nature medicine

LETTERS

https://doi.org/10.1038/s41591-018-0017-7

Clonal CD4+ T cells in the HIV-1 latent reservoir display a distinct gene profile upon reactivation

Lillian B. Cohn[1], Israel T. da Silva[2], Renan Valieris[2], Amy S. Huang[1], Julio C. C. Lorenzi[1], Yehuda Z. Cohen[1], Joy A. Pai[1], Allison L. Butler[1], Marina Caskey[1], Mila Jankovic[1,4] and Michel C. Nussenzweig[1,3,4]*

Despite suppressive combination antiretroviral therapy (ART), latent HIV-1 proviruses persist in patients. This latent reservoir is established within 48–72 h after infection, has a long half-life[1,2], enables viral rebound when ART is interrupted, and is the major barrier to a cure for HIV-1[3]. Latent cells are exceedingly rare in blood (~1 per 1×10^6 CD4+ T cells) and are typically enumerated by indirect means, such as viral outgrowth assays[4,5]. We report a new strategy to purify and characterize single reactivated latent cells from HIV-1-infected individuals on suppressive ART. Surface expression of viral envelope protein was used to enrich reactivated latent T cells producing HIV RNA, and single-cell analysis was performed to identify intact virus. Reactivated latent cells produce full-length viruses that are identical to those found in viral outgrowth cultures and represent clones of in vivo expanded T cells, as determined by their T cell receptor sequence. Gene-expression analysis revealed that these cells share a transcriptional profile that includes expression of genes implicated in silencing the virus. We conclude that reactivated latent T cells isolated from blood can share a gene-expression program that allows for cell division without activation of the cell death pathways that are normally triggered by HIV-1 replication.

To investigate the cells that contribute to the latent reservoir, we developed a method to enrich and isolate reactivated latent cells through a combination of antibody staining, magnetic enrichment, and flow cytometry[6] (latent cell capture, or LURE). Purified CD4+ T cells from donors treated with ART were activated with phytohemagglutinin (PHA), a robust in vitro latency reactivation agent[5,7], for 36 h in the presence of five potent antiretroviral drugs and a pan-caspase inhibitor. Cells expressing surface HIV-1 envelope (Env) protein were labeled with a cocktail of biotinylated anti-Env broadly neutralizing antibodies (bNAbs; 3BNC117, 10-1074, and PG16)[8–10] and enriched from suspension using magnetic beads.

Relative enrichment of the magnetically isolated Env+ cellular fraction was measured through comparison to unfractionated control cells from the same culture using flow cytometry (Fig. 1a and Supplementary Fig. 1a), and qPCR was used for comparison of HIV-1 *gag* mRNA levels between these cellular fractions (Fig. 1c). Enrichment of cell-associated HIV-1 RNA was entirely dependent on cellular activation with PHA (Supplementary Fig. 1b). Enrichment was measured in samples from ten individuals and was found to be dependent in part on the size of the latent reservoir ($r^2=0.5609$, $P=0.0127$), as measured in infectious units per million (IUPM) using viral outgrowth assays (Fig. 1d). We conclude that reactivated, latently infected cells can be enriched on the basis of surface expression of HIV-1 Env protein.

To further purify the reactivated latent cells, we used flow cytometry to sort single cells from the magnetically enriched fraction on the basis of Env staining. Individual cells expressing both *env* and *gag* were identified by the combination of surface Env staining and single-cell HIV-1 *gag* mRNA expression. The frequency of *gag* mRNA–expressing single cells in patients with high IUPMs ranged from 10 to 50% of sorted cells (Supplementary Table 1). In individuals with relatively lower IUPMs (0.49–2.43), the percent of Env+Gag+ single cells isolated varied from 0 to 4% (Supplementary Table 1).

We performed single-cell RNA sequencing (scRNA-seq) on Env+Gag+ single cells captured through LURE and on control unfractionated single cells from the same PHA-activated culture obtained from donors 603, 605, and B207. In addition, we performed scRNA-seq on activated CD4+ T cells that were productively infected with HIV-1_{YU2} (YU2) in vitro and purified through cell sorting using anti-Env antibodies (Supplementary Fig. 2). Overall, 249 cells were characterized, of which 22 cells (8.8%) were removed by quality metrics[11]. Of the 227 cells retained, 33 were YU2-infected cells, 85 were cells captured with LURE, and 109 were unfractionated control cells from the same cultures (Fig. 2a). On average, we obtained ~1,500 expressed genes per cell (Supplementary Fig. 3).

As expected, HIV reads were not detectable in the unfractionated, activated control cells (Fig. 2b). In contrast, cells captured by LURE and YU2-infected cells showed similar percentages of total mRNA reads mapping to the HIV-1 genome (3.8 and 4.5%, respectively)[12] (Fig. 2b). We conclude that reactivated latent cells captured by LURE contain RNA sequences mapping to the human genome and HIV-1 byscRNA-seq results.

We used Iterative Virus Assembler (IVA) software to reconstruct the virus from scRNA-seq reads in each individual CD4+ T cell[13]. HIV RNA recovered by scRNA-seq was dependent on proviral transcription, as determined by analysis of HIV-1 splice variants (Supplementary Fig. 4a). Fully reconstructed viruses were obtained from 26 cells infected with YU2 and from 19 cells captured by LURE (Fig. 2c and Supplementary Fig. 4b). All viruses obtained from 603 and 605 belonged to a single expanded viral clone (Fig. 2c). We identified four different viruses in B207: two were fully reconstructed, and two others were partially reconstructed (Fig. 2c). All of the fully reconstructed viruses were completely intact when analyzed by Gene Cutter software. Thus, the combination of LURE and scRNA-seq can be used to recover full-length, intact HIV-1 from single reactivated latent cells.

[1]Laboratory of Molecular Immunology, Rockefeller University, New York, NY, USA. [2]Laboratory of Computational Biology and Bioinformatics, A.C. Camargo Cancer Center (CIPE), Sao Paulo, Brazil. [3]Howard Hughes Medical Institute (HHMI), Rockefeller University, New York, NY, USA. [4]These authors contributed equally: Mila Jankovic, Michel C. Nussenzweig *e-mail: nussen@rockefeller.edu

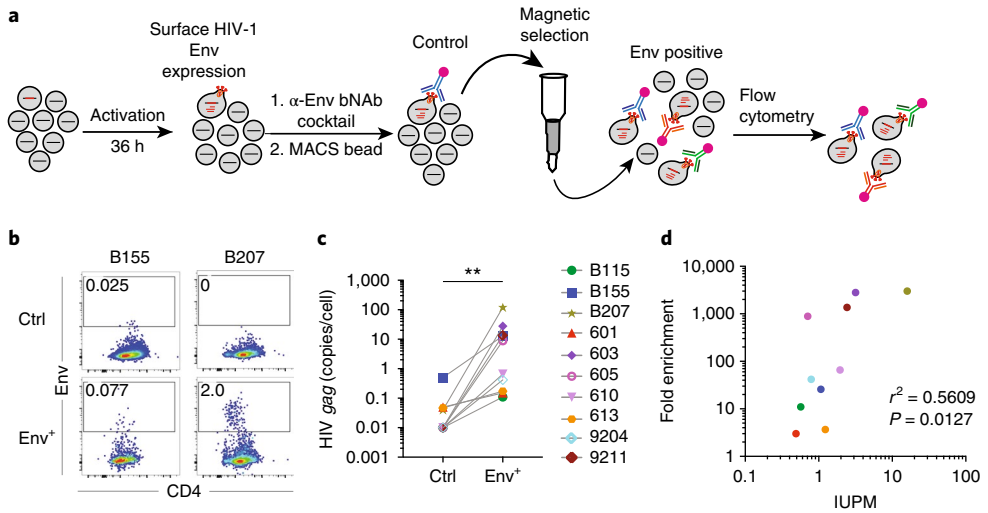

Fig. 1 | Latency capture enriches for cells producing HIV RNA. a, Schematic of LURE protocol. CD4+ T cells from ART-suppressed donors were cultured in conditioned medium with PHA, IL-2, an antiretroviral drug cocktail, and a pan-caspase inhibitor for 36 h. Cells were labeled with a biotinylated bNAb cocktail followed by phycoerythrin (PE)-conjugated streptavidin and anti-PE magnetic beads, passed over a magnetic column, and subjected to FACS analysis. **b**, Dot plots showing Env versus CD4 staining on pre-enrichment control (Ctrl; top row) and positively selected Env+ cells (bottom row) for donors B155 and B207. The gate shows the frequency of Env+ cells in each population. Two representative experiments from 15 independent experiments are shown. **c**, HIV *gag* mRNA was measured in equivalent numbers of Env+ and control cells. The graph shows the results from qPCR (limit of detection, 12.8 copies) for HIV *gag* mRNA, normalized to the number of sorted cells. Each symbol corresponds to a unique donor, with donor IDs indicated on right. $P = 0.002$, Wilcoxon matched-pairs signed rank two-tailed test. Representative data from 10 individuals from more than 30 independent experiments are shown. **d**, Fold-enrichment (for Env+ cells relative to control) in **c** compared to IUPM. Representative data from 10 individuals from more than 30 independent experiments are shown.

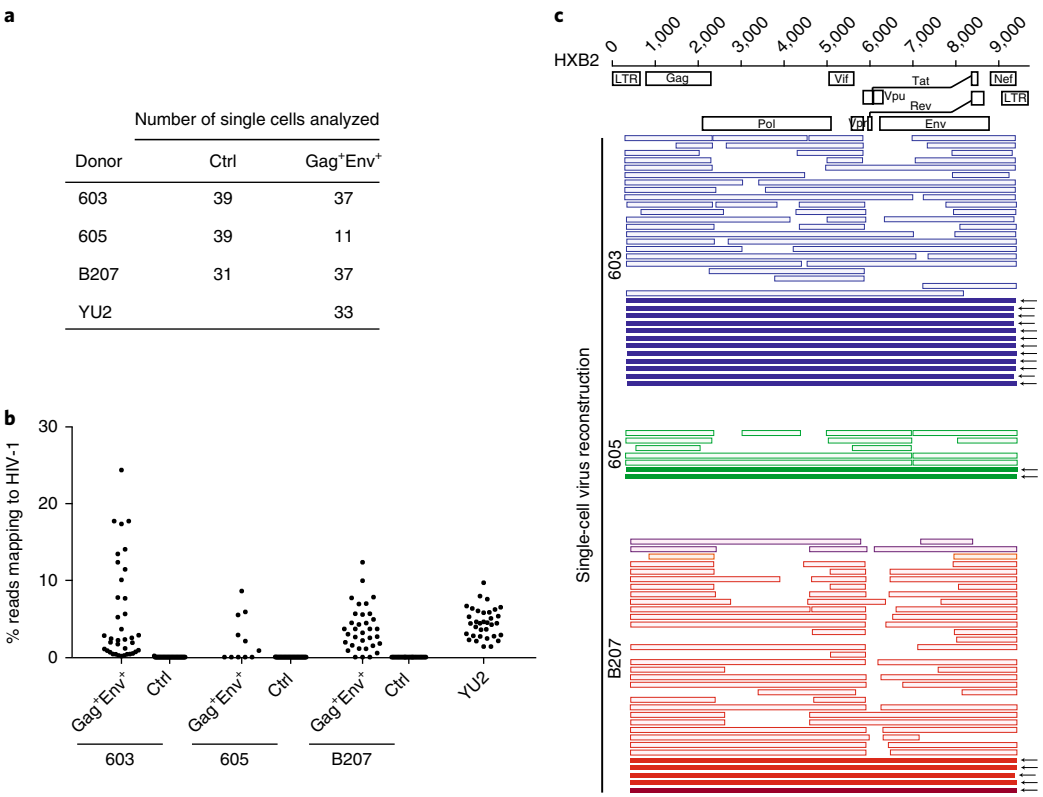

Fig. 2 | Full-length virus sequences recovered by scRNA-seq. a, Number of single cells analyzed by scRNA-seq. **b**, Fraction of reads mapping to HIV-1 in unfractionated control, LURE-purified Gag+Env+, and YU2-infected scRNA-seq libraries. **c**, Map of individual viruses reconstructed from scRNA-seq. The full-length HIV-1 genome (HXB2) is shown at the top, with the genome structure annotated below. Numbers represent individual donor. Each horizontal bar represents a single virus from an individual cell. Solid bars indicate that the entire virus was reconstructed from scRNA-seq reads. Outlined, lighter colored bars indicate incomplete genome reconstruction. Different colors indicate different sequences. For participants 603 and 605, every virus identified was identical. For B207, we identified four unique viruses, with one clone (in red) predominating. Arrows indicate fully reconstructed viruses.

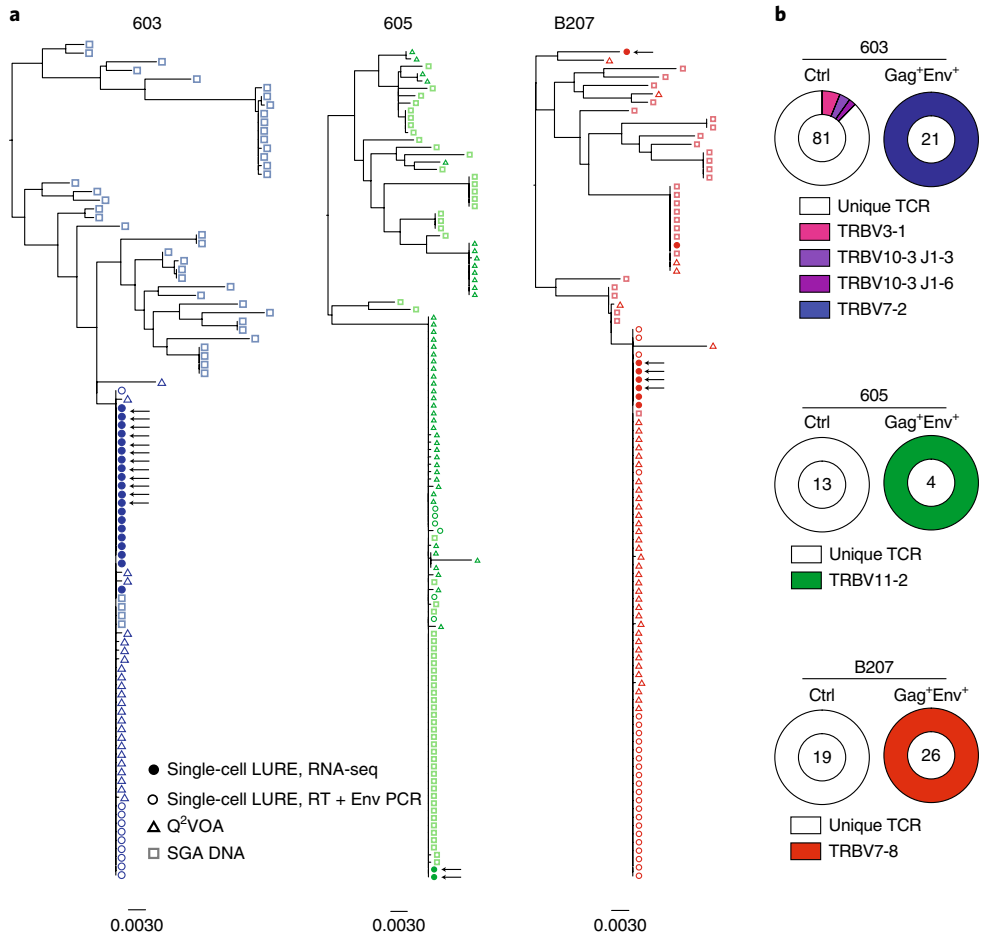

Fig. 3 | Captured cells express Env that is identical to latent virus emerging in Q²VOAs and represent expanded clones. a, Maximum likelihood phylogenetic trees compare full-length Env sequences derived from single-cell capture by LURE (solid and open circles), SGA-derived proviral DNA (open squares), and replication-competent single-cell viral outgrowth cultures (Q²VOAs) (open triangles) from participants 603, 605, and B207. Sequences from LURE cells were obtained either by recovery and assembly from RNA-seq reads (closed circles) or from reverse transcription of RNA in single cells followed by specific *env* PCR from single Gag⁺Env⁺ LURE cells (open circles). Arrows indicate confirmed full-length sequences. Scale bar represents nucleotide substitutions per site. **b**, TCR sequences recovered from scRNA-seq or amplified by PCR for control (unfractionated pre-enrichment) and Gag⁺Env⁺ LURE-purified cells. The number in the center of the pie denotes the number of cells sequenced; slices are proportional to clone size, showing unique TCRs (white slices) and clonal TCRs (colored slices). Clones were identified by their shared TCR-α and TCR-β sequences.

To determine whether the full-length viruses expressed in the purified single cells obtained by LURE correspond to the intact latent viruses that emerge in viral outgrowth assays, we compared their *env* sequences (Fig. 3a). To do so, we performed quantitative and qualitative viral outgrowth assays (Q²VOAs)[14] and *env* single-genome amplification (SGA) on DNA isolated from CD4⁺ T cells and compared these sequences to those found in the cells obtained from LURE.

Phylogenetic analysis of *env* sequences revealed that in donors 603 and B207, the *env* sequences obtained by LURE and Q²VOAs generally clustered together, were part of an expanded clone, and did not overlap significantly with sequences obtained by SGA of proviral DNA (Fig. 3a). Participant 605 had an unusual distribution of SGA-obtained proviral DNA sequences in that there was a substantial overlap with the *env* sequences found in viral outgrowth cultures. Nevertheless, the majority of LURE-derived *env* sequences belong to the major viral outgrowth clone found in Q²VOAs (Fig. 3a) in all three individuals. We conclude that the *env* sequences expressed by cells purified through LURE are typically identical to those found in viruses that emerge from latent cells in viral outgrowth cultures and therefore are replication competent.

Latent cells harboring identical replication-competent viruses may arise through T cell clonal expansion[14–22] or during a viral replicative burst when identical viruses infect a diverse group of T cells. To definitively distinguish between these possibilities, we analyzed the T cell receptor (TCR) sequences obtained from single latent cells captured by LURE. CD4⁺ T cells express unique antigen receptors produced by random TCR variable, diversity and joining gene segment (VDJ) recombination. T cells with identical TCRs are only produced by clonal expansion. As a control, we obtained TCR sequences from nearly 600 single CD4⁺ T cells from three healthy and three ART-treated, HIV-1-infected donors. We found that 99.9% of all control TCR sequences were unique, with only a single two-member clone identified in one of the six individuals (Supplementary Fig. 5). In contrast, the TCR sequences derived from the latent cells with identical proviruses captured by LURE (Figs. 2c and 3a) were entirely clonal in all three donors (Fig. 3b and Supplementary Fig. 6). The clonality was not due to T cell division in vitro, as there was no measurable T cell division in 36 h under our culture conditions (Supplementary Fig. 7). Our data demonstrates that groups of latent cells containing identical replication-competent viruses are products of CD4⁺ T cell clonal expansion in vivo.

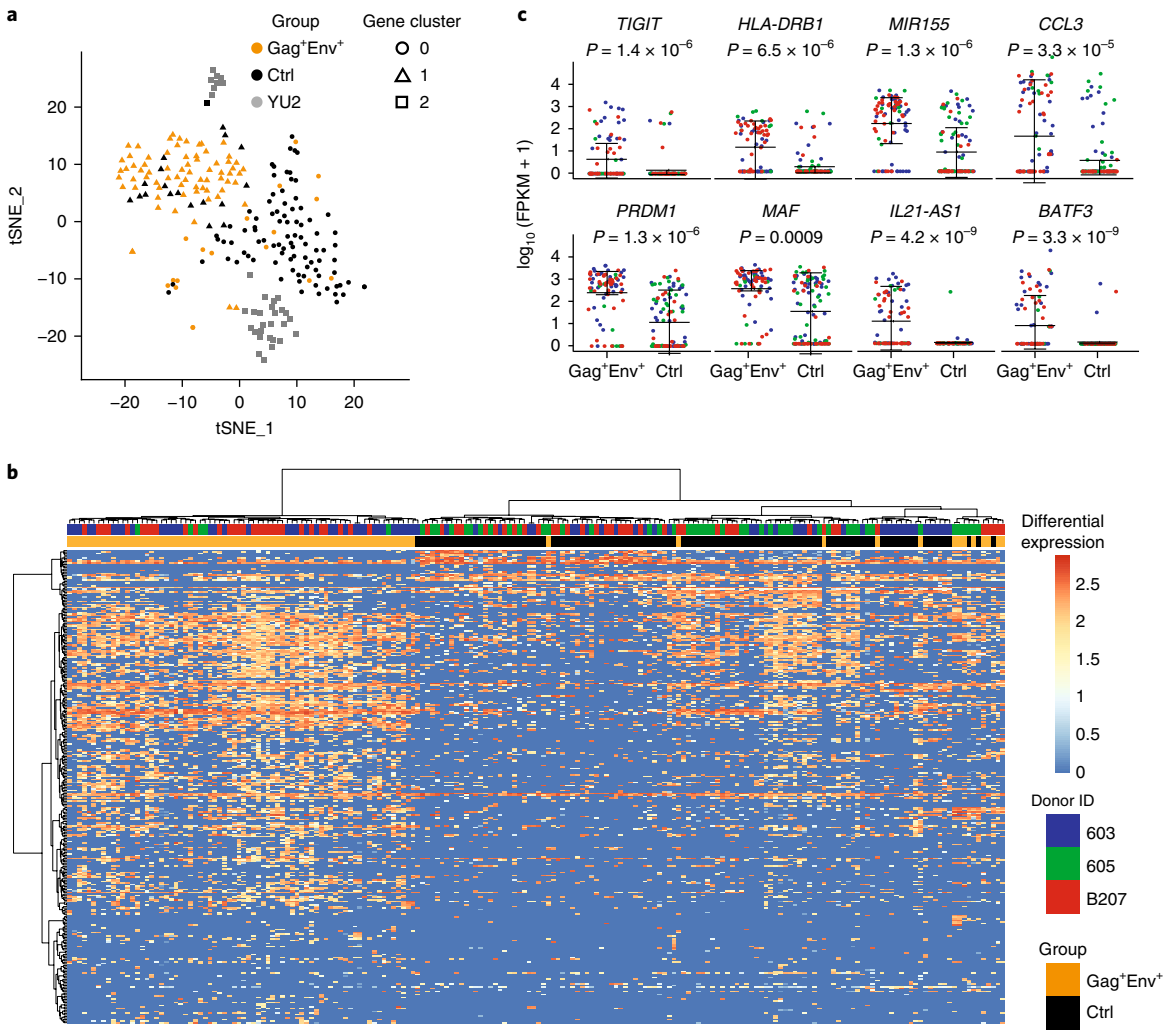

Fig. 4 | A distinct gene signature defines reactivated latent cells. a, Principal components analysis (PCA) clusters cells by group. The displayed output from Seurat t-Distributed Stochastic Neighbor Embedding (tSNE) for the three groups (control (black), Gag$^+$Env$^+$ LURE (orange) and YU2 (gray)) is shown. Each data point represents a single cell. Seurat analysis identified three distinct clusters of genes, which define three groups of cells (circles (gene cluster 0), triangles (gene cluster 1), and squares (gene cluster 2)) by performing graph-based clustering over six principal components. All data obtained from individuals 603, 605, and B207 (control and LURE cells) and YU2-infected healthy donor cells (109 control cells, 85 LURE cells, and 33 YU2-infected cells) are shown. **b**, Heat map showing unsupervised clustering of differentially expressed genes between the Gag$^+$Env$^+$ LURE-purified group (orange bars) and control unfractionated group (black bars). Cells from donor 603 are indicated in blue, 605 in green, and B207 in red along the top of the graph. Color in the heat map indicates the normalized level of expression. **c**, Graphs show expression of selected significantly differentially expressed genes in individual Gag$^+$Env$^+$ LURE-purified and control unfractionated cells from participants 603 (blue), 605 (green), and B207 (red), as determined using MAST software. All data obtained from the participants (109 control cells and 85 LURE cells) are shown. Error bars show mean and s.d. Significant differential expression was determined using the likelihood ratio test embedded in the MAST software.

To further characterize the reactivated latent cells captured by LURE, we performed single-cell transcriptome analysis and compared the results to unfractionated, PHA-stimulated control cells from the same cultures and to activated CD4$^+$ T cells productively infected with YU2. We performed hierarchical clustering through a principal-component analysis (PCA) called Seurat[23] using gene expression data from the 227 cells. This unbiased analysis identified three unique groups of genes that segregated the cells into three separate clusters. Each of these clusters was found to correspond to one of the three input groups: control, LURE, and YU2-infected cells (Fig. 4a, Supplementary Fig. 8, and Supplementary Table 2). Additional analysis that employed unsupervised clustering using all gene expression data (single-cell consensus clustering or SC3) confirmed these results from comparison of control cells and LURE-sorted cells (Supplementary Fig. 9). Thus, in PCA and unsupervised clustering, the reactivated latent cells captured by LURE cluster separately from uninfected (control) and actively infected CD4$^+$ T cells.

To further understand the transcriptional differences between the three groups of cells, we identified differentially expressed genes (DEGs) ($P < 0.01$) between reactivated latent cells and PHA-activated control cells. Using unsupervised clustering, we grouped the cells on the basis of the expression of all significantly differentially expressed genes between LURE and control cell groups ($P < 0.01$, 778 genes) (Supplementary Table 3). Irrespective of donor, reactivated cells purified through LURE generally segregated from unfractionated, activated control cells in two of three individuals (Fig. 4b), with cells from the third individual split between the LURE group and the control group. Similar results were also obtained through comparison with YU2-infected cells (Supplementary Fig. 10). We conclude that cells captured with LURE segregate from

activated control cells and productively infected cells in three different methods of analysis.

Among the 240 genes that overlapped between the PCA-identified group and DEGs ($P < 0.01$), we found a number of genes highly expressed in the isolated LURE cells compared to controls that have been shown by independent analyses to be associated with HIV-1 latency (Fig. 4d). For example, *TIGIT*[24,25] and *HLA-DR*[26] were 140- and 76-fold upregulated, respectively, in cells purified by LURE compared to control cells, and CD32a[27] was not (Fig. 4c and Supplementary Fig. 11). MiR-155, which inhibits TRIM32, prevents its interaction with HIV Tat and reinforces viral latency[28], was 368 times more highly expressed in LURE cells compared to controls. Expression of chemokine CCL3, which is reported to have HIV-1-suppressive effects[29,30], was 795 times higher in LURE cells compared to controls. Finally, a number of transcription factors were among the top 15 differentially expressed genes, including the top differentially expressed gene, *PRDM1* (1,365×). PRDM1 represses HIV-1 proviral transcription in memory CD4+ T cells by inhibition of HIV Tat[31], and its overexpression is associated with lower levels of HIV-1 transcription in elite controllers[32].

To further examine the differences between LURE and control cells, we performed enrichment analysis using the Gene Ontology database with the 240 genes that overlapped between the DEG and PCA analyses. Among the top ten most significantly enriched biological processes, eight were related to immune system function, suggesting that PHA-stimulated LURE and control cells differ in their expression of genes related to responses to pathogens. For example, LURE and control cells differ markedly in response to type I interferon and regulation of type I interferon production (Supplementary Tables 4 and 5), with control cells having higher expression of type I interferon responsive genes, such as *IFIT3*, *ISG20*, *IRF1*, *IFI6*, *RSAD2*, *STAT1*, *XAF1*, *CTNNB1*, and *UBE2L6*. Consequently, the control cells also show a higher overall expression of genes that are involved in response to viruses, such as *CCL5*, *IFIT3*, *ISG20*, *IRF1*, *SERINC5*, *IL2RA*, *RSAD2*, *DDIT4*, *STAT1*, and *PIM2*. Consistent with the altered gene expression program in reactivated latent cells, LURE and control cells show significant differences in the expression of genes that regulate transcription (Supplementary Tables 4 and 5). For example, reactivated latent cells have higher levels of expression of transcriptional regulators PRDM1, MAF, IRF4, MTDH, IKZF3, and BATF3, whereas control cells have higher expression of PIM2, STAT1, HNRNPA2B, EZR, IRF1, CTNNB1, and NFKBIZ (Supplementary Tables 4 and 5). We conclude that reactivated latent cells differ from control cells in a number of ways, many of which are related to the suppression of cellular antiviral immunity.

Our analysis is limited to three individuals and to a single reactivation agent, PHA. Examination of additional individuals and methods of latent cell reactivation may reveal additional and/or different genes and pathways involved in maintaining latency. LURE purification of reactivated latent cells requires proviral activation to induce Env protein expression on the cell surface. Therefore, LURE captures a subset of latent cells with proviruses that can be reactivated in a single round of potent T cell stimulation[33,34]. Owing to the relative resistance of some latent cells to reactivation[7], LURE mirrors the viral outgrowth assay and is unable to capture the entirety of the latent reservoir. Furthermore, our analysis is limited to circulating CD4+ T cells that express Env on the cell surface that are recognized by our antibody cocktail. Finally, some reactivated latent cells are certainly lost during the multiple processing stages involved in the LURE protocol. Thus, the cells captured by LURE represent a fraction of the circulating latent reservoir that is closely related to and that overlaps with the latent cells that emerge in traditional viral outgrowth assays. Further experiments will be required to determine whether tissue-resident latent cells have a similar gene program upon reactivation.

T cell division in response to antigen or mitogens, like PHA and HIV-1 reactivation from latency, are stimulated by shared metabolic and transcriptional pathways, including NFκB[35]. Once activated, productive HIV-1 infection typically leads to CD4+ T cell death by apoptosis or pyroptosis[36]. However, cell death after latency reactivation in vitro appears to be stochastic, with some cells being able to divide and survive after strong stimulation[19]. Our finding that latent cells can survive upon cell division in vivo confirms results from in vitro experiments[19] and is also consistent with the observation that the latent compartment contains groups of CD4+ T cells that harbor proviruses with identical *env* sequences[14,19]. Purification of reactivated latent cells by LURE and subsequent TCR sequencing provides definitive evidence that these cells arise by clonal expansion in vivo. The data is consistent with the idea that the protracted longevity of the latent compartment is due at least in part to cell division[14–22]. Finally, because the reservoir is stable over time[1,2], the finding that latent cells divide implies that they are also dying at a similar rate and that the reservoir is a dynamic compartment.

Antibody binding to Env-expressing cells in vivo leads to their accelerated clearance[37,38]. Should latent cells undergoing clonal expansion in vivo also express viral proteins, they too could be targeted for clearance by HIV-1-specific cytotoxic T cells and natural killer (NK) cells or by antibody-dependent cellular cytotoxicity.

How does a subset of latent cells divide and still survive despite expression of HIV-1? Our single-cell transcriptomic analysis of purified primary CD4+ T cells demonstrates that reactivated latent cells can express a distinct transcriptional program that includes muted responses to type I interferon and factors such as MiR-155 and PRDM1 that can suppress HIV-1 transcription[28,31,32]. We speculate that active HIV-1 suppression during CD4+ T cell division could be one of the mechanisms maintaining the latent reservoir. Further studies will be required to determine whether interference with these cellular safeguards could contribute to accelerating latent HIV-1 clearance.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41591-018-0017-7.

Received: 12 January 2018; Accepted: 14 March 2018;
Published online: 23 April 2018

References

1. Crooks, A. M. et al. Precise quantitation of the latent HIV-1 reservoir: implications for eradication strategies. *J. Infect. Dis.* **212**, 1361–1365 (2015).
2. 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 (2003).
3. Murray, A. J., Kwon, K. J., Farber, D. L. & Siliciano, R. F. The latent reservoir for HIV-1: how immunologic memory and clonal expansion contribute to hiv-1 persistence. *J. Immunol.* **197**, 407–417 (2016).
4. Henrich, T. J., Deeks, S. G. & Pillai, S. K. Measuring the size of the latent human immunodeficiency virus reservoir: the present and future of evaluating eradication strategies. *J. Infect. Dis.* **215**, S134–S141 (2017). suppl_3.
5. Spina, C. A. et al. An in-depth comparison of latent HIV-1 reactivation in multiple cell model systems and resting CD4+ T cells from aviremic patients. *PLoS Pathog.* **9**, e1003834 (2013).
6. Pape, K. A., Taylor, J. J., Maul, R. W., Gearhart, P. J. & Jenkins, M. K. Different B cell populations mediate early and late memory during an endogenous immune response. *Science* **331**, 1203–1207 (2011).
7. Laird, G. M. et al. Rapid quantification of the latent reservoir for HIV-1 using a viral outgrowth assay. *PLoS Pathog.* **9**, e1003398 (2013).
8. Scheid, J. F. et al. Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science* **333**, 1633–1637 (2011).
9. Mouquet, H. et al. Complex-type N-glycan recognition by potent broadly neutralizing HIV antibodies. *Proc. Natl. Acad. Sci. USA* **109**, E3268–E3277 (2012).
10. Walker, L. M. et al. Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science* **326**, 285–289 (2009).

11. Gaublomme, J. T. et al. Single-cell genomics unveils critical regulators of Th17 cell pathogenicity. *Cell* **163**, 1400–1412 (2015).
12. Sherrill-Mix, S., Ocwieja, K. E. & Bushman, F. D. Gene activity in primary T cells infected with HIV89.6: intron retention and induction of genomic repeats. *Retrovirology* **12**, 79 (2015).
13. Hunt, M. et al. IVA: accurate de novo assembly of RNA virus genomes. *Bioinformatics* **31**, 2374–2376 (2015).
14. Lorenzi, J. C. et al. Paired quantitative and qualitative assessment of the replication-competent HIV-1 reservoir and comparison with integrated proviralDNA. *Proc. Natl. Acad. Sci. USA* **113**, E7908–E7916 (2016).
15. Mullins, J. I. & Frenkel, L. M. Clonal expansion of human immunodeficiency virus-infected cells and human immunodeficiency virus persistence during antiretroviral therapy. *J. Infect. Dis.* **215**, S119–S127 (2017). suppl_3.
16. Cohn, L. B. et al. HIV-1 integration landscape during latent and active infection. *Cell* **160**, 420–432 (2015).
17. Wagner, T. A. et al. HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science* **345**, 570–573 (2014).
18. Maldarelli, F. et al. HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science* **345**, 179–183 (2014).
19. 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 (2017).
20. Bui, J. K. et al. Proviruses with identical sequences comprise a large fraction of the replication-competent HIV reservoir. *PLoS Pathog.* **13**, e1006283 (2017).
21. 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 (2017).
22. Simonetti, F. R. et al. Clonally expanded CD4+ T cells can produce infectious HIV-1 in vivo. *Proc. Natl. Acad. Sci. USA* **113**, 1883–1888 (2016).
23. 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).
24. 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).
25. Baxter, A. E. et al. Single-cell characterization of viral translation-competent reservoirs in HIV-infected individuals. *Cell Host Microbe* **20**, 368–380 (2016).
26. Cockerham, L. R. et al. CD4$^+$ and CD8$^+$ T cell activation are associated with HIV DNA in resting CD4$^+$ T cells. *PLoS One* **9**, e110731 (2014).
27. Descours, B. et al. CD32a is a marker of a CD4 T-cell HIV reservoir harbouring replication-competent proviruses. *Nature* **543**, 564–567 (2017).
28. Ruelas, D. S. et al. MicroRNA-155 Reinforces HIV Latency. *J. Biol. Chem.* **290**, 13736–13748 (2015).
29. Hudspeth, K. et al. Engagement of NKp30 on Vδ1 T cells induces the production of CCL3, CCL4, and CCL5 and suppresses HIV-1 replication. *Blood* **119**, 4013–4016 (2012).
30. Abdelwahab, S. F. et al. HIV-1-suppressive factors are secreted by CD4$^+$ T cells during primary immune responses. *Proc. Natl. Acad. Sci. USA* **100**, 15006–15010 (2003).
31. Kaczmarek Michaels, K. et al. Blimp-1, an intrinsic factor that represses HIV-1 proviral transcription in memory CD4$^+$ T cells. *J. Immunol.* **194**, 3267–3274 (2015).
32. de Masson, A. et al. Blimp-1 overexpression is associated with low HIV-1 reservoir and transcription levels in central memory CD4$^+$ T cells from elite controllers. *AIDS* **28**, 1567–1577 (2014).
33. Bruner, K. M. et al. Defective proviruses rapidly accumulate during acute HIV-1infection. *Nat. Med.* **22**, 1043–1049 (2016).
34. Ho, Y. C. et al. Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* **155**, 540–551 (2013).
35. Siliciano, R. F. & Greene, W. C. HIV latency. *Cold Spring Harb. Perspect. Med.* **1**, a007096 (2011).
36. Doitsh, G. & Greene, W. C. Dissecting how CD4 T cells are lost during HIV infection. *Cell Host Microbe* **19**, 280–291, https://doi.org/10.1016/j.chom.2016.02.012 (2016).
37. Lu, C. L. et al. Enhanced clearance of HIV-1-infected cells by broadly neutralizing antibodies against HIV-1 in vivo. *Science* **352**, 1001–1004 (2016).
38. Horwitz, J. A. et al. Non-neutralizing antibodies alter the course of HIV-1 infection in vivo. *Cell* **170**, 637–648.e610 (2017).

Acknowledgements

We thank all participants who contributed to this study; members of the Nussenzweig laboratory for helpful discussions, particularly E. Kara and T. Oliveira; our lab manager Z. Jankovic; L. Mesin and M. Biton for advice on scRNA-seq; A. Gazumyan for bNAb production; G. Breton for help with FACS; K. Gordon and N. Thomas for performing all FACSorting experiments; A. Han and M. Davis for TCR sequencing advice; D. Mucida, C. Rice and P. Bieniasz for helpful discussion; K. Millard for recruitment of study subjects; and M. Deal for assistance with figures. This work was supported by the Bill and Melinda Gates Foundation Collaboration for AIDS Vaccine Discovery (OPP1033115 and OPP1124068), the National Institutes of Health (NIH) Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery (CHAVI-ID) (1UM1 AI100663), BEAT-HIV Delaney Collaboratory (UM1 AI126620), the National Institute of Allergy and Infectious Diseases of the NIH (AI100148, AI037526), the Robertson Foundation, and the Rockefeller University. M.C. is supported by NIH grant U01AI118536. M.C.N. is a Howard Hughes Medical Institute (HHMI) investigator.

Author contributions

L.B.C., M.J., and M.C.N. wrote the manuscript; L.B.C., M.J., and M.C.N. designed and analyzed experiments; L.B.C. and M.J. performed LURE experiments, RNA sequencing, Q^{2}VOAs, TCR sequencing, and virus SGA; R.V. and I.T.d.S performed bioinformatics analysis of RNA-seq data; A.S.H. performed TCR sequencing and virus SGA; J.C.C.L. and Y.Z.C. performed Q^{2}VOAs; J.A.P. performed phylogenetic analysis of *env* sequencing and gene enrichment analysis; A.L.B. and M.C. performed study subject recruitment and oversaw sample collection.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41591-018-0017-7.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.C.N.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Study subjects. All study participants were recruited by the Rockefeller University Hospital, New York, New York, USA. Written, informed consent was obtained from all subjects, all relevant ethical regulations were followed, and leukapheresis was performed according to protocols approved at the Rockefeller University by the Rockefeller Internal Review Board. PBMCs were isolated by Ficoll separation and frozen in aliquots. In all cases, HIV-1-infected individuals on therapy were confirmed to be aviremic at the time of sample collection.

Latency capture protocol. $CD4^+$ T cells were isolated from ~1×10^9 PBMCs through negative selection using the Miltenyi $CD4^+$ T cell isolation kit. Cells were cultured at 2×10^6 cells/ml in R10 (RPMI supplemented with 10% heat-inactivated FCS, 10 mM HEPES, 100 U/ml penicillin–streptomycin) and 25% volume-conditioned medium. Conditioned medium was made by culturing healthy peripheral blood mononuclear cells (PBMCs) in R10 with PHA and IL-2 for 2 d; this was followed by a wash and 5 d in culture with IL-2 alone. The conditioned medium was then collected and frozen at −80 °C until use. 100 U/ml IL-2 (Peprotech), 1 µg/ml PHA (Sigma), 10 µM Z-VAD-FKM (R&D), 10 µM ritonavir, 10 µM dolutegravir, 10 µM emtricitabine, 5 µM tenofovir, and 10 µM maraviroc (all Selleckchem) were added to the medium. 36 h later, cells were labeled with 5 µg/ml each of biotinylated 3BNC117, 10-1074, and PG16, followed by Streptavidin PE (1:500, BD) and anti-PE magnetic beads (Miltenyi Biotech). Cells were then passed over a magnetic column, and bound cells were eluted for downstream analysis. For FACS sorting, cells were labeled with antibodies against the following proteins: CD1c (cat. no. 331510), CD3 (cat. no. 300430), CD4 (cat. no. 317444), CD8 (cat. no. 344726), CD14 (cat. no. 301812), CD20 (cat. no. 302318), CD32a (cat. no. 303204), and CD56 (cat. no. 318314) (all Biolegend).

***gag* bulk qPCR.** RNA was extracted from equivalent numbers of cells irrespective of enrichment. *gag* qPCR was performed using RNA-to-Ct 1-step RT–PCR mix (ThermoFisher) and previously described primers[39].

Single-cell sorting. All sorts were performed on BD FACS Aria into 96-well plates containing guanidine thiocyanate buffer (Qiagen) supplemented with 1% β-mercaptoethanol. Plates were immediately frozen on dry ice and transferred to long-term storage at −80 °C. LURE cells were gated on live, $CD1c^-CD8^-CD14^-CD20^-CD56^-CD3^+$ and were sorted based on Env staining. Control cells were gated on live, $CD1c^-CD8^-CD14^-CD20^-CD56^-$ and sorted $CD3^+$ cells.

Single-cell *gag* qPCR and *env* PCR. Nucleic acids were isolated using SPRI bead cleanup, as described[40]. RNA was reverse-transcribed into cDNA using an oligo(dT) primer. *gag* qPCR was performed on one-fifth of the cDNA[39]. Gag^+Env^+ cells were selected on the basis of the presence of cell-associated *gag* RNA measured by qPCR. Control cells were assayed for *gag* RNA, and none was detected. Nested *env* PCR was performed on one-fifth of the cDNA[14].

SGA of *env* DNA and Q^2VOAs. DNA was extracted from isolated $CD4^+$ T cells as previously described[16], and SGA of *env* DNA was performed as previously described[14]. Q^2VOAs and downstream analysis were performed and processed as previously described[14]. For quality control, Q^2VOAs were performed more than once and on samples taken at two different time-points for donor B207. IUPM calculations were performed using the data from all independent experiments using the calculator IUPMStats[41].

Clustering *env* Sequences. *env* nucleotide sequences were translation-aligned using ClustalW 2.1 with the BLOSUM cost matrix in Geneious v10.0.3. A maximum-likelihood tree was then inferred using PhyML 3.1 under the GTR model with 1,000 bootstrap replicates.

YU2 infection and sorting. $CD4^+$ T cells were activated and infected with YU2 and labeled as previously described[37]. $CD4^{lo}Env^+$ cells were sorted.

Single-cell RNA-seq. RNA-seq libraries were constructed based on Trombetta et al.[42] using primers from Islam et al.[43] Briefly, RNA was converted to full-length cDNA using oligo(dT) priming (Bio- 5′-AAT GATACGGCGACCACCGATCGTTTTTTTTTTTTTTTTTTTTTTT TTTTTTTTT-3′) and SMART template switching technology (all RNA oligonucleotides: Bio- 5′-AAUGAUACGGCGACCACCGAUNNNNNGGG-3′) followed by 24 cycles of PCR preamplification of cDNA (primer: Bio-5′-GAAT GATACGGCGACCACCGAT -3′). We used the amplified cDNA to construct standard Illumina sequencing libraries with the Nextera XT library preparation kit. Samples were sequenced by Illumina NextSeq.

RNA-seq analysis. The quality of the RNA-seq libraries was evaluated using the FastQC1 tool[44]. We used the STAR (2.4.1d)[45] aligner to map the raw paired-end reads to the reference genome GRCh37/hg19. The gene-level counts were obtained using HTSeq[44]. We performed a saturation analysis to detect the number of detected genes and filtered out the outlier cells, as described in Gaublomme et al.[11]. Briefly, we excluded cells with a number of aligned reads < 25,000 and a percentage of identified genes that was < 20% of the group maximum. Normalized expression values were calculated using the scran package[46] in Bioconductor. Heat maps and dot plots were generated in R. The gene counts were used to infer the DEGs in the data by MAST (v1.2.1)[47].

Analysis of HIV splice variants. We recovered the reads that failed to map to the human genome and mapped these reads to annotated junctions between HIV splice donors and acceptors to reconstruct the splice variants present in the scRNA-seq data.

HIV read alignment and reconstruction. We carried out HIV assembly analysis on the all reads that failed to map to the human genome by the IVA de novo assembler (v1.0.7)[13].

T cell receptor identification. TraceR[48] was used to reconstruct full-length, paired TCR sequences. TCR sequences unable to be recovered from RNA-seq reads were amplified as previously described[49].

PCA Seurat. We used the Seurat package (v1.4.0.16) to identify variable genes, principal components (PCs), clusters, and gene markers, as described[23]. Briefly, the software identifies highly variably expressed genes using a normalized *z*-score, performs linear dimensional reduction (PCA) on the filtered genes, obtains additional transcriptome PCA loading genes using projection of these principal components to the entire dataset, determines groups by clustering the t-SNE significant principal component scores on the basis of density, and performs discovery of gene markers. We also used the Improved Stochastic Ranking Evolution Strategy algorithm[50], implemented by NLopt, to find the optimal set of PCs and parameters and to find the optimal set of clusters that best correlate with each group of cells.

Single-cell consensus clustering. The SC3 tool[51] (with default settings) was used for unsupervised clustering of single cells in this study. SC3 consistently integrates different clustering solutions through a consensus approach and identifies marker genes, which are highly expressed in only one of the clusters, and distinguishes these genes from the remaining ones[51].

We have tested combinations of clustering settings (k = 2, 3, and 4) and used a quantitative measure of the diagonality of the consensus matrix to select the k (the number of clusters originally identified) in which the measure is closest to 1 (k = 3). We then used SC3 (area under the receiver operating characteristic curve (AUROC) > 0.6 and false-discovery rate (FDR) < 0.1) to identify marker genes that are highly expressed in only one of the clusters, which is distinguishable from all of the remaining clusters.

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

Data availability. The data reported in this paper is archived at the following databases: scRNA-seq data (Figs. 2 and 4) is available at National Center for Biotechnology Information Gene Expression Omnibus (GSM2801437); Env sequences (Fig. 3) are available in the GenBank database (MG196359–MG196639); TCR sequences (Supplementary Fig. 5) are available in the GenBank database (MG192535–MG193127).

References

39. Palmer, S. et al. New real-time reverse transcriptase-initiated PCR assay with single-copy sensitivity for human immunodeficiency virus type 1 RNA in plasma. *J. Clin. Microbiol.* **41**, 4531–4536 (2003).
40. Tas, J. M. et al. Visualizing antibody affinity maturation in germinal centers. *Science* **351**, 1048–1054 (2016).
41. Rosenbloom, D.I.S. et al. Designing and interpreting limiting dilution assays: general principles and applications to the latent reservoir for HIV-1. *bioRxiv* http://doi.org/10.1101/018911 (2015).
42. Trombetta, J. J. et al. Preparation of single-cell RNA-seq libraries for nextgeneration sequencing. *Curr. Protoc. Mol. Biol.* **107**, 4.22.1–17 (2014).
43. Islam, S. et al. Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* **11**, 163–166 (2014).
44. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work withhigh-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
45. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**,15–21 (2013).
46. Lun, A. T., McCarthy, D. J. & Marioni, J. C. A step-by-step workflow forlow-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res.* **5**,2122 (2016).
47. 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).

48. Stubbington, M. J. T. et al. T cell fate and clonality inference from single-celltranscriptomes. *Nat. Methods* **13**, 329–332 (2016).
49. Han, A., Glanville, J., Hansmann, L. & Davis, M. M. Linking T-cell receptorsequence to functional phenotype at the single-cell level. *Nat. Biotechnol.* **32**,684–692 (2014).
50 Runarsson, T. P. & Yao, X. Search biases in constrained evolutionaryoptimization. *IEEE Trans. on Systems, Man, and Cybernetics, Part C(Applications and Reviews)* **35**, 233–243 (2005).
51 Kiselev, V. Y. et al. SC3: consensus clustering of single-cell RNA-seq data. *Nat. Methods* **14**, 483–486 (2017).

nature research

Corresponding author(s): Michel C. Nussenzweig

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

Experimental design

1. Sample size

   Describe how sample size was determined.

   Samples were chosen based on reproducibility of the LURE assay and availability of patient material. Sample size (cell number) was determined to be adequate based on the magnitude and consistency of measurable differences between groups.

2. Data exclusions

   Describe any data exclusions.

   Single-cell RNASeq libraries were excluded based on the number of mapped reads and number of identified genes. We introduced hard lower bounds for the cutoff values (number of aligned reads>25,000; percentage of identified transcripts>20%) and only retain cells that scored above the cutoff in both cases.

3. Replication

   Describe the measures taken to verify the reproducibility of the experimental findings.

   Replicate experiments were successful. We optimized the LURE method and then tested its reproducibly on multiple individuals.

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   No randomization was performed because we took unbiased approaches when analyzing the single cell data.

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   Investigators were not blinded to patient characteristics during experiments because blinding was not relevant to our study. The nature of these experiments included negative controls from the same samples, and LURE, by definition must be performed on HIV+ individuals.

Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ Test values indicating whether an effect is present *Provide confidence intervals or give results of significance tests (e.g. P values) as exact values whenever appropriate and with effect sizes noted.*
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars in all relevant figure captions (with explicit mention of central tendency and variation)

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

Flow cytometry data analyzed using FlowJo (v10). IUPMStats was used to calculate IUPM. Phylogenetic trees were made using ClustalW 2.1 with the BLOSUM cost matrix in Geneious v10.0.3 followed by PhyML 3.1 under the GTR model with 1000 bootstraps. RNASeq libraries were analyzed using fastQC1, STAR aligner, HTSEQ and differentially expressed genes analyzed by MAST. Virus assembly was performed using IVA de novo assembler. TCR sequences were recovered using TraceR. PCA was performed using Seurat software with NLopt. Unsupervised clustering was performed by Single-cell Consensus Clustering (SC3). Statistics were calculated in Prism software.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.

No unique materials were used in this study.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

Antibodies used are as follows:
purified 3BNC117, PG16 and 10-1074 at 5ug/mL
Commercial antibodies purchased from Biolegend
CD1c: cat no. 331510 lot no. B204632 dilution. 1:200
CD3: cat no. 300430 lot no. B221391 dilution. 1:200
CD4: cat no. 317444 lot no. B235088 dilution. 1:200
CD8: cat no. 344726 lot no. B214375 dilution. 1:200
CD14: cat no. 301812 lot no. B218589 dilution. 1:200
CD20: cat no. 302318 lot no. B200970 dilution. 1:200
CD32a: cat no. 303204 lot no. B220034 dilution. 1:100
CD56: cat no. 318314 lot no. B205430 dilution. 1:200

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used. — No eukaryotic cell lines were used

b. Describe the method of cell line authentication used. — No eukaryotic cell lines were used

c. Report whether the cell lines were tested for mycoplasma contamination. — No eukaryotic cell lines were used

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use. — No eukaryotic cell lines were used

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study. — No animals used in this study.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants. — ART treated patient information is included in the supplementary information. Patients were selected to be stably treated and virologically suppressed for at least one year. We did not exclude patients based on age or sex. Healthy donors tested negative for blood borne pathogens.

nature research

Corresponding author(s): Michel C. Nussenzweig

Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data presentation

For all flow cytometry data, confirm that:

☒ 1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ 2. The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ 3. All plots are contour plots with outliers or pseudocolor plots.

☒ 4. A numerical value for number of cells or percentage (with statistics) is provided.

Methodological details

5. Describe the sample preparation.	Sample preparation listed in Methods
6. Identify the instrument used for data collection.	FACSAria II
7. Describe the software used to collect and analyze the flow cytometry data.	FACSDiva for collection and FlowJo (version 10) for analysis
8. Describe the abundance of the relevant cell populations within post-sort fractions.	Purity of Env+gag+ cells was assessed by single cell gag qPCR and results are described in main text.
9. Describe the gating strategy used.	Relevant gating strategy provided in supplemental information

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information. ☒