Salud y Hospitales de Alta Especialidad, Secretaría de Salud de México (Mexico City, Mexico)
[6] The Wistar Institute (Philadelphia, PA, USA)
[7] Division of Infectious Diseases, Children's Hospital of Philadelphia (Philadelphia, PA, USA)
[#] Present address: Institute for Medical Engineering & Science, Department of Chemistry, and Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology (Cambridge, MA, USA)
[*] Co-corresponding authors

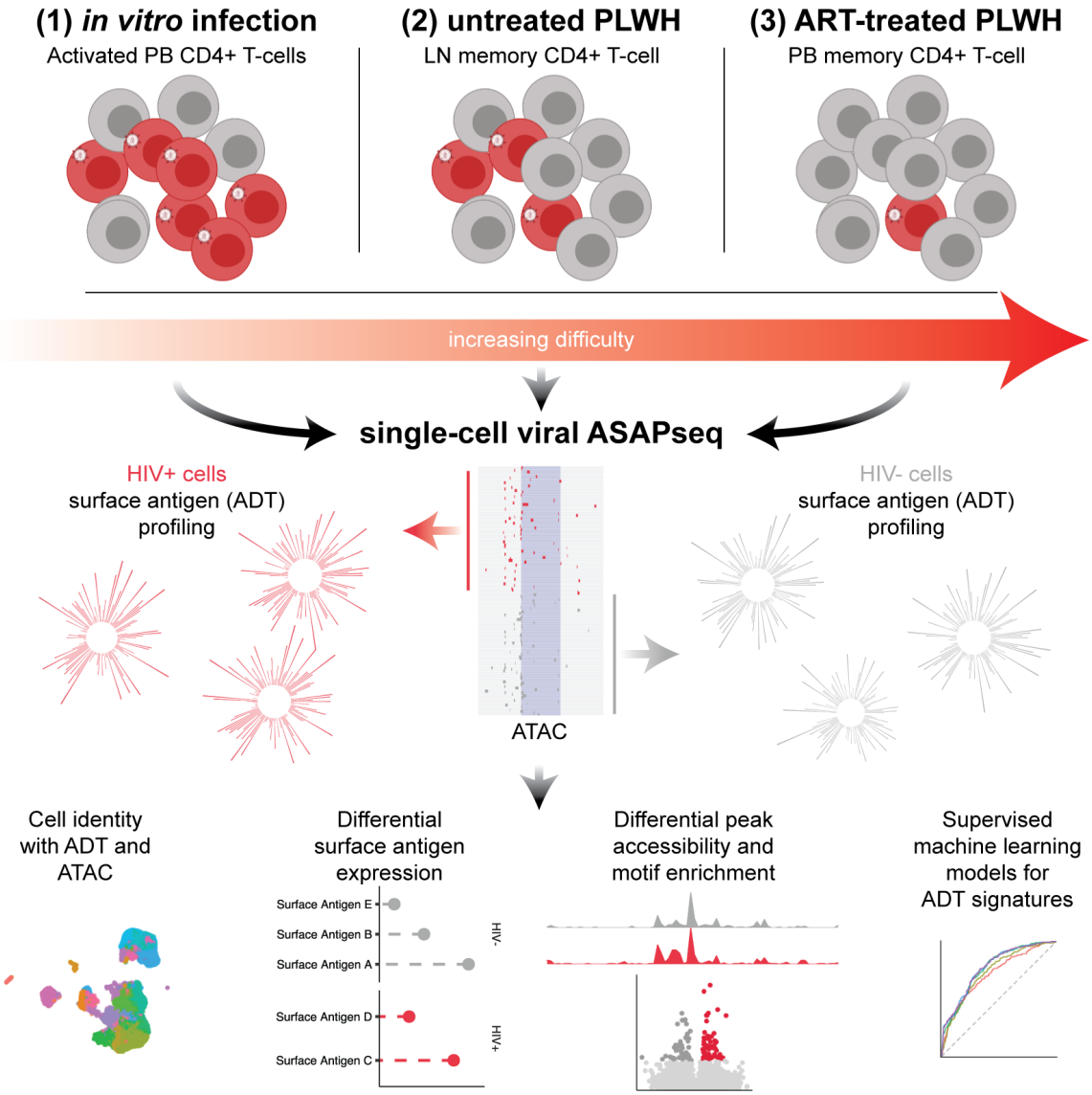
[*]Correspondence to:

Dr. Michael R. Betts, PhD
Department of Microbiology
402C Johnson Pavilion
University of Pennsylvania
betts@pennmedicine.upenn.edu | (215) 573-2773

and

Dr. Laura A. Vella, MD PhD
Division of Infectious Diseases
Room 10008, 3501 Civic Center Blvd
Children's Hospital of Philadelphia
vellal@chop.edu | (267) 250-9574

Graphical Abstract



Abstract
Defining the long-lived HIV reservoir during antiretroviral therapy remains a challenge for HIV cure research. To determine if unifying cell surface or genomic features can be identified in unperturbed HIV-1 infected CD4+ T-cells, we developed single-cell viral ASAPseq, which leverages scATACseq and surface marker profiling to define HIV-infected cells via the presence of integrated proviral DNA. Here, we employed this strategy to define infected CD4+ T-cells both *in vitro* and *ex vivo* from blood and lymph nodes of, respectively, treated and untreated people living with HIV. In each setting, we identified HIV-infected cells with both novel and known surface markers associated with provirus. We also uncovered epigenetic profiles and transcription factor motifs enriched in HIV-infected cells that more deeply define the reservoir. Our data highlight the broad heterogeneity of the HIV reservoir and create a foundation for a multiomic atlas to develop targeted elimination strategies.

Introduction

The establishment and persistence of the HIV reservoir remains one of the greatest barriers preventing a functional cure in people living with HIV (PLWH). This reservoir, primarily composed of infected long-lived memory CD4+ T cells (Chun et al., 1995, 1997; Finzi et al., 1997), resides in blood and tissue compartments throughout the body and almost invariably reactivates upon antiretroviral therapy (ART) interruption (Deleage et al., 2016; Estes et al., 2017; Pantaleo et al., 1993; Wong et al., 1997). Reservoir persistence results from the stable integration of viral DNA into the genome of a host cell, enabling infection for the lifetime of the cell and its progeny. Substantial strides have been made towards understanding the virological aspects of the HIV reservoir, including intactness, reactivation potential, and integration site landscapes (De Scheerder et al., 2019; Ho et al., 2013). However, beyond population level characterization and quantification, the cellular identity of the HIV reservoir *in vivo* remains an enigma. Defining a surface marker signature of the cellular HIV reservoir for targeting is of central importance to the HIV cure agenda.

The major impediment towards our understanding of the cellular reservoir is the challenge of identifying infected cells. On average during ART, fewer than 1/1000 CD4+ T cells are infected with HIV in the blood (Eriksson et al., 2013), the majority of which have little to no viral RNA production (Einkauf et al., 2022; McManus et al., 2019). To overcome this, the field to date has used *in vitro* models (Bradley et al., 2018; Golumbeanu et al., 2018), exogenous stimulation strategies (Cohn et al., 2018; Liu et al., 2020), and marker-targeted analysis (many studies including (Fromentin et al., 2016; Neidleman et al., 2020; Pardons et al., 2019)) to identify surface protein and/or transcriptomic properties of the HIV reservoir. These studies have collectively made inroads into the identity of infected cells but have been inherently limited by the experimental systems employed. These limitations, including 1) the heterogeneity of infected CD4+ T cells within bulk populations, 2) incomplete representation of the reservoir by sorting for viral RNA+ and/or protein-expressing cells, and 3) activation-induced malleability of both the cell surface proteome and underlying transcriptome, have obfuscated the precise characterization of infected cells and highlight the need to develop novel strategies to characterize the cellular HIV reservoir at steady state and ultimately following cure-directed strategies.

Here, we used the presence of integrated provirus as a molecular tag to directly identify individual infected cells using single-cell Assay for Transposase Accessible Chromatin (scATACseq) (Buenrostro et al., 2013) and employing a human-viral alignment strategy first developed in a chimeric antigen receptor (CAR) T-cell model (Wang et al., 2020). Further, scATACseq provides unbiased, single-cell epigenetic profiles and can therefore assign cellular identity to cells with and without detectable provirus. An additional goal of reservoir research is to identify cell surface proteins that identify or substantially enrich for infected cells. We therefore combined scATACseq with cell surface protein identity using the recently described ATAC with Select Antigen Profiling by sequencing (ASAPseq) (Mimitou et al., 2021; Swanson et al., 2021). We employed ASAPseq to study both *in vitro* and direct *ex vivo* HIV infection, including the most clinically relevant setting with peripheral blood memory CD4+ T cells from ART-treated PLWH. Our findings provide the first direct and unbiased identification of

experimentally and *in vivo* infected memory CD4+ T cells at single-cell resolution with cellular identity. With our multidimensional dataset, we highlight the heterogeneity of HIV+ cells and the variation of surface and epigenetic markers between and across infection contexts. Together, our extensive characterization of HIV infection using ASAPseq results in a multimodal single-cell dataset of cell surface protein composition, epigenetic landscape, and potential regulatory mechanisms to accelerate our understanding of the HIV reservoir for targeted therapy.

Results

ASAPseq identification of HIV-infected cells *in vitro*

To determine the potential for ASAPseq to detect HIV-infected cells, we first tested an *in vitro* primary CD4+ T-cell infection model (**Table 1**). CD4+ T cells from an HIV-uninfected individual were activated with anti-CD3, anti CD28/CD49d, and IL-2 for two days, infected with HIV-1 (molecular clone; SUMA), and rested for four subsequent days. We next performed ASAP-Seq, obtaining 7094 cells that passed quality checks for both ATAC and the antibody-derived tag (ADT) components (**Supplemental Table 1** and **Supplemental Figure 1A**). Following alignment of sequencing data to both human and viral genomes, 1323 cells (18.6%) contained an integrated HIV genome in open chromatin (HIV+, **Figure 1A-B)**. As expected, a smaller proportion of the *in vitro* infected cells expressed p24 protein by flow cytometry (5.1%, **Supplemental Figure 1B**), consistent with known per-cell discrepancies between HIV integration and p24 production (Shan et al., 2017). To confirm the specificity of provirus detection, we aligned an unrelated ASAPseq dataset of uninfected peripheral blood mononuclear cells (Mimitou et al., 2021) using our chimeric human-virus genomes (human (hg38) + SUMA HIV and hg38 + HXB2 HIV). No cells contained reads aligning to either HIV genome, demonstrating the high specificity for ASAPseq-based detection of HIV+ cells (**Supplemental Figure 1C**). Within the HIV+ CD4+ T cells, proviral reads spanned the viral genome but were most prevalent in the long-terminal repeats (LTR) as well as *gag* and *env* genes (**Supplemental Figure 1D**). Fewer reads spanned the *pol* gene consistent with previous findings (Jefferys et al., 2021). Twenty-eight percent of HIV fragments were in LTR regions, 0.01% spanned both LTR and internal (i.e. not LTR), and the remaining fragments (71%) were internal.

We next examined the phenotypic identity of HIV+ versus HIV- cells. We clustered the cells using the epigenetic data component and annotated based on imputed gene scores from chromatin accessibility and the ADT component (**Figure 1A**, **Supplemental Figures 1E, Supplemental Figures 2, 3,** and **Supplemental Table 2**). Given that PBMCs were enriched for CD4+ T-cells prior to infection, most clusters consisted of CD4+ T-cells with some contaminating B and antigen presenting (APC) cells. Clusters were separated by activated versus naive/early memory phenotypes, with the greatest number of HIV+ cells (78.5% of all HIV+ cells, **Figures 1B-D**) found within activated and/or effector memory CD4+ T cell clusters. We also observed some HIV+ cells in other CD4+ T-cell phenotypes (18.4%) and a small proportion in monocytes (1.9%). These findings reflect the known phenomenon of preferential *in vitro* HIV infection of activated CD4+ T-cells and demonstrate that this feature can be resolved on a single-cell level by ASAPseq.

In a comparison between all HIV+ and uninfected T-cells, we observed significantly higher expression of many surface markers including those associated with T-cell activation on HIV+ cells (**Supplemental Figure 4A** and **Supplemental Table 3**). The top 5 upregulated markers by π-score (Xiao et al., 2014) on HIV+ cells included CCR5, SLAM, PD-1, CD49a, and CD161 while the top 5 upregulated markers on HIV- cells were CD31, CD62L, CD55, CD27, and CD7 (**Supplemental Figure 4A**). The significant upregulation of CCR5 in HIV+ cells is consistent

with the usage of CCR5 as the co-receptor for HIV-1 infection (Alkhatib et al., 1996). The other differentially expressed surface markers in HIV+ cells reflect the known preferential infection of activated CD4+ T-cells (Zack et al., 1990; Zhang et al., 1999). We therefore asked whether any surface protein distinguished between infected and uninfected cells in the activated clusters (**Figure 1E**, **Supplemental Figure 5**, and **Supplemental Table 3**). In activated HIV+ cells, we found that CCR5, SLAM, CD2, GPR56, PD-1 were the top upregulated markers, while CD62L, CD127, CXCR5, CD49f, and CCR4 were increased in activated HIV- CD4+ cells. This finding suggests that even within the activated/effector cells, the HIV+ cells have higher expression of CCR5 and other activation markers. In the early differentiated HIV+ T-cells, we found several significantly increased expression of ICOS, CD11a, CXCR3, CD25, and CD49b while the early differentiated HIV- cells were enriched in CD62L, CD55, CD127, CD162, and CD45RA (**Supplemental Figure 4B** and **Supplemental Table 3**). These surface antigen marker findings highlight the heterogeneity of infected cells even in the activated cell population.

Differential chromatin accessibility in HIV-infected cells and surfacer marker based supervised machine learning *in vitro*

We next examined whether there was any differential genomic accessibility between activated HIV+ and activated HIV- cells. We observed heightened accessibility in several genomic regions in activated HIV+ cells (**Figures 2A-B** and **Supplemental Table 4**). Several nearest genes to these peaks have been previously implicated for HIV infection and replication including a significantly enriched peak ~1kb upstream of the *CCR5* transcription start site (**Figure 2C**), thus showing a concordance between CCR5 surface expression and gene accessibility. From these differentially accessible genomic regions, we assessed transcription factor motif enrichment. In the activated HIV- cells (**Figure 2D**), the top enriched motifs include several members of the TCF family including TCF7 (TCF-1) and LEF1, both regulators of CD4+ T-cell development and differentiation (reviewed in (Ma et al., 2012)). In the activated HIV+ cells (**Figure 2D**), the top motifs included proliferation and activation motifs in the AP-1 related family such as the JUN and FOS families (reviewed in (Foletta et al., 1998)). These findings, in combination with the surface marker profiles, are concordant with the known immunobiology of HIV-1 infection, thus showing that ASAPseq is a viable technique for ascertaining phenotypes of HIV+ cells.

Finally, as an initial step towards selectively recognizing and targeting HIV+ cells, we employed several supervised machine learning methods to determine whether any cell surface proteins could, in combination, predict infected cells. We assessed logistic regression, naive Bayes, and random forest methods using a 70/30 (training/test) split of our dataset. Using all CD4+ T-cells, we found similar results between the models with area under the receiver operating curve (ROC) (AUC) values of 0.88, 0.86, and 0.89 respectively (**Figure 2E**) suggesting a reasonable classification of HIV+ cells *in vitro* with this current ADT panel. In the logistic regression, markers with the greatest odds ratio and significance included CD2, CD49b, TCR α/β, SLAM, CD95, and CD54 for classifying HIV+ cells while CD4, NKG2D, CD62L, CD162 were the top markers for classifying HIV- cells (**Figure 2F**). For the random forest model, the top markers by importance (measured as mean decrease in accuracy when permuted) were CCR5, CD49b, CD2, SLAM, and CD95. When applied to only activated CD4+ T-cells, the random forest model performed the best with an AUC of 0.78, compared to 0.72 with logistic regression and 0.74 with

naive Bayes. Compared to the total CD4+ T-cell models, there were higher false positive rates suggesting that additional markers tailored to activated T-cells might benefit classification efforts (**Figure 2G**). The activated CD4+ T-cell random forest was driven primarily by CCR5, SLAM, CD2, CD69, and NKG2D in decreasing order of importance. Overall, these supervised machine learning models are in concordance with our differential expression testing and highlight the prominent roles of CCR5, SLAM, and CD2 as markers of HIV-1 *in vitro* infection (**Figure 2H**).

ASAPseq identification of HIV infected cells predominantly in Tfh cells from lymph nodes of untreated PLWH

We next assessed whether HIV-infected memory CD4+ T cells from the lymph nodes (LN) of viremic PLWH (**Table 1**), a tissue known to have a high burden of infected CD4+ T cells (Deleage et al., 2016; Estes et al., 2017), exhibited any bias towards specific cell surface markers or epigenetic characteristics. We enriched memory CD4+ T-cells via bead selection from inguinal (C01) or cervical (C02) lymph nodes (n = 2) and conducted ASAPseq using the same panel of 154 oligo-labeled antibodies. Alignments were made to a chimeric hg38 + HXB2 genome. In one individual (C01), we profiled a total of 2552 cells passing quality checks for ATAC and ADT datasets, including 13 HIV+ cells (0.51%). In the second individual (C02), we profiled a total of 6495 cells passing quality checks, with 41 cells (0.63%) marked as HIV+ leading to a combined chronic dataset of 9047 cells with 54 HIV+ cells (0.60%; **Supplemental Figure 6A**).The majority of sequenced fragments were internal (74%) followed by LTR only (23%) and spanning both LTR and internal (2%) (**Supplemental Figure 6B**). We also noticed a general occlusion of fragments from the *pol* region as observed from the *in vitro* dataset.

To assess the phenotype of HIV+ cells in LN, we again performed clustering into a reduced dimensional form (**Figure 3A-B**) and annotated clusters using both ADT and ATAC datasets (**Supplemental Figure 6C**, **Supplemental Figures 7-8**, and **Supplemental Table 5**). The majority of profiled cells were memory CD4+ T-cells (67.9% of all cells profiled) with the dominant cluster being CD4+ T-follicular helper (Tfh) cells (19.3% of all cells profiled) and a few contaminating clusters including APC, B, and follicular dendritic cells (**Figures 3A,C**).

Of the 54 HIV+ cells we detected, 63.0% were CD4+ Tfh cells, while the rest were identified as CD4+ T resident memory (Trm) and/or T central memory (Tcm) (14.8%), antigen presenting cells (14.8%), and CD4+ T-regulatory (Treg; 7.4%) (**Figures 3B-D**). This distribution corresponds with the known presence of the reservoir in CD4+ T-cells in germinal centers of secondary lymphoid organs, predominantly in the Tfh compartment (Pantaleo et al., 1993; Perreau et al., 2013). While there were contaminating follicular dendritic cells detected, we found no viral DNA in this cluster, even with the likely trapping of infectious virions by follicular dendritic cells (Heesters et al., 2015; Smith et al., 2001), suggesting high specificity in detecting only HIV DNA.

We next assessed whether HIV+ CD4+ T cells expressed differential surface antigens compared to HIV- CD4+ T cells. CD71, ICOS, HLA-DR, CD151, and PD-1 were preferentially upregulated on HIV+ cells, whereas CD48 was upregulated on HIV- cells (**Figure 3E**, **Supplemental Figure 9A**, and **Supplemental Table 6).** As several of these markers are known

to be elevated on Tfh cells, we next asked whether any surface marker was enriched for HIV+ versus HIV- cells within the Tfh compartment. CD71 was the only marker significantly upregulated in HIV+ Tfh cells compared to HIV- Tfh cells, suggesting that heightened activation and cycling of individual cells within the Tfh subset is associated with preferential infection (Younes et al., 2018). Within the HIV- Tfh cells, we observed higher expression of CD45 and CD69 **(Figure 3F and Supplemental Figure 9B**). Together, these analyses suggest that the HIV+ cells in lymph nodes of untreated PLWH are more activated/cycling, but that the HIV+ Tfh cell subset is otherwise quite similar to HIV- Tfh cells in the surface markers assessed.

Differential motif enrichment in HIV+ cells and surface marker based supervised machine learning from lymph nodes of untreated PLWH

To further define potential differences between the HIV+ and HIV- CD4+ T cells at the epigenetic level, we used chromVAR, specifically designed for differential motif analysis from sparse scATACseq data (Schep et al., 2017) on all HIV+ CD4+ T-cells versus HIV- CD4+ T-cells. We found several motifs that had significant bias-corrected deviations between HIV+ CD4+ cells and HIV- CD4+ cells (**Figure 4A**). In the HIV- CD4+ cells, we found significant motif signatures with increased genomic accessibility including TFE3, GATA1, and ARNTL2 (**Figure 4B**). In the HIV+ CD4+ T-cells, the most significant motifs showing increased accessibility included NFE2L1, NFE2L2, JUND, JDP2, BATF, BACH1, FOS, and JUN (**Figure 4C**). The presence of AP-1 related transcription factors (including FOS and JUN family transcription factors) motifs matched well with the observed increase in activation surface markers, as similarly observed in the *in vitro* dataset.

We sought to determine whether particular combinations of surface markers were preferentially associated with infection. Using the same approach for the *in vitro* analysis, we split all CD4+ T-cells into a 70/30 (training/test) split by their infection status. Overall, the models performed less well compared to the *in vitro* analysis. For the chronic samples, the random forest model attained an AUC value of 0.70 without downsampling, while the naive Bayes model performed with an AUC value of 0.74 (**Figure 4D**). The top markers of importance for the random forest model without downsampling include CD4, CD71, PD-1, CXCR3, CD79b, and CD45 which are generally in concordance with our differential expression testing of surface markers and suggest that the pool of infected cells in LN during chronic infection is highly heterogeneous.

ASAPseq identification of heterogeneous HIV-infected cells from peripheral blood of ART suppressed PLWH

Finally, we applied ASAPseq to blood CD4+ T-cells in the setting of ART to assess the reservoir in the most accessible compartment and in a clinically relevant context. To define the characteristics of the HIV CD4+ T-cell reservoir in the blood of ART-treated PLWH, we purified memory CD4+ T-cells from the PBMC of 3 ART-treated PLWH and performed ASAPseq. Given the rarity of infected cells during ART treatment (Chun et al., 1997) and the relative paucity of infected cells in peripheral blood compared to tissues, we substantially increased the number of memory CD4+ T-cells assessed compared to our previous studies. For all donors, we aligned the ATAC reads to autologous HIV sequences (near full length or HIV *env*) and HXB2 to increase the odds of detecting infected cells due to the prevalence of proviral mutations in

long-term PLWH (Ho et al., 2013). In one donor (A01), we profiled 24225 cells with 5 (0.02%) cells marked as HIV+ (**Supplementary Figure 10A**). This donor had previously been shown to contain 293.8 copies of total HIV DNA per million CD4+ T-cells (0.029%) (Salantes et al., 2018). In a second donor (A08), we profiled 16519 cells with 31 (0.19%) HIV+ cells. This donor had previously been shown to contain 1564.5 copies of total HIV DNA per million CD4+ T-cells (0.16%) (Salantes et al., 2018). In a third donor (B45), we profiled 10588 cells with 8 (0.08%) HIV+ cells. Altogether, the combined ART-treated dataset contained 51340 cells of which 44 (0.09%) cells were HIV+. With the autologous alignments, most aligned regions were primarily in the LTR and/or *nef* region (**Supplementary Figure 10B**).

After clustering based on the ATAC component and annotating with both ADT and ATAC components (**Supplementary Figure 10C**, **Supplementary Figures 11-12**, and **Supplemental Table 7**), we identified multiple memory CD4+ T cell subsets as well as small numbers of contaminating cells (**Figure 5A-C**). The T cell clusters separated based on epigenetic profiles associated with T-cell differentiation states; for example, central memory T-cells (Tcm) had distinct epigenetic and surface antigen profiles compared to those seen in effector/effector memory (Tem) cells (**Supplementary Figures 11-12**). HIV+ cells were found across multiple subsets including CD4+ Tcm, Th2, and transitional memory (Ttm) cells, while CD4+ Tem/effector cells had the largest contribution overall (**Figure 5B,D**). We also detected a small proportion of HIV+ cells within the contaminating APC clusters, but it is unclear whether these were in actual APCs, CD4+ T cells included in doublets with APCs, or miscalled as APCs.

In the CD4+ T-cell subset, HIV+ cells were found to be enriched in PD-1, CCR5, and CD2 surface markers (**Figure 5E**, **Supplemental Figure 13A**, and **Supplemental Table 8**). However, it is important to note that the enrichment of these markers, including PD-1, was donor specific, being primarily driven by individuals A08 and B45. In the HIV- CD4+ T-cells, we found enrichment in a multitude of markers including CD305, CD71, CD5, CD3, and CD47 (**Figure 5E**, **Supplemental Figure 13B**, and **Supplemental Table 8**). Similarly, we found that some of these markers were donor specific, indicating that the pooled set of HIV+ cells from peripheral blood of ART-treated PLWH has both interpersonal and intrapersonal heterogeneity. Finally, we performed chromVAR epigenetic analysis to compare the HIV+ CD4+ T-cells with the HIV- CD4+ T-cells. We were unable to find any motifs that were significantly enriched in either population (**Figure 5F**), but were likely underpowered for epigenetic analyses.

Discussion

A targeted elimination of the HIV reservoir requires a deep understanding of both the virologic and cellular characteristics. Prior work has yielded extensive knowledge of viral diversity, integration, clonal expansion, intactness, and latency dynamics of the viral reservoir, allowing a deeper understanding of the target for clearance. However, precise definition of cellular reservoir has proven more difficult due to the rarity of infected cells and the challenge of identifying infected cells while preserving their native state. To address this, we applied a novel single-cell genomic and bioinformatic strategy using ASAPseq to identify infected cells by the presence of integrated proviral DNA in the multiomic context of single cell epigenetic and surface antigen profiling. Using this strategy, we found that cells with proviral DNA can indeed be directly detected and individually characterized in multiple contexts of infection, including in peripheral blood of PLWH on ART. We anticipate that this strategy will have broad application for in depth analysis of the HIV reservoir within different cell populations, therapeutic interventions, and HIV cure and eradication studies (Abdel-Mohsen et al., 2020).

Numerous studies have attempted to identify cell surface markers and memory phenotypes that enrich or specifically identify CD4+ T cells harboring integrated proviral DNA. Our surface antigen profiling assessed many individual markers previously associated with HIV infection, including PD-1 (Perreau et al., 2013), CXCR5 (Perreau et al., 2013), CCR6 (Gosselin et al., 2017), CTLA-4 (McGary et al., 2017), CD69 (Cantero-Pérez et al., 2019), OX40 (Kuo et al., 2018), CD2 (Iglesias-Ussel et al., 2013), Lag-3 (Pardons et al., 2019), TIGIT (Fromentin et al., 2016; Pardons et al., 2019), CD20 (Serra-Peinado et al., 2019), and CD161 (Li et al., 2019), but did not include CD30 (Hogan et al., 2018), α4β1 (Pardons et al., 2019), and CD32a (Abdel-Mohsen et al., 2018; Descours et al., 2017), or Sialyl-LewisX (Colomb et al., 2020). In general, we found that activated CD4+ T-cells (*in vitro*), effector memory CD4+ T-cells (blood) and Tfh cells (LN) predominate the virally infected populations within their respective compartments. We also found preferential detection of provirus within activated CD4+ T-cells across infection states, as seen with the upregulation of multiple activation markers in HIV+ cells. However, multiple other memory or non-activated CD4+ T cell subpopulations contribute to the overall infected cell pool.

Importantly, between the different modalities we tested, no single previously identified cell surface marker was solely associated or predictive of the presence of integrated provirus. We do, however, find that specific combinations of markers, both by presence and relative level, may more reliably enrich for HIV-infected cells on *in vitro* infected samples. For example, CCR5 was a top differential marker between HIV+ and HIV- cells for the *in vitro* infected and ART-treated peripheral blood. However, CCR5 was not heightened on HIV+ cells in LN CD4+ T-cells of untreated PLWH, where activation and Tfh associated markers (CD71, ICOS, HLA-DR and PD-1) predominated the differentially expressed surface proteins within the total CD4 population. CD71 expression, enriched in HIV+ Tfh cells of untreated PLWH, correlates with cell cycling (Younes et al., 2018) and is a unique surface marker for ferroptosis (Feng et al., 2020), suggesting that these infected Tfh cells are dividing frequently and are potentially predisposed to cell death. Identification of cell surface markers found preferentially on HIV uninfected cells is also useful. In this regard, CD71, which we found elevated on HIV+ LN CD4+ T-cells in

untreated PLWH, was enriched on blood HIV- CD4+ T-cells from treated PLWH, possibly reflecting CD4+ T cell immune reconstitution during therapy. PD-1 was significantly enriched in all *ex vivo* and *in vitro* settings, highlighting the known association of PD-1 with the HIV reservoir (Fromentin et al., 2016; Perreau et al., 2013). We also found greater expression of CD48, a NK cell co-activating ligand for the CD244/2B4 receptor (Brown et al., 1998; Valiante and Trinchieri, 1993), on HIV- cells during untreated infection. CD48 has been previously found to be downregulated by HIV+ cells during *in vitro* infection to escape autologous NK cell mediated killing (Ward et al., 2007). Our findings, as a whole, indicate that HIV+ cells are highly heterogeneous at interpersonal and intrapersonal levels, and that each cell subset and compartment needs to be evaluated individually. Additional accumulation of infected cells from various *ex vivo* sources will be necessary to obtain better resolution of predictors of *in vivo* infection.

The epigenetic modality afforded by our ASAPseq strategy also yielded insights into potential differential gene regulation and new regulatory factors impacting HIV infection *in vitro* and *ex vivo*. Infection of *in vitro* activated CD4+ T-cells was highly associated with increased accessibility upstream of the *CCR5* transcription start site, which appears to be associated with increased CCR5 protein expression. In this same system, we further identified accessible peaks near genes previously implicated in HIV infection and replication, including *DCBLD1* and *DDIT4*. *DCBLD1* enhanced HIV replication during a RNA interference screen (Zhou et al., 2008), while *DDIT4* correlated with increased risk for HIV acquisition as part of the mTORC1 signaling pathway (Fourati et al., 2019)). In both the *in vitro* and viremic lymph node samples, we found heightened accessibility of the BATF transcription factor in HIV+ cells. BATF is a key regulator for differentiation into Th17 cells (Schraml et al., 2009) and Tfh functionality and germinal center formation (Ise et al., 2011) both known subsets targeted by HIV-1 (El Hed et al., 2010; Perreau et al., 2013). Other prominent motifs seen in HIV+ cells include POU2F1 (Oct-1) and POU2F2 (Oct-2) which have been shown to repress the HIV-1 LTR promoter (Liu and Latchman, 1997). These indicate a possible role in suppressing HIV activity and maintaining longevity of infected cells. We also detected a significant increase in NFE2L2 (NRF2) and NFE2 motif accessibility in HIV+ cells. This finding corresponds nicely with a prior report on latency reversal agents (LRA), including a compound that activates stress pathways responses upstream of NRF2 and can reverse latency *in vitro* (Timmons et al., 2020). Finally, the GATA1 motif was one of the most significant accessible motifs in HIV- cells, which is striking given its role in repressing CCR5 expression in human CD4+ T-cells (Sundrud et al., 2005). We unfortunately remained underpowered to observe epigenetic differences between HIV+ and HIV- cells in treated infection. Taken together, our epigenetic analysis provides a framework and new opportunities for understanding the complex regulation of HIV infection and ongoing processes regulating uncontrolled infection.

One major limitation of ASAPseq to identify infected cells is the requirement for the provirus to be in accessible chromatin. This may bias our surface and epigenetic profiling of virally infected cells towards an activated phenotype or selectively profile only the “active” reservoir (Einkauf et al., 2022). Proviruses become less accessible over time *in vitro* (Lindqvist et al., 2020) and after extended ART (du Chéné et al., 2007) which is a target for the cure field through epigenetic

modification with latency reversal agents (Darcis et al., 2015; Grau-Expósito et al., 2019; Leth et al., 2016; Reuse et al., 2009). This limitation is inherent in other reservoir measurement techniques (including QVOA and IPDA) which are restrained by incomplete proviral awakening via cellular stimulation or by primers being incompatible with sequence diversity, respectively (Kinloch et al., 2021). Despite this limitation, the detection of only accessible provirus by ASAPseq still offers unparalleled insight into infected cells capable of reactivation or might be actively transcribing at the proviral locus. Another limitation is the difficulty in assessing proviral intactness due to short-read sequencing and our transposase-based methodology. Alignments towards multiple proviral genes in a single cell can be used as a proxy for intactness given that most defective proviruses have large deletions in treated PLWH (Ho et al., 2013). In the *in vitro* infected cells, we found broad coverage of reads to the HIV genome and a large detection of infected cells, including those not actively expressing p24 protein. The *ex vivo* infected cells predictably yielded much sparser proviral sequence coverage. We also observed a dropoff in sequence recovery from *pol* regions, possibly explained by nucleosomal occupancy (Rafati et al., 2011) or more persistent transcription factor binding (Van Lint et al., 1994). The inclusion of scRNAseq as a third modality (Mimitou et al., 2021; Swanson et al., 2021) can assist in addressing intactness but is also limited towards accessible proviruses. Our focus on HIV DNA+ cells, regardless of intactness or transcription, is still relevant given that these cells survived clearance by the immune system and/or the short half-life associated with HIV infection kinetics (Perelson et al., 1996). Cells with defective proviruses from treated PWLH are still relevant as viral proteins can still be produced (Duette et al., 2022; Imamichi et al., 2020), which may interface with the immune system.

Overall, we have highlighted the complex heterogeneity of the HIV reservoir by using a novel genomic strategy to identify HIV infected cells at single-cell resolution with simultaneous surface antigen and epigenetic data. With this unparalleled resolution via ASAPseq compared to currently published techniques for studying the HIV reservoir, we uncovered both known and novel surface markers of HIV infection as well as accessible transcription factor motifs that may regulate or be indicative of HIV infection. Our study paves the way for future unmanipulated studies of the HIV reservoir. Furthermore, this approach can also be applied for lentiviral-based immunotherapy and other latent infections that involve nuclear, viral DNA. Our strategy and data are the start of a multiomic atlas of HIV infected cells with phenotypic and epigenetic characteristics - thus contributing towards the united goal of identifying the HIV reservoir for a targeted, functional cure of HIV.

Methods

Study approval: Persons living with chronic HIV were recruited by the Centro de Investigación en Enfermedades Infecciosas at the Instituto Nacional de Enfermedades Respiratorias (CIENI-INER) in Mexico City, Mexico. All donors provided informed consent for lymph node tissue donation in compliance with protocols set forth by the Ethics Committee and the Ethics in Research Committee of the INER (study number: B03-16) and the Institutional Review Board at the University of Pennsylvania (Philadelphia, PA). ART samples, A01 and A08, were provided from the ACTG clinical trial A5340, which was conducted with protocols set forth by the Institutional Review Boards at the University of Pennsylvania and the University of Alabama at Birmingham and was previously published (Bar et al., 2016). All donors for this study provided informed consent in compliance with protocols set forth by the respective institutional review boards. Another ART-treated PLWH, B45, was recruited from the BEAT-HIV program cohort where an apheresis sample was collected under ART.

Samples: For the *in vitro* infection model, peripheral blood mononuclear cells (PBMCs) were obtained from an HIV-negative donor apheresis from the Human Immunology Core (HIC) at the University of Pennsylvania. PBMC were cryopreserved at the time of receipt. For the chronic infection model, inguinal or cervical lymph nodes were obtained from HIV-1+ individuals prior to the introduction of ART and cryopreserved. For ART-treated studies, peripheral blood (PB) samples were acquired via apheresis and cryopreserved. Samples A01 and A08 had previously experienced therapy interruption through the ACTG A5340 trial, but were under at least 6 months of ART suppression at the time of analysis. Peripheral blood mononuclear cells and lymph node mononuclear cells (LNMCs) were prepared as previously described (Nguyen et al., 2019).

In vitro infection model: Uninfected PBMC were thawed and rested overnight in a humidified incubator at 37 °C. CD4+ T-cells were negatively enriched (STEMCELL Technologies, #19052), pelleted, and resuspended in complete RPMI media at a concentration of 2e6 cells/ml in a 6-well tissue culture plate. An activation cocktail consisting of anti-CD3 (1 µg/mL; Bio-Rad #MCA463EL), recombinant IL-2 (100 U/mL; Sigma Millipore #11011456001), and anti-CD28 + anti-CD49d (1.5ug/ml; BD #347690) were added to the cells. The cells were activated for 2 days in a humidified incubator at 37 °C. Cells were then pelleted in a 15ml tube and resuspended in 1ml of complete RPMI. HIV-1 virus stock (strain SUMA; provided by the University of Pennsylvania CFAR Virus and Reservoirs Core) was thawed briefly in a 37 °C water bath. 50ng (p24) of viral stock was added to the suspension and mixed by pipetting. Cells were then infected by spinoculation for 45 minutes at 400 x g. Cells were rested in a humidified incubator at 37 °C for 1 hour (cap loosened to promote gas exchange) before washing with complete RPMI and pelletting. Cells were resuspended at a concentration of 2e6 cells/ml and left to rest for 2 days in a 6-well tissue culture plate. Complete RPMI was added to the cells after two days. The cells were collected after an additional two days and pelleted for dead cell removal based on Annexin V using the Dead Cell Removal (Annexin V) Kit (STEMCELL Technologies, #17899).

Flow cytometric verification of HIV-1 infection: Staining and flow cytometry were based on a previously published protocol (Kuri-Cervantes et al., 2020). Approximately 1.5 million cells from the *in vitro* infection culture were spun down at 400 x g for 5 min and resuspended in 45ul of PBS. Live/dead staining was performed using 5µl of a 1:60 dilution stock of prepared Live/Dead Fixable Aqua Dead Cell Stain (Invitrogen). Cells were stained for 5 minutes in the dark at room temperature. A staining cocktail with FACS buffer and CD8 BV570 (BioLegend #301038, clone RPA-T8) was added for a 10 minute stain in the dark at room temperature. 1ml FACS buffer was added and the cells were spun down at 400 x g for 5 min. Cells were permeabilized with 250µl of BD Cytofix/Cytoperm solution (BD #554714) for 18 minutes in the dark at room temperature. 1ml of BD Perm/Wash Buffer (BD #554714) was added and cells were spun down at 600 x g for 5 min. After supernatant was discarded, cells were resuspended in staining solution containing anti-p24 FITC (Beckman Coulter #6604665; clone KC57) and BD Perm/Wash Buffer for a final staining volume of 50µl. Cells were stained in the dark for 1 hour at room temperature. Cells were washed with 1ml of BD Perm/Wash Buffer and fixed with 350ul of 1% paraformaldehyde. Data were acquired on a BD FACS Symphony A5 cytometer and analyzed using FlowJo (Version 10, BD).

Memory CD4$^+$ T-cell enrichment: Memory CD4+ T-cells were enriched by negative selection using the EasySep Human Memory CD4+ T Cell Enrichment Kit (STEMCELL Technologies #19157) following the recommended protocol. After collecting enriched memory CD4+ T-cells, the cells were spun down at 400 x g for 5 min and resuspended in 500µl PBS. Cells were counted with trypan blue staining using a Countess II (Invitrogen) prior to beginning the ASAPseq protocol.

ASAPseq - cell preparation and staining: Buffer and cell preparation were performed as previously published (Mimitou et al., 2021) and as summarized below. 5e5 to 1e6 cells were resuspended in 22.5µl of Staining Buffer and incubated with 2.5µl of TruStain FcX (BioLegend #422302) for 10 minutes on ice. One test of the TotalSeq-A Human Universal Cocktail, V1.0 (BioLegend #399907) was prepared according to the manufacturer's protocol with ASAPseq Staining Buffer. 25µl of the antibody cocktail was added to the cells for 30 minutes on ice. Cells were washed with 1ml Staining Buffer and pelleted (all spins at 400g for 5 minutes with centrifuge set at 10C) for 3 total washes. Cells were then resuspended in 450µl of PBS. 30ul of 16% paraformaldehyde was added to fix the cells for 10 minutes at RT with occasional swirling and inversion. The fixation reaction was quenched with 25.26µl of 2.5M glycine solution. Cells were then washed and pelleted with 1ml of ice-cold PBS for a total of two washes. Fixed cells were lysed using 100µl of OMNI lysis buffer for 3 minutes on ice. Cells were then washed and pelleted with 1ml Wash Buffer (spin at 500g for 5 min with centrifuge set at 10C). After supernatant was removed, cells were resuspended in at least 150µl (depending on original cell input) of 1X Nuclei Buffer (10X Genomics) and strained to remove aggregates using a 40µm FlowMi strainer (Sigma #BAH136800040-50EA). Cells were counted using trypan blue staining to verify successful permeabilization (>95% trypan blue positive staining). Cells were diluted as needed according to the scATACseq Chip H protocol (10X Genomics).

ASAPseq - library preparation: Single-cell droplets were generated using the Chromium platform and a scATACseq Chip H kit (10X Genomics). Modifications mentioned in the original ASAPseq protocol were performed to allow for capture of antibody-derived tags (ADT). Library preparation for the ATAC libraries were performed as recommended by 10X Genomics. ADT libraries were prepared as recommended by the ASAPseq protocol with 3µl of the silane bead elution used as additional input to the ADT indexing PCR.

ASAPseq - library sequencing: Sequencing runs were performed on NextSeq 550 or NovaSeq 6000 platforms (Illumina) with a target of at least 10,000 reads/cell for ADT libraries and 25,000 reads/cell for ATAC libraries.

Initial ATAC data processing: Libraries sequenced via the NovaSeq were filtered for potential index hopping by using index-hopping-filter (10X Genomics; v1.1) in ATAC mode. Filtered reads were then processed the same as other libraries sequenced with the NextSeq. A chimeric genome was built using cellranger-atac mkref (10X Genomics; v2.0.0) with the default hg38 genome (10X Genomics; refdata-cellranger-arc-GRCh38-2020-A-2.0.0) along with either SUMA_TF1 (for *in vitro*), HXB2 (for chronic), or autologous sequences + HXB2 (for ART). The motifs.pfm file from the default hg38 genome reference was copied over to the new chimeric genome to assist with quality control during downstream processing by cellranger. The reads were aligned and counted to their respective chimeric genome using cellranger-atac count. Raw fastq files and processed cellranger-atac files are deposited in the NCBI Gene Expression Omnibus (GEO) under accession number GSE199727. After alignment, barcodes pertaining to multiplets were selected using AMULET (Thibodeau et al., 2021) for downstream filtering. Fragment files from cellranger-atac were loaded into ArchR (v.1.0.2; commit 48dccf) (Granja et al., 2021) for downstream analysis. Arrow files were created for each sample. For *in vitro* and ART-treated samples, cells were filtered for TSS Enrichment $\geqq$ 8 while chronic samples were filtered for TSS Enrichment $\geqq$ 6. Barcodes that were selected by AMULET as being multiplets were then filtered out of the Arrow file. Samples, by level, continued in the ArchR pipeline. Briefly, iterative latent semantic indexing, Harmony (for batch effect correction as needed), cluster generation (via Seurat) and UMAP generation were performed. All specific parameters and settings are documented in our code (https://github.com/betts-lab/asapseq-hiv-art).

Initial ADT data processing: Sequencing reads were converted to fastq files and demultiplexed using bcl2fastq (Illumina). Reads were then inputted into kallisto bustools (Melsted et al., 2021) to generate cell barcode by feature matrices. These matrices were loaded into R and checked for empty droplets (i.e. background noise) using the emptyDrops function from the DropletUtils (Lun et al., 2019) package with a lower bound of 500 UMIs. Barcodes with false discovery rates < 0.01 were kept as genuine cells. This filtered matrix was then loaded into Seurat (v4.0.5) (Hao et al., 2021) for centered log ratio transformation with a scale factor of 10000. Harmony (Korsunsky et al., 2019) was used for batch correction if multiple 10X GEM wells were used for the same individual. A wilcoxon test (right-tailed) was used to select for features with a count distribution that was significantly different than background signal detected from isotype control antibodies, as analyzed similarly in a prior report (Swanson et al., 2021). A feature was not included in downstream analyses if $p > 0.05$ in at least 4 isotype controls.

Identification of HIV+ cells: We built a custom Python pipeline (hiv-haystack; https://github.com/betts-lab/hiv-haystack) built and derived from epiVIA (Wang et al., 2020) to identify proviral reads from the BAM output file of cellranger-atac count. BAM records are parsed and selected into three groups: 1) both mates aligning to provirus, 2) both mates in host with a soft align clip, and 3) one mate in provirus and one mate in host. In scenario 1, the original alignments were kept as bona fide proviral alignments. If there was a soft clip present, we aligned it to the host genome using bwa mem as done in epiVIA to assess if an integration site is present. In scenario 2, any host soft clips were checked for exact matching to LTR sequences to allow for integration site detection. This was only allowed if a LTR alignment was reported (checked with BLAST against HXB2 5' LTR). In scenario 3, the proviral mate was saved as a proviral fragment. If either mate had a soft clip, integration site checking is performed similar to scenarios 1 and 2. In all scenarios, only exact integration sites with exact nucleotide position are kept for downstream analyses. Our program also differs from epiVIA in that hiv-haystack allows for alignments to multiple sequences (from near full-length genomes and single genome sequencing of HIV Env) which enhances the detection of donor-derived sequences.

Downstream data processing: After initial ATAC and ADT processing, barcodes that passed quality control metrics in both modalities were selected for downstream analyses. Clusters were annotated from a base panel of ADT expression patterns, imputed gene scores from ATAC, and differentially expressed markers present in any given cluster over all other cells. UMAP embeddings in lower dimension were generated using ArchR's latent semantic indexing implementation. Manual annotations were then used to subset cells for downstream ADT differential expression analysis between HIV+ and HIV- cells using Seurat FindMarkers (using "DESeq2" for the findMarkerMethod parameter) to limit false discoveries (Squair et al., 2021) and the subset of features that were significantly different from background expression. For downstream ATAC differential expression analysis, manual clusters were grouped based on CD4+/- and HIV+/- (and by memory phenotype for the *in vitro* experiment) and pseudo-bulk replicates were generated. Clusters with less than 0.5% of the population were filtered out due to annotation difficulty. Peaks were called using the default 501-bp fixed-width method with ArchR and MACS2 (Zhang et al., 2008). Differential peaks were assessed using a wilcoxon test as implemented by ArchR's getMarkerFeatures with bias parameters set to "c(TSSEnrichment", "log10(nFrags"))" as recommended by ArchR. Motif annotations of differential peaks were added using addMotifAnnotations from the CISBP dataset. For experiments with lower numbers of HIV+ cells, we used ArchR's implementation of chromVAR to determine differential motifs. getMarkerFeatures was called to determine differentially expressed motifs using the chromVAR deviation scores (i.e. z-score).

Supervised machine learning: Samples were split into a 70% training and 30% testing dataset using the caTools package in R. Logistic regression models were developed using the glm function with parameters "family = binomial(link = 'logit')". Naive Bayesian models were constructed using the naivebayes package in R. Random forest models were constructed using the randomForest package in R with ntree = 4000. Classifier performances (true positive rate

and false positive rate) were calculated using the ROCR package in R and plotted using ggplot2.

NFL SGS of proviral DNA from resting CD4+ T cells: DNA was extracted from multiples of 4 million resting CD4+ T cells according to the manufacturer's instructions (QIAamp DNA Mini and Blood Mini Kit, QIAGEN). Amplification of NFL genomes was performed by limiting dilution, nested PCR using Platinum Taq High Fidelity Polymerase (Life Technologies, Thermo Fisher Scientific) as previously described (Bar et al., 2016), adapted for NFL genomes with primary forward and reverse primers, respectively 5′-AAATCTCTAGCAGTGGCGCCCGAACAG-3′ and 5′-TGAGGGATCTCTAGTTACCAGAGTC-3′ followed by nested forward and reverse primers, respectively 5′-GCGGAGGCTAGAAGGAGAGAGATGG-3′ and 5′-GCACTCAAGGCAAGCTTTATTGAGGCTTA-3′. Amplicons of appropriate size were directly sequenced on the Illumina MiSeq platform, inspected for evidence of priming from multiple templates, stop codons, large deletions, or introduction of PCR error in early cycles; a threshold of 85% identity at each nucleotide position was used.

HIV custom sequence annotation: HIV custom sequences from untreated and treated PLWH were loaded into the Gene Cutter tool (HIV LANL Database) for alignment to HIV genes/regions. The output was then loaded into our custom tool, called geneCutterParser (https://github.com/wuv21/geneCutterParser), to extract the coordinate locations for any alignments. The coordinate locations were then used for alignment graphs in the supplemental figures.

Graphics: All figures were made in R with the following packages: gridExtra, ggplot2 (Wickham, 2016), ArchR, and patchwork. Graphical abstract cell icons were made with Biorender.com. All code to produce graphs can be found in the study GitHub repository.

Author contributions:
VHW, KJB, LAV, and MRB conceptualized experiments. PMDRE, FTR, MGN, YALV, SAR, GRT, PT, LJM, and KJB recruited and organized sample collection and cell preparation that were used for the study. VHW and JLMN performed the ASAPseq and flow experiments. JJ, FM, and KJB conducted and compiled the consensus viral sequencing. VHW and SN performed the ASAPseq data analysis. VHW, LAV, and MRB wrote the manuscript.

Acknowledgements
We thank the individuals who donated samples to our study. We would also like to thank members of the Betts, Vella, and Bar labs for their assistance, Andrew Wells and James Pippin for assistance with NovaSeq sequencing, the University of Pennsylvania Next-Generation Sequencing Core, and the University of Pennsylvania Human Immunology Core. Additional thanks to Wenliang Wang, Golnaz Vahedi, and Brandon Keele for helpful discussions regarding proviral alignment and software. MRB is supported by the NIH grant (U19-A1-149680-02). Both MRB and KJB are supported by the NIH grant (P01-AI31338). Both MRB and LAV are further supported by the Penn Center for AIDS Research (P30-AI045008). LAV is supported by NIH grant K08-AI136660. VHW is supported by NIH grant (T32-AI007632). MRB, KJB, PT, and LJM are supported by UM-1AI164570 (BEAT-HIV Collaboratory) which is co-supported by the National Institute of Allergies and Infectious Diseases (NIAID), the National Institute of Mental Health (NIMH), the National Institute of Neurological Disorders and Stroke (NINDS), the National Institute on Drug Abuse (NIDA), and the Robert I. Jacobs Fund of The Philadelphia Foundation. LJM is also supported by the Herbert Kean, M.D., Family Professorship. CIENI-INER is supported by the Mexican Government (Programa Presupuestal P016; Anexo 13 del Decreto del Presupuesto de Egresos de la Federación).

Conflict of interest:
MRB is a consultant for Interius BioTherapeutics. No other conflicts are reported by the authors.

Tables

Table 1: Donor information

Stage	Donor	Sex	Age (y)	Tissues profiled**	Plasma viral load (log copies /ml)
in vitro infection	ND492	M	37	PB	-
chronic infection	C01	M	33	Inguinal LN	6.8
	C02	M	33	Cervical LN	4.88
ART-treated infection	A01	M	*	PB	<1.7
	A08	M	*	PB	<1.7
	B45	M	43	PB	<1.7

* Age was not reported at the participant level for the original study. However, the age range was reported as between 34 to 52 years.
** PB = peripheral blood; LN = lymph node

Figure Legends

Figure 1: ASAPseq identification of HIV-infected cells in vitro. (A) UMAP representation of ATAC component colored by manually annotated cell phenotypes. (B) UMAP representation of ATAC component colored by detection of HIV reads. (C) Absolute count of cell numbers based on annotated clusters. Red color denotes HIV+ cells while gray color denotes HIV- cells. (D) Percentage of HIV+ cells found in each annotated cluster. (E) Surface enrichment profile in activated HIV+ cells (red) versus activated HIV- cells (gray). Activated cells are defined as the combination of CD4 Activated/effector, CD4 Activated #1, and CD4 Activated #2 clusters.

Figure 2: Differential chromatin accessibility in HIV-infected cells and surfacer marker based supervised machine learning in vitro. (A) Volcano plot showing differentially accessible peaks between activated HIV+ versus activated HIV- cells. (B) Top 15 significant peaks in activated HIV- versus activated HIV+ cells. Peaks are ranked by π-score (-log10(FDR) * log2(FC)). Y-axis labels denote nearest TSS and/or if the peak is in a gene (marked by asterisk *) and numbers in parentheses indicate distance to the nearest TSS. Negative numbers indicate that the nearest TSS is upstream of the peak. (C) (top) Aggregated and normalized ATAC signal between activated HIV- (gray) and activated HIV+ cells (red) at the *CCR5* locus. (middle) Binarized ATAC signal at a single-cell (showing random 750 cells for each group). (bottom) Gene map for the genomic region shown: chr3:46360853-46380854 (centered on *CCR5*). Orange genes are in positive orientation while blue genes are in negative orientation. (D) Significant motif enrichments found in the differential peaks (ordered by significance) are shown in activated HIV- cells or activated HIV+ cells. (E) Area under the receiver operating curve (ROC) (AUC) plots for multiple supervised machine learning models for all CD4+ T-cells. RF = random forest. Ratios for different RF models indicate the number of HIV- cells used for each HIV+ cell (i.e. 5:1 ratio meant that the HIV- cells in the training dataset were randomly downsampled to get only 5 times the number of HIV+ cells in the training dataset). (F) Significant coefficients for the logistic regression model shown in (E). Positive odds ratio (red) indicates a marker has more weight for HIV+ cells while negative odds ratio (gray) indicates a marker has more weight for HIV- cells. (G) AUC plots for multiple supervised machine learning models for activated CD4+ T-cells. (H) Proportion of HIV- (gray) and HIV+ (red) activated CD4+ T-cells with different surface marker combinations. In the bottom dot plot, a black dot indicates that the combination is positive for that marker, which was determined by the thresholds indicated with the dotted line on the ridge plots. These ridge plots display the expression distribution of activated HIV+ CD4+ T-cells (red), activated HIV- CD4+ T-cells (gray), and total cells in the *in vitro* dataset (blue) to help with gating.

Figure 3: ASAPseq identification of HIV infected cells predominantly in Tfh cells from lymph nodes of untreated PLWH. (A) UMAP representation of ATAC component colored by manually annotated cell phenotypes. (B) UMAP representation of ATAC component colored by detection of HIV reads. (C) Absolute count of cell numbers based on annotated clusters. Red color denotes HIV+ cells while grey color denotes HIV- cells. (D) Percentage of HIV+ cells found in each annotated cluster. (E) Surface enrichment profile in HIV+ CD4+ T-cells versus HIV- CD4+ T-cells. (F) Same as shown in (E) but for HIV+ Tfh cells versus HIV- Tfh cells.

Figure 4: Differential motif enrichment in HIV+ cells and surface marker based supervised machine learning from lymph nodes of untreated PLWH. (A) Volcano plot of differentially enriched motifs from chromVAR, with a threshold of false discovery rate (FDR) < 0.05 indicating statistical significance. Mean difference values greater than 0 indicate an enrichment in HIV+ cells, while negative mean difference values indicate enrichment in HIV- cells. These specific and significant motif enrichments (ordered by FDR) are shown for (B) CD4+ HIV- T-cells and C) CD4+ HIV+ T-cells. (D) Area under the receiver operating curve (AUC) plots for various supervised machine learning methods. RF = random forest. Numbering format for RF models is the same as explained for Figure 2.

Figure 5: ASAPseq identification of heterogeneous HIV-infected cells from peripheral blood of ART suppressed PLWH. (A) UMAP representation of ATAC component colored by manually annotated cell phenotypes. (B) UMAP representation of ATAC component colored by detection of HIV reads. (C) Absolute count of cell numbers based on annotated clusters. Red color denotes HIV+ cells while gray color denotes HIV- cells. (D) Percentage of HIV+ cells found in each annotated cluster. (E) Surface enrichment profile in HIV+ CD4+ T-cells (red) versus HIV- CD4+ T-cells (gray). (F) Volcano plot of motif false discovery rate and mean difference in deviations between HIV+ CD4+ T-cells and HIV- CD4+ T-cells using chromVAR.

Figure 1

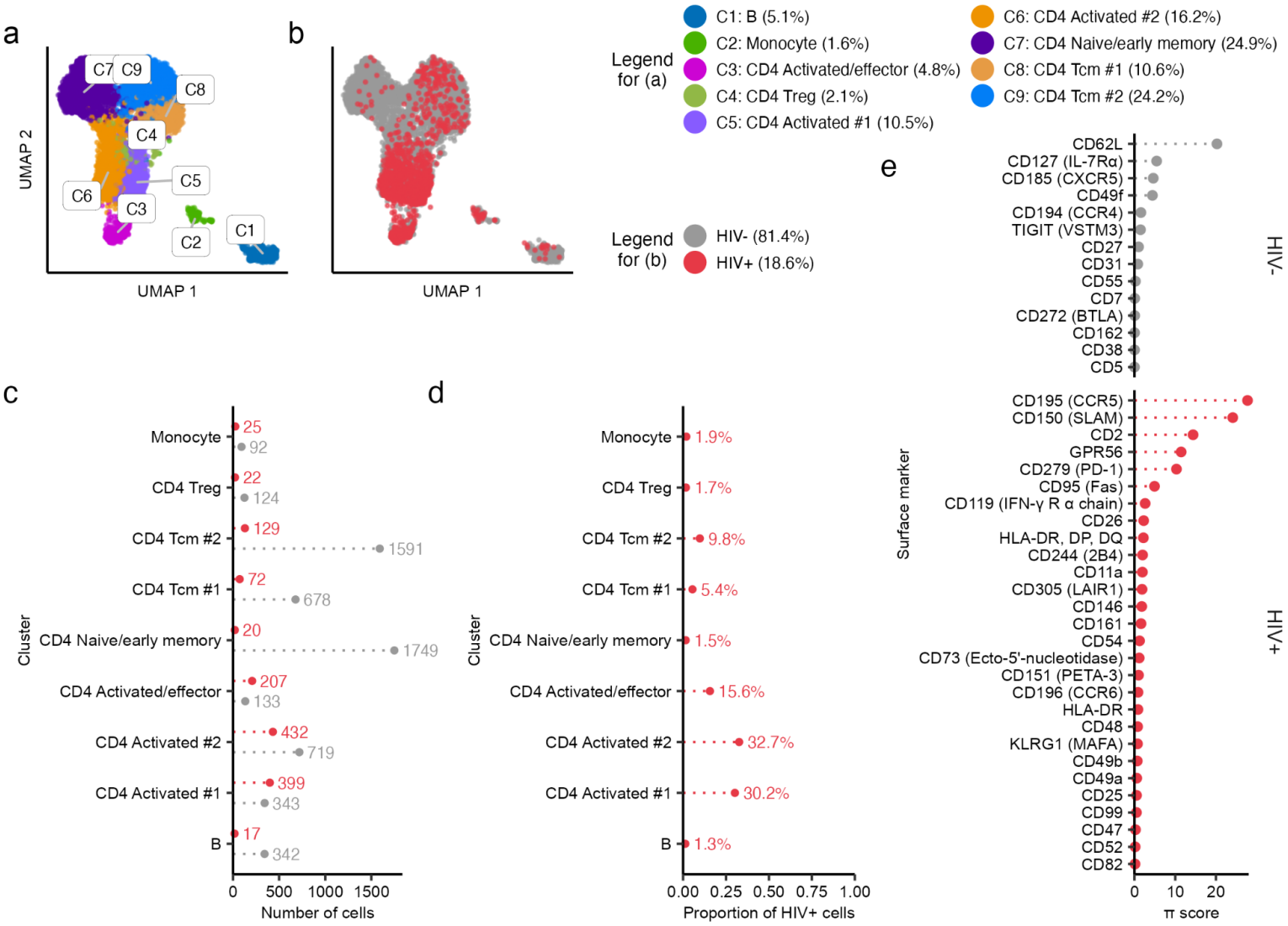

a
b
UMAP 2
UMAP 1
C1
C2
C3
C4
C5
C6
C7
C8
C9
Legend for (a)
C1: B (5.1%)
C2: Monocyte (1.6%)
C3: CD4 Activated/effector (4.8%)
C4: CD4 Treg (2.1%)
C5: CD4 Activated #1 (10.5%)
C6: CD4 Activated #2 (16.2%)
C7: CD4 Naive/early memory (24.9%)
C8: CD4 Tcm #1 (10.6%)
C9: CD4 Tcm #2 (24.2%)
Legend for (b)
HIV- (81.4%)
HIV+ (18.6%)
c
Cluster
Monocyte
CD4 Treg
CD4 Tcm #2
CD4 Tcm #1
CD4 Naive/early memory
CD4 Activated/effector
CD4 Activated #2
CD4 Activated #1
B
25
92
22
124
129
1591
72
678
20
1749
207
133
432
719
399
343
17
342
0
500
1000
1500
Number of cells
d
1.9%
1.7%
9.8%
5.4%
1.5%
15.6%
32.7%
30.2%
1.3%
0.00
0.25
0.50
0.75
1.00
Proportion of HIV+ cells
e
Surface marker
HIV-
HIV+
CD62L
CD127 (IL-7Rα)
CD185 (CXCR5)
CD49f
CD194 (CCR4)
TIGIT (VSTM3)
CD27
CD31
CD55
CD7
CD272 (BTLA)
CD162
CD38
CD5
CD195 (CCR5)
CD150 (SLAM)
CD2
GPR56
CD279 (PD-1)
CD95 (Fas)
CD119 (IFN-γ R α chain)
CD26
HLA-DR, DP, DQ
CD244 (2B4)
CD11a
CD305 (LAIR1)
CD146
CD161
CD54
CD73 (Ecto-5'-nucleotidase)
CD151 (PETA-3)
CD196 (CCR6)
HLA-DR
CD48
KLRG1 (MAFA)
CD49b
CD49a
CD25
CD99
CD47
CD52
CD82
0
10
20
π score

Figure 2

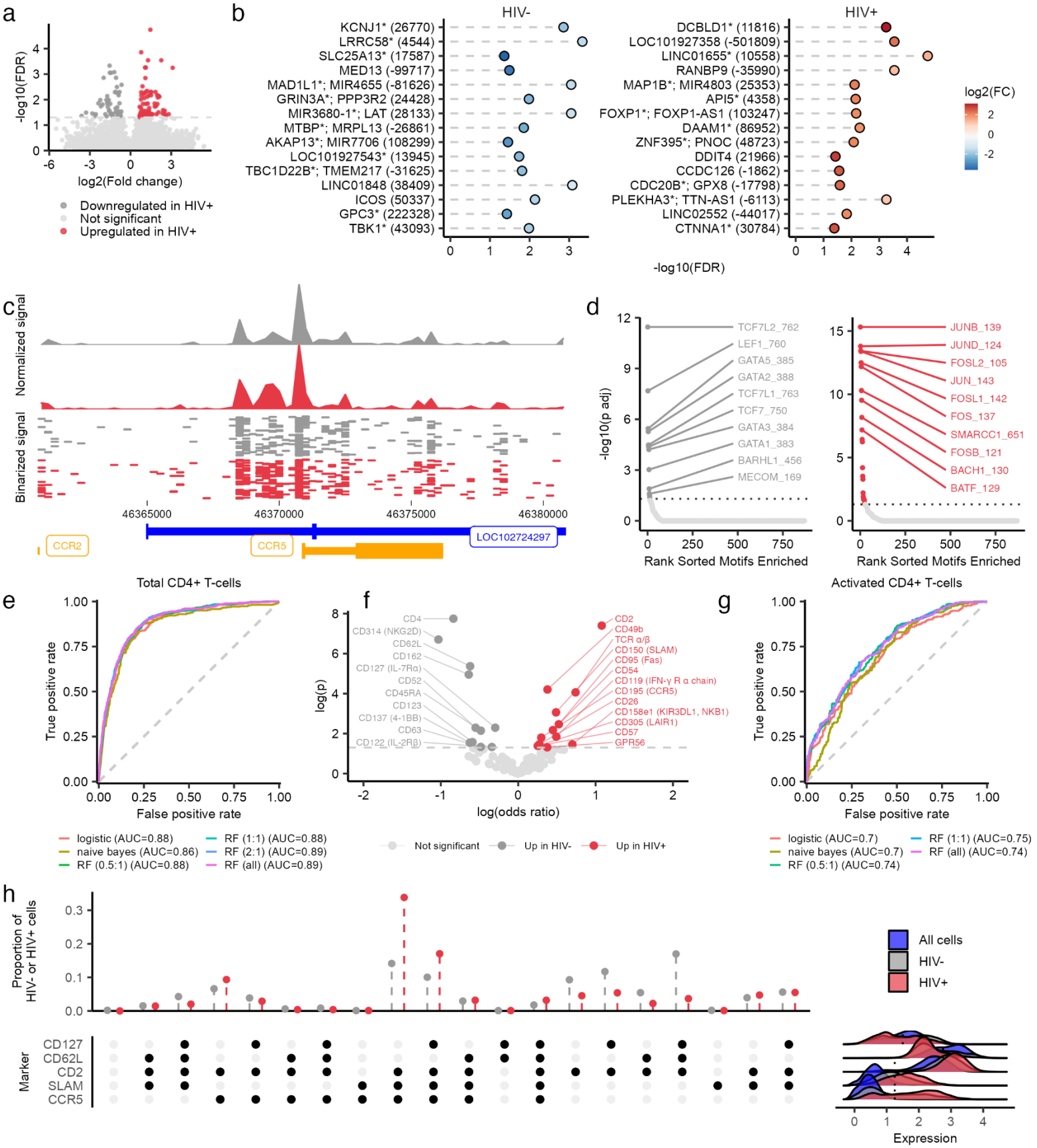

a
-log10(FDR)
log2(Fold change)
Downregulated in HIV+
Not significant
Upregulated in HIV+
b
HIV-
KCNJ1* (26770)
LRRC58* (4544)
SLC25A13* (17587)
MED13 (-99717)
MAD1L1*; MIR4655 (-81626)
GRIN3A*; PPP3R2 (24428)
MIR3680-1*; LAT (28133)
MTBP*; MRPL13 (-26861)
AKAP13*; MIR7706 (108299)
LOC101927543* (13945)
TBC1D22B*; TMEM217 (-31625)
LINC01848 (38409)
ICOS (50337)
GPC3* (222328)
TBK1* (43093)
HIV+
DCBLD1* (11816)
LOC101927358 (-501809)
LINC01655* (10558)
RANBP9 (-35990)
MAP1B*; MIR4803 (25353)
API5* (4358)
FOXP1*; FOXP1-AS1 (103247)
DAAM1* (86952)
ZNF395*; PNOC (48723)
DDIT4 (21966)
CCDC126 (-1862)
CDC20B*; GPX8 (-17798)
PLEKHA3*; TTN-AS1 (-6113)
LINC02552 (-44017)
CTNNA1* (30784)
-log10(FDR)
log2(FC)
c
Normalized signal
Binarized signal
46365000
46370000
46375000
46380000
CCR2
CCR5
LOC102724297
d
-log10(p adj)
TCF7L2_762
LEF1_760
GATA5_385
GATA2_388
TCF7L1_763
TCF7_750
GATA3_384
GATA1_383
BARHL1_456
MECOM_169
JUNB_139
JUND_124
FOSL2_105
JUN_143
FOSL1_142
FOS_137
SMARCC1_651
FOSB_121
BACH1_130
BATF_129
Rank Sorted Motifs Enriched
Rank Sorted Motifs Enriched
e
Total CD4+ T-cells
True positive rate
False positive rate
logistic (AUC=0.88)
naive bayes (AUC=0.86)
RF (0.5:1) (AUC=0.88)
RF (1:1) (AUC=0.88)
RF (2:1) (AUC=0.89)
RF (all) (AUC=0.89)
f
log(p)
log(odds ratio)
CD4
CD314 (NKG2D)
CD62L
CD162
CD127 (IL-7Rα)
CD52
CD45RA
CD123
CD137 (4-1BB)
CD63
CD122 (IL-2Rβ)
CD2
CD49b
TCR α/β
CD150 (SLAM)
CD95 (Fas)
CD54
CD119 (IFN-γ R α chain)
CD195 (CCR5)
CD26
CD158e1 (KIR3DL1, NKB1)
CD305 (LAIR1)
CD57
GPR56
Not significant
Up in HIV-
Up in HIV+
g
Activated CD4+ T-cells
True positive rate
False positive rate
logistic (AUC=0.7)
naive bayes (AUC=0.7)
RF (0.5:1) (AUC=0.74)
RF (1:1) (AUC=0.75)
RF (all) (AUC=0.74)
h
Proportion of HIV- or HIV+ cells
All cells
HIV-
HIV+
Marker
CD127
CD62L
CD2
SLAM
CCR5
Expression

Figure 3

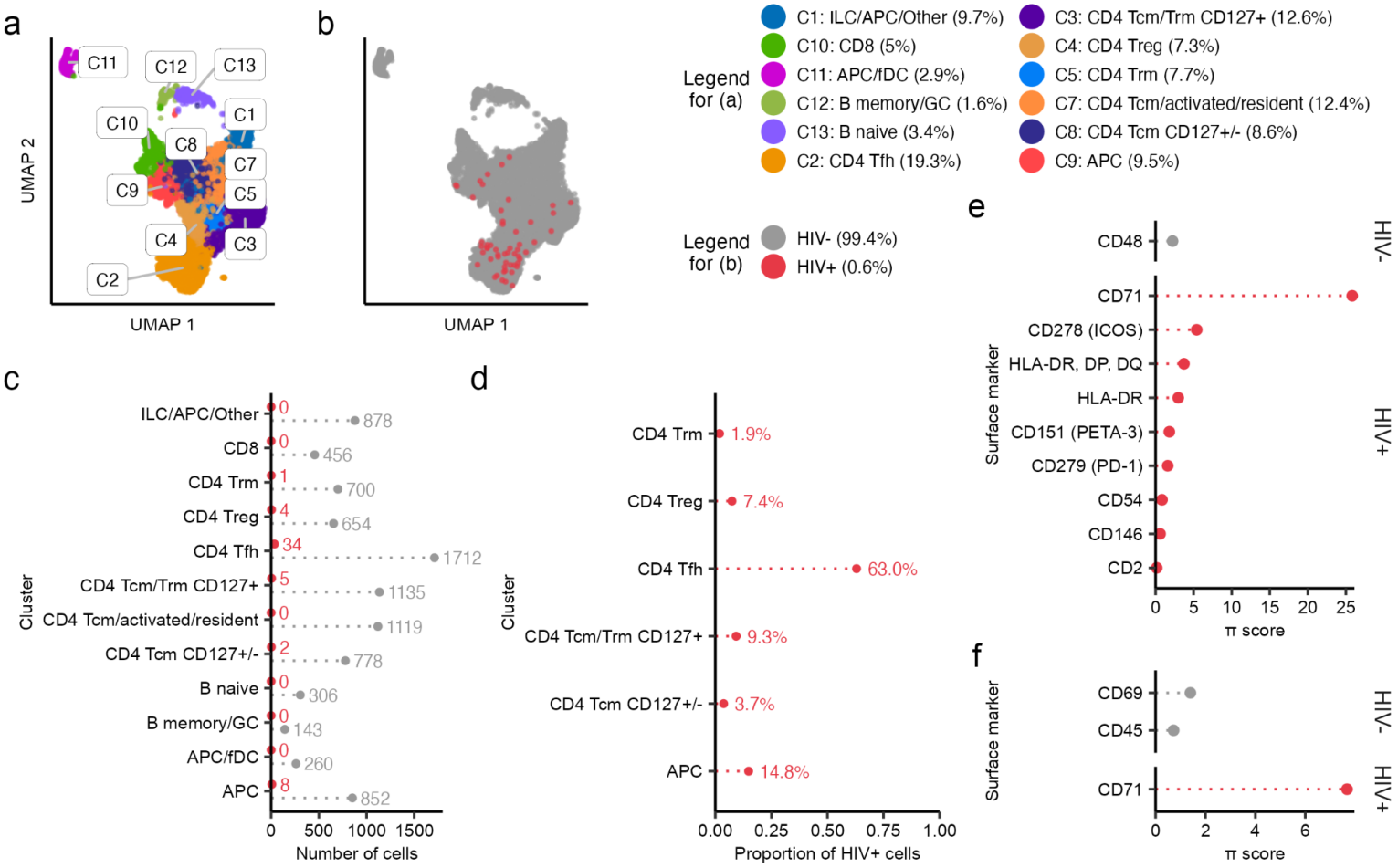

a
C11
C12
C13
C10
C1
C8
C7
C9
C5
C4
C3
C2
UMAP 2
UMAP 1
b
UMAP 1
Legend for (a)
C1: ILC/APC/Other (9.7%)
C10: CD8 (5%)
C11: APC/fDC (2.9%)
C12: B memory/GC (1.6%)
C13: B naive (3.4%)
C2: CD4 Tfh (19.3%)
C3: CD4 Tcm/Trm CD127+ (12.6%)
C4: CD4 Treg (7.3%)
C5: CD4 Trm (7.7%)
C7: CD4 Tcm/activated/resident (12.4%)
C8: CD4 Tcm CD127+/- (8.6%)
C9: APC (9.5%)
Legend for (b)
HIV- (99.4%)
HIV+ (0.6%)
c
Cluster
ILC/APC/Other 0 878
CD8 0 456
CD4 Trm 1 700
CD4 Treg 4 654
CD4 Tfh 34 1712
CD4 Tcm/Trm CD127+ 5 1135
CD4 Tcm/activated/resident 0 1119
CD4 Tcm CD127+/- 2 778
B naive 0 306
B memory/GC 0 143
APC/fDC 0 260
APC 8 852
0 500 1000 1500
Number of cells
d
Cluster
CD4 Trm 1.9%
CD4 Treg 7.4%
CD4 Tfh 63.0%
CD4 Tcm/Trm CD127+ 9.3%
CD4 Tcm CD127+/- 3.7%
APC 14.8%
0.00 0.25 0.50 0.75 1.00
Proportion of HIV+ cells
e
Surface marker
CD48
CD71
CD278 (ICOS)
HLA-DR, DP, DQ
HLA-DR
CD151 (PETA-3)
CD279 (PD-1)
CD54
CD146
CD2
HIV-
HIV+
0 5 10 15 20 25
π score
f
Surface marker
CD69
CD45
CD71
HIV-
HIV+
0 2 4 6
π score

Figure 4

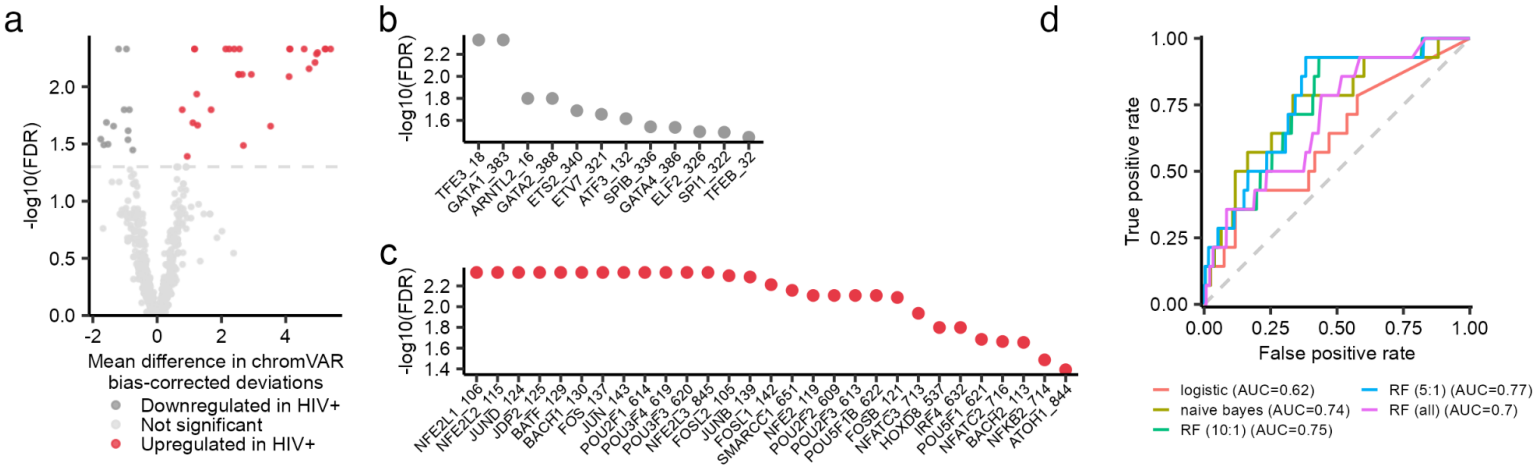
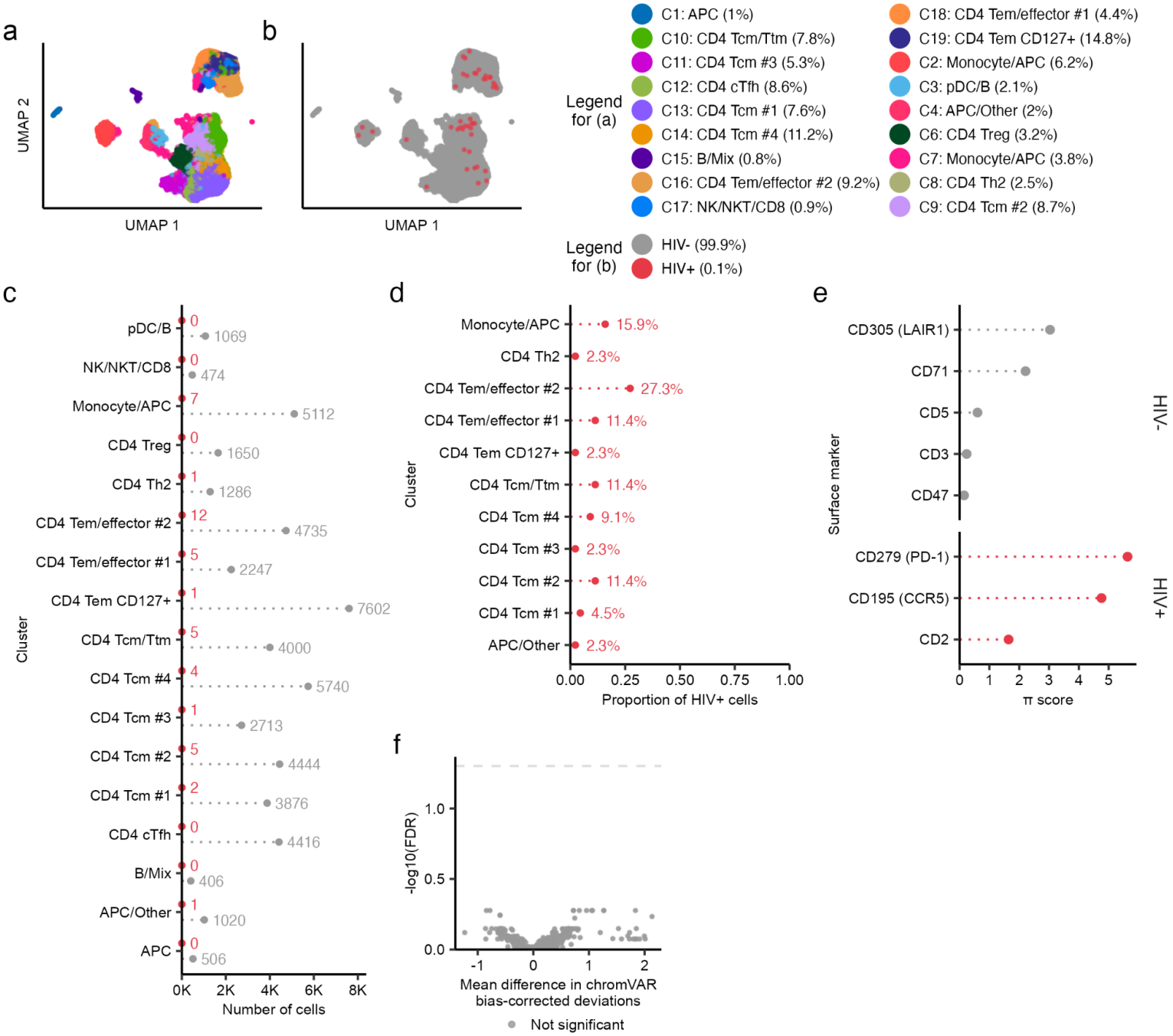

a
-log10(FDR)
Mean difference in chromVAR bias-corrected deviations
Downregulated in HIV+
Not significant
Upregulated in HIV+
b
-log10(FDR)
TFE3_18
GATA1_383
ARNTL2_16
GATA2_388
ETS2_340
ETV7_321
ATF3_132
SPIB_336
GATA4_386
ELF2_326
SPI1_322
TFEB_32
c
-log10(FDR)
NFE2L1_106
NFE2T2_115
JUND_124
JDP2_125
BATF_129
BACH1_130
FOS_137
JUN_143
POU2F1_614
POU3F4_619
POU5F3_620
NFE2L3_845
FOSL2_105
JUNB_139
FOSL1_142
SMARCC1_651
NFE2_119
POU2F2_609
POU2F3_613
POU6F1B_622
FOSB_121
NFATC3_713
HOXD8_537
IRF4_632
POU5F1_621
NFATC2_716
BACH2_113
NFKB2_714
ATOH1_844
d
True positive rate
False positive rate
logistic (AUC=0.62)
naive bayes (AUC=0.74)
RF (10:1) (AUC=0.75)
RF (5:1) (AUC=0.77)
RF (all) (AUC=0.7)

Figure 5


a
b
UMAP 2
UMAP 1
UMAP 1
Legend for (a)
C1: APC (1%)
C10: CD4 Tcm/Ttm (7.8%)
C11: CD4 Tcm #3 (5.3%)
C12: CD4 cTfh (8.6%)
C13: CD4 Tcm #1 (7.6%)
C14: CD4 Tcm #4 (11.2%)
C15: B/Mix (0.8%)
C16: CD4 Tem/effector #2 (9.2%)
C17: NK/NKT/CD8 (0.9%)
C18: CD4 Tem/effector #1 (4.4%)
C19: CD4 Tem CD127+ (14.8%)
C2: Monocyte/APC (6.2%)
C3: pDC/B (2.1%)
C4: APC/Other (2%)
C6: CD4 Treg (3.2%)
C7: Monocyte/APC (3.8%)
C8: CD4 Th2 (2.5%)
C9: CD4 Tcm #2 (8.7%)
Legend for (b)
HIV- (99.9%)
HIV+ (0.1%)
c
Cluster
pDC/B 0 1069
NK/NKT/CD8 0 474
Monocyte/APC 7 5112
CD4 Treg 0 1650
CD4 Th2 1 1286
CD4 Tem/effector #2 12 4735
CD4 Tem/effector #1 5 2247
CD4 Tem CD127+ 1 7602
CD4 Tcm/Ttm 5 4000
CD4 Tcm #4 4 5740
CD4 Tcm #3 1 2713
CD4 Tcm #2 5 4444
CD4 Tcm #1 2 3876
CD4 cTfh 0 4416
B/Mix 0 406
APC/Other 1 1020
APC 0 506
0K 2K 4K 6K 8K
Number of cells
d
Cluster
Monocyte/APC 15.9%
CD4 Th2 2.3%
CD4 Tem/effector #2 27.3%
CD4 Tem/effector #1 11.4%
CD4 Tem CD127+ 2.3%
CD4 Tcm/Ttm 11.4%
CD4 Tcm #4 9.1%
CD4 Tcm #3 2.3%
CD4 Tcm #2 11.4%
CD4 Tcm #1 4.5%
APC/Other 2.3%
0.00 0.25 0.50 0.75 1.00
Proportion of HIV+ cells
e
Surface marker
CD305 (LAIR1)
CD71
CD5
CD3
CD47
HIV-
CD279 (PD-1)
CD195 (CCR5)
CD2
HIV+
0 1 2 3 4 5
π score
f
-log10(FDR)
1.0
0.5
0.0
-1 0 1 2
Mean difference in chromVAR bias-corrected deviations
Not significant

References

Abdel-Mohsen, M., Kuri-Cervantes, L., Grau-Exposito, J., Spivak, A.M., Nell, R.A., Tomescu, C., Vadrevu, S.K., Giron, L.B., Serra-Peinado, C., Genescà, M., et al. (2018). CD32 is expressed on cells with transcriptionally active HIV but does not enrich for HIV DNA in resting T cells. Sci. Transl. Med. *10*.

Abdel-Mohsen, M., Richman, D., Siliciano, R.F., Nussenzweig, M.C., Howell, B.J., Martinez-Picado, J., Chomont, N., Bar, K.J., Yu, X.G., Lichterfeld, M., et al. (2020). Recommendations for measuring HIV reservoir size in cure-directed clinical trials. Nat. Med. *26*, 1339–1350.

Alkhatib, G., Combadiere, C., Broder, C.C., Feng, Y., Kennedy, P.E., Murphy, P.M., and Berger, E.A. (1996). CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. Science *272*, 1955–1958.

Bar, K.J., Sneller, M.C., Harrison, L.J., Justement, J.S., Overton, E.T., Petrone, M.E., Salantes, D.B., Seamon, C.A., Scheinfeld, B., Kwan, R.W., et al. (2016). Effect of HIV Antibody VRC01 on Viral Rebound after Treatment Interruption. N. Engl. J. Med. *375*, 2037–2050.

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 Rep. *25*, 107–117.e3.

Brown, M.H., Boles, K., van der Merwe, P.A., Kumar, V., Mathew, P.A., and Barclay, A.N. (1998). 2B4, the natural killer and T cell immunoglobulin superfamily surface protein, is a ligand for CD48. J. Exp. Med. *188*, 2083–2090.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Cantero-Pérez, J., Grau-Expósito, J., Serra-Peinado, C., Rosero, D.A., Luque-Ballesteros, L., Astorga-Gamaza, A., Castellví, 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.

du Chéné, I., Basyuk, E., Lin, Y.-L., Triboulet, R., Knezevich, A., Chable-Bessia, C., Mettling, C., Baillat, V., Reynes, J., Corbeau, P., et al. (2007). Suv39H1 and HP1gamma are responsible for chromatin-mediated HIV-1 transcriptional silencing and post-integration latency. EMBO J. *26*, 424–435.

Chun, T.W., Finzi, D., Margolick, J., Chadwick, K., Schwartz, D., and Siliciano, R.F. (1995). In vivo fate of HIV-1-infected T cells: quantitative analysis of the transition to stable latency. Nat. Med. *1*, 1284–1290.

Chun, T.W., Carruth, L., Finzi, D., Shen, X., DiGiuseppe, J.A., Taylor, H., Hermankova, M., Chadwick, K., Margolick, J., Quinn, T.C., et al. (1997). Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. Nature *387*, 183–188.

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., et al. (2018). Clonal CD4+ T cells in the HIV-1 latent reservoir display a distinct gene profile upon reactivation. Nat. Med. *24*, 604–609.

Colomb, F., Giron, L.B., Kuri-Cervantes, L., Adeniji, O.S., Ma, T., Dweep, H., Battivelli, E., Verdin, E., Palmer, C.S., Tateno, H., et al. (2020). Sialyl-LewisX Glycoantigen Is Enriched on Cells with Persistent HIV Transcription during Therapy. Cell Rep. *32*, 107991.

Darcis, G., Kula, A., Bouchat, S., Fujinaga, K., Corazza, F., Ait-Ammar, A., Delacourt, N., Melard, A., Kabeya, K., Vanhulle, C., et al. (2015). An In-Depth Comparison of Latency-Reversing Agent Combinations in Various In Vitro and Ex Vivo HIV-1 Latency Models Identified Bryostatin-1+JQ1 and Ingenol-B+JQ1 to Potently Reactivate Viral Gene Expression. PLoS Pathog. *11*, e1005063.

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

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.e7.

Descours, B., Petitjean, G., López-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.

Duette, G., Hiener, B., Morgan, H., Mazur, F.G., Mathivanan, V., Horsburgh, B.A., Fisher, K., Tong, O., Lee, E., Ahn, H., et al. (2022). The HIV-1 proviral landscape reveals Nef contributes to HIV-1 persistence in effector memory CD4+ T-cells. J. Clin. Invest.

Einkauf, K.B., Osborn, M.R., Gao, C., Sun, W., Sun, X., Lian, X., Parsons, E.M., Gladkov, G.T., Seiger, K.W., Blackmer, J.E., et al. (2022). Parallel analysis of transcription, integration, and sequence of single HIV-1 proviruses. Cell.

El Hed, A., Khaitan, A., Kozhaya, L., Manel, N., Daskalakis, D., Borkowsky, W., Valentine, F., Littman, D.R., and Unutmaz, D. (2010). Susceptibility of human Th17 cells to human immunodeficiency virus and their perturbation during infection. J. Infect. Dis. *201*, 843–854.

Eriksson, S., Graf, E.H., Dahl, V., Strain, M.C., Yukl, S.A., Lysenko, E.S., Bosch, R.J., Lai, J., Chioma, S., Emad, F., et al. (2013). Comparative analysis of measures of viral reservoirs in HIV-1 eradication studies. PLoS Pathog. *9*, e1003174.

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. Nat. Med. *23*, 1271–1276.

Feng, H., Schorpp, K., Jin, J., Yozwiak, C.E., Hoffstrom, B.G., Decker, A.M., Rajbhandari, P., Stokes, M.E., Bender, H.G., Csuka, J.M., et al. (2020). Transferrin Receptor Is a Specific Ferroptosis Marker. Cell Rep. *30*, 3411–3423.e7.

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 *278*, 1295–1300.

Foletta, V.C., Segal, D.H., and Cohen, D.R. (1998). Transcriptional regulation in the immune

system: all roads lead to AP-1. J. Leukoc. Biol. *63*, 139–152.

Fourati, S., Ribeiro, S.P., Blasco Tavares Pereira Lopes, F., Talla, A., Lefebvre, F., Cameron, M., Kaewkungwal, J., Pitisuttithum, P., Nitayaphan, S., Rerks-Ngarm, S., et al. (2019). Integrated systems approach defines the antiviral pathways conferring protection by the RV144 HIV vaccine. Nat. Commun. *10*, 863.

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 Pathog. *12*, e1005761.

Golumbeanu, M., Cristinelli, S., Rato, S., Munoz, M., Cavassini, M., Beerenwinkel, N., and Ciuffi, A. (2018). Single-Cell RNA-Seq Reveals Transcriptional Heterogeneity in Latent and Reactivated HIV-Infected Cells. Cell Rep. *23*, 942–950.

Gosselin, A., Wiche Salinas, T.R., Planas, D., Wacleche, V.S., Zhang, Y., Fromentin, R., Chomont, N., Cohen, É.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.

Granja, J.M., Corces, M.R., Pierce, S.E., Bagdatli, S.T., Choudhry, H., Chang, H.Y., and Greenleaf, W.J. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nat. Genet. *53*, 403–411.

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

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29.

Heesters, B.A., Lindqvist, M., Vagefi, P.A., Scully, E.P., Schildberg, F.A., Altfeld, M., Walker, B.D., Kaufmann, D.E., and Carroll, M.C. (2015). Follicular Dendritic Cells Retain Infectious HIV in Cycling Endosomes. PLoS Pathog. *11*, e1005285.

Ho, Y.-C., Shan, L., Hosmane, N.N., Wang, J., Laskey, S.B., Rosenbloom, D.I.S., 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 Pathog. *14*, e1006856.

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., 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. Proc. Natl. Acad. Sci. U. S. A. *117*, 3704–3710.

Ise, W., Kohyama, M., Schraml, B.U., Zhang, T., Schwer, B., Basu, U., Alt, F.W., Tang, J., Oltz, E.M., Murphy, T.L., et al. (2011). The transcription factor BATF controls the global regulators of class-switch recombination in both B cells and T cells. Nat. Immunol. *12*, 536–543.

Jefferys, S.R., Burgos, S.D., Peterson, J.J., Selitsky, S.R., Turner, A.-M.W., James, L.I., Tsai, Y.-H., Coffey, A.R., Margolis, D.M., Parker, J., et al. (2021). Epigenomic characterization of latent HIV infection identifies latency regulating transcription factors. PLoS Pathog. *17*, e1009346.

Kinloch, N.N., Ren, Y., Conce Alberto, W.D., Dong, W., Khadka, P., Huang, S.H., Mota, T.M., Wilson, A., Shahid, A., Kirkby, D., et al. (2021). HIV-1 diversity considerations in the application of the Intact Proviral DNA Assay (IPDA). Nat. Commun. *12*, 165.

Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.-R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods *16*, 1289–1296.

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.e5.

Kuri-Cervantes, L., Pampena, M.B., and Betts, M.R. (2020). Phenotypic Characterization of SLex+ and CLA+ CD4+ T Cells. STAR Protoc *1*, 100154.

Leth, S., Schleimann, M.H., Nissen, S.K., Højen, J.F., Olesen, R., Graversen, M.E., Jørgensen, S., Kjær, A.S., Denton, P.W., Mørk, A., et al. (2016). Combined effect of Vacc-4x, recombinant human granulocyte macrophage colony-stimulating factor vaccination, and romidepsin on the HIV-1 reservoir (REDUC): a single-arm, phase 1B/2A trial. Lancet HIV *3*, e463–e472.

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*.

Lindqvist, B., Svensson Akusjärvi, S., Sönnerborg, A., Dimitriou, M., and Svensson, J.P. (2020). Chromatin maturation of the HIV-1 provirus in primary resting CD4+ T cells. PLoS Pathog. *16*, e1008264.

Liu, Y.Z., and Latchman, D.S. (1997). The octamer-binding proteins Oct-1 and Oct-2 repress the HIV long terminal repeat promoter and its transactivation by Tat. Biochem. J *322 (Pt 1)*, 155–158.

Liu, R., Yeh, Y.-H.J., Varabyou, A., Collora, J.A., Sherrill-Mix, S., Talbot, C.C., Jr, Mehta, S., Albrecht, K., Hao, H., Zhang, H., et al. (2020). Single-cell transcriptional landscapes reveal HIV-1-driven aberrant host gene transcription as a potential therapeutic target. Sci. Transl. Med. *12*.

Lun, A.T.L., Riesenfeld, S., Andrews, T., Dao, T.P., Gomes, T., participants in the 1st Human Cell Atlas Jamboree, and Marioni, J.C. (2019). EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. Genome Biol. *20*, 63.

Ma, J., Wang, R., Fang, X., and Sun, Z. (2012). β-catenin/TCF-1 pathway in T cell development and differentiation. J. Neuroimmune Pharmacol. *7*, 750–762.

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.e5.

McManus, W.R., Bale, M.J., Spindler, J., Wiegand, A., Musick, A., Patro, S.C., Sobolewski, M.D., Musick, V.K., Anderson, E.M., Cyktor, J.C., et al. (2019). HIV-1 in lymph nodes is maintained by cellular proliferation during antiretroviral therapy. J. Clin. Invest. *130*, 4629–4642.

Melsted, P., Booeshaghi, A.S., Liu, L., Gao, F., Lu, L., Min, K.H.J., da Veiga Beltrame, E., Hjörleifsson, K.E., Gehring, J., and Pachter, L. (2021). Modular, efficient and constant-memory single-cell RNA-seq preprocessing. Nat. Biotechnol. *39*, 813–818.

Mimitou, E.P., Lareau, C.A., Chen, K.Y., Zorzetto-Fernandes, A.L., Hao, Y., Takeshima, Y., Luo, W., Huang, T.-S., Yeung, B.Z., Papalexi, E., et al. (2021). Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. Nat. Biotechnol.

Neidleman, J., Luo, X., Frouard, J., Xie, G., Hsiao, F., Ma, T., Morcilla, V., Lee, A., Telwatte, S., Thomas, R., et al. (2020). Phenotypic analysis of the unstimulated in vivo HIV CD4 T cell reservoir. Elife *9*.

Nguyen, S., Deleage, C., Darko, S., Ransier, A., Truong, D.P., Agarwal, D., Japp, A.S., Wu, V.H., Kuri-Cervantes, L., Abdel-Mohsen, M., et al. (2019). Elite control of HIV is associated with distinct functional and transcriptional signatures in lymphoid tissue CD8(+) T cells. Sci. Transl. Med. *11*.

Pantaleo, G., Graziosi, C., Demarest, J.F., Butini, L., Montroni, M., Fox, C.H., Orenstein, J.M., Kotler, D.P., and Fauci, A.S. (1993). HIV infection is active and progressive in lymphoid tissue during the clinically latent stage of disease. Nature *362*, 355–358.

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 Pathog. *15*, e1007619.

Perelson, A.S., Neumann, A.U., Markowitz, M., Leonard, J.M., and Ho, D.D. (1996). HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. Science *271*, 1582–1586.

Perreau, M., Savoye, A.-L., De Crignis, E., Corpataux, J.-M., Cubas, R., Haddad, E.K., De Leval, L., Graziosi, C., and Pantaleo, G. (2013). Follicular helper T cells serve as the major CD4 T cell compartment for HIV-1 infection, replication, and production. J. Exp. Med. *210*, 143–156.

Rafati, H., Parra, M., Hakre, S., Moshkin, Y., Verdin, E., and Mahmoudi, T. (2011). Repressive LTR nucleosome positioning by the BAF complex is required for HIV latency. PLoS Biol. *9*, e1001206.

Reuse, S., Calao, M., Kabeya, K., Guiguen, A., Gatot, J.-S., Quivy, V., Vanhulle, C., Lamine, A., Vaira, D., Demonte, D., et al. (2009). Synergistic activation of HIV-1 expression by deacetylase inhibitors and prostratin: implications for treatment of latent infection. PLoS One *4*, e6093.

Salantes, D.B., Zheng, Y., Mampe, F., Srivastava, T., Beg, S., Lai, J., Li, J.Z., Tressler, R.L.,

Koup, R.A., Hoxie, J., et al. (2018). HIV-1 latent reservoir size and diversity are stable following brief treatment interruption. J. Clin. Invest. *128*, 3102–3115.

Schep, A.N., Wu, B., Buenrostro, J.D., and Greenleaf, W.J. (2017). chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat. Methods *14*, 975–978.

Schraml, B.U., Hildner, K., Ise, W., Lee, W.-L., Smith, W.A.-E., Solomon, B., Sahota, G., Sim, J., Mukasa, R., Cemerski, S., et al. (2009). The AP-1 transcription factor Batf controls T(H)17 differentiation. Nature *460*, 405–409.

Serra-Peinado, C., Grau-Expósito, 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.

Shan, L., Deng, K., Gao, H., Xing, S., Capoferri, A.A., Durand, C.M., Rabi, S.A., Laird, G.M., Kim, M., Hosmane, N.N., et al. (2017). Transcriptional Reprogramming during Effector-to-Memory Transition Renders CD4+ T Cells Permissive for Latent HIV-1 Infection. Immunity *47*, 766–775.e3.

Smith, B.A., Gartner, S., Liu, Y., Perelson, A.S., Stilianakis, N.I., Keele, B.F., Kerkering, T.M., Ferreira-Gonzalez, A., Szakal, A.K., Tew, J.G., et al. (2001). Persistence of infectious HIV on follicular dendritic cells. J. Immunol. *166*, 690–696.

Squair, J.W., Gautier, M., Kathe, C., Anderson, M.A., James, N.D., Hutson, T.H., Hudelle, R., Qaiser, T., Matson, K.J.E., Barraud, Q., et al. (2021). Confronting false discoveries in single-cell differential expression. Nat. Commun. *12*, 5692.

Sundrud, M.S., Vancompernolle, S.E., Eger, K.A., Bruno, T.C., Subramaniam, A., Mummidi, S., Ahuja, S.K., and Unutmaz, D. (2005). Transcription factor GATA-1 potently represses the expression of the HIV-1 coreceptor CCR5 in human T cells and dendritic cells. Blood *106*, 3440–3448.

Swanson, E., Lord, C., Reading, J., Heubeck, A.T., Genge, P.C., Thomson, Z., Weiss, M.D.A., Li, X.-J., Savage, A.K., Green, R.R., et al. (2021). Simultaneous trimodal single-cell measurement of transcripts, epitopes, and chromatin accessibility using TEA-seq. Elife *10*.

Thibodeau, A., Eroglu, A., McGinnis, C.S., Lawlor, N., Nehar-Belaid, D., Kursawe, R., Marches, R., Conrad, D.N., Kuchel, G.A., Gartner, Z.J., et al. (2021). AMULET: a novel read count-based method for effective multiplet detection from single nucleus ATAC-seq data. Genome Biol. *22*, 252.

Timmons, A., Fray, E., Kumar, M., Wu, F., Dai, W., Bullen, C.K., Kim, P., Hetzel, C., Yang, C., Beg, S., et al. (2020). HSF1 inhibition attenuates HIV-1 latency reversal mediated by several candidate LRAs In Vitro and Ex Vivo. Proc. Natl. Acad. Sci. U. S. A. *117*, 15763–15771.

Valiante, N.M., and Trinchieri, G. (1993). Identification of a novel signal transduction surface molecule on human cytotoxic lymphocytes. J. Exp. Med. *178*, 1397–1406.

Van Lint, C., Ghysdael, J., Paras, P., Jr, Burny, A., and Verdin, E. (1994). A transcriptional regulatory element is associated with a nuclease-hypersensitive site in the pol gene of human

immunodeficiency virus type 1. J. Virol. *68*, 2632–2648.

Wang, W., Fasolino, M., Cattau, B., Goldman, N., Kong, W., Frederick, M.A., McCright, S.J., Kiani, K., Fraietta, J.A., and Vahedi, G. (2020). Joint profiling of chromatin accessibility and CAR-T integration site analysis at population and single-cell levels. Proc. Natl. Acad. Sci. U. S. A. *117*, 5442–5452.

Ward, J., Bonaparte, M., Sacks, J., Guterman, J., Fogli, M., Mavilio, D., and Barker, E. (2007). HIV modulates the expression of ligands important in triggering natural killer cell cytotoxic responses on infected primary T-cell blasts. Blood *110*, 1207–1214.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer).

Wong, J.K., Günthard, H.F., Havlir, D.V., Zhang, Z.Q., Haase, A.T., Ignacio, C.C., Kwok, S., Emini, E., and Richman, D.D. (1997). Reduction of HIV-1 in blood and lymph nodes following potent antiretroviral therapy and the virologic correlates of treatment failure. Proc. Natl. Acad. Sci. U. S. A. *94*, 12574–12579.

Xiao, Y., Hsiao, T.-H., Suresh, U., Chen, H.-I.H., Wu, X., Wolf, S.E., and Chen, Y. (2014). A novel significance score for gene selection and ranking. Bioinformatics *30*, 801–807.

Younes, S.-A., Talla, A., Pereira Ribeiro, S., Saidakova, E.V., Korolevskaya, L.B., Shmagel, K.V., Shive, C.L., Freeman, M.L., Panigrahi, S., Zweig, S., et al. (2018). Cycling CD4+ T cells in HIV-infected immune nonresponders have mitochondrial dysfunction. J. Clin. Invest. *128*, 5083–5094.

Zack, J.A., Arrigo, S.J., Weitsman, S.R., Go, A.S., Haislip, A., and Chen, I.S. (1990). HIV-1 entry into quiescent primary lymphocytes: molecular analysis reveals a labile, latent viral structure. Cell *61*, 213–222.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zhang, Z., Schuler, T., Zupancic, M., Wietgrefe, S., Staskus, K.A., Reimann, K.A., Reinhart, T.A., Rogan, M., Cavert, W., Miller, C.J., et al. (1999). Sexual transmission and propagation of SIV and HIV in resting and activated CD4+ T cells. Science *286*, 1353–1357.

Zhou, H., Xu, M., Huang, Q., Gates, A.T., Zhang, X.D., Castle, J.C., Stec, E., Ferrer, M., Strulovici, B., Hazuda, D.J., et al. (2008). Genome-scale RNAi screen for host factors required for HIV replication. Cell Host Microbe *4*, 495–504.