

Single-cell immune profiling reveals the impact of antiretroviral therapy on HIV-1-induced immune dysfunction, T cell clonal expansion, and HIV-1 persistence *in vivo*

Jack A. Collora¹, Delia Pinto-Santini², Siavash Pasalar², Neal Ravindra^{3,4}, Carmela Ganoza⁵, Javier Lama⁵, Ricardo Alfaro⁶, Jennifer Chiarella⁷, Serena Spudich⁷, David van Dijk^{3,4}, Ann Duerr², Ya-Chi Ho^{1*}

¹Department of Microbial Pathogenesis, Yale University School of Medicine, New Haven, CT 06519, USA

²Vaccine and Infectious Disease & Public Health Science Divisions, Fred Hutchinson Cancer Research Center, Seattle, MA, 98109, USA

³Department of Internal Medicine (Cardiology), Yale School of Medicine, New Haven, CT 06520, USA

⁴Department of Computer Science, Yale University, New Haven, CT 06520, USA

⁵Asociación Civil Impacta Salud y Educación, Lima, 15063, Perú

⁶Centro de Investigaciones Tecnológicas Biomédicas y Medioambientales (CITBM), Lima, 07006, Perú

⁷Department of Neurology, Yale University School of Medicine, New Haven, CT 06519, USA

*Correspondence: ya-chi.ho@yale.edu

Abstract

Despite antiretroviral therapy (ART), HIV-1 persists in proliferating T cell clones that increase over time. To understand whether early ART affects HIV-1 persistence *in vivo*, we performed single-cell ECCITE-seq and profiled 89,279 CD4⁺ T cells in paired samples during viremia and after immediate versus delayed ART in six people in the randomized interventional Sabes study. We found that immediate ART partially reverted TNF responses while delayed ART did not. Antigen and TNF responses persisted despite immediate ART and shaped the transcriptional landscape of CD4⁺ T cells, HIV-1 RNA⁺ cells, and T cell clones harboring them (clone_{HIV-1}⁺). Some HIV-1 RNA⁺ cells reside in the most clonally expanded cytotoxic T cell populations (GZMB and GZMK Th1 cells). Clone_{HIV-1}⁺ were larger in clone size, persisted despite ART, and exhibited transcriptional signatures of antigen, cytotoxic effector, and cytokine responses. Using machine-learning algorithms, we identified markers for HIV-1 RNA⁺ cells and clone_{HIV-1}⁺ as potential therapeutic targets. Overall, by combining single-cell immune profiling and T cell expansion dynamics tracking, we identified drivers of HIV-1 persistence *in vivo*.

Despite effective antiretroviral therapy (ART), HIV-1-infected CD4⁺ T cells not only persist lifelong^{1,2} but also proliferate *in vivo*³⁻⁷. More than 50% of the HIV-1 latent reservoir is maintained by clonal expansion of infected cells³⁻⁵. The proliferation of HIV-1-infected cells is a major barrier to cure. The proliferation of HIV-1-infected cells is driven by antigen stimulation⁸⁻¹¹, homeostatic proliferation¹², and HIV-1-driven proliferation gene expression^{6,7,13}. Understanding the immune programs promoting the clonal expansion of HIV-1-infected cells is critical for stopping the proliferation of HIV-1-infected cells.

ART suppresses HIV-1 plasma viral load to clinically undetectable levels. Early ART initiated within 6 months of infection, as opposed to delayed ART initiated after 6 months of infection, significantly reduces chronic immune activation¹⁴. However, early ART may not fully restore immune function to the levels in uninfected individuals¹⁵⁻¹⁷. ART neither kills infected cells nor inhibits HIV-1 viral protein production from the existing infected cells. Therefore, HIV-1-infected cells can continue to produce viral antigens^{18,19} and induce immune activation and exhaustion^{19,20}. We postulate that both antigen stimulation and systemic immune activation promote the proliferation of HIV-1-infected cells. Given that people treated with immediate ART have a shorter duration of HIV-1 viremia, shorter duration of overt HIV-1 antigen exposure, and lower levels of immune activation¹⁴, we test our hypothesis by comparing the immune programs of CD4⁺ T cells, HIV-1-infected cells, and T cell clonal expansion dynamics in people treated with immediate ART versus delayed ART. In particular, we want to examine the drivers of HIV-1 persistence both for the HIV-1-infected cells and the T cell clones harboring them, given that T cells having the same T cell receptor (TCR) sequence respond to the same antigen stimulation and immune activation. We propose to identify dysregulated immune programs that promote the persistence of HIV-1-infected cells and identify cellular markers of HIV-1 RNA⁺ cells and T cell clones harboring them for therapeutic targeting.

Several challenges prevent mechanistic understanding of HIV-1 persistence in people with HIV-1. First, inducible HIV-1-infected cells in ART-treated individuals are extremely rare, accounting for <0.001% of CD4⁺ T cells in the peripheral blood²¹⁻²³. Second, there are no known cellular markers that exclusively select for heterogeneous HIV-1-infected T cells, given the differences²⁴ in polarization²⁵, activation state²⁶, memory differentiation²⁷⁻²⁹, and immune exhaustion states³⁰. Third, while HIV-1-infected cells can be isolated using viral markers such as HIV-1 RNA expression¹³ or Env protein expression³¹, these methods require *ex vivo* activation and thus cannot evaluate the *in vivo* status of infected cells. Advancements in single-cell technologies enabled high-dimensional immune profiling to dissect the heterogeneous states of immune cells, high-resolution capture of rare cells, T cell clonality, and identification of upstream drivers of immune dysregulation³²⁻³⁷. Further, computational techniques including supervised and unsupervised machine learning, network analysis, and statistical methodologies enable confident identification of higher-fidelity predictors of different cellular states from the sparse and highly complex single-cell multi-modal data³⁸⁻⁴³.

Here, we combined a translational approach using paired and longitudinally archived blood samples from a prospective, randomized, and interventional Sabes study, single-cell multi-omics immune profiling, and machine learning analysis to capture HIV-1-infected cells without *ex vivo* activation. We tracked T cell clonal expansion dynamics over time, identified T cells harboring the rare HIV-1-infected cells, identified immune programs driving HIV-1 persistence, and identified cellular markers for HIV-1 RNA⁺ cells and T cell clones harboring them as therapeutic targets.

Results

Single-cell multi-omics of CD4⁺ T cells from the randomized and interventional Sabes study identifies the impact of immediate versus delayed ART on HIV-1-induced immune dysfunction

To understand why people treated with immediate versus delayed ART have different immune dysfunction¹⁵⁻¹⁷, we compared the immune profiles and T cell clonal expansion dynamics of CD4⁺ T cells from people with acute HIV-1 infection treated with immediate versus delayed ART. We used longitudinally archived blood samples from the Sabes Study in which HIV-1 infection was prospectively tested monthly for early detection and treatment stratification^{44,45}. Using detailed histories of HIV-1 testing, including date, type and result of test, we calculated the estimated date of detectable infection (EDDI) using an established algorithm⁴⁶. We examined the single-cell transcriptional landscape of six people with acute HIV-1 infection treated either with immediate ART (three individuals receiving ART 22 – 47 days after EDDI) versus delayed ART (three individuals receiving ART 187 – 207 days after EDDI) (Fig. 1a, Supplementary Fig. 1, Supplementary Table 1). Paired blood samples during viremia (20 – 48 days after EDDI) and after one year of ART (172 – 384 days of suppression) from each of the

six individuals were profiled to examine the cellular transcriptional landscape. Two biological sex-matched uninfected individuals were recruited for comparison.

To characterize the heterogeneous functional states of CD4⁺ T cells, to track T cell clonal expansion dynamics, and to identify the rare HIV-1-infected cells, we used Expanded CRISPR compatible Cellular Indexing of Transcriptomes and Epitopes by Sequencing (ECCITE-seq⁴⁷) to profile the cellular transcriptome, surface protein markers, TCR sequence, and HIV-1 RNA within the same single cells (Fig. 1b). A total of 52,473 CD4⁺ T cells during viremia, 36,806 CD4⁺ T cells during suppressive ART, and 33,406 CD4⁺ T cells from uninfected individuals were examined by single-cell ECCITE-seq (Supplementary Table 2). Batch effects were removed using Seurat⁴² integration, as shown in dimensionality reduction plots by Potential of Heat-diffusion for Affinity-based Trajectory Embedding (PHATE)⁴⁸(Supplementary Fig. 2). The single-cell transcriptional landscape of all 122,685 cells [having a median of 4,144 unique molecular identifiers (UMI) and 1,345 genes per cell] were visualized using PHATE⁴⁸ (Fig. 1c)(Supplementary Table 3).

We identified 14 transcriptome defined clusters. The inferred CD4⁺ T cell phenotypes was determined based on transcriptional markers and surface protein expression (Fig. 1c, 1d, Supplementary Fig. 3 for gene expression violin plots for each cluster, Supplementary Table 4 for genes and proteins that infer the phenotype of each cluster, and Supplementary Table 5 for a list of differentially expressed genes). These included as one interferon (IFN) response cluster, three Th1 clusters, one Th17 cluster, two Treg clusters, one proliferating cell cluster, two naïve cell clusters, one *KLRB1* (CD161) cluster, and one CXCR5 (homing) cluster. Using surface CD45RA and CCR7 protein expression (determined by >90% expression levels of isotype ECCITE-seq antibody controls), we examined the T cell memory phenotype (naïve, central memory, effector memory, and effector)⁴⁹ of each single cell (Fig. 1e). These clusters showed distinct CD4⁺ T cell effector profiles (Fig. 1f, Supplementary Fig. 4). Importantly, in addition to classic Th1 cells expressing high levels of *TBX21* (Tbet), *IFNG*, *TNF* (tumor necrosis factor), *GZMB* (granzyme B), and *CCL5* (RANTES), we found that the largest population of Th1 cells was the GZMK Th1 cell subset expressing high levels of *GZMK* (granzyme K), *TNF*, and *CCL5* but not *GZMB*. The cytotoxic CD4⁺ T cells expressing granzyme K, previously under-appreciated in HIV-1-infection, have been shown to be important anti-tumor effectors in bladder cancer⁵⁰ and in SARS-CoV-2 infection⁵¹. In order to identify the underlying phenotype of proliferating cells that can be masked by cell cycle-related gene signatures, we performed sub-clustering of the proliferating cell cluster. Sub-clustering of proliferating cells revealed a subset of cells in viremic and suppressed conditions with increased *GZMK* expression which was absent in uninfected conditions (Supplementary Fig. 5). This indicates that a subset of *GZMK* expressing cells expressed proliferation and cell-cycle related genes and may have high proliferation potential.

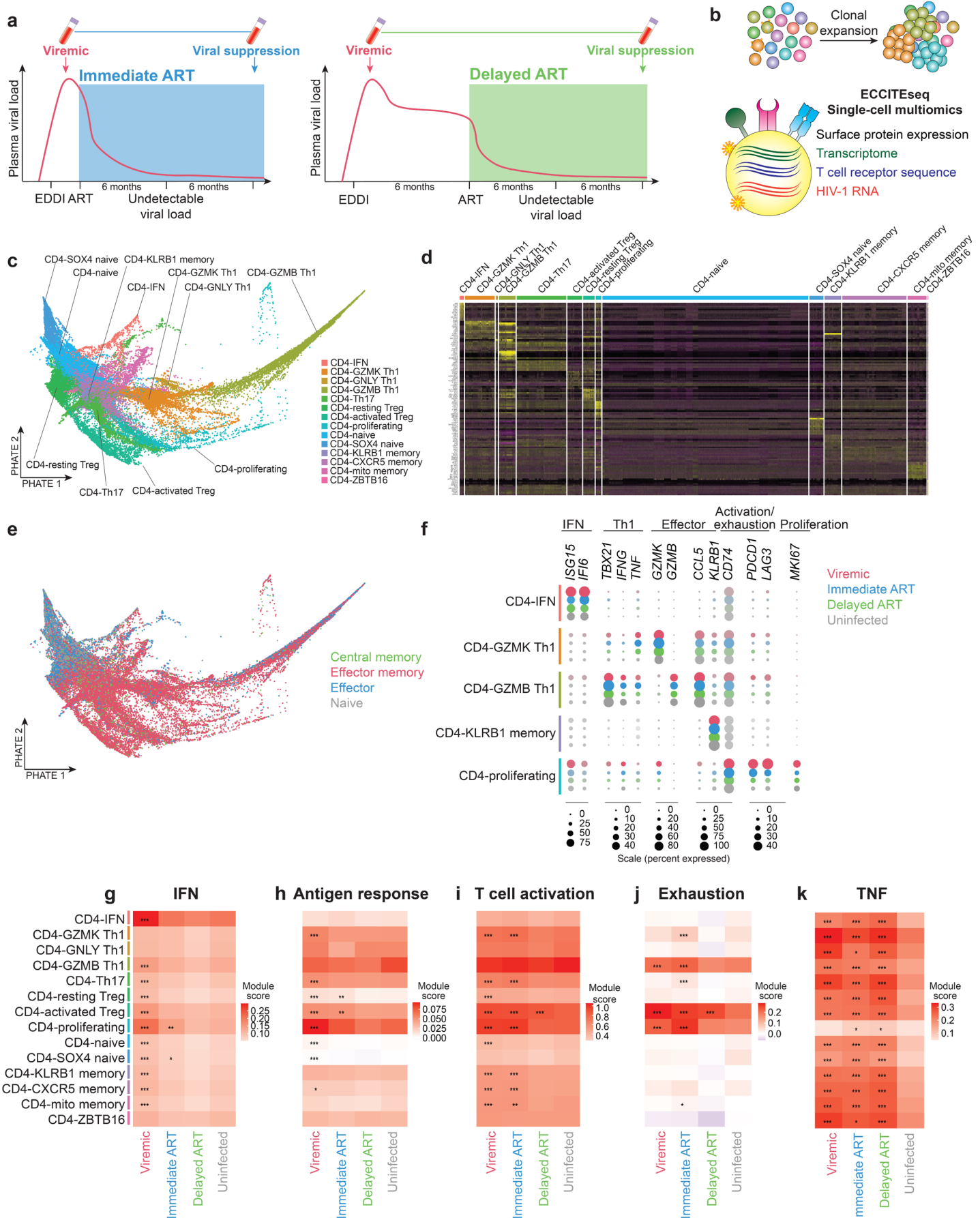


Fig. 1. Single-cell immune profiling of CD4⁺ T cells of paired longitudinal blood samples during viremia and after immediate versus delayed initiation of suppressive ART. **a**, Study design. Participants enrolled the Sabes study were prospectively tested monthly for HIV-1 infection using antibody detection, antigen detection, and viral RNA quantification. Paired CD4⁺ T

cells during acute viremia and after suppressive ART from six people with HIV-1 enrolled in Sabes and subsequently followed in the Merlin study were used for immune profiling. Among them, three individuals received immediate ART and three individuals received delayed ART (24 weeks after EDDI). CD4⁺ T cells from two sex-matched uninfected individuals were used as controls. **b**, Single-cell ECCITE-seq captures surface protein expression, cellular transcriptome ($n = 52,473, 13,569, 23,237,$ and $33,406$ cells in the viremic, immediate ART, delayed ART, and uninfected conditions respectively), HIV-1 RNA, and TCR sequence within the same single cell, allowing clonal expansion dynamics tracking in longitudinally archived samples. **c**, PHATE plot of all 122,685 CD4⁺ T cells from HIV-1-infected and uninfected individuals depicts 14 clusters of CD4⁺ T cells defined by transcriptome. Clusters were labeled with inferred cell type based on transcriptome signatures and surface protein expression. **d**, Heatmap of top 10 differentially expressed genes in each cluster. Heat map marker genes were identified by Wilcoxon rank-sum test with a cutoff of 0.25 average log fold change and an adjusted P value cutoff of 0.05. **e**, PHATE plot of memory phenotypes defined by surface CD45RA and CCR7 expression. CD45RA and CCR7 positivity was determined by >90th percentile of the expression level of isotype barcoded surface protein staining antibody controls. **f**, The expression profile of genes involved in type I IFN responses, Th1 polarization, cytotoxic effector functions, T cell activation, and exhaustion phenotypes of CD4⁺ T cell clusters of interest. Color intensity indicates level of expression. Dot size indicates the percentage of cells expressing transcript. **g–k**, Mean module scores of predefined gene sets involved in different immune responses during viremia, viral suppression after immediate ART, viral suppression after delayed ART, and uninfected conditions. **g**, IFN response module was defined by top 30 genes correlating with *ISG15*. **h**, Antigen response module was defined by Goldrath antigen response⁵². **i**, T cell activation module was defined by top 30 genes correlating with *CD74* (HLA-DR antigens-associated invariant chain) expression. **j**, Exhaustion module was defined by top 30 genes correlated with *PDCD1* (PD-1) expression. **k**, TNF response module was defined by Phong TNF response⁵³. P values were determined by Wilcoxon rank-sum test comparing HIV-1 infected conditions and uninfected conditions. *, $P < 0.01$; **, $P < 0.001$; ***, $P < 0.0001$.

Immediate ART limits TNF responses but not ongoing antigen response in CD4⁺ T cells

To identify differences of HIV-1-induced immune dysregulation during viremia, and suppression after immediate vs delayed treatment initiation, we examined T cell functional states in each T cell subset by comparing the gene expression profile with pre-defined gene modules, or group of functionally related and co-regulated genes (Fig. 1g – 1k). Using gene set module scoring⁵⁴, we found that IFN responses, particularly in CD4-IFN cells, were upregulated during viremia and returned to baseline after viral suppression (Fig. 1g). Antigen responses and T cell responses peaked during viremia, reflecting overt antigen exposure and T cell activation during acute infection (Fig. 1h, 1i). However, immune exhaustion persisted in some subsets of cells (Fig. 1j), while TNF responses (Fig. 1k) continued in all subsets of cells despite immediate ART or delayed ART.

In addition to using pre-defined gene sets of immune regulatory pathways, we wanted to identify co-regulated genes that can distinguish the immune responses between viremia, immediate ART, delayed ART, and uninfected conditions using *de novo* gene set identification. Using weighted correlation network analysis (WGCNA), we identified 39 modules across individuals which were merged into eight consensus modules based on gene overlap (Supplementary Fig. 6). From eight consensus gene modules we found three modules which differentiated between different conditions across all participants: IFN responses (Fig. 2a–2d), antigen responses (Fig. 2e–2h), and TNF responses (Fig. 2i–2l)(see Supplementary Table 6 for the gene features of each module). The module names of these *de novo* identified gene sets were determined by gene ontology or pathway enrichment identified by Enrichr^{55,56}. We compared gene-gene correlation levels between different conditions (Fig. 2a, 2e, 2f, and Supplementary Fig. 7) and examined the module scores on PHATE (Supplementary Fig. 8) and in CD4⁺ T cell clusters that were most affected (Fig. 2b, 2f, 2i, Supplementary Fig. 9). In these gene modules, we identified the enriched pathways of these co-regulated genes (Fig. 2c, 2g, 2k) and upstream regulators of these genes (Fig. 2d, 2h, 2l) using Ingenuity Pathway Analysis (IPA)⁵⁷.

The IFN response module was characterized by 25 IFN-stimulated genes (ISGs) such as *ISG15*, *IFI6*, *IFI44L*, *IRF7*, *MX1*, *IFITM2*, *IFITM3*, *STAT1*, *OAS1*, *OAS3*, and *XAF1* (Fig. 2a). Consistent with analysis using pre-defined IFN gene sets (Fig. 1g), the upregulation of IFN response module (Fig. 2a) returned to baseline after one year of suppressive ART (Fig. 2b, Supplementary Fig. 8a, Supplementary Fig. 9a). These IFN-stimulated genes were upregulated in acute HIV-1 viremia (Fig. 2b) but also in hyperacute HIV-1 viremia³⁴, SIV infection⁵⁸, SARS-CoV-2 infection⁵⁹, and influenza infection⁶⁰, reflecting a shared type I IFN response to acute viral infections. The pathways and upstream regulators are IFN signaling and type I IFNs (Fig. 2c and 2d). Overall, type I IFN responses were induced upon acute HIV-1 infection and returned to baseline after suppressive ART.

The antigen response module included genes involved in T cell exhaustion (*TOX*, *TOX2*, *PRDM1* (encoding BLIMP-1), *LAG3*, and *TIGIT*), MHC class II (*CD74* (encoding MHCII invariant chain), *HLA-DRB1*, and *HLA-DRB5*), *ICOS*, and cytotoxic effector molecules (*GZMA*, *GZMH*, and *PFN1*) (Fig. 2e). This module of 439 genes provides a broader scope of antigen responses and T cell activation than the predefined gene sets (Fig. 1h, 1i). We found that immediate ART slightly reduced antigen responses and T cell activation during HIV-1 infection compared with delayed ART but did not restore antigen responses and T cell activation to levels in uninfected individuals (Fig. 2f, Supplementary Fig. 8b, Supplementary Fig. 9b). The immune pathways involved in this module include Th1 and Th2 activation, T cell differentiation, antigen presentation, and immune exhaustion (Fig. 2g). The upstream regulators driving this persistent T cell activation involved cytokine dysregulation [interleukin-2 (IL-2), TNF, IFN γ , IL-27, tumor growth factor- β (TGF β), IL-15], lipopolysaccharide (LPS), and TCR stimulation.

The TNF response module included *DUSP1*, *DUSP2*, *JUN*, *JUNB*, *FOS*, *FOSB*, *NFKBIA*, *ZFP36*, and *BHLHE40* (Fig. 2i). These TNF response genes were upregulated during viremia and persisted during delayed viral suppression (Fig. 2j, Supplementary Fig. 3c). We found that immediate ART significantly reduced TNF responses relative to delayed ART, but not to levels in uninfected individuals (Fig. 2i, 2j, Supplementary Fig. 8c, Supplementary Fig. 9c). The immune pathways involved in this module included glucocorticoid receptor signaling, TNF receptor signaling, OX40 signaling, IL-10 signaling, and JAK/STAT signaling (Fig. 2k). The upstream cytokine regulators driving persistent TNF responses included IL-1 β , TNF, and TGF β (Fig. 2j). Of note, some TNF-response genes (*DUSP1*, *FOS*, *FOSB*, *JUN*, *JUNB*, *NFKBIA*) were reported to be upregulated in coronavirus disease 2019 (COVID-19) patients⁶¹, and in severe COVID-19 but not in severe influenza infection⁶⁰. Further upregulation of TNF/IL-1 β -driven inflammatory response was found in hyperacute HIV-1 infection³⁴ and in chronically infected untreated HIV-1 infected individuals⁶². Overall, we found that immediate ART slightly reduced antigen responses and T cell activation and significantly reduced TNF responses compared to delayed ART, although this effect is partial. People with HIV-1, despite immediate or delayed ART, may continue to have TNF responses and chronic inflammation.

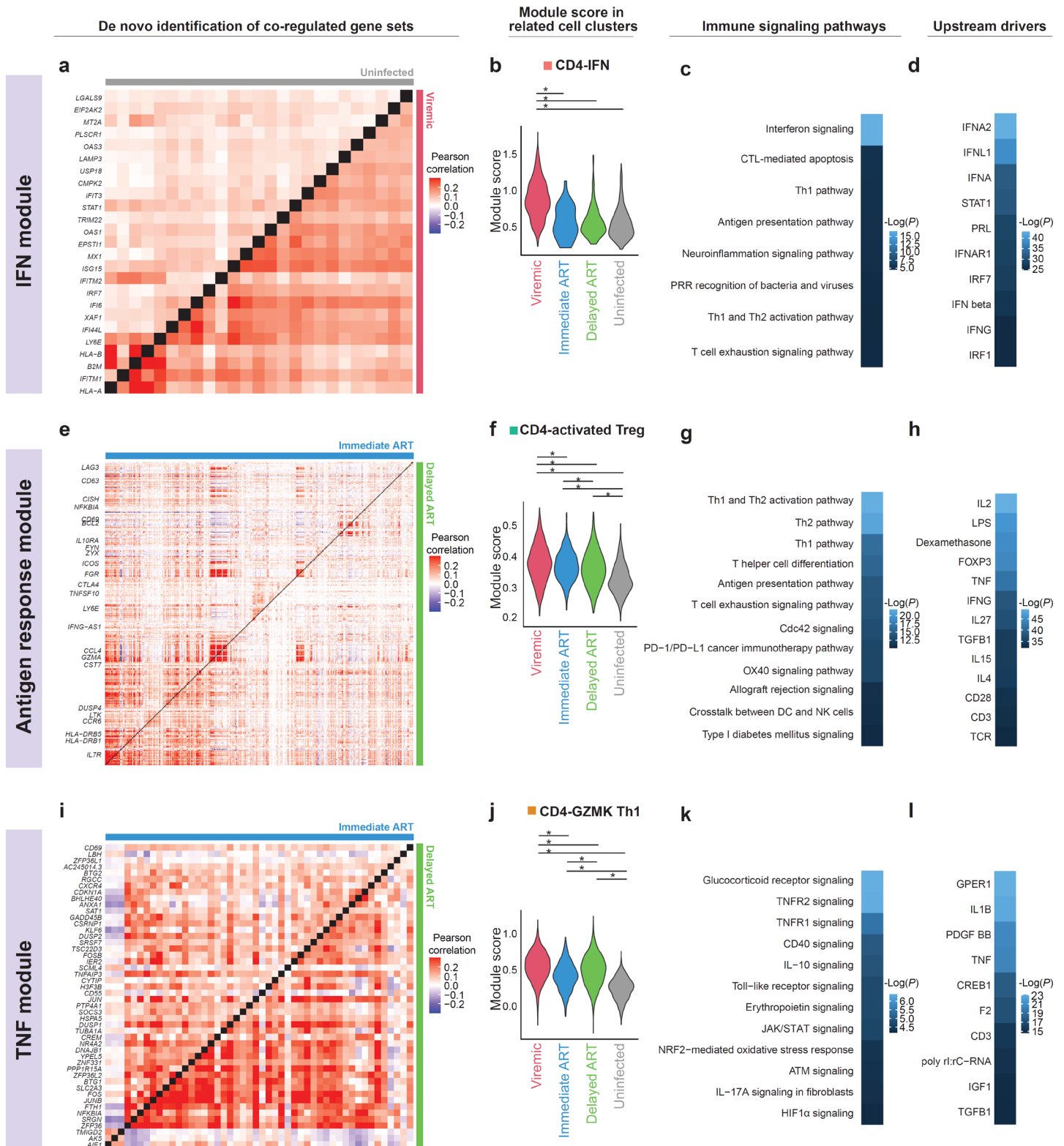


Fig. 2. *De novo* identification of co-regulated gene signatures of CD4⁺ T cells during viremia, after immediate ART suppression, after delayed ART suppression, and in uninfected conditions. Weighted correlation network analysis (WGCNA) on all T cells identified a total of three consensus gene modules across all participants that can differentiate between the four conditions. These three gene modules identified *de novo* were IFN responses (a – d, 25 genes), antigen responses (e – h, 348 genes), and TNF responses (i – l, 48 genes). **a, e, i,** Heatmaps of Pearson correlation coefficients of gene expression levels between different conditions. **b, f, j,** Violin plots of module score in clusters showing most prominent differences, such as IFN module score in CD4-IFN cells (n = 1,133 cells), antigen response module score in CD4-activated Tregs (n = 3,073 cells), and TNF module score in GZMK Th1 cells (n = 7,973). *P* values were determined by Wilcoxon rank-sum test. **c, g, k,** Immune pathways significantly enriched in respective modules. **d, h, l,** Upstream regulators driving the transcriptional landscape in respective modules. Immune pathways and upstream regulators were identified using IPA. *P* values were calculated by Fisher's exact test. *, *P* < 0.05.

HIV-1 RNA⁺ cells have diverse T cell phenotypes

To determine the transcriptional landscape of HIV-1-infected cells, we mapped the transcriptome to autologous HIV-1 genomes and identified 90 HIV-1 RNA⁺ cells, including 81 from viremic samples and 9 from ART-suppressed samples (Fig. 3a). The low number of HIV-1 RNA⁺ cells identified during viremia (0.15%) and after suppressive ART (0.02%) was consistent with the low frequency of HIV-1-infected cells. In these HIV-1 RNA⁺ cells, there were comparable levels of HIV-1 RNA transcripts (normalized to cellular transcripts) in each single cell in viremic samples versus ART-suppressed samples (Supplementary Fig. 10), suggesting that HIV-1-infected cells expressed detectable levels of HIV-1 RNA *in vivo* despite suppressive ART⁶³. Of note, cells that are negative for HIV-1 RNA are not necessarily uninfected cells – HIV-1 infected cells in latency or cells harboring transcription-defective HIV-1 may not express HIV-1 RNA. Given the low frequency of total HIV-1 DNA levels in CD4⁺ T cells (<0.1%)^{22,23}, >99% of cells that do not have detectable levels of HIV-1 RNA are presumably uninfected. Although HIV-1 RNA positivity does not equal replication competence, these are induced HIV-1-infected cells that are actively producing viral products *in vivo*. There were zero reads mapped to HIV-1 genomes in uninfected individuals. We first examined the normalized frequency of HIV-1 RNA⁺ cells in respect to the size of the respective cluster. We found that a median of 0.42% of GZMK Th1 cells harbored HIV-1 RNA⁺ cells, followed by KLRB1 cells (0.40%), GZMB Th1 cells (0.25%), Th17 cells (0.22%), mitochondria-high cells (0.21%), activated Tregs (0.15%), CXCR5 cells (0.06%), and naïve cells (0.03%)(Fig. 3b). We next examined the number of HIV-1 RNA⁺ cells in each cluster regardless of the size of the cluster. We found that HIV-1 RNA⁺ cells resided in GZMK Th1 cells (14/90 cells, 15.6%), Th17 cells (14/90 cells, 15.6%), CXCR5 memory cells (11/90 cells, 12.2%), naïve cells (10/90 cells, 11.1%), and KLRB1 memory cells (10/90 cells, 11.1%)(see also Supplementary Fig. 11). As previously reported, *KLRB1* (CD161) cells⁶⁴, Th1 cells²⁵, Th17 cells⁶⁵, CXCR5⁺ cells^{65,66}, and naïve cells²⁹ may harbor HIV-1-infected cells, although we found that naïve cells were significantly dis-enriched for HIV-1 RNA⁺ cells (Supplementary Fig. 11a). Overall, we identified GZMK Th1 cells as a previously unappreciated population that harbors HIV-1 RNA⁺ cells.

HIV-1 RNA⁺ cells are enriched in exhausted and effector memory cells during viremia

Using DNA-barcoded surface protein staining in ECCITE-seq, we examined the memory T cell subsets that were enriched with HIV-1 RNA⁺ cells. We found that the majority of HIV-1 RNA⁺ cells were effector memory cells (CD45RA⁻ CCR7⁻)²⁸ during viremia (67/81 cells, 82.7%) and under suppressive ART (7/9 cells, 77.8%), although central memory cells, effector cells, and naïve cells also harbored HIV-1 RNA⁺ cells. The majority of HIV-1 RNA⁺ cells express activation markers during viremia (49/81 cells, 60.5%), but not under suppressive ART (0/9 cells, 0%). We found that around half of HIV-1 RNA⁺ cells (48/81 cells, 59.3%) expressed exhaustion markers (either PD-1 or TIGIT) during viremia and during suppressive ART (4/9, 44.4%). We next examined the proportion of HIV-1 RNA⁺ cells in each subset. We found that effector memory cells, activated cells (expressing CD25 or HLA-DR), and exhausted cells (expressing PD-1 or TIGIT) harbored more HIV-1 RNA⁺ cells during viremia (Fig. 3c–3e). When we compared the proportion of HIV-1 RNA⁺ cells versus HIV-1 RNA⁻ cells across T cell subsets, we found that HIV-1 RNA⁺ cells were enriched in effector memory cells, dis-enriched in naïve cells during viremia ($P < 0.05$)(Supplementary Fig. 11b), and enriched in exhausted T cells ($P < 0.05$)(Supplementary Fig. 11d). There is a trend of higher frequency of HIV-1 RNA⁺ cells in activated cells but it did not reach statistical significance (Supplementary Fig. 11c). Consistent with previous studies, we showed that effector memory cell markers²⁸, T cell activation markers²⁶, and immune exhaustion markers^{30,31} may enrich for a subset of HIV-1-infected cells during viremia.

TNF and homeostasis cytokine responses shape the transcriptional landscape of HIV-1 RNA⁺ cells during viral suppression

We next examined the gene expression profile of HIV-1 RNA⁺ cells relative to HIV-1 RNA⁻ cells at the same time point both during viremia and after viral suppression. We ranked genes using expression fold change and conducted gene set enrichment analysis (GSEA)⁶⁷ with predefined gene sets⁶⁸. HIV-1 RNA⁺ cells upregulated viral infection-related genes previously identified in the context of HIV-1 infection and hepatitis B virus infection, including *IFI16*, *IFI27*, *STAT1*, *GZMK*, *GZMA*, *CCL5*, *CCR5*, *CD28*, *PSMB8*, and *PSMB9* (Fig. 3f, 3g). After ART suppression, the transcriptome of HIV-1 RNA⁺ cells were dominated by cytokine and cytokine receptor interactions, such as cytokines (*TNF*, *TGFB1*, *TNFSF13B*, *TNFSF14*, *IL7*, and *IL15*) and cytokine receptors (*IL10RA*, *IL21R*, *IL2RB*, *IL15RA*, *IL21R*, *IFNAR2*, and *IFNGR1*) (Fig. 3h), suggesting that cytokine dysregulation (such as TNF and TGF β , Fig. 2k, 2l) and homeostasis common γ chain cytokines (such as IL7 and IL15) during suppressive ART may shape the transcriptome of HIV-1 RNA⁺ cells.

To identify cellular markers that consistently distinguish HIV-1 RNA⁺ cells from HIV-1 RNA⁻ cells from the same individual, we employed a statistical bootstrapping approach to compare HIV-1 RNA⁺ cells to cluster and individual matched pools of HIV-1 RNA⁻ cells repeatedly for 10,000 times. We identified a total of 22 markers (Fig 3i–t, Supplementary Fig. 12, Supplementary Table 7), including molecules related to cytotoxic effectors *GZMK*, *PRF1* (perforin), and *CST7* (Fig 3i-k). Further, we identified *CFLAR* (c-FLIP)(Fig 3l), a caspase-8 inhibitor that provides resistance to apoptosis, that can be upregulated by Tat expression⁶⁹. We also identified MHC II molecules (*CD74*, *HLA-DRB1*, *HLA-DRB5*, and *HLA-DPA1*) (Fig 3m–n, Supplemental Fig. 12 b–d), *CLIC1* (Fig. 3o)(NCC27, a chloride channel involved in cell cycle regulation⁷⁰), *NEAT1* (Fig. 3p)(a long noncoding RNA that restricts Rev-dependent HIV-1 RNA export⁷¹), and *C9orf16* (Fig. 3q)(of unknown function). Finally, we identified several latency-related genes including *AHNAK* (Fig. 3r, a histone methyltransferase⁷²), *CD99* (Fig 3s, upregulated by Tat⁷³), and *CD52* (Fig. 3t, a marker of HIV-1 infection during ART⁷⁴).

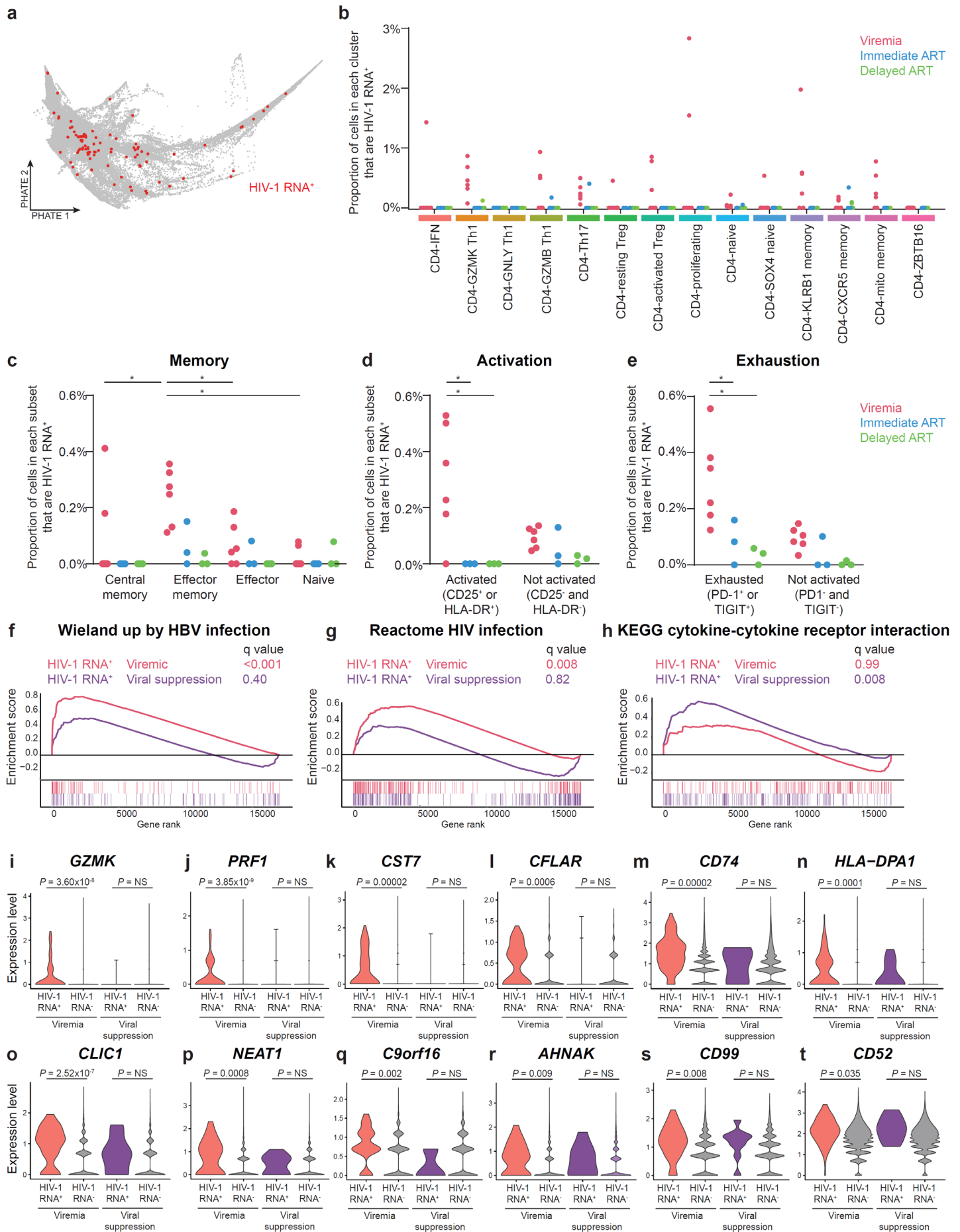


Fig. 3. The transcriptional landscape of HIV-1 RNA⁺ cells. **a**, PHATE plot with 90 HIV-1 RNA⁺ cells (including 81 from the viremic condition and 9 from suppressed conditions) labeled in red. **b – e**, Proportion of cells that are HIV-1 RNA⁺ in transcriptome-defined clusters (**b**), memory cell differentiation states (**c**), T cell activation states (**d**), and T cell exhaustion states (**e**). *P* values were determined by two-way ANOVA followed by Tukey's multiple comparisons test. Surface protein positive staining was determined by >90th percentile of the expression level of isotype barcoded surface protein staining antibody controls. *P* values were determined by two-way ANOVA followed by Tukey's multiple comparisons test. **f – h**, Identification of immune pathway enrichment in HIV-1 RNA⁺ cells versus HIV-1 RNA⁻ cells during viremia (**f, g**) and after viral suppression (**h**) using gene set enrichment analysis (GSEA). **i – t**, Markers of HIV-1 RNA⁺ cells identified by bootstrapping. The transcriptome of HIV-1 RNA⁺ cells was compared with HIV-1 RNA⁻ cells (*n* = 52,392 and 36,797, for viremic and viral suppression respectively) in cluster and individual-matched pools repeatedly for 10,000 times. Genes were selected based on being positive in >95% of test. *P* values were determined by Wilcoxon rank-sum test for 22 genes and corrected by Benjamini-Hochberg procedure. Y axis shows Ln expression levels. *, *P* < 0.05.

The upstream regulators of T cell clonal expansion are antigen stimulation and cytokine dysregulation

TCR sequences can act as barcodes for CD4⁺ T cell clones denoting antigen specificity, due to the sequence diversity generated by genomic rearrangement of the T cell receptor locus. T cells having the same TCR sequence likely originate from the same T cell clone and respond to the same antigen presentation. By examining the T cell repertoire at both viremic and viral suppression time points, we identified cells that were captured as clones (TCR sequence captured two or more times in the same individual) at either time point (Fig. 4a) and tracked the same clones over time. We identified TCR sequences at both the bulk and the single-cell level in separate pools. Bulk TCR sequencing enabled us to achieve a greater depth of TCR capture than single-cell alone and increased our ability to identify clonal T cells and to determine a more accurate clonal frequency. When we examined the transcriptome-defined T cell clusters, we found that GZMB Th1 cells and GZMK Th1 cells were significantly more clonal than other clusters (Fig. 4b, Supplementary Fig. 13). When we examined the memory T cell subsets, we found effector memory cells have the largest proportion of clonal cells (Fig. 4c). To identify cellular transcriptional programs that correlate with T cell clone size, we identified genes having expression levels that positively correlated with T cell clone size by Pearson correlation. Using GSEA, we found that these clone-size correlated genes were strongly involved in antigen responses such as cytotoxic effector expression (*GZMA*, *GZMB*, *GZMK*, *PRF1*), *CCL5*, *KLRG1*, *CD74*, and *BHLHE40* in viremic, immediate ART, delayed ART, and uninfected conditions (Fig. 4d). In order to identify regulators of T cell clone size, we used elastic net regression⁷⁵ to learn the relationship between specific genes and clone size during viremia, immediate ART, delayed ART, and uninfected conditions. By identifying genes having high clone-size determination scores, we found that T cell exhaustion, T cell activation, and antigen stimulation genes were the underlying transcriptional program that correlated with T cell clone size during delayed ART initiation (as opposed to immediate ART initiation)(Fig. 4e). The upstream regulators of this transcriptional program were LPS, T cell receptor signaling, and cytokine dysregulation (IFN γ , TNF, TGF β , and IL-21) (Fig. 4f).

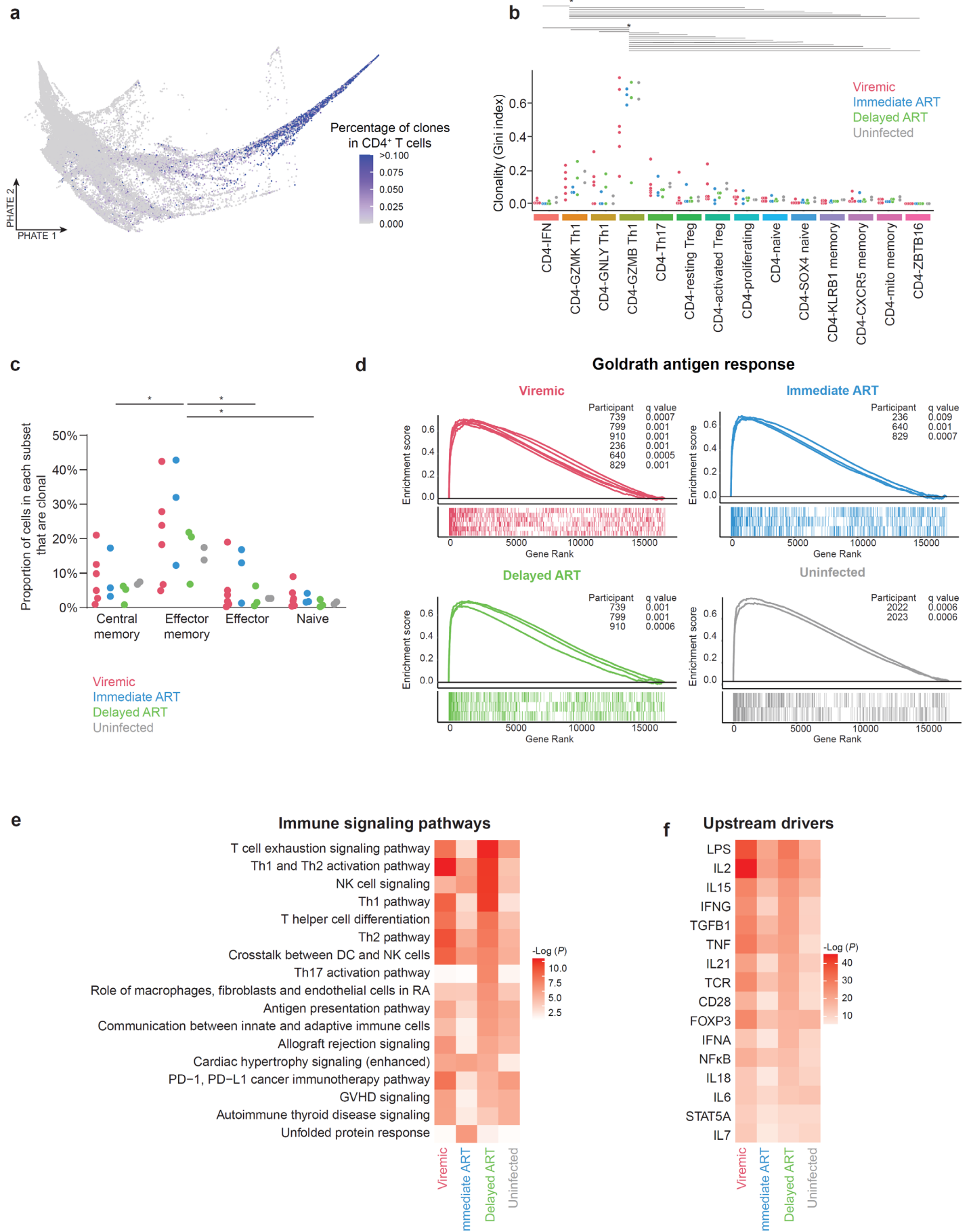


Fig. 4. The transcriptional landscape and regulators of T cell clones. **a**, PHATE plot with cells colored by clone size (109,307 cells with TCRs captured) determined by bulk TCR sequencing. **b**, Level of clonality in each transcriptome-defined CD4⁺ T

cell clusters measured by Gini Index of cells in that cluster. *P* values were determined by two-way ANOVA followed by Tukey's multiple comparisons test. *, *P* < 0.05. **c**, Proportion of T cell clones in different memory cell subsets. *P* values were determined by two-way ANOVA followed by Tukey's multiple comparisons test. *, *P* < 0.05. **d**, Pearson correlation of gene expression levels and T cell clone size (Log₂) determined the rank of genes having expression levels correlating with T cell clone size. **e**, Using elastic net regression, we identified genes that correlated with clone size during viremia, immediate ART, delayed ART, and uninfected conditions. Using these genes for IPA, we identified immune pathways (**e**) and upstream regulators (**f**) that determined T cell clone size. *P* values in **e** and **f** were determined by Fisher's exact test.

T cell clones harboring HIV-1 RNA⁺ cells remain stable over time

We examined how HIV-1 infection and ART initiation affects T cell clones harboring HIV-1 RNA⁺ cells (clone_{HIV-1⁺}) versus T cell clones that do not harbor HIV-1 RNA⁺ cells (clone_{HIV-1⁻}). We examined clone_{HIV-1⁺} in addition to HIV-1 RNA⁺ cells (Fig. 3) because T cell clones having the same TCR sequence may proliferate upon the same antigen stimulation, and the immune drivers of the proliferation of these clones may also induce the proliferation of HIV-1 RNA⁺ cells within them. Of note, each clone_{HIV-1⁺} (27 clones, with bulk sizes of 2-1225) only had one or two HIV-1 RNA⁺ cells, either because HIV-1 infection happened after T cell clonal expansion or because other HIV-1-infected cells did not express high levels of HIV-1 RNA without *ex vivo* stimulation. By using shared TCR sequences to track T cell clones, we examined T cell clone size (Fig. 5a–b), T cell expansion dynamics (Fig. 5c–d), and the immune phenotypes of clone_{HIV-1⁺} (Fig. 5e–m), during viremia and after viral suppression. We found that clone_{HIV-1⁺} were significantly larger in clone size than clone_{HIV-1⁻}, during viremia and after viral suppression (Fig. 5a). Immediate versus delayed ART did not significantly change the size of clone_{HIV-1⁺} (Fig. 5b). Despite immediate ART, the size of clone_{HIV-1⁺} was larger than clone_{HIV-1⁻}, suggesting that clone_{HIV-1⁺} had preferential proliferation despite immediate ART.

We next examined the change in T cell clone size during viremia and after viral suppression (Fig. 5c). We found that the majority of clone_{HIV-1⁺} remained stable in size and none contracted over time (Fig. 5d), unlike clone_{HIV-1⁻} which may contract over time (Fig. 5d). We then examined the immune phenotype of clones harboring HIV-1 RNA⁺ cells (Fig. 5e – 5m). We found that the immune phenotype of clone_{HIV-1⁺}, as characterized by cluster, did not significantly change over time (Fig. 5f, 5i, 5l), showing few differential expressed T cell function-related genes (Supplementary Fig. 14). These clones were mainly GZMB Th1 cells in these three represented clones. These cells are mainly effector memory cells, and their memory phenotypes did not significantly change over time (Fig. 5g, 5j, 5m).

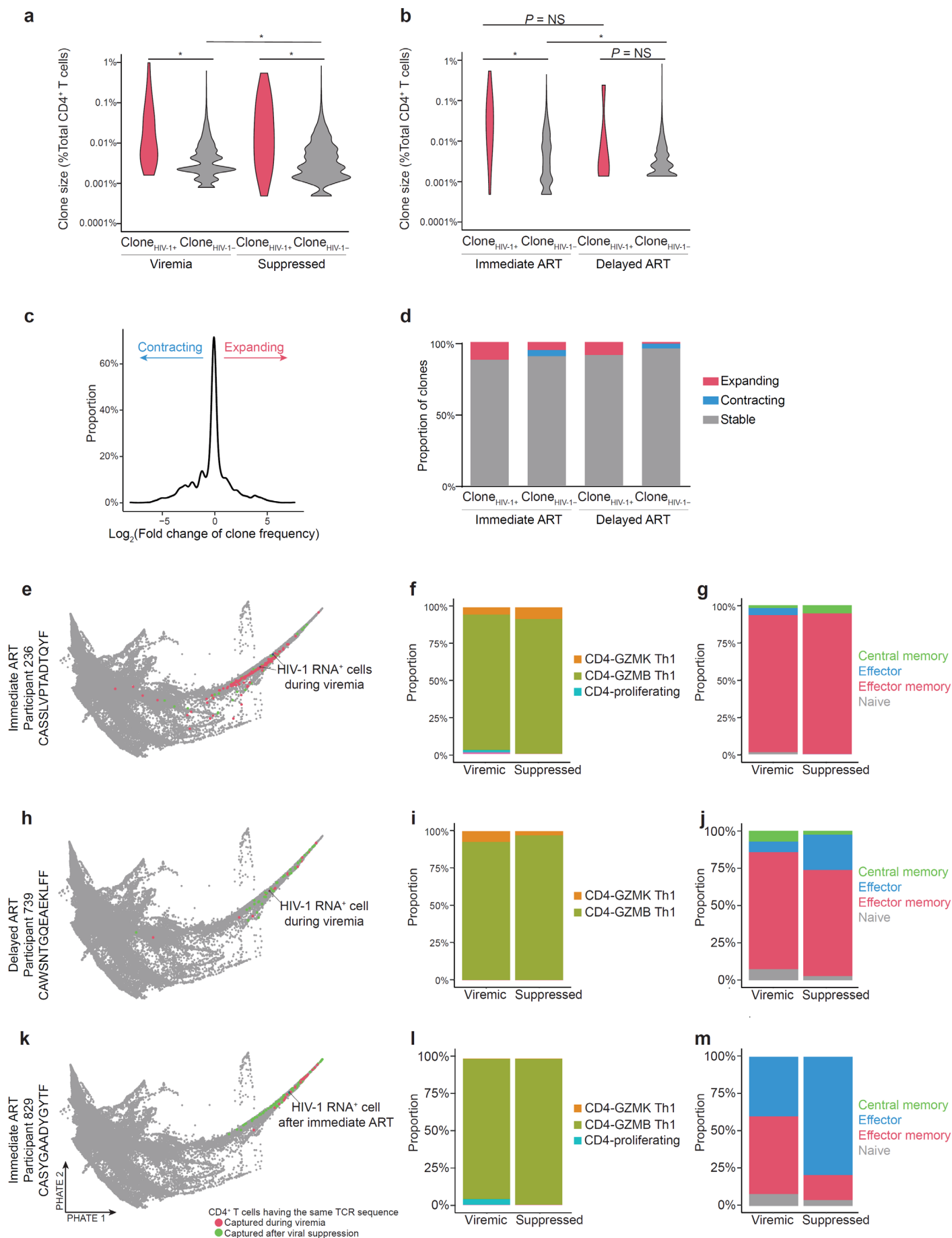


Fig. 5. The transcriptional landscape and clonal expansion dynamics of T cell clones harboring HIV-1 RNA⁺ cells. **a–b**, Clone size of T cells harboring HIV-1 RNA⁺ cells (clone_{HIV-1+}, n = 27, 20, and 7 for clones from viremia and suppression,

immediate ART, and delayed ART respectively) versus T cells that do not harbor HIV-1 RNA⁺ cells (clone_{HIV-1}⁻, n = 12,421, 5,658, and 6,763 for clones from viremia and suppression, immediate ART, and delayed ART respectively respectively). *P* values were determined by Wilcoxon rank-sum test followed by Benjamini-Hochberg procedure. *, *P* < 0.05. **c**, Log₂ fold change of normalized clone size between the viremic and virally suppressed samples. **d**, T cell expansion dynamics in clone_{HIV-1}⁺ versus clone_{HIV-1}⁻. Significant expansion and contraction was determined by having *P* < 0.05 calculated by Fisher's exact test **e**, **h**, **k**, PHATE plots of three representative T cell clones harboring HIV-1 RNA⁺ cells (n = 204, 54, and 108 for **e**, **h**, and **k** respectively). **f**, **i**, **l**, Transcriptome-defined cluster of T cell clones harboring HIV-1 RNA⁺ cells during viremia and after viral suppression in three representative clones. **g**, **j**, **m**, Memory phenotype of clone_{HIV-1}⁺ during viremia and after viral suppression in three representative clones.

Cytotoxic T cell response, antigen response, and cytokine response, shapes the transcription landscape of clone_{HIV-1}⁺

We wanted to determine whether clone_{HIV-1}⁺ (13 clones with 2 or more cells captured, 4–201 cells in each clone, out of a total of 537 cells, including 276 during viremia and 261 during viral suppression) had a distinct transcriptional signature compare with cells not in clone_{HIV-1}⁺ (a total of 78,199 cells, including 45,706 during viremia and 32,493 during viral suppression)(Fig. 6a). To compare the transcriptome profiles of the disproportionate numbers of clone_{HIV-1}⁺ (537 cells) with non-clone_{HIV-1}⁺ (78,199 cells), we employed single-cell identity definition using random forests and recursive feature elimination (scrFE)⁴³ with bootstrapping to identify key genes that were necessary and sufficient to differentiate clone_{HIV-1}⁺ from non-clone_{HIV-1}⁺. We identified 100 genes (Supplementary Table 8) that were as effective as the 5,000 most highly variable genes for differentiating clone_{HIV-1}⁺ from non-clone_{HIV-1}⁺ (Fig. 6b, 6c). The number 100 was selected to optimize for the maximum recall without reducing precision in differentiating clone_{HIV-1}⁺ from non-clone_{HIV-1}⁺ (Supplementary Fig. 15). The majority of the 100 genes identified across all cells were shared between viremic and suppressed conditions suggesting a persistent transcriptional phenotype differentiates clones_{HIV-1}⁺ from non-clone_{HIV-1}⁺ (Fig. 6d). By examining the differential gene expression profiles of these 100 key genes, we found that the genes that are differentially expressed in clones_{HIV-1}⁺ versus clones_{HIV-1}⁻ were involved in cytotoxic T cell function (*GZMA*, *GZMB*, *GZMH*, *GNLY*, *NKG7*, *CCL4*, *CCL5*, *ZNF683*, *ZEB2*, *CTSW*, *CST7*, and *FGFBP2*)⁷⁶ and antigen responses (*HLA-DRB1*, *HLA-DPA1*, *HLA-DPB1*, and *FCRL6*) and had lower levels of central memory T cell genes (*TCF7*, *SELL* (CD62L), *IL7R*, and *LEF1*). To ensure that we were differentiating differences between T cell clones with versus without HIV-1, as opposed to cluster differences such as GZMB Th1 versus non-GZMB Th1, we identified genes that could distinguish clone_{HIV-1}⁺ versus cluster-proportion-matched non-clone_{HIV-1}⁺ pools (Supplementary Table 8). Among them, 39 genes were significantly upregulated in clone_{HIV-1}⁺ (Fig. 6e, Supplementary Table 9). Using gene ontology analysis, we found that these genes were involved in IFN γ responses, cytokine responses, T cell activation, antigen presentation, and viral processes (Fig. 6f – 6p). Among them, we identified cytotoxic CD4⁺ T cell genes (*ZEB2*⁷⁷, *CCL4*, *ZNF683*)(Fig. 6g – 6i) and antigen response genes (Fig. 6l – 6p). Surprisingly, we identified upregulation of genes that may promote HIV-1 infection, such as *LGALS1* (Galectin-1)(promoting HIV-1 attachment⁷⁸) and *EZR* (Ezrin)(promoting HIV-1 infection⁷⁹). Overall, we found that clone_{HIV-1}⁺ demonstrated gene signatures of antigen responses, cytotoxic T cell responses, and cytokine signaling pathways.

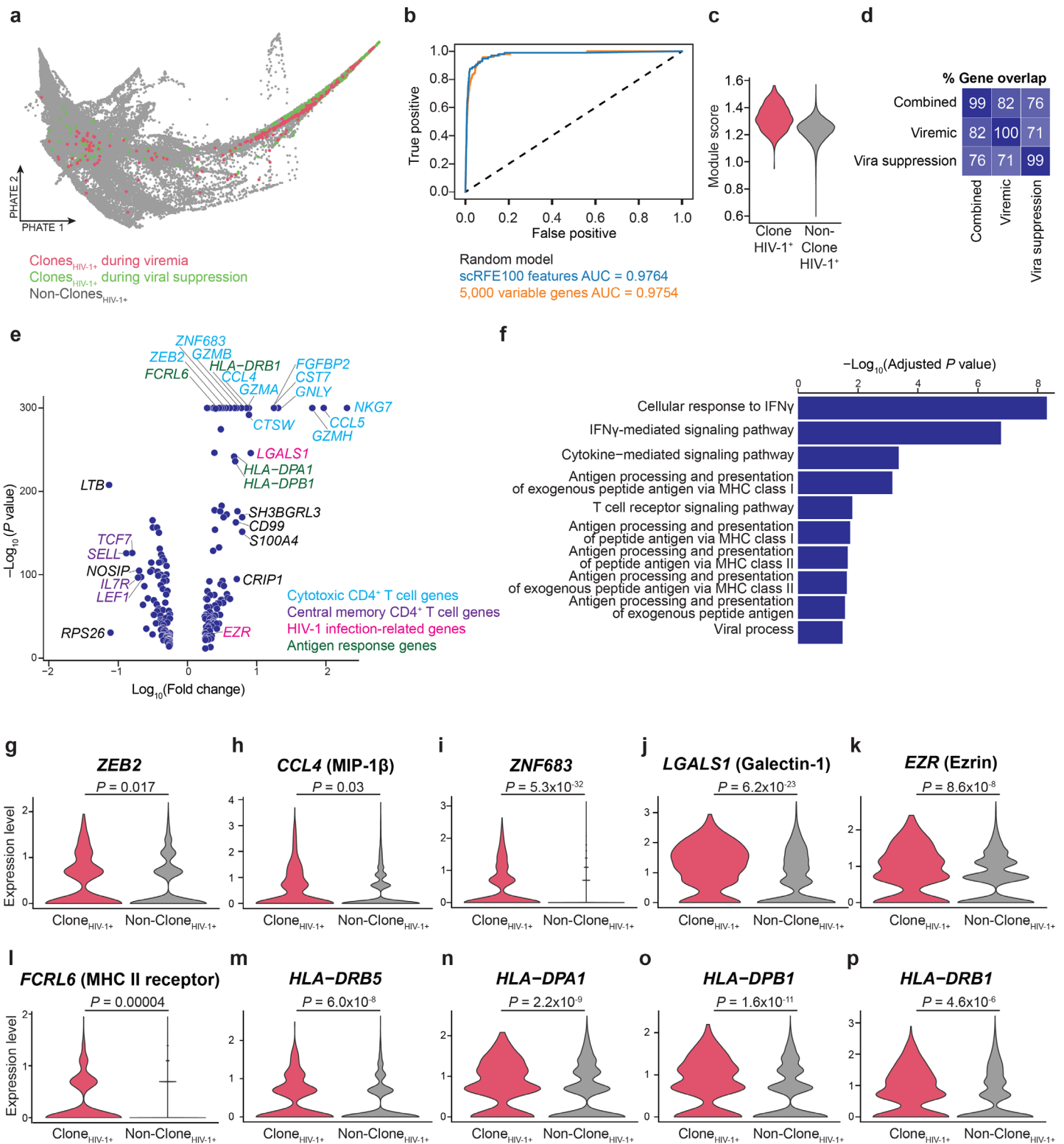


Fig. 6. Cellular markers of T cell clones harboring HIV-1 RNA⁺ cells (clone_{HIV-1+}) (13 clones, 537 cells) versus non-clone_{HIV-1+} (78,199 cells). **a**, Cells in clones (≥ 2 cells) harboring at least one HIV-1 RNA⁺ cell. **b**, Binary classification random forest model performance. Single-cell identity definition using random forests and recursive feature elimination (scRFE) identified key genes that were necessary and sufficient to distinguish clone_{HIV-1+} versus non-clone_{HIV-1+}. We used bootstrapping (comparing the single-cell transcriptome of clone_{HIV-1+} to that of random sampling of non-clone_{HIV-1+} with 10,000 iterations) to overcome the disproportionate rarity of clone_{HIV-1+}. We used the top 100 key genes from these iterations to train a model and tested how confident that these 100 genes, as opposed to top 5,000 highly variable genes, or random chance, can differentiate cells in clone_{HIV-1+} versus non-clone_{HIV-1+}. AUC, area under curve. **c**, Module score of the 100 key genes in clone_{HIV-1+} versus non-clone_{HIV-1+}. **d**, Gene overlap of the 100 key genes identified by scRFE bootstrapping analysis when trained on all cells, cells from the viremic time point, and cells from the suppressed time point. **e**, Volcano plot of differential gene expression analysis of the 100 key genes comparing clone_{HIV-1+} to non-clone_{HIV-1+} using differential expression testing

in Seurat (Wilcoxon rank-sum test). **f**, Gene ontology enrichment of genes ($n = 39$) from the 100 key genes that are significantly upregulated in clone_{HIV-1+} relative to a cluster proportion matched pool of non-clone_{HIV-1+} using Enrichr. **g-p** Gene expression differences of key genes upregulated in clones containing an HIV-1 RNA⁺ cell relative to a cluster proportion matched pool of clone_{HIV-1-} ($n = 573$ and $3,305$ for clone_{HIV-1+} and non-clone_{HIV-1+}, respectively). Y axis shows natural log expression levels.

Discussion

We found immune drivers that promote HIV-1 persistence and cellular markers as potential therapeutic targets. First, we found that an ongoing TNF response was the major immune dysfunction in delayed versus immediate ART (Supplementary Fig. 7–8), shaped the transcriptional program of HIV-1 RNA⁺ cells (Fig. 3h), and was an upstream regulator shaping T cell clonal expansion (Fig. 4f). Second, we found that cytotoxic CD4⁺ T lymphocytes, particularly those expressing *GZMK* (granzyme K) and *GZMB* (granzyme B), harbored HIV-1 RNA⁺ cells (Fig. 3b) and clone_{HIV-1+} (Fig. 4). Third, we found that T cell clones harboring HIV-1 RNA⁺ cells were larger in clone size (Fig. 5a). Fourth, using machine learning algorithms, we identified 22 markers for HIV-1 RNA⁺ cells (Fig. 3) and 39 upregulated genes for clone_{HIV-1+} that distinguished HIV-1 RNA⁺ cells from HIV-1 RNA⁻ cells and clone_{HIV-1+} from non-clone_{HIV-1+}. Altogether, we found that HIV-1 resides in cytotoxic T cells which are naturally proliferative. These cytotoxic CD4⁺ T cells continue to be impacted by antigen stimulation and TNF responses from viremia through viral suppression.

Our study highlights that immune dysfunction is established rapidly during acute HIV-1 infection. Unless initiated in hyperacute infection³⁴, early ART initiation could not fully revert chronic immune activation. By identifying co-regulated genes during HIV-1 infection, we found that acute viremia induced IFN, antigen, and TNF responses (Fig. 2, Supplementary Fig. 7–9). Both immediate and delayed ART reduced IFN responses to the level in uninfected individuals, but antigen responses and TNF responses persisted (Supplementary Fig. 7). Importantly, reducing TNF responses was the most prominent impact that immediate ART provides relative to delayed ART, although immediate ART did not fully restore TNF response to the level in uninfected individuals (Supplementary Fig. 7 and 8), particularly on *GZMK* Th1 cells (Fig. 1k, 2j)(see below). Immediate versus delayed ART did not significantly change the cell subsets harboring HIV-1 RNA⁺ cells (Fig. 3) and the clonal expansion dynamics of T cell clones in CD4⁺ T cells (Fig. 4) and in clone_{HIV-1+} (Fig. 5). Pathway analysis suggested that both antigen responses and TNF signaling shaped the landscape of HIV-1 RNA⁺ cells (Fig. 3h) and T cell clonal expansion (Fig. 4e). Of note, despite the strong effect of antigen responses on T cell clonal expansion (Fig. 4d), common γ chain cytokines that promote homeostatic proliferation (such as IL-2 and IL-15) shaped the transcriptional landscape of HIV-1 RNA⁺ cells (Fig. 3h) and T cell clones (Fig. 4f), suggesting that both antigen stimulation and homeostatic proliferation¹² contribute to HIV-1 persistence. Strategies that can reduce chronic antigen stimulation in people with HIV-1^{80,81}, or strategies targeting HIV-1-induced cytokine dysregulation⁶², should be explored to eliminate these immune drivers that promote the persistence of HIV-1-infected cells.

Antigen-responding CD4⁺ T cells are known to be an important source of the HIV-1 latent reservoir⁹⁻¹¹. Antigen-responding CD4⁺ T cells have distinct functional states, such as cytolytic granule secretion (such as granzyme B, granzyme K, and perforin), cytokine production (such as IL-2, IFN γ , and TNF), and antiviral chemokine production (such as *CCL3*, *CCL4*, and *XCL1*). Cytotoxic CD4⁺ T cells are critical effectors responding to HIV-1⁸²⁻⁸⁶, cytomegalovirus^{87,88}, SARS-CoV-2, and tuberculosis⁸⁹ infections. We have previously found that HIV-1 RNA⁺ cells after *ex vivo* stimulation exhibit Th1 polarization and upregulate antiviral chemokines *CCL3*, *CCL4*, and *XCL1*¹³. Our study identified an under-appreciated role of *GZMB* and *GZMK*-expressing cytotoxic CD4⁺ T cells as a source of HIV-1 reservoir. First, we found that *GZMK* Th1 CD4⁺ T cells harbored HIV-1 RNA⁺ cells (Fig. 3b). The *GZMK* Th1 cells were phenotypically distinct from *GZMB* Th1 cells (Fig. 1). Importantly, *GZMK*-expressing CD4⁺ T cells were recently found to be important in SARS-CoV-2 infection⁵⁹ and in anti-tumor cytotoxicity⁵⁰. Second, we found that *GZMB* and *GZMK* Th1 T cells exhibited strong antigen responses (Fig. 1h), were clonally expanded (Fig. 4b, Supplementary Fig. 5f), were enriched of HIV-1 RNA⁺ cells (Fig. 3b), and were the major T cell phenotype of clone_{HIV-1+} (Fig. 5). Third, cytotoxic CD4⁺ T cell markers were among the key cellular markers that could distinguish HIV-1 RNA⁺ cells from HIV-1 RNA⁻ cells and clone_{HIV-1+} from non-clone_{HIV-1+}, such as *GZMK*, *CST7*, and *CCL4* (Fig. 3i, 3k, 6e). We postulate that despite ART, persistent antigen stimulation and TNF responses induce the proliferation of HIV-1-infected cytotoxic CD4⁺ T cells and promote HIV-1 persistence.

Specific targeting of HIV-1-infected cells requires viral or cellular markers that can distinguish HIV-1-infected cells from uninfected cells. Studying HIV-1-infected cells from people with HIV-1 provides the most clinically

relevant understanding^{13,19,21,23,80,90-92}. We identified HIV-1 RNA⁺ cells and clone_{HIV-1}⁺ without *ex vivo* stimulation and identified genes that marked these cells. Using machine-learning analysis, we found that HIV-1 RNA⁺ cells and clone_{HIV-1}⁺ can be distinguished by a distinct transcriptional phenotype. While the majority of identified markers were consistent with T cell activation during viremia, a subset of markers were previously unreported, such as *GZMK*, *CLIC1*, *NEAT1*, and *C9orf16* for HIV-1 RNA⁺ cells (Fig. 3) and *ZEB2*, *ZNF683*, and *LGALS1* for clone_{HIV-1}⁺ (Fig. 6). The combination of markers may define useful populations for studying HIV-1-infected cells in the future. These cells, though not latent at time of capture, are a source of persistent immune activation through HIV-1 transcriptional activity. Our finding raises the possibility of targeting HIV-1 RNA⁺ cells and clone_{HIV-1}⁺ more precisely.

Acknowledgements. We thank all study participants. We thank Alex K. Shalek, Sam Kazer, Joseph Craft, Steven Kleinstein, Nicolas Chomont, and Michael Betts for their insightful suggestions. This work is supported by Yale Top Scholar, Rudolf J. Anderson Fellowship, NIH R01 AI141009, NIH R61 DA047037, NIH R37 AI147868, NIH R01 DA051906, NIH UM1 DA051410, NIH CHEETAH P50 AI150464, NIH BEAT-HIV Delaney Collaboratory UM1 AI126620, Gilead HIV-1 Research Scholar Grant, American Foundation for AIDS Research (amfAR 110029-67-RGRL), Lupus Research Alliance – Celgene, and NIH T32 AI055403 (J.A.C.). The Sabes and MERLIN studies were supported by NIH/NIDA R01 DA032106 and NIH/NIDA R01 DA040532 (A.D.). We gratefully acknowledge ART drug donation from Merck & Co. and Gilead Sciences Inc.

Author contributions. J.A.C. and Y.-C.H. designed the experiments, performed analyses, and wrote the manuscript. J.A.C. performed experiments and bioinformatic analyses. J.A.C. performed machine learning analyses with help from N.P. and D.V.D. D.P.-S., S.P., C.G., R.A., J.L., J.C., S.S., and A.C.D. recruited study participants and processed clinical samples. A.D. designed the Sabes and MERLIN studies and assumed oversight for study conduct and analysis.

Competing Interests. The authors have no competing interests.

References

- 1 Siliciano, J. D. *et al.* Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4⁺ T cells. *Nat Med* **9**, 727-728, doi:10.1038/nm880 (2003).
- 2 Crooks, A. M. *et al.* Precise Quantitation of the Latent HIV-1 Reservoir: Implications for Eradication Strategies. *J Infect Dis* **212**, 1361-1365, doi:10.1093/infdis/jiv218 (2015).
- 3 Bui, J. K. *et al.* Proviruses with identical sequences comprise a large fraction of the replication-competent HIV reservoir. *PLoS Pathog* **13**, e1006283, doi:10.1371/journal.ppat.1006283 (2017).
- 4 Hosmane, N. N. *et al.* Proliferation of latently infected CD4(+) T cells carrying replication-competent HIV-1: Potential role in latent reservoir dynamics. *J Exp Med* **214**, 959-972, doi:10.1084/jem.20170193 (2017).
- 5 Lorenzi, J. C. *et al.* Paired quantitative and qualitative assessment of the replication-competent HIV-1 reservoir and comparison with integrated proviral DNA. *Proc Natl Acad Sci U S A* **113**, E7908-E7916, doi:10.1073/pnas.1617789113 (2016).
- 6 Maldarelli, F. *et al.* HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science* **345**, 179-183, doi:10.1126/science.1254194 (2014).
- 7 Wagner, T. A. *et al.* HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science (New York, N. Y.)* **345**, 570-573, doi:10.1126/science.1256304 [doi] (2014).
- 8 Douek, D. C. *et al.* HIV preferentially infects HIV-specific CD4⁺ T cells. *Nature* **417**, 95-98, doi:10.1038/417095a (2002).
- 9 Simonetti, F. R. *et al.* Antigen-driven clonal selection shapes the persistence of HIV-1 infected CD4⁺ T cells in vivo. *J Clin Invest*, doi:10.1172/jci145254 (2020).
- 10 Simonetti, F. R. *et al.* Clonally expanded CD4⁺ T cells can produce infectious HIV-1 in vivo. *Proc Natl Acad Sci U S A* **113**, 1883-1888, doi:10.1073/pnas.1522675113 (2016).
- 11 Mendoza, P. *et al.* Antigen-responsive CD4⁺ T cell clones contribute to the HIV-1 latent reservoir. *J Exp Med* **217**, doi:10.1084/jem.20200051 (2020).
- 12 Chomont, N. *et al.* HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nat Med* **15**, 893-900, doi:10.1038/nm.1972 (2009).

- 13 Liu, R. *et al.* Single-cell transcriptional landscapes reveal HIV-1-driven aberrant host gene transcription as a potential therapeutic target. *Sci Transl Med* **12**, doi:10.1126/scitranslmed.aaz0802 (2020).
- 14 Jain, V. *et al.* Antiretroviral Therapy Initiated Within 6 Months of HIV Infection Is Associated With Lower T-Cell Activation and Smaller HIV Reservoir Size. *Journal of Infectious Diseases* **208**, 1202-1211, doi:10.1093/infdis/jit311 (2013).
- 15 Ndhlovu, Z. M. *et al.* Augmentation of HIV-specific T cell function by immediate treatment of hyperacute HIV-1 infection. *Sci Transl Med* **11**, doi:10.1126/scitranslmed.aau0528 (2019).
- 16 Schuetz, A. *et al.* Initiation of ART during early acute HIV infection preserves mucosal Th17 function and reverses HIV-related immune activation. *PLoS Pathog* **10**, e1004543, doi:10.1371/journal.ppat.1004543 (2014).
- 17 Sereti, I. *et al.* Persistent, Albeit Reduced, Chronic Inflammation in Persons Starting Antiretroviral Therapy in Acute HIV Infection. *Clin Infect Dis* **64**, 124-131, doi:10.1093/cid/ciw683 (2017).
- 18 Halvas, E. K. *et al.* HIV-1 viremia not suppressible by antiretroviral therapy can originate from large T cell clones producing infectious virus. *J Clin Invest* **130**, 5847-5857, doi:10.1172/jci138099 (2020).
- 19 Pollack, R. A. *et al.* Defective HIV-1 Proviruses Are Expressed and Can Be Recognized by Cytotoxic T Lymphocytes, which Shape the Proviral Landscape. *Cell Host Microbe* **21**, 494-506.e494, doi:10.1016/j.chom.2017.03.008 (2017).
- 20 Bengsch, B. *et al.* Epigenomic-Guided Mass Cytometry Profiling Reveals Disease-Specific Features of Exhausted CD8⁺T Cells. *Immunity* **48**, 1029-1045.e1025, doi:10.1016/j.immuni.2018.04.026 (2018).
- 21 Ho, Y. C. *et al.* Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* **155**, 540-551, doi:10.1016/j.cell.2013.09.020 (2013).
- 22 Bruner, K. M. *et al.* Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nat Med* **22**, 1043-1049, doi:10.1038/nm.4156 (2016).
- 23 Bruner, K. M. *et al.* A quantitative approach for measuring the reservoir of latent HIV-1 proviruses. *Nature* **566**, 120-125, doi:10.1038/s41586-019-0898-8 (2019).
- 24 Neidleman, J. *et al.* Phenotypic analysis of the unstimulated in vivo HIV CD4 T cell reservoir. *Elife* **9**, doi:10.7554/eLife.60933 (2020).
- 25 Lee, G. Q. *et al.* Clonal expansion of genome-intact HIV-1 in functionally polarized Th1 CD4⁺ T cells. *J Clin Invest* **127**, 2689-2696, doi:10.1172/JCI93289 (2017).
- 26 Lee, E. *et al.* Memory CD4⁺ T-Cells Expressing HLA-DR Contribute to HIV Persistence During Prolonged Antiretroviral Therapy. *Front Microbiol* **10**, 2214, doi:10.3389/fmicb.2019.02214 (2019).
- 27 Gantner, P. *et al.* Single-cell TCR sequencing reveals phenotypically diverse clonally expanded cells harboring inducible HIV proviruses during ART. *Nature Communications* **11**, 4089, doi:10.1038/s41467-020-17898-8 (2020).
- 28 Hiener, B. *et al.* Identification of Genetically Intact HIV-1 Proviruses in Specific CD4(+) T Cells from Effectively Treated Participants. *Cell Rep* **21**, 813-822, doi:10.1016/j.celrep.2017.09.081 (2017).
- 29 Venanzi Rullo, E. *et al.* Persistence of an intact HIV reservoir in phenotypically naive T cells. *JCI Insight* **5**, doi:10.1172/jci.insight.133157 (2020).
- 30 Fromentin, R. *et al.* CD4⁺ T Cells Expressing PD-1, TIGIT and LAG-3 Contribute to HIV Persistence during ART. *PLoS Pathog* **12**, e1005761, doi:10.1371/journal.ppat.1005761 (2016).
- 31 Cohn, L. B. *et al.* Clonal CD4(+) T cells in the HIV-1 latent reservoir display a distinct gene profile upon reactivation. *Nat Med* **24**, 604-609, doi:10.1038/s41591-018-0017-7 (2018).
- 32 Nguyen, S. *et al.* Elite control of HIV is associated with distinct functional and transcriptional signatures in lymphoid tissue CD8(+) T cells. *Sci Transl Med* **11**, doi:10.1126/scitranslmed.aax4077 (2019).
- 33 Hoang, T. N. *et al.* Baricitinib treatment resolves lower-airway macrophage inflammation and neutrophil recruitment in SARS-CoV-2-infected rhesus macaques. *Cell*, doi:10.1016/j.cell.2020.11.007.
- 34 Kazer, S. W. *et al.* Integrated single-cell analysis of multicellular immune dynamics during hyperacute HIV-1 infection. *Nat Med* **26**, 511-518, doi:10.1038/s41591-020-0799-2 (2020).
- 35 Shalek, A. K. *et al.* Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498**, 236-240, doi:10.1038/nature12172 (2013).
- 36 Yost, K. E. *et al.* Clonal replacement of tumor-specific T cells following PD-1 blockade. *Nat Med* **25**, 1251-1259, doi:10.1038/s41591-019-0522-3 (2019).
- 37 Wilk, A. J. *et al.* A single-cell atlas of the peripheral immune response in patients with severe COVID-19. *Nature Medicine* **26**, 1070-1076, doi:10.1038/s41591-020-0944-y (2020).

- 38 Pappalardo, J. L. *et al.* Transcriptomic and clonal characterization of T cells in the human central nervous system. *Sci Immunol* **5**, doi:10.1126/sciimmunol.abb8786 (2020).
- 39 Torang, A., Gupta, P. & Klinke, D. J., 2nd. An elastic-net logistic regression approach to generate classifiers and gene signatures for types of immune cells and T helper cell subsets. *BMC Bioinformatics* **20**, 433, doi:10.1186/s12859-019-2994-z (2019).
- 40 Hu, Y. *et al.* A machine learning approach for the identification of key markers involved in brain development from single-cell transcriptomic data. *BMC Genomics* **17**, 1025, doi:10.1186/s12864-016-3317-7 (2016).
- 41 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559, doi:10.1186/1471-2105-9-559 (2008).
- 42 Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e1821, doi:10.1016/j.cell.2019.05.031 (2019).
- 43 Park, M., Vorperian, S., Wang, S. & Pisco, A. O. Single-cell identity definition using random forests and recursive feature elimination. *bioRxiv*, 2020.2008.2003.233650, doi:10.1101/2020.08.03.233650 (2020).
- 44 Lama, J. R. *et al.* Design Strategy of the Sabes Study: Diagnosis and Treatment of Early HIV Infection Among Men Who Have Sex With Men and Transgender Women in Lima, Peru, 2013-2017. *Am J Epidemiol* **187**, 1577-1585, doi:10.1093/aje/kwy030 (2018).
- 45 Lama, J. R. *et al.* Clinical and Immunologic Outcomes after Immediate or Deferred Antiretroviral Therapy Initiation during Primary HIV Infection: The Sabes Randomized Clinical Study. *Clin Infect Dis*, doi:10.1093/cid/ciaa167 (2020).
- 46 Grebe, E. *et al.* Interpreting HIV diagnostic histories into infection time estimates: analytical framework and online tool. *BMC Infect Dis* **19**, 894, doi:10.1186/s12879-019-4543-9 (2019).
- 47 Mimitou, E. P. *et al.* Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. *Nat Methods* **16**, 409-412, doi:10.1038/s41592-019-0392-0 (2019).
- 48 Moon, K. R. *et al.* Visualizing structure and transitions in high-dimensional biological data. *Nat Biotechnol* **37**, 1482-1492, doi:10.1038/s41587-019-0336-3 (2019).
- 49 Sallusto, F., Lenig, D., Förster, R., Lipp, M. & Lanzavecchia, A. Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature* **401**, 708-712, doi:10.1038/44385 (1999).
- 50 Oh, D. Y. *et al.* Intratumoral CD4(+) T Cells Mediate Anti-tumor Cytotoxicity in Human Bladder Cancer. *Cell* **181**, 1612-1625.e1613, doi:10.1016/j.cell.2020.05.017 (2020).
- 51 Zhang, J.-Y. *et al.* Single-cell landscape of immunological responses in patients with COVID-19. *Nature Immunology* **21**, 1107-1118, doi:10.1038/s41590-020-0762-x (2020).
- 52 Goldrath, A. W., Luckey, C. J., Park, R., Benoist, C. & Mathis, D. The molecular program induced in T cells undergoing homeostatic proliferation. *Proc Natl Acad Sci U S A* **101**, 16885-16890, doi:10.1073/pnas.0407417101 (2004).
- 53 Phong, M. S. *et al.* p38 mitogen-activated protein kinase promotes cell survival in response to DNA damage but is not required for the G(2) DNA damage checkpoint in human cancer cells. *Mol Cell Biol* **30**, 3816-3826, doi:10.1128/mcb.00949-09 (2010).
- 54 Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189-196, doi:10.1126/science.aad0501 (2016).
- 55 Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128, doi:10.1186/1471-2105-14-128 (2013).
- 56 Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90-97, doi:10.1093/nar/gkw377 (2016).
- 57 Krämer, A., Green, J., Pollard, J., Jr. & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30**, 523-530, doi:10.1093/bioinformatics/btt703 (2014).
- 58 Sandler, N. G. *et al.* Type I interferon responses in rhesus macaques prevent SIV infection and slow disease progression. *Nature* **511**, 601-605, doi:10.1038/nature13554 (2014).
- 59 Hadjadj, J. *et al.* Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science* **369**, 718-724, doi:10.1126/science.abc6027 (2020).
- 60 Lee, J. S. *et al.* Immunophenotyping of COVID-19 and influenza highlights the role of type I interferons in development of severe COVID-19. *Science Immunology* **5**, eabd1554, doi:10.1126/sciimmunol.abd1554 (2020).
- 61 Wen, W. *et al.* Immune cell profiling of COVID-19 patients in the recovery stage by single-cell sequencing. *Cell Discovery* **6**, 31, doi:10.1038/s41421-020-0168-9 (2020).

- 62 Beyer, M. *et al.* Tumor-necrosis factor impairs CD4(+) T cell-mediated immunological control in chronic viral infection. *Nat Immunol* **17**, 593-603, doi:10.1038/ni.3399 (2016).
- 63 Yukl, S. A. *et al.* HIV latency in isolated patient CD4(+) T cells may be due to blocks in HIV transcriptional elongation, completion, and splicing. *Sci Transl Med* **10**, doi:10.1126/scitranslmed.aap9927 (2018).
- 64 Li, X. *et al.* CD161(+) CD4(+) T Cells Harbor Clonally Expanded Replication-Competent HIV-1 in Antiretroviral Therapy-Suppressed Individuals. *mBio* **10**, doi:10.1128/mBio.02121-19 (2019).
- 65 Pardons, M. *et al.* Single-cell characterization and quantification of translation-competent viral reservoirs in treated and untreated HIV infection. *PLOS Pathogens* **15**, e1007619, doi:10.1371/journal.ppat.1007619 (2019).
- 66 Niessl, J. *et al.* Persistent expansion and Th1-like skewing of HIV-specific circulating T follicular helper cells during antiretroviral therapy. *EBioMedicine* **54**, doi:10.1016/j.ebiom.2020.102727 (2020).
- 67 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).
- 68 Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739-1740, doi:10.1093/bioinformatics/btr260 (2011).
- 69 Gibellini, D. *et al.* HIV-1 Tat protein concomitantly down-regulates apical caspase-10 and up-regulates c-FLIP in lymphoid T cells: a potential molecular mechanism to escape TRAIL cytotoxicity. *J Cell Physiol* **203**, 547-556, doi:10.1002/jcp.20252 (2005).
- 70 Valenzuela, S. M. *et al.* The nuclear chloride ion channel NCC27 is involved in regulation of the cell cycle. *J Physiol* **529 Pt 3**, 541-552, doi:10.1111/j.1469-7793.2000.00541.x (2000).
- 71 Zhang, Q., Chen, C. Y., Yedavalli, V. S. & Jeang, K. T. NEAT1 long noncoding RNA and paraspeckle bodies modulate HIV-1 posttranscriptional expression. *mBio* **4**, e00596-00512, doi:10.1128/mBio.00596-12 (2013).
- 72 Boehm, D. *et al.* SMYD2-Mediated Histone Methylation Contributes to HIV-1 Latency. *Cell Host Microbe* **21**, 569-579 e566, doi:10.1016/j.chom.2017.04.011 (2017).
- 73 Lee, E., Kim, Y. R., Lee, M. K. & Lee, I. S. Human immunodeficiency virus-1 Tat positively regulates the human CD99 gene via DNA demethylation. *Korean Journal of Microbiology* **44**, 277-281 (2008).
- 74 Vojdeman, F. J. *et al.* Brief Report: CD52 Expression on CD4+ T Cells in HIV-Positive Individuals on cART. *J Acquir Immune Defic Syndr* **77**, 217-220, doi:10.1097/qai.0000000000001568 (2018).
- 75 Zhou, H. H., Trevor. Regularization and variable selection via the Elastic Net. *Journal of the Royal Statistical Society. Series B (Methodological)* **67**, 301-320 (2005).
- 76 Patil, V. S. *et al.* Precursors of human CD4(+) cytotoxic T lymphocytes identified by single-cell transcriptome analysis. *Sci Immunol* **3**, doi:10.1126/sciimmunol.aan8664 (2018).
- 77 Dominguez, C. X. *et al.* The transcription factors ZEB2 and T-bet cooperate to program cytotoxic T cell terminal differentiation in response to LCMV viral infection. *J Exp Med* **212**, 2041-2056, doi:10.1084/jem.20150186 (2015).
- 78 Ouellet, M. *et al.* Galectin-1 Acts as a Soluble Host Factor That Promotes HIV-1 Infectivity through Stabilization of Virus Attachment to Host Cells. *The Journal of Immunology* **174**, 4120-4126, doi:10.4049/jimmunol.174.7.4120 (2005).
- 79 Roy, N. H., Lambelé, M., Chan, J., Symeonides, M. & Thali, M. Ezrin is a component of the HIV-1 virological presynapse and contributes to the inhibition of cell-cell fusion. *J Virol* **88**, 7645-7658, doi:10.1128/jvi.00550-14 (2014).
- 80 Yeh, Y. J. *et al.* Filgotinib suppresses HIV-1-driven gene transcription by inhibiting HIV-1 splicing and T cell activation. *J Clin Invest*, doi:10.1172/jci137371 (2020).
- 81 Hunt, P. W. *et al.* Valganciclovir reduces T cell activation in HIV-infected individuals with incomplete CD4+ T cell recovery on antiretroviral therapy. *J Infect Dis* **203**, 1474-1483, doi:10.1093/infdis/jir060 (2011).
- 82 Appay, V. *et al.* Characterization of CD4⁺ CTLs Ex Vivo. *The Journal of Immunology* **168**, 5954-5958, doi:10.4049/jimmunol.168.11.5954 (2002).
- 83 Rosenberg, E. S. *et al.* Vigorous HIV-1-Specific CD4⁺ T Cell Responses Associated with Control of Viremia. *Science* **278**, 1447-1450, doi:10.1126/science.278.5342.1447 (1997).
- 84 Johnson, S. *et al.* Cooperativity of HIV-Specific Cytolytic CD4 T Cells and CD8 T Cells in Control of HIV Viremia. *J Virol* **89**, 7494-7505, doi:10.1128/jvi.00438-15 (2015).

- 85 Buggert, M. *et al.* Limited immune surveillance in lymphoid tissue by cytolytic CD4+ T cells during health and HIV disease. *PLoS Pathogens* **14**, e1006973, doi:10.1371/journal.ppat.1006973 (2018).
- 86 Soghoian, D. Z. *et al.* HIV-Specific Cytolytic CD4 T Cell Responses During Acute HIV Infection Predict Disease Outcome. *Science Translational Medicine* **4**, 123ra125-123ra125, doi:10.1126/scitranslmed.3003165 (2012).
- 87 van Leeuwen, E. M. *et al.* Emergence of a CD4+CD28- granzyme B+, cytomegalovirus-specific T cell subset after recovery of primary cytomegalovirus infection. *J Immunol* **173**, 1834-1841, doi:10.4049/jimmunol.173.3.1834 (2004).
- 88 Casazza, J. P. *et al.* Acquisition of direct antiviral effector functions by CMV-specific CD4+ T lymphocytes with cellular maturation. *J Exp Med* **203**, 2865-2877, doi:10.1084/jem.20052246 (2006).
- 89 Canaday, D. H. *et al.* CD4(+) and CD8(+) T cells kill intracellular Mycobacterium tuberculosis by a perforin and Fas/Fas ligand-independent mechanism. *J Immunol* **167**, 2734-2742, doi:10.4049/jimmunol.167.5.2734 (2001).
- 90 Jiang, C. *et al.* Distinct viral reservoirs in individuals with spontaneous control of HIV-1. *Nature*, doi:10.1038/s41586-020-2651-8 (2020).
- 91 Bender, A. M. *et al.* The Landscape of Persistent Viral Genomes in ART-Treated SIV, SHIV, and HIV-2 Infections. *Cell Host Microbe* **26**, 73-85.e74, doi:10.1016/j.chom.2019.06.005 (2019).
- 92 Wang, Z. *et al.* Expanded cellular clones carrying replication-competent HIV-1 persist, wax, and wane. *Proc Natl Acad Sci U S A* **115**, E2575-e2584, doi:10.1073/pnas.1720665115 (2018).