nature communications

Article

https://doi.org/10.1038/s41467-025-57368-7

Tracking HIV persistence across T cell lineages during early ART-treated HIV-1-infection using a reservoir-marking humanized mouse model

Received: 8 May 2024

Accepted: 19 February 2025

Published online: 06 March 2025

Check for updates

Namita Satija[1,6], Foramben Patel[1,6], Gerrit Schmidt[1], Donald V. Doanman[1], Manav Kapoor[2,3], Annalena La Porte[1], Ying-Chih Wang[2], Kenneth M. Law[1,5], Anthony M. Esposito[4], Kimaada Allette[2], Kristin G. Beaumont[2], Robert P. Sebra[2] & Benjamin K. Chen[1]

Human immunodeficiency virus (HIV) infection depletes CD4 T-cells, and long-term persistence of latent virus prevents full clearance of HIV even in the presence of effective antiretroviral therapy (ART), Here we present the HIV-1-induced lineage tracing (HILT) system, a model that irreversibly marks infected cells within a humanized mouse model, which detects rare latently infected cells. Immunodeficient mice transplanted with genetically modified hematopoietic stem cells develop a human immune system, in which CD4 T-cells contain a genetic switch that permanently labels cells infected by HIV-1 expressing cre-recombinase. Through single-cell RNA sequencing of HILT-marked cells during acute infection and post-ART treatment, we identify distinct CD4+ T-cell transcriptional lineages enriched in either active or latent infections. Comparative gene expression analysis highlights common pathways modulated in both states, including EIF2, Sirtuin, and protein ubiquitination. Critical regulators of these pathways, including *JUN*, *BCL2*, and *MDM2*, change to opposite directions in the two states, highlighting gene expression programs that may support HIV persistence across T-cell lineages and states.

Effective antiretroviral therapy (ART) in individuals infected with Human Immunodeficiency Virus-1 (HIV-1) halts active viral replication, preserves the immune system, and restores life expectancy[1–3]. However, a cure for the infection remains elusive due to the persistence of latently HIV-1-infected cells[4,5]. Latent cells carry quiescent HIV-1 provirus integrated into their genome, which can be reactivated upon stimulation. Most of the latent reservoir resides in the $CD4^+$ T-cell compartment, though there are no definitive reservoir markers. $CD4^+$ T-cells are highly diverse and can be subdivided based on their differentiation and activation status. Persistent HIV has been identified within central memory (TCM), effector memory (TEM), transitional memory (TTM), and stem-cell memory (TSCM) CD4+ T-cell populations[6–8]. Additionally, immune checkpoint molecules like PD-1, TIGIT, and LAG-3 are also observed to be enriched on HIV-infected cells[9,10]. Latent provirus has been found to be enriched in the Th17 subset or cells expressing homing receptors CXCR3 and CCR6[11–15]. Additional insights into the latent pool come from studying which CD4+ T-cell subsets and cellular states support infection. Not all the

[1]Division of Infectious Diseases, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [2]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [3]Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [4]Department of Biology, New Jersey City University, New Jersey City, NJ, USA. [5]Present address: Lexeo Therapeutics, New York, NY, USA. [6]These authors contributed equally: Namita Satija, Foramben Patel. e-mail: benjamin.chen@mssm.edu

states and subsets support HIV-1 infection to the same degree. Resting peripheral blood CD4+ T-cells, for instance, are resistant to active viral infection[11,16–18] whereas CCR4 + CCR6+ and CXCR3 + CCR6 + CD4+ T-cells are highly permissive to HIV-1 infection[19,20]. The reservoir is established rapidly after infection. ART treatment initiated as early as three days after acute infection still gives rise to a long-lived reservoir[21,22]. Once infected, both host and virus factors control CD4+ T-cell survival. Over time, particularly with ART, defective proviruses may accumulate, and only a fraction of infected cells carry a full-length, replication-competent provirus[23,24]. An emerging picture of the latent reservoir is that it is not comprised of static cellular composition that persists over many years. It is maintained through the clonal expansion of T-cells that proliferate within ART-treated patients[25]. This proliferation can be mediated by antigen-driven proliferation activation[26,27], homeostatic proliferation[6,28], or proliferation activated by HIV integration[29–31]. Given that the reservoir is established rapidly and can persist in clonally derived T-cell lineages, it remains unclear which T-cell states are compatible with harboring the reservoir.

Advancements in high-throughput single-cell omics have enabled comprehensive characterization of HIV infection at the cellular level, facilitating the exploration of the latent reservoir through single-cell annotation and transcriptional analysis. Latency has been associated with cells with greater proliferative potential and survival signaling[25,32,33]. In addition to central memory cells, multiple studies have implicated CD4 cytotoxic T-cells as a significant reservoir cell type due to their proliferative potential in response to antigen[34,35]. While no single marker or surface phenotype can define latently infected cells, single-cell data augments our capacity to describe the complexity and heterogeneity of the latent reservoir.

Latently infected cell line models and ex-vivo interrogation of CD4 + T-cells from patient samples have also provided valuable insights into latency[36–40]. These studies probe cellular states that are compatible with quiescent HIV, but they do not reflect the selective pressures that infected T-cells encounter in human lymphoid tissues in vivo. On the other hand, humanized mouse models can support the development of near-complete immune lineages that faithfully recapitulate many aspects of HIV-1 pathogenesis[41–43].

Here, we introduce a novel humanized mouse model enabling genetic tracing of latent and productive HIV-1 infection in T-cells. This HIV-induced lineage tracing (HILT) model employs a fluorescent switch triggered by HIV-1 infection, persisting even if the provirus becomes silenced, allowing permanent marking of HIV-infected cells in the mouse and enabling lineage tracing of HIV infection history. Through flow sorting of HILT-marked cells and conducting scRNA-Seq profiling, we find that diverse CD4+ T-cell lineages support acute HIV infection and we observe enrichment of persistent HIV in certain memory T-cell subsets in ART-treated animals. Transcriptomic analysis identifies key pathways and genes associated with HIV latency and persistence during ART treatment across many T cell lineages, offering potential drug targets for altering latency status or infected cell survival.

Results

A two-component HIV-1-induced lineage tracing (HILT) system

To study infected and latent cell populations, we developed a genetic system to mark HIV-1 infected cells irreversibly. The approach utilizes a two-component, Cre-lox-based recombination system to report on the HIV-1 infection status of cells. The system involves a lentiviral construct (Lenti-RG) carrying a “red-to-green” Cre-lox-based cassette (Fig. 1A). This cassette contains the dsRed gene flanked by two LoxP sites, blocking eGFP translation until exposure to Cre recombinase triggers GFP expression[44]. The Cre recombinase is expressed from an HIV-1 clone (NL-CreI), replacing nef. Nef expression is restored with an internal ribosome entry site (IRES) (Fig. 1B). This HIV-1 construct is called HIV-1 NL-CreI. It has been employed here in a version carrying an NL4-3 envelope (X4-tropic) or a clade B transmitted founder envelope from HIV RHPA (R5-tropic). To test the efficiency of the two-component system, we transduced the Jurkat T-cell line with the Lenti-RG virus and analyzed the cells for dsRed expression using flow cytometry (Fig. 1C). Upon transduction with the Lenti-RG vector 39% of cells were positive for dsRed expression (Fig. 1D). This was followed by infection of the transduced cells with HIV-1 NLCreI (X4-tropic) and a phenotypic switch from red to green was observed in 3% of the total Lenti-RG-transduced population. The red-to-green switch was blocked by the reverse transcriptase inhibitor, Azidothymidine (AZT), indicating that productive HIV infection is required for the phenotypic switch (Fig. 1D). During the acute infection of Jurkat cells, the upregulation of GFP occurs relatively early in infection, the expected decrease in dsRed expression was not apparent at 2 days post-infection. In a separate infection 3 days post infection, dsRed is downregulated in the GFP positive fraction (Fig. S1), suggesting that the decay of dsRed occurs after GFP is upregulated.

The red-to-green switch enriches cells that carry HIV-1 proviral DNA

We next tested if the switched GFP+ cells were enriched for HIV-1 NLCreI proviral DNA. We infected Lenti-RG-transduced Jurkat cells with HIV-1 NLCreI and sorted 5000 unmarked, dsRed+/GFP- (marked targets), and GFP+ (switched) cells each (Fig. S1A). Genomic DNA extracted from sorted populations was directly measured for HIV-Cre DNA content by qPCR assay detecting the cre gene. Two different cre-specific PCR primer probes (Table S1) were used to measure the relative level of enrichment of HIV-1 NL-CreI in unmarked, marked (dsRed +/GFP-), and switched (GFP+) samples in comparison to the control genomic locus, RNaseP. We observed a >100-fold enrichment of HIV-1 NL-CreI proviral DNA in DNA isolated from sorted GFP+ cells in comparison to sorted dsRed+/GFP- or sorted unmarked cells (Fig. 1E). This data supports the idea that the HILT system effectively enriches for rare HIV-1 DNA+ infected cells.

The HILT-switched cells infected in humanized mice exhibit enrichment of HIV DNA

To measure the enrichment of HIV-1 proviral DNA in HILT-marked, GFP + cells in humanized mice, we used HIV NL-CreI (RHPA) to infect 3 huPBL mice engrafted with Lenti-RG transduced human peripheral blood mononuclear cells (PBMC) and 3 mice with a replication-competent GFP-expressing HIV, HIV GFP (RHPA) for comparison (Fig. 1F). Infection of humanized huPBL mice with the HIV RHPA Cre-I virus results in high viral titers of up to 10^7 copies/ml at peak viremia and showed a similar range of infectivity to the HIV-GFP virus(Fig. 1G). At 21 dpi following peak viremia, splenocytes from these mice were collected and sorted into GFP+, dsRed+/GFP-, and double negative (DN) fractions (Fig. 1H). The DNA extracted from sorted cells was measured for the enrichment of HIV DNA in the GFP+ fractions by DNA PCR assay. A 1600-fold enrichment for HIV-Cre DNA was observed in sorted GFP+ cells compared to dsRed+/GFP- cells. We note that the dsRed+/GFP- fraction was not enriched in HIV DNA as compared to the unmarked fraction, indicating to us that the GFP provides a strong indicator of an infection with an intact cre-expressing provirus (Fig. 1I). These results support that in humanized mice the switched GFP+ cells are highly enriched in HIV NL-CreI provirus. The low background DNA PCR signal in the unswitched cells also indicates that we are not missing a large percentage of silent integration events where there is insufficient cre expressed from an integrated provirus that does not activate the red-to-green switch.

https://doi.org/10.1038/s41467-025-57368-7

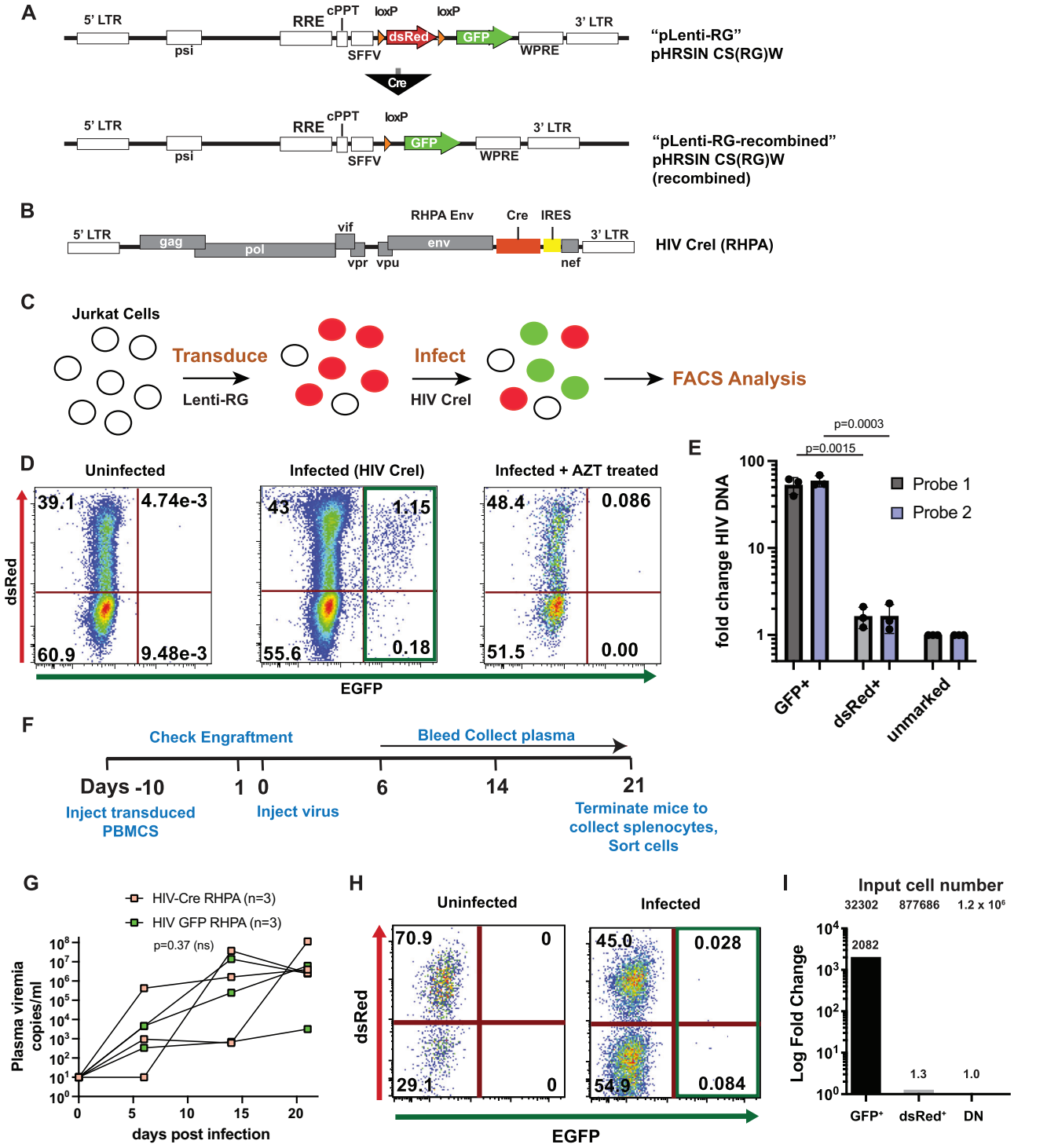
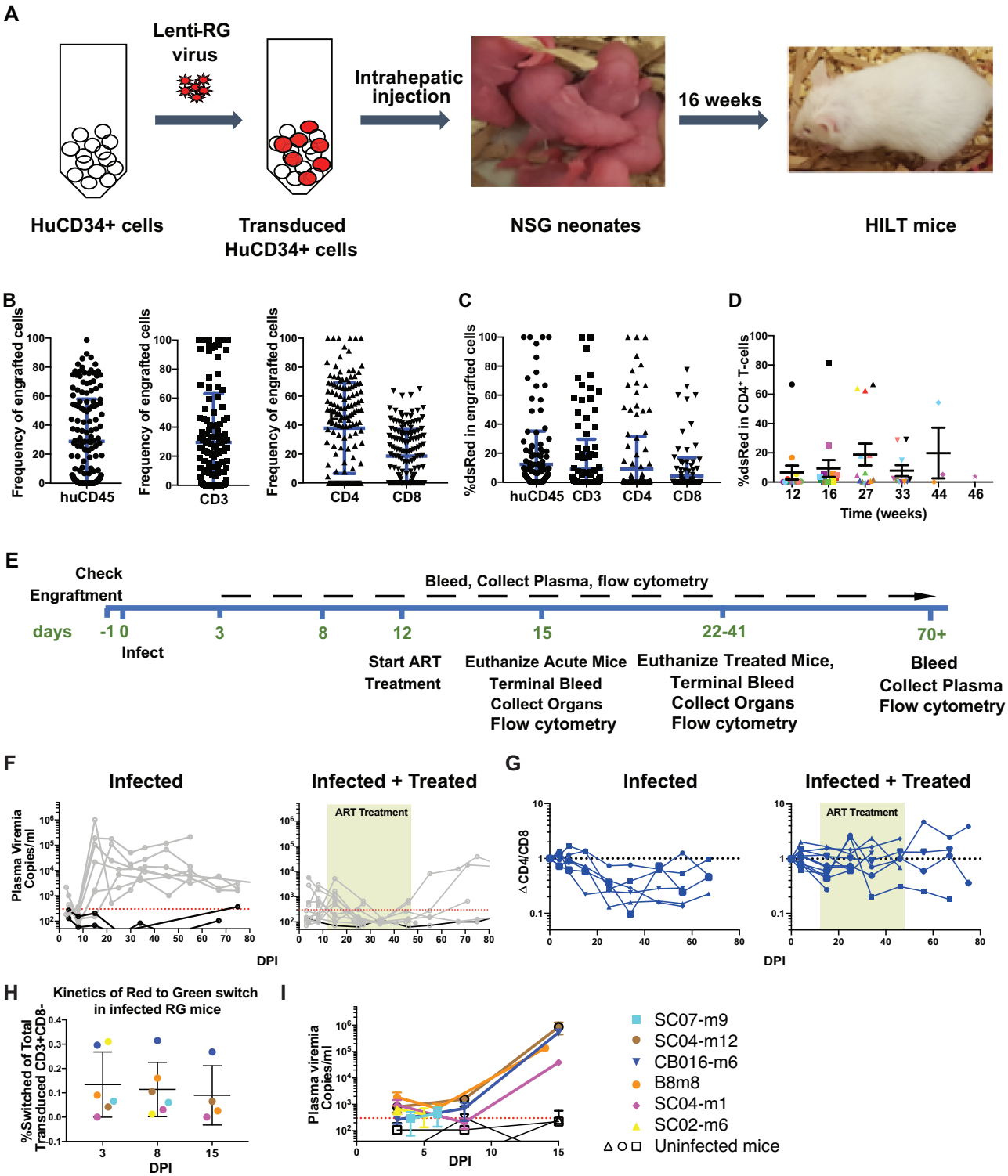


Fig. 1 | HIV-1 induced lineage tracing (HILT) system enables the tracking of infected cells both in vitro and in vivo. A Lentiviral vector pHR SIN CS(RG)W (Lenti-RG) carries a floxed dsRed gene followed by a GFP gene, which switches from red to green in the presence of Cre. **B** HIV-1 NL-CreI (RHPA), an HIV-1 clone that carries cre in place of nef, and nef expression is restored with an internal ribosome entry site (HIV- Cre). The virus carries an Env gene from a transmitted founder virus, RHPA. **C** Schematic representation of infection. Jurkat T cells were transduced with Lenti-RG and infected with HIV-Cre. **D** Flow cytometry plots of red to green switch in vitro in Jurkat cells after infection with NL Cre I and treated with AZT. **E** Bar plots represent log fold change of HIV Cre DNA in GFP+, dsRed + (GFP-), relative to double negative (DN) cells. The signals from two different primer-probe sets are illustrated in blue and black. Data represent mean +/- SD from $n = 3$ cells separate infections, Unpaired t test (**F**) Experimental timeline for infection of transduced HuPBL mice with HIV-Cre and HIV-GFP virus. (**G**) Plasma viremia of HIV-CreI monitored over time in infected Hu-PBL mice in comparison to a HIV-expressing HIV, NL-GI ($n = 3$ infected vs $n = 3$ uninfected), unpaired t test, $p = 0.37$ (ns). **H** Splenocytes from HIV-Cre infected HuPBL Lenti-RG mice were sorted at 21 dpi for GFP+, dsRed + (GFP-), and DN cells, and (**I**) Sorted cells from infected huPBL mice (M5 and M6) were flow sorted and pooled by phenotype, yielding 32,302 GFP+ cells, 877,686 dsRed+ cells, and 1.2 ×10^6 double negative (DN) cells to make DNA for quantitative PCR analysis. Q-PCR for HIV DNA shows log-fold change relative to the double negative population. Bar plots representing high enrichment of HIV-Cre DNA in the GFP+ switched cells in HuPBL mice.

https://doi.org/10.1038/s41467-025-57368-7



Humanized mice with dsRed expression in immune lineages (HILT mice)

Human CD34+ hematopoietic stem cell engrafted mouse models (huHSC) support the study of long-term HIV-1 infection as well as latency during ART therapy[42,45,46]. A strength of this model is the development of diverse hematopoietic cell lineages, including naïve and memory T-cell subsets that are susceptible to HIV. Human cord blood-derived CD34+ cells were transduced by Lenti-RG as described by Wang et al.[47] and injected into 0 to 3 days old neonates of NSG (NOD.Cg-Prkdc^{scid} Il2rg^{tm1Wjl}/SzJ) mice (Fig. 2A). At 12–16 weeks diverse human immune lineages developed from transduced CD34+ cells (Fig. 2B). An average of 28.75% were huCD45 cells. 29.54% of these cells were CD3+ T-cells. The distribution of CD4 and CD8 in the CD3 compartment was 37.9% and 18.62%, respectively (Fig. 2B, $n = 183$). Engrafted human cells were found in the peripheral blood, the spleen, lung, and bone marrow (Fig. S3A, B, C, D).

Transduction of the immune lineages with Lenti-RG was detected by measuring the dsRed positive population in huCD45, CD3, CD4, and CD8 compartments in the peripheral blood (Fig. 2C). dsRed+ cells were observed in all the lineages. Since CD4+ T-cells are the major targets for HIV-1 infection, we selected well-engrafted mice with more than 20% dsRed+ CD4 T-cells to challenge HIV-1 NL-CreI. Lineage-marked,

Fig. 2 | Strategy for generating HILT mice. A Newborn NSG mice are injected with Lenti-RG transduced cord blood-derived HuCD34+ hematopoietic stem cells and are engrafted with human immune cell lineages in 16 weeks. **B** Percentages of total peripheral blood cells in reconstituted NSG mice measuring huCD45, CD3, CD4, or CD8, respectively, at week 16 ($n = 184$ mice, data represented as mean +/− SD). **C** Frequency of dsRed+ cells in huCD45, CD3, CD4 and CD8 cells. ($n = 184$ mice, data represent mean +/− SD). **D** The percentage of engrafted CD4+ dsRed+ T cells throughout the lifetime of the mice ($n = 14$ mice, data represented as mean +/- SD), with each individual mouse represented as the same color filled, with a different symbol at each time point. (**E**) Experimental timeline of engraftment and infection in days. 17 well-engrafted mice were infected with HIV-Cre. At 12 dpi, mice were either treated with antiretroviral (ART) therapy or left untreated, and plasma viremia was determined for each mouse over the course of the infection. **F** Plasma viremia (copies/ml) for infected mice as determined by RT-qPCR assay specific for HIV-1 ($n = 7$). Plasma viremia (copies/ml) for infected and treated mice as determined by RT-qPCR assay specific for HIV-1 ($n = 10$). The red dotted line depicts the limit of detection (~300 copies/ml) for the RT-qPCR assay. The olive-green shaded box shows the duration of ART treatment. **G** The change in CD4/CD8 ratio was plotted over time following infection. The CD4/CD8 ratio changes over time in infected and ART-treated mice. The black dotted line represents the fractional deviation from the starting CD4/CD8 ratio, set to 1 for each animal. **H** Graph showing percentage switching normalized to total transduced (dsRed) targets in peripheral blood for infected HILT mice ($n = 6$, bars indicate mean +/− SD). **I** Plasma viremia (copies/ml) for infected mice as determined by RT-qPCR assay specific for HIV-1. (Infected mice $n = 6$, uninfected mice, $n = 3$). The dotted line indicates the background detection level of the assay.

dsRed+ cells were detected over time within the CD4 compartment in engrafted mice (Fig. 2D), suggesting variability of dsRed expression. Additionally, our observations revealed uniform dsRed expression in various organs beyond peripheral blood, indicating a consistent distribution of dsRed throughout each organ (Fig. S3H).

HILT mice recapitulate the features of HIV-1 pathogenesis

To validate that HILT mice can be used to study long-term HIV-1 infection as well as latency, we infected HILT mice with HIV NLCreI (RHPA), treated with ART, and monitored plasma viremia over time (Fig. 2E). To study HIV-1 persistence in HILT mice, 17 mice were infected in 3 independent experiments. Seven mice out of 17 were followed for acute infection, while ten were ART-treated starting day 12 dpi. Seven acutely infected mice exhibited a peak viral load at around 15 dpi with a mean of 1.5 x 10^5 copies/ml. By day 25, the viral load was down to a mean of 6.6 x 10^3 copies/ml. Two mice in the cohort were followed for up to 105 days. They showed sustained plasma viremia throughout this period (Fig. 2F).

Ten mice out of 17 infected were ART-treated starting day 12 post-infection, and their plasma viral load followed over time. Viral loads were suppressed in treated animals about ten days after treatment started, with viremia falling below the limit of detection of the assay (Fig. 2F). At 49 dpi, the mice that were taken off treatment exhibited a rebound in viremia (Fig. 2F). The viral rebound was confirmed in 3 surviving mice out of 10 that were treated. The HIV-I NL Cre-I infected HILT mice show key features of infection, including sustained virus infection, suppression by ART, and viral rebound after ART interruption, suggesting that they are suitable for studying HIV-1 persistence.

HIV-1 NL-CreI infection led to CD4 T-cell depletion, as measured by CD4/CD8 ratio changes from the baseline (Fig. 2G). CD4 counts dropped below the baseline (baseline set to 1 for day 0 dpi) as the infection progressed, showing an overall decrease in CD4 and an increase in the CD8 population. In contrast, the mean CD4/CD8 ratio in uninfected mice did not fall below the baseline and trended towards greater than one throughout the time course (Fig. S3J-L).

GFP expression is induced by infection with HIV-Cre in HILT mice

While establishing the features of HIV-1 infection in HILT mice, we also examined if the infection could be followed in these mice using the red-to-green switch as an in vivo proxy for HIV infection status. For this study, six well-engrafted mice with more than 20% dsRed expression in the CD4+ compartment (Fig. S3H, I) were infected in 4 independent experiments to determine red to green switch in vivo in HILT mice. The infected mice were bled to determine plasma viremia, and the HIV-Cre-activated red-to-green switch was monitored in their peripheral blood (Fig. 2H, I). Switched green cells were observed in peripheral blood during acute infection and arose with plasma viremia. The average plasma viremia at 14-15 dpi (terminal time point) was 3.7 x 10^5 copies/ml. Mice were euthanized at 15 dpi, and spleen, lung, and bone marrow were harvested to measure the levels of switched cells. The red-to-green switch in CD3 + CD8- cells in peripheral blood, spleen, lungs, and bone marrow was observed at 15 dpi (Fig. S3H). Additionally, a red-to-green switch was not observed in the CD8+ T-cell compartment in the infected mice, confirming the high specificity of switching due to infection in the CD4 + T-cell compartment (Fig. S3H). Red-to-green switch was not observed in any of the uninfected mice. On average, 0.071%, 0.073%, 0.076%, and 0.13% CD3 + CD8- cells switched in peripheral blood, spleen, lung, and bone marrow, respectively (Fig. S3I, Table S2). In summary, HILT mice show the capacity to reveal HIV infectivity through the red-to-green switch mechanism.

Single-cell RNA sequencing identifies HIV-1 transcripts in HILT mice

Human CD4+ T-cells are composed of diverse subsets and activation states that can be partly resolved by their distinct transcriptional states[32,48]. To assess the transcriptomic states of infected CD4 T-cells and their persistence at single cell level in the HILT-marked huHSC mice, scRNA-seq was performed on four types of sorted CD3 + CD8- splenocytes: (1) dsRed+, (2) Unmarked, (3) GFP+ and (4) Mixed cells (dsRed+ and GFP+) of three acutely infected mice, two 10-days treated mice, one 29-days treated mice, and two uninfected mice (Fig. S4A, B).

After sequencing, reads were analyzed with the Cell Ranger pipeline[48] using default parameters and aligned to our custom reference genome, which incorporated six regions on HIV clone, NL-CreI (RHPA), based on HIV splice donor (D1 and D4) and acceptors (A1 and A7) (Fig. 3A)[49]. Across these datasets from infected and treated mice, we observed a wide range of HIV transcript counts to over 1000 in a total of 989 cells (Fig. 3B). In this analysis we grouped HIV RNA+ cells based on the number of HIV transcripts detected, defining HIV RNA+ low cells as containing 2-10 HIV UMIs and HIV RNA+ high cells as containing ≥11 HIV UMIs. There was a total of 479 HIV+ low and 510 HIV + high cells observed throughout our dataset. When looking at the overall distribution of HIV transcripts in acutely infected versus ART-treated mice, nearly all of the HIV RNA+ cells were from acutely infected mice. With 315, 369, and 301 HIV RNA+ cells in all three acute mice and 4 HIV RNA+ cells in one treated mouse (T1) (Fig. 4E). No HIV RNA+ cells were detected in other treated or uninfected mice (T2, T3, U1, and U2).

We profiled 47,850 CD4 T-cells from eight mice, including 19,986 from acutely infected, 10,391 from treated, and 17,473 from uninfected mice, regardless of the presence or absence of HIV transcripts in the cells (Table S3). Batch effects were corrected using fastMNN[50]. After individual QC, 40,782 cells were integrated, clustered, and visualized based on a manual annotation, predicted CD4 T-cell type, the number of HIV transcripts, experimental conditions, and individual mouse IDs, respectively, using Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) (Fig. 3C–G). Unsupervised clustering revealed fourteen clusters that were annotated based on the expression of marker genes selected from literature (Fig. 3C, Table S4)[51–53]. We identified helper CD4 T-

https://doi.org/10.1038/s41467-025-57368-7

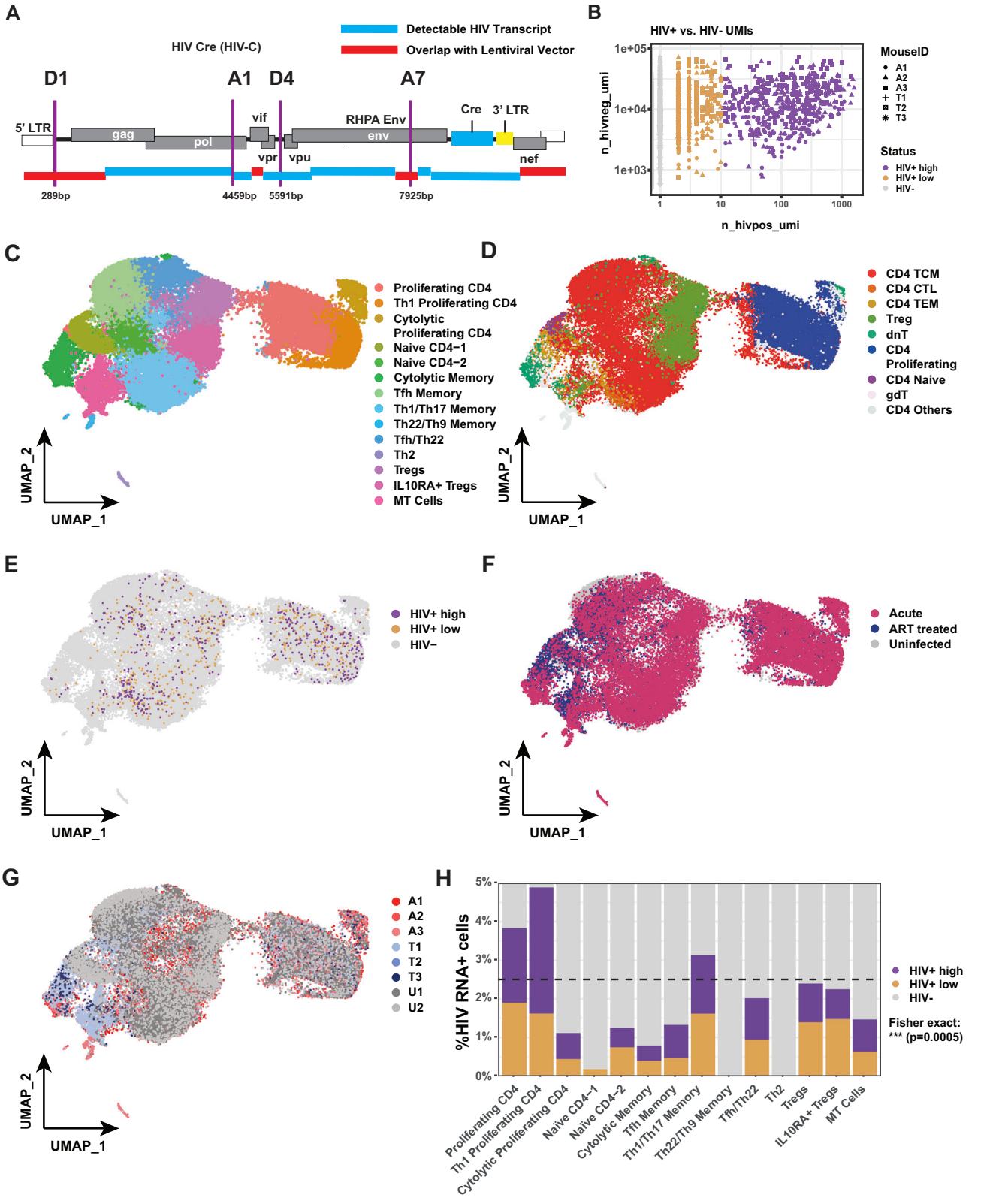


Fig. 3 | Single-cell RNA sequencing identifies HIV-1 transcripts in HILT mice. **A** Segmented HIV Cre I genome with HIV transcript and overlap annotations (blue and red) used in the 10x Genomics Cell Ranger Count custom reference. Locations of splice donors (D1 and D4) and acceptors (A1 and A7) within the genome of HIV Cre I are determined based on splice sites within the HIV-1 genome53. **B** Representative scatter plot of total HIV UMIs detected in cells across acutely infected and treated mice in proportion to total cellular UMI counts. Cells with >2 transcripts were defined as HIV RNA+ (n = 989). HIV RNA+ low with a range of 2 to 10 HIV transcripts per cell (n = 479) and HIV RNA+ high encompassing cells with >11 transcripts (n = 510). UMAP of all cells (n = 40,782). Cells were colored based on identified cell type (**C**), predicted cell type (**D**), HIV RNA status (**E**), experimental condition (**F**), and mouse ID (**G**). (Mouse ID: A1-A3 acutely infected mice, T1-T3 ART-treated mice, U1-U2 uninfected mice) (**H**) Bar plots representing HIV-1 RNA+ high and HIV RNA+ low cells across identified CD4 T cell subsets. The black dotted line shows the average distribution of HIV RNA+ cells (two-sided Fisher exact test: $p = 0.0005$).

https://doi.org/10.1038/s41467-025-57368-7

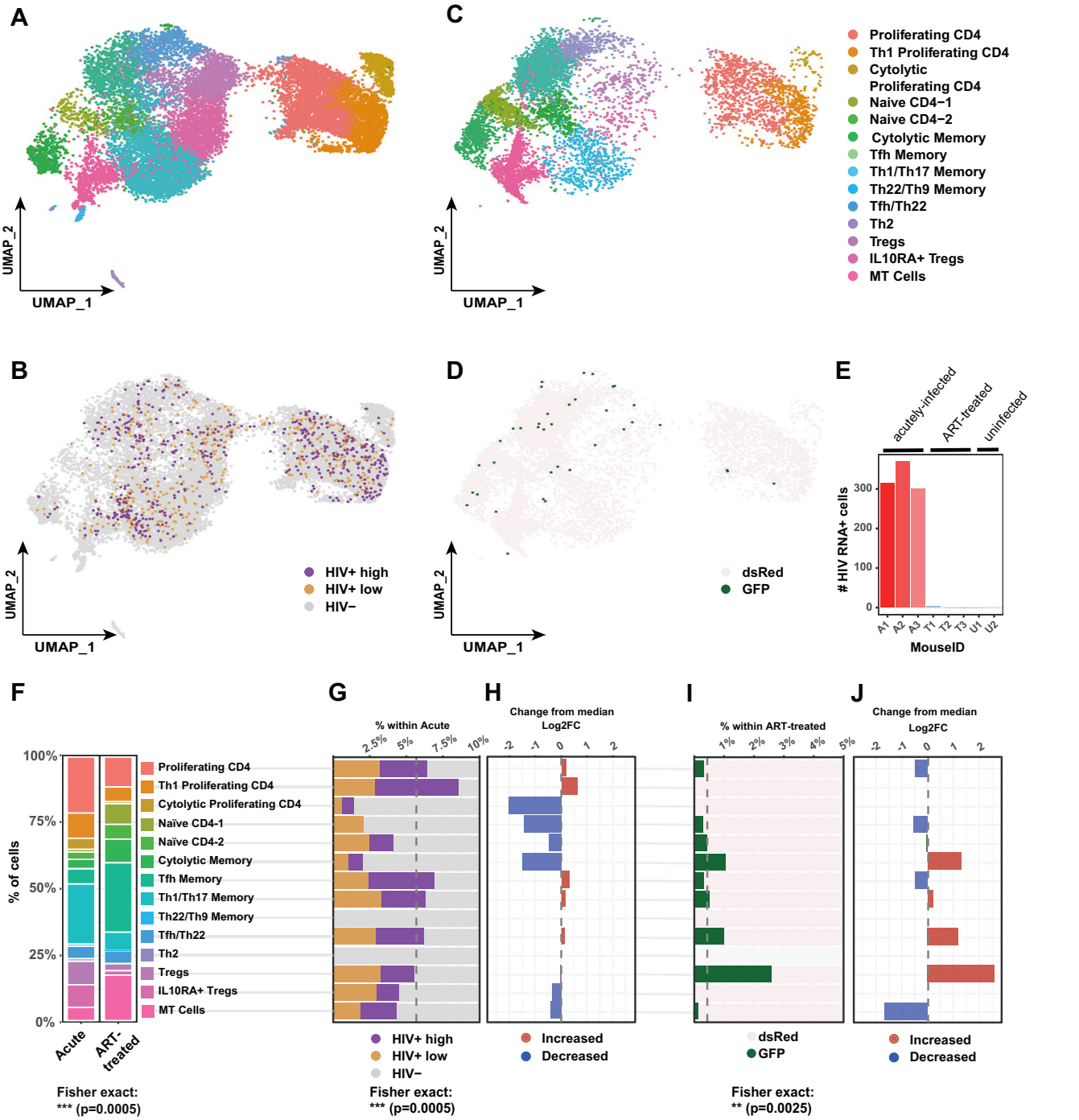


Fig. 4 | Distribution of HIV Transcript positive cells and HILT marking of putative latent cells across CD4 T-cell subsets in acutely infected and ART-treated humanized mice. Distribution of HIV-1 RNA across diverse CD4 T-cell clusters in acutely infected HILT mice and GFP+ cells are enriched among clusters within ART-treated mice (**A**) UMAP of cells from acutely infected mice A1-3 ($n = 17{,}346$) highlighting cell types. **B** UMAP highlighting the location of HIV RNA+ cells with high or low transcript detection ($n = 985$). **C** UMAP of cells for representative ART-treated mice T1/T2 ($n = 8503$) highlighting identified cell types. **D** UMAP highlighting the location of HILT-marked GFP+, putative latently infected cells ($n = 37$). **E** Number of HIV RNA+ cells isolated from each mouse in the pooled analysis, acutely infected (A1-A3), ART-treated (T1-T3), and uninfected mice (U1-U2). **F** Proportion of cell types within acutely infected and ART-treated mice datasets (two-sided Fisher exact test, $p = 0.0005$). The black dotted lines show the average percentage GFP+ across all T cells. **G** Distribution of HIV RNA+ cells within identified cell types within the acute dataset (two-sided Fisher exact test: $p = 0.0005$). **H** The deviation from the mean % HIV RNA+ cells among all cells, represented as a log2-fold change for each cluster. **I** Distribution of HILT-marked GFP+ cells within identified cell types (two-sided Fisher exact test: $p = 0.0025$). The black dotted lines show the average percentage of GFP+ across all T cells. Within the ART-treated mice, the percentage of GFP+ (putative latent) is represented for each cluster. **J** The deviation from the mean % GFP+ cells among all cells for each cluster represented as a log2-fold change.

cells, memory cells, naïve cells, proliferating cells, and regulatory T-cells (Tregs) and subsets within these larger groups. Using multimodal reference mapping to a reference PBMC dataset[54], we validated those designations and predicted nine distinct CD4 T-cell types (Fig. 3D). This reference mapping approach grouped the T cell types into a smaller number of predicted cell phenotypes.

HIV RNA+ cells were dispersed across most predicted CD4 T-cell types (Fig. 3D, E, Fig. S4D). HIV RNA+ cells were found in all predicted

CD4 T-cell subtypes except naïve and γδT cells, which were low in abundance (CD4 Naïve $n = 97$, γδT $n = 104$) and only found in five and four of all eight mice, respectively (Fig. S4D, Table S5). When comparing the UMAP distribution of cells by conditions (Fig. 3F, Fig. S5) or by individual engrafted mice (Fig. 3G), the overall distribution of cell types was broadly maintained.

We also examined the distribution of HIV RNA+ cells in manually annotated cell types and observed that both HIV RNA+ high and HIV RNA+ low groups were present and dispersed throughout most CD4 T-cells (Fig. 3H). We observed HIV RNA+ cells in all subtypes except Th22/Th9 memory cells and Th2 cells. We note that both these cell types were scarce within our datasets (Th22/Th9 memory cells $n = 146$, Th2 $n = 155$) and only found in three and four of all eight mice. We identified the highest number of HIV RNA+ cells (4.8%) in Th1 Proliferating CD4 T-cells (Fig. 3H, Table S6, Fisher test $p = 0.0005$). These results are consistent with the preference of HIV for active replication in activated, proliferating T cells, though low and high RNA expression is observed across a broad distribution of T cell lineages.

HIV-1 transcripts and transcriptionally latent GFP+ cells are enriched in T-cell clusters and present across heterogeneous CD4 T-cell clusters in both acutely infected and ART-treated mice

A focused analysis was conducted on cells from acutely infected mice (A1-A3). By narrowing our analysis to these datasets, we mapped the distribution of HIV RNA+ cells across cell types within acute infection (Fig. 4A, B). In mouse #A1, flow sorting enriched for GFP+ and dsRed+/GFP- cells, which were processed together and labeled as mixed cells (Fig. S4A, B). In mice #A2 and #A3, GFP+ cells and dsRed+/GFP- cells were processed as separate samples but sequenced together. In total, we analyzed 17,346 cells distributed across 14 identified cell types.

By annotating cells with the HIV RNA expression status, we assessed whether HIV expression of both high and low was associated with distinct CD4 T-cell types. Cells were distributed throughout the UMAP (Fig. 4B). We observed both HIV RNA+ high and HIV RNA+ low cells in a majority of identified cell types, except for low abundant Th2 and Th22/Th9 Memory cells, indicating that most, if not all, CD4 T-cell states can support active HIV infection (Fig. 4F). HIV RNA+ cells were detected almost exclusively in the acutely infected mice, and not the ART-treated or uninfected mice (Fig. 4E). The percentage of HIV RNA+ cells within the cell types was significantly heterogeneous (Fisher test, $p = 0.0005$), ranging from the highest 8.5% in Th1 Proliferating CD4 T-cells and the lowest 1.3% in cytolytic proliferating CD4 T-cells (Fig. 4F, Table S7). We evaluated the representation of acutely infected cells within each subcluster, and found increased representation in proliferating CD4, Th1 proliferating CD4 and Tfh memory cells, and decreased representation within cytolytic proliferating CD4, two naïve cell clusters, and cytolytic memory T cell clusters (Fig. 4G, H).

Within cells collected from our 15dpi acutely infected datasets, HIV RNA transcripts were detected in dsRed+ HILT-marked CD4 + T-cells with a high range of detected HIV-specific transcripts (Fig. 3B). Surprisingly, we did not detect HIV RNA in GFP+ sorted cells from those same datasets. We interpret this as indicating that at 15 days post infection, replicating viruses do not continue to express functional cre-recombinase after and that the replicating virus represents HIV that carries non-functional cre. We note however, that since we observed robust enrichment of HIV-Cre DNA within GFP+ cells from acutely infected humanized mice (Fig. 1I), we also conclude that HILT-marked GFP+ cells are highly enriched with HIV that are functionally defined by having expressed an active cre enzyme and then persisted in a transcriptionally latent state.

Next, we examined HILT-marked, GFP+ cells in ART-treated mice at 10- and 29-days post-ART initiation, when plasma viremia was undetectable. Flow-sorted GFP+ cells were isolated from two ART-treated mice (T1, T2) (Fig. S4A, B) and sequenced alongside dsRed+/GFP- sorted cells of the same mouse. Within the analyzed 8503 cells we observed up to 2.3% GFP+ ($n = 37$) cells across a majority of the thirteen clusters, notably no Th2 cells were identified within the ART-treated datasets (Fig. 4D, F, I, J). This observation could be explained by the low number of GFP+ cells as well as the rarity or lack of some of these observed cell types within treated mice (Fig. 4C–E). Furthermore, we observed a significant difference in the distribution of these cells across the cell types (Fisher, $p = 0.0025$) with enrichment of GFP+ cells within Tregs, cytolytic memory, and Tfh/Th22 cells (Fig. 4F, I, J, Table S7) and lower representation in proliferating cells, naïve CD4-1, and Tfh memory cells (Fig. 4H, J).

Notably, some of these cell types were significantly more abundant within this dataset in direct comparison to the acute data sets (A1-A3) (Fisher test, $p = 0.0005$) (Fig. 4F, Table S7). The proportion of proliferating cell types (Proliferating CD4, and Th1-Proliferating), Tregs (Tregs, IL10RA+ Tregs), as well as certain CD4 T-helper cell types Th2, Th1/Th17 Memory Th22/Th9 Memory cells were significantly increased within the acutely infected mice. Interestingly, we found an inverse effect in Cytolytic Memory and Tfh Memory proportions as well as naïve CD4 T-cells. Furthermore, the proportion of Tfh/Th22 cells was not significantly changed between both conditions. When analyzing these results, however, we note that the total number of cells analyzed from the acutely infected was almost twice the amount in the ART-treated dataset. These results suggest that the Infection with HIV-Cre and the GFP + HILT phenotype that marks cells as transcriptionally latent cells infected with HIV-Cre, are decreased in some cell types and enriched in others and can persist in a broad range CD4 T-cell types in our humanized mouse system.

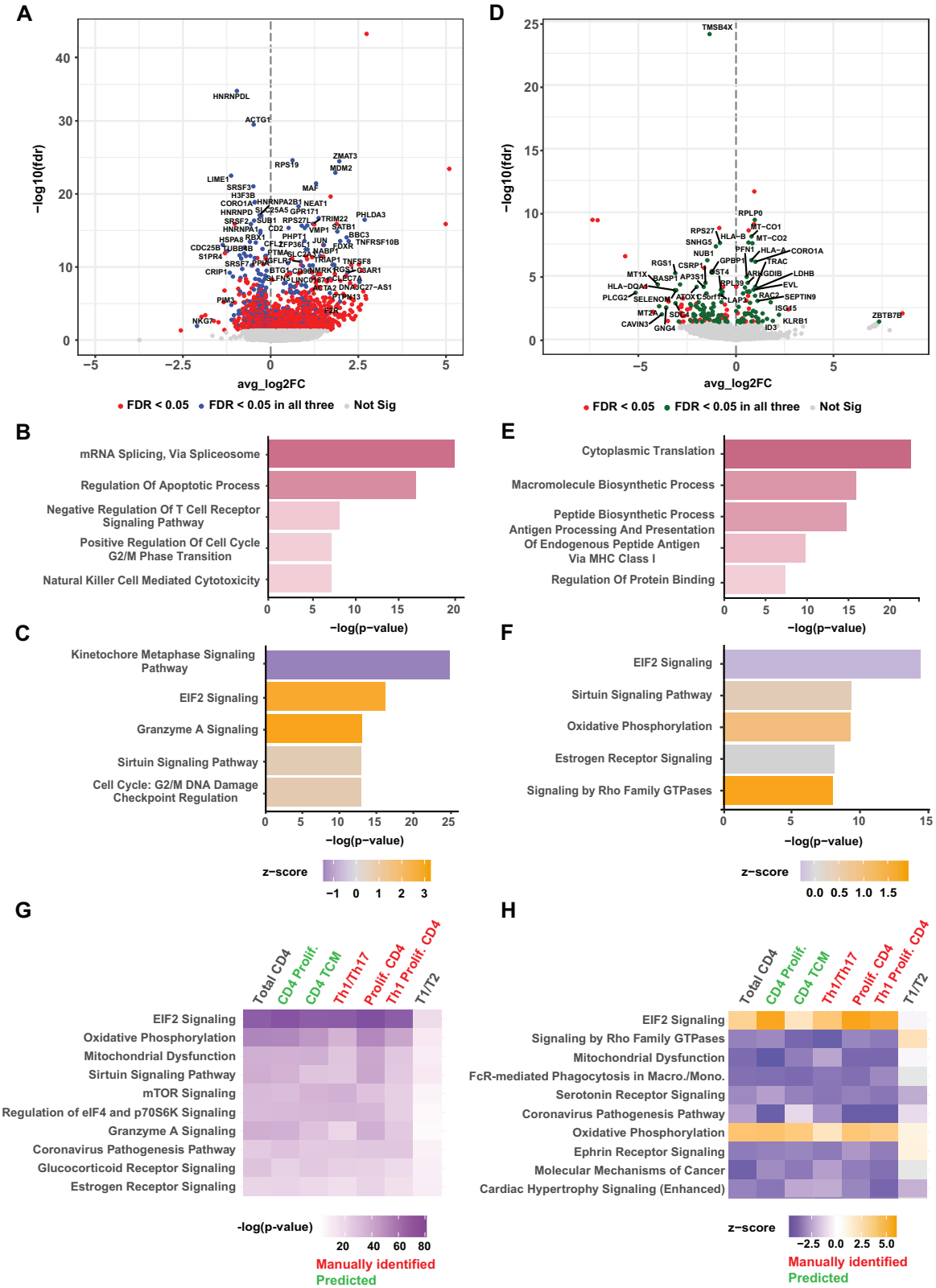
Acute and ART-treated datasets reveal distinct transcriptional patterns in overlapping pathways

We next examined datasets of acutely infected mice to identify transcriptional signatures associated with high HIV-1 RNA expression. A differential gene expression (DGE) analysis between two populations (HIV RNA+ high vs. HIV RNA-) was conducted initially on all CD4 T-cell subsets, independent of cell type designation, within three acutely infected mice datasets (A1-A3). The results of mouse #A3 are displayed as a volcano plot (Figs. 5A, S6A, B). We observed many highly downregulated RNA binding proteins in HIV RNA+ high cells, including HNRNPs and SRS proteins involved in RNA metabolism. Notably, HNRNPs have distinct interactions in viral synthesis and HIV-1 mRNA splicing[55,56], while several genes of the SRS family are associated with the downregulation of HIV-1 replication through control of alternative splicing of the viral genome[55,57–60]. In comparison, Genes like *JUN*, *MDM2*, and *BAX* displayed a positive log-fold change across all acutely infected mice (Figs. 5A, S6A, B), reflecting alterations in T-cell activation (*JUN*), proliferation (*MDM2*), and apoptosis (*BAX*). Notably, BAX, a well-studied proapoptotic protein, is activated by Vpr[61,62].

We then focused on the top 145 common significant DEGs found in all three acute datasets in the Enrichr gene set enrichment analysis[63–65]. HSPA8 and HNRNPA2B1 were connected gene sets relating to mRNA splicing and Regulation of the Apoptotic Process. *JUN* and *PHLDA3*, which were upregulated in Fig. 5A, were also identified to be part of the apoptotic regulation gene set and the positive regulation of cell cycle G2/M phase transition (Fig. 5B). To examine predicted pathway perturbations that resemble pathways changed during acute infection, we employed Ingenuity Pathway Analysis (IPA) (Fig. 5C). Most significant was the Kinetochore metaphase signaling pathway, associated with cell cycle regulation. In addition, EIF2 Signaling, Granzyme A Signaling, Sirtuin Signaling, and G2/M DNA Damage checkpoint regulation in the cell cycle were among the top pathways by p-value and were upregulated during acute infection.

We analyzed GFP+ cells surviving ART treatment for 10 or 29 days, all of which were HIV RNA-negative. DGE analysis compared GFP+ and GFP- cells within the treated mice dataset (#T1-T2). Upregulation of

https://doi.org/10.1038/s41467-025-57368-7



MHC Class-I genes (*HLA-A*, *HLA-B*) and downregulation of Class-II MHC gene (*HLA-DQA*) were observed. Additionally, ribosomal genes associated with translation (e.g., *RPS* and *RPL*) exhibited high log fold change in GFP+ cells (Fig. 5D), linked with cytoplasmic translation and macromolecule biosynthetic processes in transcriptionally repressed cells. In ART-treated mice, gene set enrichment analysis of 117 genes (Fig. 5E) revealed significant pathways related to biosynthesis of molecules in translation and antigen processing and presentation via MHC class-I, with higher significance compared to our analysis of acute infection.

IPA of the ART-treated dataset revealed a high degree of overlap of the top pathways in the acutely infected mouse. The EIF2 Signaling pathway was upregulated in acutely infected cells and downregulated in GFP+ cells. While the Sirtuin pathway, Oxidative phosphorylation,

Fig. 5 | Differential gene expression analysis of cells within acute and ART-treated datasets reveals distinct transcriptional patterns in overlapping pathways. A Volcano Plot for DGE analysis of HIV RNA+ high vs HIV RNA- cells in 15 dpi acutely infected mouse ID A3. Red-colored dots represent any differentially expressed gene <0.05 FDR, while blue color dots represent genes with <0.05 FDR found in all three datasets of acutely infected mice (common genes). **B** Enrichr gene set enrichment analysis of common genes within acutely infected mice displaying top 5 gene sets by -log p-value (one-sided Fisher's exact test, Benjamini-Hochberg multiple testing correction). **C** IPA analysis of the top 5 enriched canonical pathways by p-value (one-sided Fisher's exact test, Benjamini-Hochberg multiple testing correction) and colored by z-score. **D** Volcano Plot for HILT marked dsRed-/GFP+ cells vs HILT marked dsRed+/GFP- cells within 10-days and 29-days ART-treated mice T1/T2. Red dots represent any differentially expressed gene <0.05 FDR threshold, while green color dots represent genes <0.05 FDR found in all mice. **E** Enrichr gene set enrichment analysis of common genes within ART-treated mice displaying top 5 gene sets by -log p-value (one-sided Fisher's exact test, Benjamini-Hochberg multiple testing correction). **F** IPA shows the top 5 enriched canonical pathways by p-value (one-sided Fisher's exact test, Benjamini-Hochberg multiple testing correction) and colored by z-score in ART-treated mice. Pathway comparison analysis of total CD4 cells and subsets of CD4 cells sorted by -log pvalue (one-sided Fisher's exact test, Benjamini-Hochberg multiple testing correction) (**G**) and z-score (**H**). HIV RNA+ high cells ($n = 508$) from acutely infected mice vs uninfected dsRed+/GFP- cells ($n = 7504$) from uninfected mice U1/U2 with HILT-marked GFP+ cells vs HILT marked dsRed+/GFP- cells within 10-days and 29-days ART-treated mice T1/T2 were used.

and Signaling by Rho family GTPases show positive z-scores with higher p-values (Fig. 5F). To determine the extent to which initial differential expression analysis is influenced by differences in the infection rate among T cell subpopulations, we examined whether the same pathways are affected within CD4 T-cell subsets within populations with the highest number of HIV RNA+ high cells. In this analysis, we examined two predicted cell types from reference map annotation and three manually annotated Seurat cell clusters (Fig. 5G, H). Comparative analysis of HIV RNA+ high cells vs. uninfected cells showed that the top pathways remained highly significant across all T-cell subsets, with p-values much greater significance than within the acute mice themselves (Fig. 5C). This indicates that the main pathways enriched in acutely infected cells are common to different T cell subsets. Across T cell subsets, the EIF2 signaling pathway was consistently the most significant, with high p-values and positive z-scores across subsets except in the treated condition (Fig. 5G, H). Oxidative phosphorylation and Sirtuin signaling pathways were also consistently dysregulated across subsets, indicating their involvement in acute infection and transcriptionally latently infected cells.

Top DGE pathways in acutely infected humanized mice resemble those identified in acutely infected human patient-derived cells

To assess the comparability of humanized mouse T-cells to those in humans, we validated our pathway analysis results against a study by Collora et al., 2022[34], which used ECCITE-Seq to profile 267 HIV-1 RNA+ cells from six human patient samples. Implementing their DGE results within our analysis, we observed remarkably similar pathways with high statistical significance when compared to total CD4 T-cell populations from acutely infected humanized mice (Fig. 6A, B). The top pathways, including EIF2 Signaling, Oxidative Phosphorylation, Sirtuin Signaling, and Protein Ubiquitination, were consistently top-ranked according to p-value. Our model system yielded p-values of much greater significance, due to higher numbers of HIV RNA+ cells from our acutely infected in vivo samples (Fig. 6A). Despite this difference, the z-score directionality of top pathways remained consistent, further highlighting the utility of the HILT system (Fig. 6B).

Acutely infected T cells and latently infected T cells perturb similar differentially expressed pathways but frequently modulate them in opposite directions

We next conducted a focused comparative pathway analysis directly comparing all three acute and treated datasets separately (Fig. 6C, Fig. S7A). In our comparison analysis, we again found the EIF2 signaling and Sirtuin pathways, with higher significance in HIV RNA+ high cells in acute infection and in GFP+ cells in treated mice. When we examined the expression of chosen genes associated with three selected pathways EIF2, Sirtuin, and protein ubiquitination, some of the genes were differentially expressed in acute versus ART-treated conditions (Fig. 6D–F). In the EIF2 pathway, *BCL2*, *RPS27L*, and *AGO2* were highly downregulated in treated GFP+ cells. Whereas *HNRPA1*, *EIF2S1*, *EIF3M*, *PPP1CA*, *PTBP1* were downregulated in HIV RNA+ high cells in acute Infection (Fig. 6D). *JUN*, *GADD45A*, *NAMPT*, *XPC*, and *BAX* showed diminished expression in the treated condition in the Sirtuin pathway (Fig. 6E). In protein ubiquitination, which is related to cell cycle, genes like *MDM2*, *DNAJB9*, and *USP9Y*, showed enhanced expression in acute infection, while in GFP+ cells of treated mice, expression of these genes is suppressed (Fig. 6F). Besides these top three pathways, we also looked at genes associated with Kinetochore metaphase signaling, Coronavirus pathogenesis, and inhibition of ARE-mediated mRNA degradation pathways (Fig. S7 B–D). Overall, it is reassuring that in the changes we observe, the top gene expression changes associated with acute infection are driven in opposite direction in latency. We suggest that a comparison of these data sets elucidates key transcriptional differences that may predispose cells towards active versus latent HIV infection.

Discussion

It is important to identify and study latent cells without perturbation to understand the mechanisms that maintain HIV-1 latent reservoirs. Therefore, we developed a two-component HILT system, a powerful tool for studying HIV-1 infection dynamics, latency, and persistence within various cell populations. Our system employs a Cre-lox-based recombination mechanism to irreversibly mark and track HIV-1-infected and latent cells.

The lentiviral construct Lenti-RG, carrying a red-to-green, Cre-lox-based genetic switch, has proven effective in reporting the infection status of cells. Upon exposure to Cre recombinase, the cassette switches from expressing dsRed to eGFP, allowing for the tracking of infected cells and revealing productive HIV-1 infection. Cre recombinase derived from a full-length HIV-1 clone, inserted in place of nef, ensures that only actively infected cells trigger the switch, as productive infection is required for Cre expression.

We validated the efficiency of the HILT system through various experiments, in Jurkat T-cells transduced with Lenti-RG, a notable proportion of cells switched from red to green upon subsequent HIV-1 infection. Importantly, this switch was sensitive to reverse transcriptase inhibition, confirming the dependence on productive infection for the phenotype change. Further studies in humanized mice demonstrated significant enrichment of HIV-1 proviral DNA in switched GFP+ cells compared to unmarked cells, highlighting the ability of the HILT system to enrich infected cells in vivo. We have thus succeeded in developing a system that can enrich for HIV-infected cells without perturbing the cells which has allowed us to use a single cell transcriptomics to identify gene signatures that are enriched in acutely infected and latently infected T cells.

Single-cell RNA sequencing approaches resolve heterogeneous cellular populations at a single-cell level. In our humanized mice we find that human CD4 T-cell subsets resemble CD4 T-cell populations in humans[66]. In the context of HIV-1 infection, studies have utilized primary cell latency models and in vivo patient samples[31,32,34,67]. Cohn et al.

https://doi.org/10.1038/s41467-025-57368-7

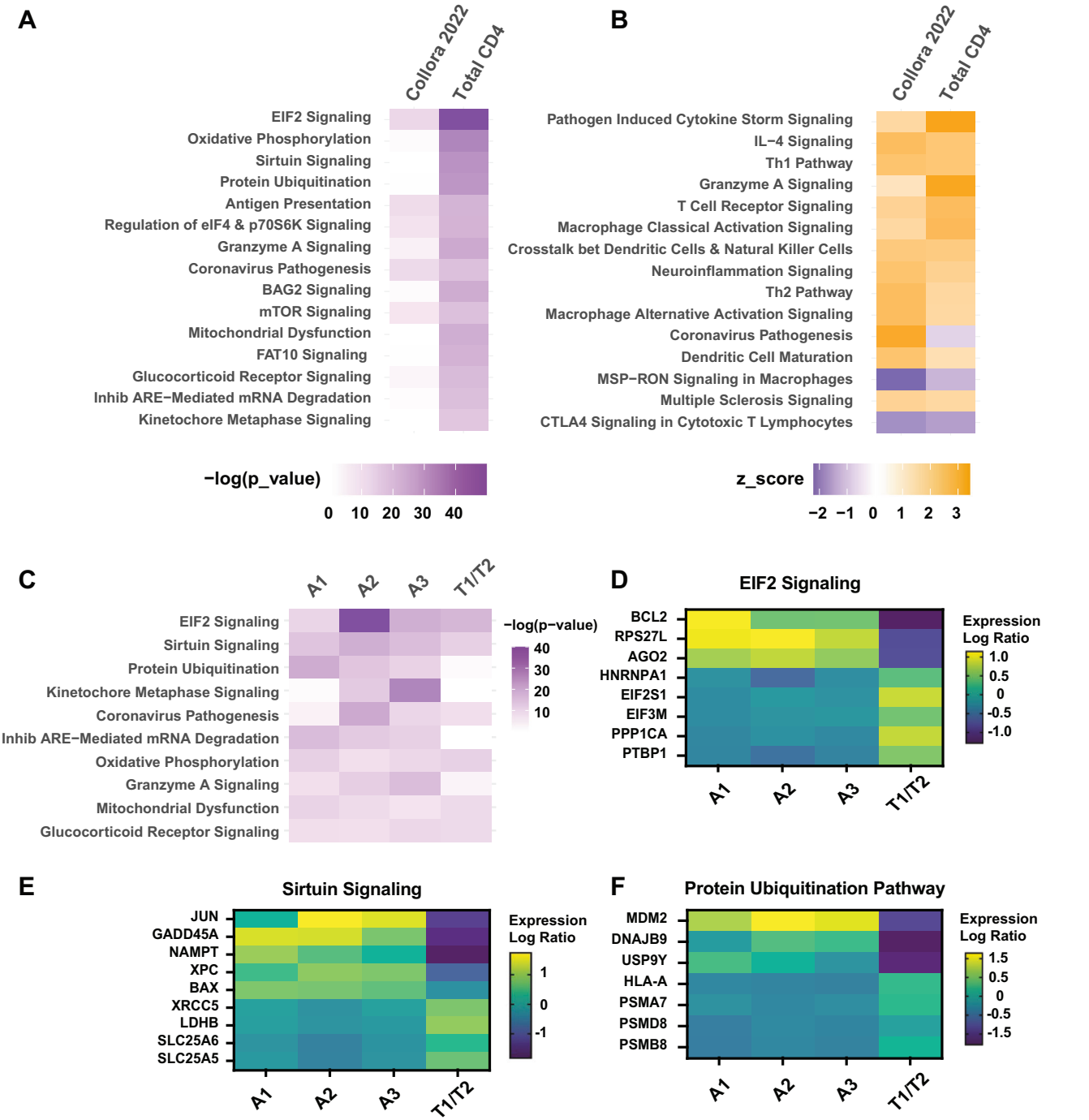


Fig. 6 | Comparison analysis of pathways in acutely infected human T cells and humanized mice comparing acutely infected with latently infected T cells. IPA pathway comparison analysis of between HIV-1 RNA+ cells of Collora et al., 2022 dataset and HIV-1 RNA+ cells from Total CD4 T-cells in our dataset represented in heatmap showing -log *P*-value (one-sided Fisher's exact test, Benjamini-Hochberg multiple testing correction) (**A**), and Z-score (**B**). **C** Comparison of top canonical pathways ranked by -log p-value across acutely infected mice A1-A3 and ART-treated mice T1/T2 (one-sided Fisher's exact test, Benjamini-Hochberg multiple testing correction). Expression log ratio represented in heatmaps of selected genes within EIF2 signaling (**D**), Sirtuin Signaling (**E**), and Protein Ubiquitination Pathway (**F**) across DEG of HIV RNA+ high vs. HIV RNA- cells within three acute (A1-A3) and GFP+ vs dsRed+/GFP- cells of two ART-treated mice datasets (T1/T2). The negative expression log ratio is colored blue, while the high is yellow, indicating a positive expression log ratio.

(2015) investigated cells harboring reactivatable CD4 T cell HIV-1 reservoirs by sequencing them after ex vivo activation. Unlike models that identify latent cells by reactivating them, the HILT humanized mouse model identifies unperturbed latent cells, marking cells with hidden provirus without exogenous stimuli. When comparing DGE pathways that define active vs latent infection in vivo, the analysis convergently identified the same pathways, however the directionality of the pathways was often perturbed in opposite directions. We suggest that these gene expression differences may define key transcriptomic features that regulate active vs latent infection in vivo. We also find that HIV infection is enriched in proliferating cells yet not uniquely restricted to specific CD4 T-cell subtypes as defined by transcriptomic profile. During ART persistent HIV expression is decreased in proliferating cells and enriched in memory T cell

populations including granzyme expressing cytolytic memory cells as has been described in T cells from people living with HIV.

Our investigation identified persistent CD4+ T-cells at ten- and 29-days post ART, despite plasma viremia being below detection limits. This differs from patients with long-term ART control, as studied by Finzi et al., 1999[68]. Rapid viral decay kinetics occur within the first few days of therapy, followed by slower kinetics, with DNA-positive cells decaying more slowly than RNA-positive cells. This study focuses on cells surviving the initial decay phase, on their way to establishing latent reservoirs. These cells show diminished HIV RNA expression but cluster in memory T-cell lineages consistent with the ability of cells harboring HIV to differentiate and undergo homeostatic proliferation. Future studies will include longer ART treatment in humanized mice, allowing observation for up to a year.

DGE analysis revealed transcriptional signatures associated with HIV-1 infection and latency. Our pathway analysis highlights the significance of EIF2 signaling and Sirtuin pathways and protein ubiquitination pathways across various infection and treatment contexts. In acutely infected cells, genes related to mRNA splicing and apoptotic regulation were differentially expressed. EIF2 signaling, Sirtuin, and Protein Ubiquitination pathways were associated with both acute infection and latency, suggesting important roles in regulating viral persistence. HIV and other viruses can hijack the integrated stress response (ISR) to facilitate replication, inhibit cell growth, and induce cell death signaling[69–71]. Upregulation of EIF2 pathways in acutely infected mice may be linked to viral protein translation and replication. HIV-1 Tat proteins inhibits SIRT1, hyperactivating T-cells through AP1[72]. Key genes in these analyses include BCL2, JUN, and MDM2 that were oppositely regulated in acute and latent infection help to define key nodes that may be the focus of latency reversal approaches. The pathway analysis data indicates that JUN is upregulated during acute infection and downregulated during ART treatment, indicating that it is regulated transcriptionally when performing its central role in controlling HIV-1 transcription.

A limitation of the system we have observed is that the cre-recombinase activity encoded by HIV is not maintained over weeks of infection, which may be expected to be deleted over time, given the fitness burden of the additional cre and IRES sequences in the virus. However, since we are able to identify persistent GFP-marked cells after ART treatment, we estimate that we are preferentially isolating T cells that were marked within the first rounds of infection of the humanized mice. Another limitation of the system is that we are examining human cells that have developed within a xeno-environment of the mouse, where T cells may develop some degree of graft versus host disease. The ability of the system to recapitulate many diverse T cell lineages, including naïve and memory T cell subsets, along with corroboration of key differentially expressed genes with a single cell transcriptomic dataset from human provides some confidence that the gene pathways that we identify here may be physiologically relevant. Future studies are needed to validate which, if any, of these key pathways may be manipulated to modulate HIV latency in vivo.

An additional limitation of humanized mouse models, in general, is that the evolution of the reservoir can only be followed for the lifespan of the model, which is less than a year. In contrast, the estimated half-life of the reservoir in persons with HIV has been suggested to be $t_{1/2} = 44$ months[68]. It does, however, provide a genetically tractable model system where we can enrich for the reservoir to study its origins, monitor its response to therapies designed to eradicate it, and compare the results to those measured in human studies.

In summary, the HILT model provides a versatile platform for investigating HIV-1 infection dynamics, latency, and persistence in various experimental settings. Genetic marking, flow cytometry, and single-cell transcriptomics have offered a comprehensive view of infected and latent cell populations, revealing HIV persistence across broad, transcriptionally defined subsets. The system allows tracking of HIV acute infection in T cell lineages and shifts in populations that occur during ART. The data obtained here show similarity to data from patients yet enable collection of abundant data under genetically controlled conditions. The analysis of common pathway perturbations with opposing directionality in acute and latent infection may enforce latency and provide potential targets for therapeutic interventions for latency reversal to eradicate the HIV latent reservoir.

Methods

The research complies with all relevant ethical regulations; Human tissues used in this research are determined by consultation with the Icahn School of Medicine Program for the Protection of Human Subjects Institutional Review Board as being exempt. Human stem cells were obtained from a commercial stem cell resource (Stem Cell Technologies, Cat# 70008 Mixed Cord Blood CD34) and an institution that approved the study protocol. Animal experiments were conducted under IACUC protocol #IACUC-2014-0097, adhering to the Guide for the Care and Use of Laboratory Animals.

Virus constructs

The HIV-1 NL-CreI construct (Fig. 1B) expresses cre recombinase in place of early viral gene nef. The expression of nef in the construct has been restored downstream by inserting an internal ribosome entry site[73,74]. The envelopes of these constructs are either X4-tropic - NL4-3 or R5 tropic - pRHPA4259 clone 7 (SVPB14), which is a clade B primary envelope obtained through the NIH AIDS Reagent Program (ARP) from BH. Hahn and J.F. Salazar-Gonzalez[73]. pHRSIN CS (RG)W construct (Fig. 1A) was made by cloning the LoxP-dsRed-LoxP-GFP from construct pMSCV-loxp-dsRed-loxp-eGFP-Puro-WPRE[44] into pHRSIN CSGW[75] in place of GFP by using enzymes BamHI and SbfI.

Virus Production

293 T cells were transfected with NLCreI using transfection reagent Polyjet (Signagen) following the manufacturer's instructions. Forty-eight hours post-transfection, virus particles were concentrated by centrifugation through a 6% Optiprep solution and at 54,000 *g* for 2 hours. Quantification of virus particles was done using p24 ELISA assay. For Lenti-RG lentiviral production, third-generation plasmids pMDL gag/pol RRE, pREV, and VSV-G expressing plasmids[76] were cotransfected with pHR_SIN_CS(RG)W at a ratio of 2:1:1, respectively, and virus particles were collected 48 hours post-transfection.

Cell culture and infection

Jurkat Clone E6 T cell line (The NIH HIV Reagent Program, ARP-177) was used for in vitro infection studies. All human cells, peripheral blood mononuclear cells, or cord blood-derived hematopoietc stem cells were obtained from deidentified donors (IRB expempt) and included both male and female donors. Sex was not tracked as an independent variable. Primary human peripheral blood mononuclear cells (PBMC) were isolated from buffy coat/leukopaks from the New York Blood Center (NYBC) from normal healthy blood donors who have consented to providing blood products anonymously for research purposes. To ensure the privacy and confidentiality of donors, the NYBC provides no personal identifying information associated with these samples. The use of these samples was compliant with all relevant ethical human participants research guidelines and regulations. PBMC were isolated by ficoll density gradient centrifugation, followed by the isolation of CD4+ T-cells by magnetic bead separation (Miltenyi). Primary CD4+ T-cells were activated by co-culturing with irradiated feeder PBMCS in the presence of phytohemagglutinin (PHA; 2 µg/ml, Sigma) and IL-2 (50 IU/ml, Roche) for 48–72 hours. Activated cells were spinoculated with pHRSINCS(RG)W virus. 48–72 hours post-infection, Jurkat T-cells were analyzed by flow cytometry (Attune NxT ThermoFisher Scientific, FlowJo, BD Biosciences).

Humanized Mice

NOD-NSG Il2rg^{NULL} (NOD.Cg-Prkdc^{scid} Il2rg^{tm1Wjl}/SzJ, NSG, Strain #005557) mice (The Jackson Laboratory, JAX.org) were maintained at the Center for Comparative Medicine and Surgery (CCMS) at Icahn School of Medicine at Mount Sinai according to guidelines established by the Institutional Animal Committee. All experiments were performed with authorization from the Institutional Review Board and the Institutional Animal Care and Use Committee (IACUC) at Icahn School of Medicine at Mount Sinai. For HILT mice, 0-3 days-old neonatal NSG mice were non-lethally irradiated (100 cGy) and injected intrahepatically with 1 x 10^5 human cord blood-derived CD34+ cells (Stemcell Technologies, Cat# 70008) that had been transduced with pHR_SIN CS (RG)W[47]. 16 weeks post-injection, 100–150 ul blood was obtained from mice by facial cheek bleed to check for engraftment of human lineages. For HuPBL mice, 6-8 weeks old mice were injected with Lenti-RG transduced activated PBMC. 10 days post-injection, 100–150 ul of blood was obtained from mice by facial cheek bleed to check for engraftment of human lineages. Briefly, cells were stained with huCD45, msCD45, huCD3, huCD4, and huCD8 antibodies and analyzed by flow cytometry (Fig. S2). Mice with greater than 10% huCD45, greater than 20% huCD3, greater than 20% CD4+ T-cells, and greater than 20% dsRed+ of the CD4 were used for further studies.

Infection in Humanized mice

In the huHSC model, mice were injected intravenously with 250 ng NLCreI (RHPA) virus into the retro-orbital sinus. Mice were bled periodically starting 3 days post-infection (dpi) to check for the presence of GFP+ cells in the CD3 + CD8- compartment. In the HuPBL model, mice were injected with 250 ng NLCreI (RHPA) virus via an intraperitoneal route.

Plasma Viral Load

Plasma from the infected mice was collected from the periodic bleeds, and viral load (copies/ml) was measured as described in Mohri et al., 2008[77]. Briefly, blood samples from mice were obtained by facial cheek bleed periodically, and plasma was separated from cells after centrifugation twice at 900 x *g* for 12 mins and 1200 x *g* for 9 min. RNA was extracted from 100 µl plasma and used in RT-qPCR for plasma viremia determination.

Antiretroviral treatment of mice

At 12 dpi, infected mice were placed under ART treatment to study the latent population as previously described by Halper-Stromberg et al.[45]. Briefly, tenofovir disproxil-fumerate (TDF, Gilead) and Emtricitabine (FTC, Gilead) tablets (crushed into powder), and Raltegravir (Merck) powder were used to manufacture mouse chow (TestDiet 5BXG, ½" irradiated pellets) along with Amoxicillin (0.12%). This food was left in the mouse cages to feed ad libitum for the treatment (39 days).

Cell isolation and sorting

Spleens from infected mice were harvested and processed by homogenizing the spleen through 40um sterile filters. The cells were washed with 2 mM EDTA containing PBS and treated with a lysing buffer ACK (Life Technologies) to remove RBC. The remaining cells were stained with antibodies described in the engraftment panel and sorted as described. Cells were sorted to enrich for GFP+ infected cells as described in the results section. In 15 dpi acutely infected cells, GFP + HIV-infected cells were pooled with dsRed + (GFP-) and unmarked cells for 10x Chromium scRNA-Seq sample processing. 10 d and 29 d ART-treated GFP+ latently infected cells were sorted into carrier C57BL/6 mouse (Jackson Laboratory) B cells and processed for scRNA-Seq sample processing. Unmarked and dsRed + (GFP-) bystander cells were sorted and sequenced separately. The bulk sorted cells were counted after trypan blue staining for estimating the proportion of viable cells and introduced into the 10x Genomics Chromium platform workflow for high throughput scRNAseq library preparation (V2 3' GEX protocol). For sequencing, batched libraries were run on an Illumina NovaSeq S4 flowcell (2 ×100 nt paired-end read length) with a target of 50,000-100,000 reads per cell.

Single-cell RNA sequencing data processing and analysis

Raw read data were first analyzed by 10x Cell Ranger Count (v6.1.2) using default parameters[48]. In order to identify HIV sequences, pHR_SIN_CS(RG)W and NLCreI_11036 reference sequences were added to 10x Genomics' standard human and mouse reference genome builds (v.2020-A) with annotations that include SFFV/dsRed/EGFP/WPRE regions on pHR SIN_CS(RG)W and six regions on NLCreI_11036 based on HIV splice donor and acceptors[49]. Each region was annotated as a gene to enable CellRanger to call the respective UMIs individually. These reads then underwent an additional quality control (QC) filtering, which included the removal of adapter sequences, read trimming, and mapping to the genome within the Cell Ranger interface.

Analysis was conducted using Seurat (v4.3.0)[54,78] in R (v4.2.0)[79]. Cells with more than 5% of mouse gene UMIs were removed, keeping cells with >95% human gene transcripts. Standard QC based on individual mitochondrial gene cutoffs (5%,10%, or 15%) and overall minimum of 200 and individual maximum of unique feature cell gene counts ranges were applied. We utilized the multimodal reference mapping workflow from the Satija Lab to annotate the cells in our dataset in relation to a comprehensive Cellular Indexing of Transcriptomes and Epitopes by Sequencing (CITE-seq) derived reference PBMC dataset[54].

Principal component analysis (PCA) was performed to identify variable genes, and Lenti-RG and HIV-specific genes were manually removed. Using canonical immune cell marker genes, clusters with high expression of non-T-cell markers were removed, and the datasets were re-normalized and clustered. Cells not predicted as any CD4 T-cell type but expressing T-cell markers were determined as "CD4 Others" and kept in the analysis. Integration was performed using the fastMNN function within SeuratWrappers (v.0.3.1)[50].

Differential gene expression (DGE) analysis was conducted independent of cluster identity using the FindMarkers function of Seurat (v4.1.0)[54,78] and Model-based analysis of single-cell transcriptomics (MAST)[80]. MAST identifies differentially expressed genes (DEGs) between two groups of cells using a hurdle mode, which utilizes a two-part discrete and continuous model tailored to scRNAseq gene expression data. Multiplicity correction was performed with the Benjamini-Hochberg method[80,81]. These results were displayed with a false discovery rate cutoff (FDR) < 0.05 using EnhancedVolcano (v1.16.0)[82].

Defining HIV-Cre-positive cells

HIV-Cre-positive cells were defined as containing >2 transcripts UMI for any HIV transcript defined in our custom reference genome. Due to the wide range of HIV-specific UMIs, we described three categories as the following based on the observed distribution: HIV transcript negative (0-1), HIV transcript low (2-10), and HIV transcript high (11+).

Gene set enrichment analysis

Based on our DGE analysis results, we implemented a restricted list of the top 145 common differentially expressed genes across all three acutely infected datasets and the top 117 common differentially expressed genes from our treated dataset in the web-based Gene set enrichment analysis platform Enrichr[63–65]. Results were focused on gene ontology processes.

Ingenuity pathway analysis

Differential gene expression (DGE) analysis of dsRed+(GFP-) cells, GFP + HIV infected cells, and GFP + ART-treated cells from acutely infected and treated mice was analyzed with Ingenuity Pathway

Analysis (QIAGEN Inc., https://digitalinsights.qiagen.com/IPA)[83]. Datasets were analyzed using a fdr<0.05 threshold value. Comparison analysis, canonical pathways, pathway analysis, network comparison and upstream regulators were employed to identify top pathways.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All single-cell RNA sequencing raw data files and processed data have been deposited at GEO and are publicly available. Source data are provided with this paper.

Code availability

All original code has been deposited on GitHub and is publicly available (https://github.com/benjamin-k-chen/manuscript-satija).

References

1. Ho, D. D. et al. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* **373**, 123–126 (1995).
2. Perelson, A. S., Neumann, A. U., Markowitz, M., Leonard, J. M. & Ho, D. D. HIV-1 dynamics in vivo: virion clearance rate, infected cell lifespan, and viral generation time. *Science* **271**, 1582–1586 (1996).
3. Palella, F. J. Jr. et al. Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. *HIV Outpatient Study Investig. N. Engl. J. Med.* **338**, 853–860 (1998).
4. Sengupta, S. & Siliciano, R. F. Targeting the latent reservoir for HIV-1. *Immunity* **48**, 872–895 (2018).
5. Davenport, M. P. et al. Functional cure of HIV: the scale of the challenge. *Nat. Rev. Immunol.* **19**, 45–54 (2019).
6. Chomont, N. et al. HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nat. Med.* **15**, 893–900 (2009).
7. Buzon, M. J. et al. HIV-1 persistence in CD4+ T cells with stem cell-like properties. *Nat. Med.* **20**, 139–142 (2014).
8. Jaafoura, S. et al. Progressive contraction of the latent HIV reservoir around a core of less-differentiated CD4(+) memory T Cells. *Nat. Commun.* **5**, 5407 (2014).
9. Fromentin, R. et al. CD4+ T cells expressing PD-1, TIGIT and LAG-3 contribute to HIV persistence during ART. *PLoS Pathog.* **12**, e1005761 (2016).
10. Evans, V. A. et al. Programmed cell death-1 contributes to the establishment and maintenance of HIV-1 latency. *AIDS* **32**, 1491–1497 (2018).
11. Sun, Z. et al. Intrarectal transmission, systemic infection, and CD4+ T cell depletion in humanized mice infected with HIV-1. *J. Exp. Med.* **204**, 705–714 (2007).
12. Sun, H. et al. Th1/17 polarization of CD4 T cells supports HIV-1 persistence during antiretroviral therapy. *J. Virol.* **89**, 11284–11293 (2015).
13. Wacleche, V. S. et al. New insights into the heterogeneity of Th17 subsets contributing to HIV-1 persistence during antiretroviral therapy. *Retrovirology* **13**, 59 (2016).
14. Gosselin, A. et al. HIV persists in CCR6+CD4+ T cells from colon and blood during antiretroviral therapy. *AIDS* **31**, 35–48 (2017).
15. Khoury, G. et al. Human immunodeficiency virus persistence and T-cell activation in blood, rectal, and lymph node tissue in human immunodeficiency virus-infected individuals receiving suppressive antiretroviral therapy. *J. Infect. Dis.* **215**, 911–919 (2017).
16. Zack, J. A. et al. HIV-1 entry into quiescent primary lymphocytes: molecular analysis reveals a labile, latent viral structure. *Cell* **61**, 213–222 (1990).
17. Zack, J. A., Haislip, A. M., Krogstad, P. & Chen, I. S. Incompletely reverse-transcribed human immunodeficiency virus type 1 genomes in quiescent cells can function as intermediates in the retroviral life cycle. *J. Virol.* **66**, 1717–1725 (1992).
18. Pierson, T. C. et al. Molecular characterization of preintegration latency in human immunodeficiency virus type 1 infection. *J. Virol.* **76**, 8518–8531 (2002).
19. Gosselin, A. et al. Peripheral blood CCR4+CCR6+ and CXCR3+CCR6+CD4+ T cells are highly permissive to HIV-1 infection. *J. Immunol.* **184**, 1604–1616 (2010).
20. Alvarez, Y. et al. Preferential HIV infection of CCR6+ Th17 cells is associated with higher levels of virus receptor expression and lack of CCR5 ligands. *J. Virol.* **87**, 10843–10854 (2013).
21. Whitney, J. B. et al. Rapid seeding of the viral reservoir prior to SIV viraemia in rhesus monkeys. *Nature* **512**, 74–77 (2014).
22. Shiau, S., Abrams, E. J., Arpadi, S. M. & Kuhn, L. Early antiretroviral therapy in HIV-infected infants: can it lead to HIV remission? *Lancet HIV* **5**, e250–e258 (2018).
23. Bruner, K. M. et al. Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nat. Med.* **22**, 1043–1049 (2016).
24. Bruner, K. M. et al. A quantitative approach for measuring the reservoir of latent HIV-1 proviruses. *Nature* **566**, 120–125 (2019).
25. Liu, R., Simonetti, F. R. & Ho, Y. C. The forces driving clonal expansion of the HIV-1 latent reservoir. *Virol. J.* **17**, 4 (2020).
26. Douek, D. C. et al. HIV preferentially infects HIV-specific CD4+ T cells. *Nature* **417**, 95–98 (2002).
27. Simonetti, F. R. et al. Clonally expanded CD4+T cells can produce infectious HIV-1 in vivo. *Proc. Natl Acad. Sci.* **113**, 1883–1888 (2016).
28. Bosque, A., Famiglietti, M., Weyrich, A. S., Goulston, C. & Planelles, V. Homeostatic proliferation fails to efficiently reactivate HIV-1 latently infected central memory CD4+ T cells. *PLoS Pathog.* **7**, e1002288 (2011).
29. Maldarelli, F. et al. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science* **345**, 179–183 (2014).
30. Wagner, T. A. et al. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science* **345**, 570–573 (2014).
31. Cohn, LillianB. et al. HIV-1 integration landscape during latent and active infection. *Cell* **160**, 420–432 (2015).
32. Bradley, T., Ferrari, G., Haynes, B. F., Margolis, D. M. & Browne, E. P. Single-cell analysis of quiescent hiv infection reveals host transcriptional profiles that regulate proviral latency. *Cell Rep.* **25**, 107–117.e103 (2018).
33. Clark, I. C. et al. HIV silencing and cell survival signatures in infected T cell reservoirs. *Nature* **614**, 318–325 (2023).
34. Collora, J. A. et al. Single-cell multiomics reveals persistence of HIV-1 in expanded cytotoxic T cell clones. *Immunity* **55**, 1013–1031 e1017 (2022).
35. Weymar, G. H. J. et al. Distinct gene expression by expanded clones of quiescent memory CD4(+) T cells harboring intact latent HIV-1 proviruses. *Cell Rep.* **40**, 111311 (2022).
36. Swiggard, W. J. et al. Human immunodeficiency virus type 1 can establish latent infection in resting CD4+ T cells in the absence of activating stimuli. *J. Virol.* **79**, 14179–14188 (2005).
37. Saleh, S. et al. CCR7 ligands CCL19 and CCL21 increase permissiveness of resting memory CD4+ T cells to HIV-1 infection: a novel model of HIV-1 latency. *Blood* **110**, 4161–4164 (2007).
38. Bosque, A. & Planelles, V. Induction of HIV-1 latency and reactivation in primary memory CD4+ T cells. *Blood* **113**, 58–65 (2009).
39. Yang, H. C. et al. Small-molecule screening using a human primary cell model of HIV latency identifies compounds that reverse latency without cellular activation. *J. Clin. Invest.* **119**, 3473–3486 (2009).
40. Lassen, K. G., Hebbeler, A. M., Bhattacharyya, D., Lobritz, M. A. & Greene, W. C. A flexible model of HIV-1 latency permitting evaluation of many primary CD4 T-cell reservoirs. *PLoS One* **7**, e30176 (2012).

41. Nischang, M. et al. Humanized mice recapitulate key features of HIV-1 infection: a novel concept using long-acting antiretroviral drugs for treating HIV-1. *PLoS One* **7**, e38853 (2012).
42. Arainga, M., Su, H., Poluektova, L. Y., Gorantla, S. & Gendelman, H. E. HIV-1 cellular and tissue replication patterns in infected humanized mice. *Sci. Rep.* **6**, 23513 (2016).
43. Gruell, H. & Klein, F. Progress in HIV-1 antibody research using humanized mice. *Curr. Opin. HIV AIDS* **12**, 285–293 (2017).
44. Koo, B. K. et al. Controlled gene expression in primary Lgr5 organoid cultures. *Nat. Methods* **9**, 81–83 (2011).
45. Halper-Stromberg, A. et al. Broadly neutralizing antibodies and viral inducers decrease rebound from HIV-1 latent reservoirs in humanized mice. *Cell* **158**, 989–999 (2014).
46. Audige, A. et al. Long-term leukocyte reconstitution in NSG mice transplanted with human cord blood hematopoietic stem and progenitor cells. *BMC Immunol.* **18**, 28 (2017).
47. Wang, C. X. et al. Rapamycin relieves lentiviral vector transduction resistance in human and mouse hematopoietic stem cells. *Blood* **124**, 913–923 (2014).
48. Zheng, G. X. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
49. Leon-Rivera, R., Morsey, B., Niu, M., Fox, H. S. & Berman, J. W. Interactions of monocytes, HIV, and ART identified by an innovative scRNAseq pipeline: pathways to reservoirs and HIV-associated comorbidities. *mBio* **11**, e01037–20 (2020).
50. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36**, 421–427 (2018).
51. Ding, J. et al. Characterisation of CD4+ T-cell subtypes using single cell RNA sequencing and the impact of cell number and sequencing depth. *Sci. Rep.* **10**, 19825 (2020).
52. Andreatta, M. et al. A CD4(+) T cell reference map delineates subtype-specific adaptation during acute and chronic viral infections. *Elife* **11**, e76339 (2022).
53. Ciucci, T. et al. The emergence and functional fitness of memory CD4(+) T cells require the transcription factor Thpok. *Immunity* **50**, 91–105.e104 (2019).
54. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e3529 (2021).
55. Lund, N. et al. Differential effects of hnRNP D/AUF1 isoforms on HIV-1 gene expression. *Nucleic Acids Res.* **40**, 3663–3675 (2012).
56. Kutluay, S. B. et al. Genome-wide analysis of heterogeneous nuclear Ribonucleoprotein (hnRNP) binding to HIV-1 RNA reveals a key role for hnRNP H1 in alternative viral mRNA splicing. *J Virol.* **93**, e01048–19 (2019).
57. Bhattarai, K. & Holcik, M. Diverse roles of heterogeneous nuclear ribonucleoproteins in viral life cycle. *Front. Virol.* **2**, 1044652 (2022).
58. Wong, R. W., Balachandran, A., Ostrowski, M. A. & Cochrane, A. Digoxin suppresses HIV-1 replication by altering viral RNA processing. *PLoS Pathog.* **9**, e1003241 (2013).
59. Jacquenet, S., Decimo, D., Muriaux, D. & Darlix, J. L. Dual effect of the SR proteins ASF/SF2, SC35 and 9G8 on HIV-1 RNA splicing and virion production. *Retrovirology* **2**, 33 (2005).
60. Wang, J. et al. Multiple functions of heterogeneous nuclear ribonucleoproteins in the positive single-stranded RNA virus life cycle. *Front. Immunol.* **13**, 989298 (2022).
61. Chandrasekar, A. P., Cummins, N. W. & Badley, A. D. The Role of the BCL-2 family of proteins in HIV-1 pathogenesis and persistence. *Clin Microbiol. Rev.* **33**, e00107–19 (2019).
62. Andersen, J. L. et al. HIV-1 Vpr-induced apoptosis is cell cycle dependent and requires Bax but not ANT. *PLoS Pathog.* **2**, e127 (2006).
63. Chen, E. Y. et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinforma.* **14**, 128 (2013).
64. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
65. Xie, Z. et al. Gene set knowledge discovery with Enrichr. *Curr. Protoc.* **1**, e90 (2021).
66. Stuart, T. et al. Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e1821 (2019).
67. Golumbeanu, M. et al. Single-Cell RNA-Seq reveals transcriptional heterogeneity in latent and reactivated HIV-infected cells. *Cell Rep.* **23**, 942–950 (2018).
68. Finzi, D. et al. Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nat. Med.* **5**, 512–517 (1999).
69. Mendes, E. A., Tang, Y. & Jiang, G. The integrated stress response signaling during the persistent HIV infection. *iScience* **26**, 108418 (2023).
70. Jiang, G. et al. HIV exploits antiviral host innate GCN2-ATF4 signaling for establishing viral replication early in infection. *mBio* **8** (2017).
71. Mekdad, H. E. et al. Characterization of the interaction between the HIV-1 Gag structural polyprotein and the cellular ribosomal protein L7 and its implication in viral nucleic acid remodeling. *Retrovirology* **13**, 54 (2016).
72. Murray, J. M. et al. HIV dynamics linked to memory CD4+ T cell homeostasis. *PLoS One* **12**, e0186101 (2017).
73. Li, M. et al. Human immunodeficiency virus type 1 env clones from acute and early subtype B infections for standardized assessments of vaccine-elicited neutralizing antibodies. *J. Virol.* **79**, 10108–10125 (2005).
74. Law, K. M. et al. In vivo HIV-1 cell-to-cell transmission promotes multicopy micro-compartmentalized infection. *Cell Rep.* **15**, 2771–2783 (2016).
75. Demaison, C. et al. High-level transduction and gene expression in hematopoietic repopulating cells using a human immunodeficiency [correction of imunodeficiency] virus type 1-based lentiviral vector containing an internal spleen focus forming virus promoter. *Hum. Gene Ther.* **13**, 803–813 (2002).
76. Dull, T. et al. A third-generation lentivirus vector with a conditional packaging system. *J. Virol.* **72**, 8463–8471 (1998).
77. Mohri, H. & Markowitz, M. In vitro characterization of multidrug-resistant HIV-1 isolates from a recently infected patient associated with dual tropism and rapid disease progression. *J. Acquir Immune Defic. Syndr.* **48**, 511–521 (2008).
78. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
79. R-Core-Team. R: A Language and Enviroment for Statistical Computing (2022).
80. Finak, G. et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 (2015).
81. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B-Stat. Methodol.* **57**, 289–300 (1995).
82. Blighe, K., Rana, S., Lewis, M. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling (2022).
83. Kramer, A., Green, J., Pollard, J. Jr & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30**, 523–530 (2014).

Acknowledgements

The work in this study was supported by NIH/NIAID AI116191 and AI162223 to BKC. We thank Avi Ma'ayan at the Icahn School of Medicine at Mount Sinai (ISMMS) for advice and Enrichr program. We thank Talia Swartz, Schahram Akbarian, Alyssa Wilson, and Brad Rosenberg at the

Icahn School of Medicine at Mount Sinai (ISMMS) for experimental and computational advice. Flow sorting was performed with the support of the ISMMS Dean's Flow Cytometry. Computational and data resources were provided by Scientific Computing and Data at the Icahn School of Medicine at Mount Sinai and supported by the Clinical and Translational Science Awards (CTSA) grant UL1TR004419 from the National Center for Advancing Translational Sciences. Research reported in this publication was also supported by the Office of Research Infrastructure of the National Institutes of Health under award number S10OD026880 and S10OD030463. Figure S4A was generated with Biorender under license: TN27M7KTKC.

Author contributions

B.K.C., N.S., F.P., and A.E. conceived, developed, and tested the HILT strategies. B.K.C., N.S., K.L., F.P., and A.L. designed and developed the HuHSC model system and HIV infection approaches. G.S., D.V.D., and M.K. contributed equally to bioinformatic analyses in this work. Single cell sequencing was planned and performed by K.A. and Y.-C.W., and overseen by K.G.B. and R.S. Single-cell sequencing processing, software development, analysis was performed by G.S., M.K., N.S., F.P., and D.V.D. N.S., B.K.C., G.S., F.P., and D.V.D. were involved in drafting and editing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-57368-7.

Correspondence and requests for materials should be addressed to Benjamin K. Chen.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.