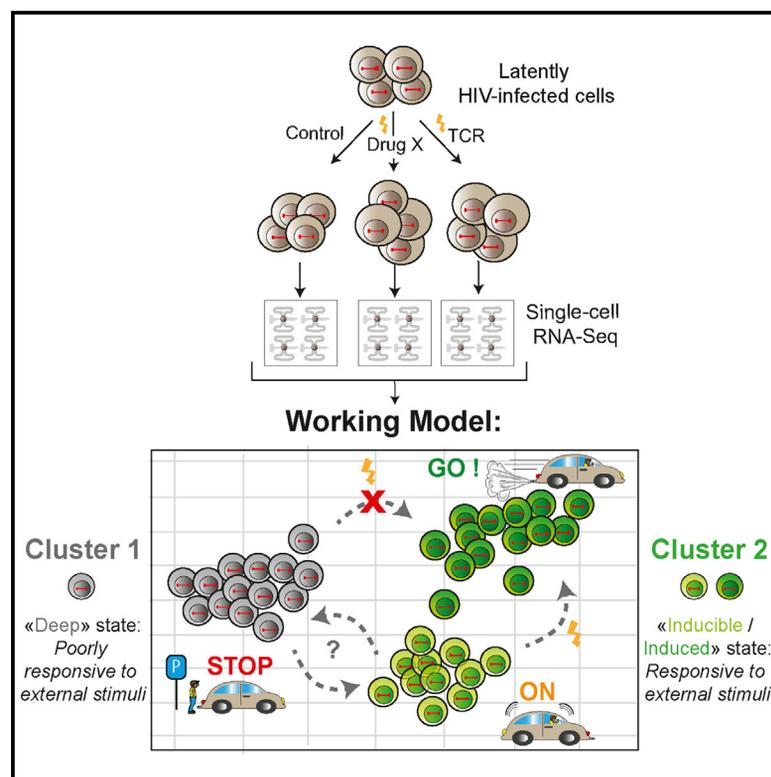# Cell Reports

Single-Cell RNA-Seq Reveals Transcriptional Heterogeneity in Latent and Reactivated HIV-Infected Cells

Graphical Abstract

Authors

Monica Golumbeanu, Sara Cristinelli, Sylvie Rato, Miguel Munoz, Matthias Cavassini, Niko Beerenwinkel, Angela Ciuffi

Correspondence

niko.beerenwinkel@bsse.ethz.ch (N.B.), angela.ciuffi@chuv.ch (A.C.)

In Brief

HIV latency hampers HIV cure. The shock-and-kill strategy aims at reactivating HIV expression to purge the latent reservoir of HIV-infected cells. However, latently infected cells do not respond equally to stimulation. Golumbeanu et al. use single-cell RNA-seq to characterize cell heterogeneity and identify transcriptional features leading to reactivation success.

Highlights

- Latent and reactivated HIV-infected cells are heterogeneous

- Single-cell RNA-seq identified two distinct clusters in latent and reactivated cells

- Cell clusters displayed differential susceptibility to HIV reactivation

- A 134-gene-specific transcriptional signature identified the inducible cell

Data and Software Availability

GSE111727

CellPress

Single-Cell RNA-Seq Reveals Transcriptional Heterogeneity in Latent and Reactivated HIV-Infected Cells

Monica Golumbeanu,[1,2] Sara Cristinelli,[3] Sylvie Rato,[3] Miguel Munoz,[3] Matthias Cavassini,[4] Niko Beerenwinkel,[1,2,*] and Angela Ciuffi[3,5,*]

[1]Department of Biosystems Science and Engineering, ETH Zurich, Basel 4058, Switzerland
[2]SIB Swiss Institute of Bioinformatics, Basel 4058, Switzerland
[3]Institute of Microbiology, Lausanne University Hospital and University of Lausanne, Lausanne 1011, Switzerland
[4]Service of Infectious Diseases, Department of Medicine, Lausanne University Hospital and University of Lausanne, Lausanne 1011, Switzerland
[5]Lead Contact
*Correspondence: niko.beerenwinkel@bsse.ethz.ch (N.B.), angela.ciuffi@chuv.ch (A.C.)
 https://doi.org/10.1016/j.celrep.2018.03.102

SUMMARY

Despite effective treatment, HIV can persist in latent reservoirs, which represent a major obstacle toward HIV eradication. Targeting and reactivating latent cells is challenging due to the heterogeneous nature of HIV-infected cells. Here, we used a primary model of HIV latency and single-cell RNA sequencing to characterize transcriptional heterogeneity during HIV latency and reactivation. Our analysis identified transcriptional programs leading to successful reactivation of HIV expression.

INTRODUCTION

Early on during the course of natural infection, HIV establishes a reservoir of persisting infected cells. Two types of viral reservoirs have been described so far: anatomical sanctuaries and a latent reservoir characterized by the absence of viral particle production (Chun et al., 2015; Eisele and Siliciano, 2012). These reservoirs are not eliminated by antiretroviral therapy (ART), as viruses can rebound upon ART cessation and are thus considered a major barrier to viral eradication and cure (Davey et al., 1999; Whitney et al., 2014). Although many cell types can be infected by HIV and may constitute part of the latent reservoir, most studies have focused on the CD4+ T cell latent reservoir. Two hypotheses exist for the establishment of the CD4+ T cell latent reservoir: (1) direct infection of resting cells, and (2) infection of activated cells that revert to a memory resting phenotype (Chavez et al., 2015; van der Sluis et al., 2013). Diverse mechanisms have been associated with HIV latency, including transcriptional repression (availability and location of transcription factors, epigenetic regulation, chromatin environment of the provirus) and post-transcriptional blocks (nuclear export, translation) (Abbas and Herbein, 2012; Cary et al., 2016; Ciuffi et al., 2015; Ciuffi and Telenti, 2013; Huang et al., 2007; Lassen et al., 2006; Li et al., 2012; Mbonye and Karn, 2017; Mohammadi et al., 2014). Multiple CD4+ T cell subsets can be infected with HIV and be present at multiple cellular states, either resting or

activated (Baxter et al., 2018; Kulpa and Chomont, 2015; Shan et al., 2017). All these factors contribute to the complexity and heterogeneity of the HIV latency reservoir (Boritz et al., 2016; Chun et al., 2015).

One approach toward an HIV cure, the so-called "shock and kill" strategy, aims at reactivating HIV particle production from the latent cell so that it will die upon virus-mediated cytotoxicity or be killed by cytotoxic CD8+ T lymphocytes, thereby purging the latent reservoirs (Darcis et al., 2017). Multiple latency reversing agents (LRAs) have been tested for their ability to reactivate HIV expression in latently infected cells albeit with limited success (Darcis et al., 2015; Ho et al., 2013; Mohammadi et al., 2014; Spina et al., 2013). Some histone deacetylase inhibitors (HDACis), such as vorinostat (SAHA), proved successful in inducing viral transcription, but failed to induce successful protein expression and viral particle release (Blazkova et al., 2012; Bullen et al., 2014; Ho et al., 2014; Mohammadi et al., 2014). Other stimuli leading to T cell receptor (TCR)-mediated cellular activation have been shown to induce successful viral particle production (Darcis et al., 2015; Spina et al., 2013). Despite displaying the most potent phenotype, TCR-mediated HIV reactivation occurs only in a fraction of cells, i.e., some cells were successfully induced while some others remained unresponsive and thus not induced, confirming cellular heterogeneity of the latent reservoir (Ho et al., 2013, 2014). The mechanisms responsible for this differential reactivation potential and thus for the inducible phenotype upon exposure to LRA and TCR have yet to be elucidated.

The recent advent of single-cell sequencing technologies has been a major asset for the field of virology, allowing to study the heterogeneity of cellular response to viral infection (Ciuffi et al., 2016). Single-cell approaches have opened new perspectives in HIV research, such as the characterization of HIV replication cycle delays in individual cells (Holmes et al., 2015), the study of cellular heterogeneity of the latent reservoir (Baxter et al., 2016; Yucha et al., 2017), and assessing the heterogeneity of cellular response to LRAs (Passaes et al., 2017). Although technical and computational challenges are still abundant in the field of single-cell sequencing, numerous efforts exist to alleviate the burden of technical biases, and a multitude of statistical
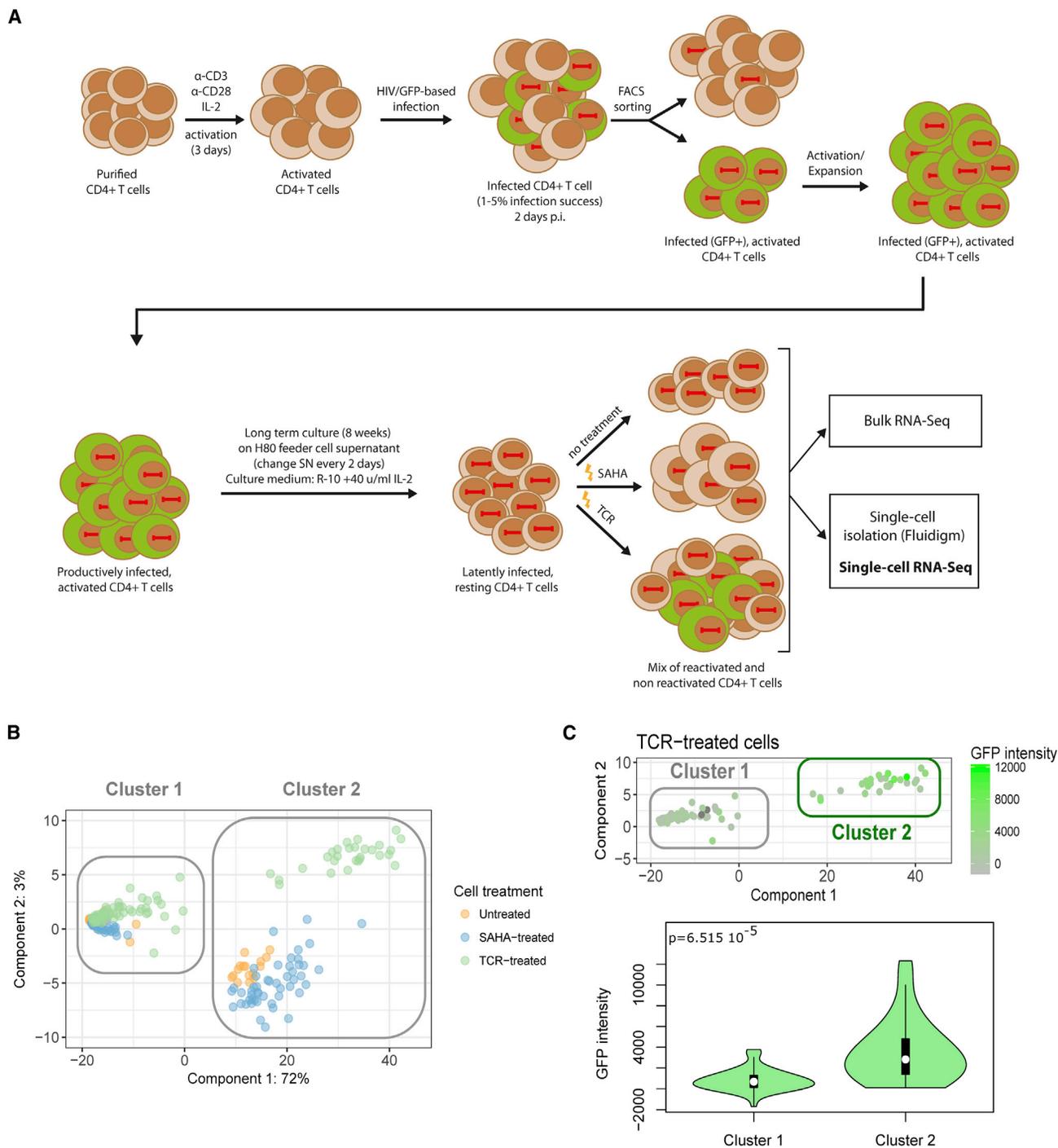
Figure 1. Latency and Reactivation at the Single-Cell Level

(A) Experimental design overview. Primary CD4[+] T cells were first activated by TCR stimulation and infected with a HIVGFP/VSV-G virus. Two days post-infection, successfully infected GFP[+] cells were sorted by FACS and further expanded in culture. Cells were then cultured for 8 weeks on a minimal medium, in presence of H80 cell culture supernatant to promote cell survival and generate latently infected cells. Cells were then either left untreated or exposed to SAHA or to TCR treatment before single-cell isolation and single-cell RNA sequencing. After single-cell isolation of TCR-stimulated cells, individual cells were imaged by fluorescence microscopy in order to assess GFP expression. For each condition, bulk and single-cell RNA sequencing were performed.

(B) Principal component analysis (PCA) of single-cell gene expression profiles segregates cells in two major clusters for each condition. Each dot represents the gene expression profile of a single cell. The colors indicate the various experimental conditions: untreated (orange), SAHA-treated (blue), and TCR-treated (green). The axis labels indicate the percentages of explained variance corresponding to the represented principal component.

(legend continued on next page)

methods for the analysis of single-cell data are continuously being published (Bacher and Kendziorski, 2016; Brennecke et al., 2013; Buettner et al., 2015; Rato et al., 2017).

In this work, we aimed at exploring the cellular heterogeneity and characterizing the transcriptomic profile of latent and reactivated HIV-infected cells at the single-cell level. For this purpose, we employed a previously established HIV latency model that uses human primary CD4$^+$ T cells (Mohammadi et al., 2014) and exposed them to different reactivation conditions. Our data revealed that latently infected cells are transcriptionally heterogeneous, separating in two distinct cell clusters, and transcriptional profiles correlate with the susceptibility to cellular activation and HIV expression reactivation, thereby allowing to identify features of the HIV-inducible cell.

RESULTS AND DISCUSSION

Latent and Reactivated Cells Are Heterogeneous, Separating in Two Distinct Clusters

We used a previously described primary model of HIV latency that consists in infecting activated human primary CD4$^+$ T cells with a GFP-based HIV vector, sorting the GFP$^+$ cells by fluorescence-activated cell sorting (FACS), and culturing the infected cells for a long time to allow cell reversion to a resting, latent phenotype (Mohammadi et al., 2014; Sahu et al., 2006; Tyagi et al., 2010). Latently infected cells were either left untreated or exposed to SAHA or to TCR stimulation, followed by single-cell isolation and single-cell RNA-sequencing (scRNA-seq) analysis (Figure 1A). Bulk RNA-seq was also performed as control. The addition of spike-in control sequences allowed calibrating the abundances of sequencing reads for the single-cell samples across conditions.

After preprocessing of bulk and single-cell RNA sequencing data, filtering of low-quality cells (Supplemental Experimental Procedures; Figures S1A–S1D; Table S1), and normalization, we assessed the heterogeneity of single-cell host gene expression profiles for all three experimental conditions with principal component analysis (PCA). Viral transcripts were not used in the PCA to avoid any bias in cell clustering. The first two principal components explained 75% of the total variance and spatially separated the cells into two major groups, named cluster 1 and cluster 2 (Figure 1B; Table S2). These two clusters were observed in every condition tested (i.e., untreated, SAHA-treated, or TCR-treated) and remained stable also with alternative normalization and dimensionality reduction methods (Supplemental Experimental Procedures; Figure S2).

Cell Clusters Correlate with HIV Reactivation Potential

In order to investigate the correlation between the cellular gene expression profile (and thus the cell cluster) and the successful induction of HIV expression at the protein level, we quantified virally encoded GFP expression, used as a surrogate of viral protein expression, at the single-cell level by fluorescence microscopy and performed subsequent image analysis. As previously shown, SAHA-treated cells did not result in effective GFP expression, consistent with possible presence of post-transcriptional blocks (Table S2) (Mohammadi et al., 2014). In contrast, successful GFP expression was observed in a fraction of the TCR-treated cells. GFP expression intensity was reported for each TCR-treated cell, revealing different levels of GFP intensity between the two clusters, with cells from cluster 1 systematically presenting lower GFP expression than cells from cluster 2 (two-sample t test: p < 10^{-4}, Figure 1C; Table S2). Altogether, these results suggest that the TCR-treated cell population is heterogeneous and displays different HIV reactivation phenotypes, induced and non-induced, overlapping with transcriptionally defined cell clusters. These data are consistent with a model in which TCR-treated cluster-1 cells are similar to untreated or SAHA-treated cluster-1 cells, which are poorly responsive and thus not successfully induced. Conversely, TCR-treated cluster-2 cells correspond to induced cells and could be distinguished from cluster 2 of untreated and SAHA-treated cells, which likely correspond to inducible cells (Figure 1B).

We performed a comparative analysis of HIV transcription and global gene expression for the two cellular clusters. For the three tested conditions, HIV transcript levels were consistently higher in cluster 2 than in cluster 1 (Figure S3A). In TCR-treated cells, virally encoded GFP protein expression was significantly correlated with the number of viral transcripts per cell (Spearman correlation 0.49, Spearman rank correlation test: p < 10^{-3}). Furthermore, cells in cluster 1 have globally a reduced number of expressed genes compared to cluster 2 (Figure S3B). All together, these data are consistent with a cellular model where latent cells display different degrees of resting depths that correlate with their level of global HIV transcription and their response to extracellular stimuli. In this sense, cells from cluster 1 are in a deeper resting state, more difficult to activate upon TCR stimulation, and where HIV expression is also more difficult to induce and reactivate. Cells from cluster 2 are in a less deep resting state and are more responsive to cellular activation and to HIV expression reactivation.

A 134-Gene-Specific Transcriptional Signature Identifies the Inducible Latent Cell

We further compared the two identified cellular subpopulations for each condition in terms of differential expression (DE) of their transcriptional profiles (Figure S4). Globally DE genes between cluster 1 and cluster 2 in the three tested conditions are enriched in metabolism, gene expression, disease, immune system, and DNA repair, processes that are consistent with a cellular state more prone to activation in cluster 2 compared to cluster 1 (Figure S4B; Tables S3 and S4).

(C) The two isolated cellular clusters for the TCR-treated condition show different levels to HIV reactivation response. The upper panel recapitulates the PCA result for TCR-treated cells, where in addition each cell is color-coded according to its measured GFP expression intensity. The corresponding distributions of GFP expression intensity for each cluster are depicted in the lower panel by a violin plot, showing the frequency of observation (violin shape), the median (white dot), the interquartile range (black box) and the outlier boundaries defined as beyond 1.5 of the interquartile range in both directions (vertical bars). GFP expression intensity is lower in cluster 1 than in cluster 2 (two-sample t test: p = 6.515*10^{-5}).
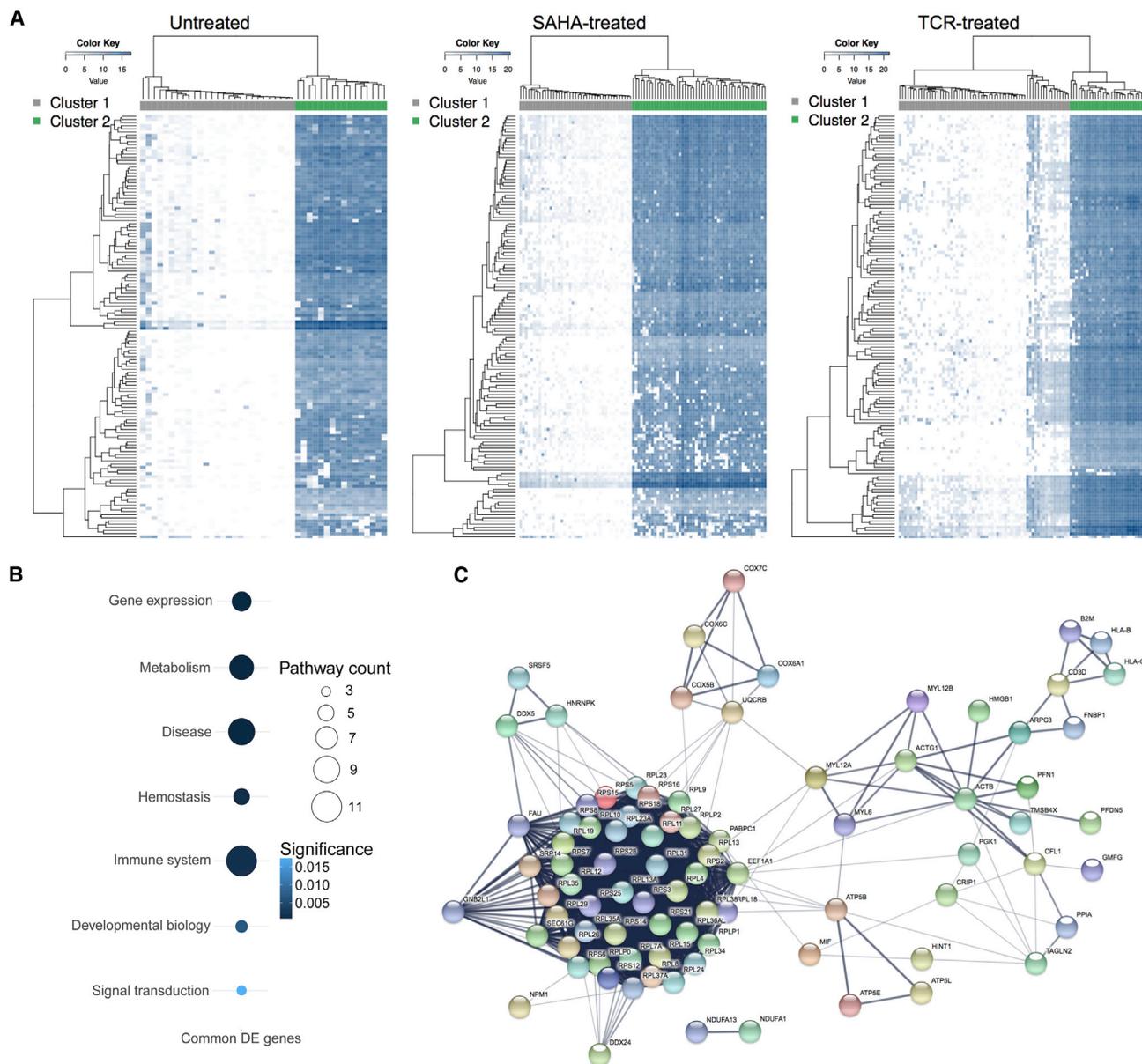See also Figures S1 and S2 and Tables S1 and S2.

Figure 2. Analysis of Differentially Expressed Genes between Cluster 1 and Cluster 2
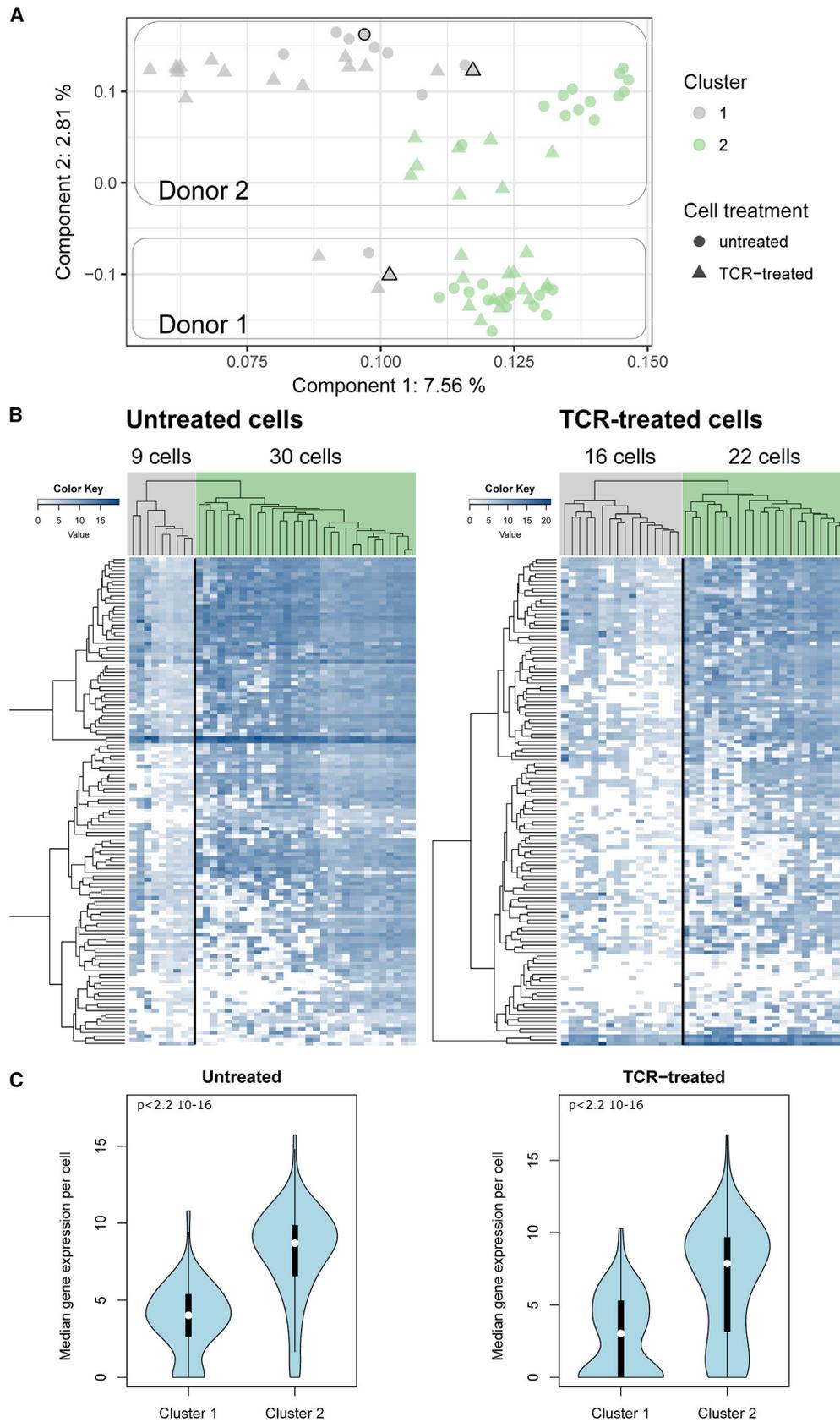
(A) Expression heatmaps for the 134 common differentially expressed (DE) genes across conditions. Each heatmap corresponds to a different treatment condition and displays the log2-transformed expression values of DE genes across cells. Every row of the heatmaps corresponds to a DE gene and each column to a cell. Heatmap rows and columns are grouped by hierarchical clustering and the columns are color-coded according to the corresponding cluster (gray for cluster 1 and green for cluster 2). Expression intensity is color-coded from low (white) to high (blue) expression.

(B) Enrichment analysis result for the 134 DE genes. Enriched pathways were grouped according to the Reactome hierarchy into categories associated to biological processes indicated on the left side of the figure. The size of the circles is proportional to the number of enriched pathways in the corresponding category. The color of each circle corresponds to a significance score equal to the geometric mean of all the corrected p values attributed to the pathways included in each category. Reactome categories are ranked according to their p value.

(C) Network of functional interactions among DE genes. STRING database network analysis was performed for the 134 DE genes. Each node corresponds to a gene and only the connected genes are represented. The edges of the networks correspond to existing experimental and database evidence for gene interaction. See also Figures S3 and S4 and Tables S3 and S4.

Although there are many DE genes in each condition tested, we found overall 134 DE genes differentially expressed between the distinct two cell clusters across all three conditions (Figure S4A; Table S3). With the exception of the Metazoa_SRP gene, the other commonly identified DE genes were upregulated in cluster 2 as compared to cluster 1 (Figure 2A). Enrichment analysis of the 134 common DE genes between cluster 1 and cluster 2 also confirmed the hypothesis that cluster 1 and cluster 2 reflect two

A

B

Untreated cells

9 cells   30 cells

Color Key

0 5 10 15
Value

TCR-treated cells

16 cells   22 cells

Color Key

0 5 10 15 20
Value

C

Untreated

p<2.2 10-16

TCR−treated

p<2.2 10-16

(legend on next page)

distinct cellular states with different impact on cellular activation potential and HIV reactivation efficiency. With ribosomal proteins representing over 48.5% of the 134 DE genes, the resulting enriched pathways corresponded to processes related to the metabolism of RNA and protein, electron transport, RNA splicing, immune system, HIV infection, and translational regulation (Figure 2B; Table S4). This further argues that the cells in cluster 2 display a higher metabolic activity (higher activity of cellular machinery and higher gene expression, including viral genes) as compared to the cells in cluster 1. In contrast, no enrichment in apoptosis genes was found, discarding the hypothesis that cells in cluster 1 are apoptotic. An additional analysis using the STRING database online resource (Szklarczyk et al., 2017), applied to the 134 common DE genes, revealed a strongly connected network of functional interactions and enrichment of viral processes, translational regulation, RNA and protein metabolism, as well as cell activation (Figure 2C; Table S4). Thus, the set of 134 DE genes represents a unique transcriptional signature, hallmarking cells from cluster 1 opposed to cluster 2, and a very specific signature for predisposition to successful activation, hence for the HIV-inducible cell. These 134 genes constitute key players of the cellular machinery, suggesting that the two clusters correspond to different cellular activation stages. On one hand, the cells in cluster 1 are in a deep resting state, where the applied stimulation is not able to successfully reactivate them in the given time window of treatment. On the other hand, the cells in cluster 2 are more responsive to stimulation, due to the increased potential of the existing cellular machinery. These two states of activation also reflect the capability of viral reactivation, with cells in cluster 2 being inducible, thus able to successfully reactivate HIV expression, as opposed to cells in cluster 1. The gene expression profiling of the 134 gene-specific signature and the PCA analysis of the TCR-treated cells highlights a subset of 16 cells in the cluster 1 with an intermediate phenotype, which is consistent with cells transitioning between the two clusters (Figures 2B and S4C).

The 134-Gene-Specific Signature Also Identifies Two Cell Clusters *In Vivo*

To explore cellular heterogeneity *in vivo* and to validate the 134-gene signature identified using the primary HIV latency *in vitro* model, we performed a similar analysis using primary CD4$^+$ T cells isolated from HIV$^+$ individuals. As for the *in vitro* model, resting cells from HIV$^+$ individuals were either not treated or TCR-treated before single-cell isolation and single-cell RNA-seq (Supplemental Experimental Procedures; Figures S1E and

S1F; Tables S1 and S2). Principal component analysis applied to the genome-wide expression profiles of single cells isolated from the two HIV$^+$ donors and followed by clustering also identified two groups of cells for each donor in both the untreated and TCR-treated condition (Figure 3A). We performed an additional analysis, using hierarchical clustering and focusing only on the expression of the 134-gene-specific signature identified in the HIV latency model. Similar results were obtained for most of the single cells, suggesting that the 134-gene signature is able to discriminate cell clusters (Figure 3). As in the *in vitro* model, expression of the 134 genes was higher in cluster 2 as compared to cluster 1 (Figures 3B and 3C). These results confirm the presence of the previously observed cellular heterogeneity *in vivo*, using cells isolated from HIV$^+$ individuals, thereby excluding the possibility that the two cell clusters identified *in vitro* are an artifact of the HIV latency model used.

Working Model

Immune cells are heterogeneous and consist of different cellular subpopulations such as central memory, effector memory, and circulating follicular helper cells, each of which are present at different cell-cycle stages (Banga et al., 2016; Baxter et al., 2018; Kulpa and Chomont, 2015). Consequently, these cells may not respond uniformly to stimulation (Mohammadi et al., 2015). This study used single-cell resolution (single-cell RNA sequencing) to highlight and characterize the transcriptional heterogeneity of HIV-infected cells in the context of latency and reactivation in a primary HIV latency model, without any *a priori* knowledge about the pre-existing cell heterogeneity. We identified two major cell subpopulations with different HIV reactivation potential, characterized by a set of 134 markers. These two cell clusters also display differential levels of global gene expression and metabolic activity, consistent with one cluster of cells (cluster 2) being more prone to cellular activation and HIV reactivation (Besnard et al., 2016). Our data suggest a model in which cells may transition between the two clusters (Figure 1B, untreated and SAHA-treated cells), alternating between a poorly responsive and a responsive and hence inducible cell state. When cells are confronted to strong external signals such as TCR-mediated stimulation, cluster-1 cells will remain non-induced (Figure 1B, TCR-treated cells in cluster 1) while cluster-2 cells will be successfully induced (Figure 1B, TCR-treated cells in cluster 2). Although HIV integration sites were recently shown to impact HIV reactivation, our single-cell data suggest that the cellular environment can also contribute to HIV reactivation success (Chen et al., 2017; Ciuffi et al., 2017; Mohammadi et al., 2014).

Figure 3. Expression of Selected 134 DE Genes across Untreated and TCR-Treated Primary CD4$^+$ T Cells Isolated from HIV$^+$ Individuals

(A) Principal component analysis and clustering of single cells from the two donors according to their gene expression profiles. The axis labels indicate the percentages of explained variance by the respective principal component. The first principal component separates the cell clusters, while the second principal component separates the cells from the two donors. The two clusters (except for the three highlighted cells, black border) are equivalent to clusters identified by hierarchical clustering of the cells using the expression of the 134 DE gene signature.

(B) Each heatmap corresponds to a different treatment condition, every row corresponds to a DE gene, and each column represents a cell. Genes and cells are grouped according to hierarchical clustering. For each condition, the 134 DE genes separate the cells in two major groups with contrasting expression.

(C) Distribution of the median gene expression per cell for the two previously identified clusters per condition using a violin plot, showing the frequency of observation (violin shape), the median (white dot), the interquartile range (black box), and the outlier boundaries defined as beyond 1.5 of the interquartile range in both directions (vertical bars). Cells in cluster 1 display a significantly lower expression than cells in cluster 2 (two-sample t test: $p < 2.2*10^{-16}$).

See also Figure S1 and Table S1.

The gene signature identified in this study should provide a valuable tool to facilitate the identification of successful LRA, able to stimulate HIV expression in all the resting cells regardless of their phenotypes, and to help identify potential biomarkers of inducible cells.

EXPERIMENTAL PROCEDURES

Ethics Statement
Blood donors and patients included in the Swiss HIV Cohort Study have all provided written informed consent forms for research use, which were approved and validated by local ethics committees.

Primary HIV Latency Model
The latency model was generated as described previously (Mohammadi et al., 2014) (Figure 1A). Briefly, CD4$^+$ T cells were purified from uninfected donors' buffy coat (human CD4$^+$ T Cell enrichment kit; Stem Cell Technologies #17952), resuspended at 10^6 cells/mL, and activated through TCR-mediated stimulation (25 μL/mL ImmunoCult Human CD3/CD28 T Cell Activator; Stem Cell Technologies #10971) and 100 U/mL IL-2 (RD #202-IL). Three days post-TCR activation, cells (2 million [mio]) were transduced with 1.2 × 10^5 TU of HIVGFP/VSV-G in presence of 5 μg/mL polybrene by spinoculation (3 hr, 1,500 × g, 25°C). After 3 days, cells were sorted by FACS based on GFP expression and further expanded. Infected GFP$^+$ cells were finally allowed to revert to a resting phenotype by long-term culture in Latency Medium (50% R-10/50% H80 feeder cell supernatant supplemented with 40 U/mL IL-2) for 8 weeks. For reactivation, 2 million cells were either left untreated or incubated with 0.5 μM SAHA for 24 hr, or exposed to anti-CD3/CD28/IL-2 for TCR stimulation for 48 hr as described above. Cells were then washed and collected for bulk or single-cell RNA-seq. On one hand, cells (1 mio) were resuspended in 400 μL lysis buffer of the ZR-Duet DNA/RNA miniprep (Zymo Research #D7001) and processed for RNA extraction according to manufacturer's instructions and used for bulk RNA-seq (Illumina HiSeq Ribo-Zero TruSeq stranded [str] RNA-seq). On the other hand, cells were resuspended at 0.5–1 × 10^6 cells/mL in R-10 (0.45 μm filtered) and processed for single-cell isolation and cDNA synthesis on a small 5–10 μm fluidigm plate (fluidigm C1 Single Cell AutoPrep System, Clontech SMARTer Ultra Low RNA kit for Illumina sequencing), followed by single-cell RNA-seq (Illumina Nextera XT DNA Sample Preparation). High-throughput sequencing (100 cycles, single end) was performed on Illumina HiSeq2500 (University of Lausanne, Genomic Technology Facility).

Quantitative Analysis of GFP Expression in TCR-Treated Cells
After single-cell capture on the Fluidigm C1 IFC plate (5–10 μm), each capture chamber was visually analyzed by microscopy and pictures were captured with a Zeiss Axiovert 200 M fluorescence microscope (Plan-Neofluar 20X lens) equipped with a Roper Scientific CoolSnap HQ camera. Pictures in bright field and FITC channel were taken with MetaMorph 6.3 software, and analyzed using ImageJ 1.50b software. The corrected total cell fluorescence was calculated as previously described (McCloy et al., 2014).

Raw data are freely available in Zenodo repository https://doi.org/105281/zenodo.1204334.

Cells from HIV-Infected Individuals
Blood was collected from two HIV-infected individuals participating in the Swiss HIV Cohort Study (http://www.shcs.ch) that were on antiretroviral therapy for more than 3 years, with undetectable viremia and a CD4 cell count above 300 for more than 1 year. Blood (~25 mL) was directly collected in 4 CPT tubes (BD Vacutainer CPT; BD Biosciences #362753) and processed for peripheral blood mononuclear cell (PBMC) isolation according to manufacturer's instructions. Resting CD4$^+$ T cells were purified by negative selection and magnetic separation using the human CD4$^+$ T Cell enrichment kit supplemented with anti-HLA-DR, anti-CD25, and anti-CD69 (Stem Cell Technologies #19052/#17962) and resuspended in R-10 (0.45 μm filtered). Cell samples were either left untreated (control cells) or were activated by T cell receptor

(TCR)-mediated stimulation for 48 hr in a 48-well plate and used for single-cell RNA-seq as described above.

Computational Analysis of Bulk and Single-Cell RNA-Seq Data
A standard pipeline for preprocessing RNA-seq data was used to filter and align the sequencing reads from both single-cell and bulk RNA-seq experiments (Supplemental Experimental Procedures; Zenodo_Data S1 to S3 available in Zenodo repository https://doi.org/105281/zenodo.1204334).

Based on the Software tools *scran* (Lun et al., 2016a) and *scater* (McCarthy et al., 2017) applied to the gene expression read counts for each single-cell dataset, we designed a customized filtering approach to discard low-quality cells in the single-cell RNA sequencing datasets. The filtering criteria were the total number of reads associated to genes, the proportion of mitochondrial reads, and the proportion of ERCC spike-ins reads (Figure S1C).

For the data produced with the experimental latency model, single-cell read counts were normalized using linear size factors calculated from the ERCC spike-ins with the package *scran* (Lun et al., 2016a). In absence of ERCC spike-ins, single-cell read counts for the single-cell RNA sequencing data from HIV$^+$ individuals were normalized using a normalization procedure specifically designed for single-cell data without spike-ins measurements, based on computing cell-specific size factors by considering random pools of cells (Lun et al., 2016b). Normalized read counts were log2-transformed.

Principal component analysis (PCA) (Pearson, 1901) was performed on the normalized single-cell read counts from the latency model, followed by k-means clustering in order to identify subpopulations of cells. We validated the two identified cellular subpopulations through a comprehensive study employing other normalization and dimension reduction methods (Supplemental Experimental Procedures). For the cells isolated from HIV$^+$ donors, PCA was performed on the normalized read counts and was followed by model-based clustering of the cells from each donor and treatment condition by using the R package *mclust* (Scrucca et al., 2016). One outlier cell was removed from the analysis following outlier detection.

Differential expression analysis was performed with the software MAST (Finak et al., 2015). Only genes expressed in at least one cell (expression level >0) per cluster were considered. Following the statistical test implemented in MAST, genes with a Benjamini-Hochberg corrected p value <0.01 were considered differentially expressed.

Enrichment analysis was conducted on the identified DE genes using the pathways defined in the Reactome database (Fabregat et al., 2016). For this purpose, a hypergeometric test was performed using a background list consisting of genes expressed at least in one cell and condition and the resulting p values were corrected with the Benjamini-Hochberg method.

DATA AND SOFTWARE AVAILABILITY

The accession number for the RNA-seq data reported in this paper is GEO: GSE111727.

Filtered and normalized gene expression read counts for the single-cell RNA-seq data from the latency model and from the patient samples are compressed into rdata libraries that can be directly accessed using the statistical software R and available in Zenodo repository https://doi.org/105281/zenodo.1204334 (Zenodo_Data_S2 and S3).

Raw data on image analysis to assess GFP expression intensity are freely available in Zenodo repository https://doi.org/105281/zenodo.1204334.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and four tables and can be found with this article online at https://doi.org/10.1016/j.celrep.2018.03.102.

ACKNOWLEDGMENTS

REFERENCES

Abbas, W., and Herbein, G. (2012). Molecular understanding of HIV-1 latency. Adv. Virol. *2012*, 574967.

Bacher, R., and Kendziorski, C. (2016). Design and computational analysis of single-cell RNA-sequencing experiments. Genome Biol. *17*, 63.

Banga, R., Procopio, F.A., Noto, A., Pollakis, G., Cavassini, M., Ohmiti, K., Corpataux, J.M., de Leval, L., Pantaleo, G., and Perreau, M. (2016). PD-1(+) and follicular helper T cells are responsible for persistent HIV-1 transcription in treated aviremic individuals. Nat. Med. *22*, 754–761.

Baxter, A.E., Niessl, J., Fromentin, R., Richard, J., Porichis, F., Charlebois, R., Massanella, M., Brassard, N., Alsahafi, N., Delgado, G.-G., et al. (2016). Single-cell characterization of viral translation-competent reservoirs in HIV-infected individuals. Cell Host Microbe *20*, 368–380.

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

Besnard, E., Hakre, S., Kampmann, M., Lim, H.W., Hosmane, N.N., Martin, A., Bassik, M.C., Verschueren, E., Battivelli, E., Chan, J., et al. (2016). The mTOR complex controls HIV latency. Cell Host Microbe *20*, 785–797.

Blazkova, J., Chun, T.W., Belay, B.W., Murray, D., Justement, J.S., Funk, E.K., Nelson, A., Hallahan, C.W., Moir, S., Wender, P.A., and Fauci, A.S. (2012). Effect of histone deacetylase inhibitors on HIV production in latently infected, resting CD4(+) T cells from infected individuals receiving effective antiretroviral therapy. J. Infect. Dis. *206*, 765–769.

Boritz, E.A., Darko, S., Swaszek, L., Wolf, G., Wells, D., Wu, X., Henry, A.R., Laboune, F., Hu, J., Ambrozak, D., et al. (2016). Multiple origins of virus persistence during natural control of HIV infection. Cell *166*, 1004–1015.

Brennecke, P., Anders, S., Kim, J.K., Kołodziejczyk, A.A., Zhang, X., Proserpio, V., Baying, B., Benes, V., Teichmann, S.A., Marioni, J.C., and Heisler, M.G. (2013). Accounting for technical noise in single-cell RNA-seq experiments. Nat. Methods *10*, 1093–1095.

Buettner, F., Natarajan, K.N., Casale, F.P., Proserpio, V., Scialdone, A., Theis, F.J., Teichmann, S.A., Marioni, J.C., and Stegle, O. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. Nat. Biotechnol. *33*, 155–160.

Bullen, C.K., Laird, G.M., Durand, C.M., Siliciano, J.D., and Siliciano, R.F. (2014). New ex vivo approaches distinguish effective and ineffective single agents for reversing HIV-1 latency in vivo. Nat. Med. *20*, 425–429.

Cary, D.C., Fujinaga, K., and Peterlin, B.M. (2016). Molecular mechanisms of HIV latency. J. Clin. Invest. *126*, 448–454.

Chavez, L., Calvanese, V., and Verdin, E. (2015). HIV latency is established directly and early in both resting and activated primary CD4 T cells. PLoS Pathog. *11*, e1004955.

Chen, H.C., Martinez, J.P., Zorita, E., Meyerhans, A., and Filion, G.J. (2017). Position effects influence HIV latency reversal. Nat. Struct. Mol. Biol. *24*, 47–54.

Chun, T.-W., Moir, S., and Fauci, A.S. (2015). HIV reservoirs as obstacles and opportunities for an HIV cure. Nat. Immunol. *16*, 584–589.

Ciuffi, A., and Telenti, A. (2013). State of genomics and epigenomics research in the perspective of HIV cure. Curr. Opin. HIV AIDS 8, 176–181.

Ciuffi, A., Mohammadi, P., Golumbeanu, M., di Iulio, J., and Telenti, A. (2015). Bioinformatics and HIV latency. Curr. HIV/AIDS Rep. *12*, 97–106.

Ciuffi, A., Rato, S., and Telenti, A. (2016). Single-cell genomics for virology. Viruses *8*, 123.

Ciuffi, A., Cristinelli, S., and Rato, S. (2017). Single-virus tracking uncovers the missing link between HIV integration site location and viral gene expression. Nat. Struct. Mol. Biol. *24*, 8–11.

Darcis, G., Kula, A., Bouchat, S., Fujinaga, K., Corazza, F., Ait-Ammar, A., Delacourt, N., Melard, A., Kabeya, K., Vanhulle, C., et al. (2015). An in-depth comparison of latency-reversing agent combinations in various in vitro and ex vivo HIV-1 latency models identified bryostatin-1+JQ1 and ingenol-B+JQ1 to potently reactivate viral gene expression. PLoS Pathog. *11*, e1005063.

Darcis, G., Van Driessche, B., and Van Lint, C. (2017). HIV latency: should we shock or lock? Trends Immunol. *38*, 217–228.

Davey, R.T., Jr., Bhat, N., Yoder, C., Chun, T.-W., Metcalf, J.A., Dewar, R., Natarajan, V., Lempicki, R.A., Adelsberger, J.W., Miller, K.D., et al. (1999). HIV-1 and T cell dynamics after interruption of highly active antiretroviral therapy (HAART) in patients with a history of sustained viral suppression. Proc. Natl. Acad. Sci. USA *96*, 15109–15114.

Eisele, E., and Siliciano, R.F. (2012). Redefining the viral reservoirs that prevent HIV-1 eradication. Immunity *37*, 377–388.

Fabregat, A., Sidiropoulos, K., Garapati, P., Gillespie, M., Hausmann, K., Haw, R., Jassal, B., Jupe, S., Korninger, F., McKay, S., et al. (2016). The Reactome pathway Knowledgebase. Nucleic Acids Res. *44* (*D1*), D481–D487.

Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A.K., Slichter, C.K., Miller, H.W., McElrath, M.J., Prlic, M., et al. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol. *16*, 278.

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

Ho, Y.C., Laird, G.M., and Siliciano, R.F. (2014). Measuring reversal of HIV-1 latency ex vivo using cells from infected individuals. Proc. Natl. Acad. Sci. USA *111*, 6860–6861.

Holmes, M., Zhang, F., and Bieniasz, P.D. (2015). Single-cell and single-cycle analysis of HIV-1 replication. PLoS Pathog. *11*, e1004961.

Huang, J., Wang, F., Argyris, E., Chen, K., Liang, Z., Tian, H., Huang, W., Squires, K., Verlinghieri, G., and Zhang, H. (2007). Cellular microRNAs contribute to HIV-1 latency in resting primary CD4+ T lymphocytes. Nat. Med. *13*, 1241–1247.

Kulpa, D.A., and Chomont, N. (2015). HIV persistence in the setting of antiretroviral therapy: when, where and how does HIV hide? J. Virus Erad. *1*, 59–66.

Lassen, K.G., Ramyar, K.X., Bailey, J.R., Zhou, Y., and Siliciano, R.F. (2006). Nuclear retention of multiply spliced HIV-1 RNA in resting CD4+ T cells. PLoS Pathog. *2*, e68.

Li, M., Kao, E., Gao, X., Sandig, H., Limmer, K., Pavon-Eternod, M., Jones, T.E., Landry, S., Pan, T., Weitzman, M.D., and David, M. (2012). Codon-usage-based inhibition of HIV protein synthesis by human schlafen 11. Nature *491*, 125–128.

Lun, A.T., Bach, K., and Marioni, J.C. (2016a). Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. Genome Biol. *17*, 75.

Lun, A.T.L., McCarthy, D.J., and Marioni, J.C. (2016b). A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. F1000Res. *5*, 2122.

Mbonye, U., and Karn, J. (2017). The molecular basis for human immunodeficiency virus latency. Annu. Rev. Virol. *4*, 261–285.

McCarthy, D.J., Campbell, K.R., Lun, A.T.L., and Wills, Q.F. (2017). Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics *33*, 1179–1186.

McCloy, R.A., Rogers, S., Caldon, C.E., Lorca, T., Castro, A., and Burgess, A. (2014). Partial inhibition of Cdk1 in G 2 phase overrides the SAC and decouples mitotic events. Cell Cycle *13*, 1400–1412.

Mohammadi, P., di Iulio, J., Muñoz, M., Martinez, R., Bartha, I., Cavassini, M., Thorball, C., Fellay, J., Beerenwinkel, N., Ciuffi, A., and Telenti, A. (2014). Dynamics of HIV latency and reactivation in a primary CD4+ T cell model. PLoS Pathog. *10*, e1004156.

Mohammadi, P., Ciuffi, A., and Beerenwinkel, N. (2015). Dynamic models of viral replication and latency. Curr. Opin. HIV AIDS *10*, 90–95.

Passaes, C.P.B., Bruel, T., Decalf, J., David, A., Angin, M., Monceaux, V., Muller-Trutwin, M., Noel, N., Bourdic, K., Lambotte, O., et al.; ANRS RHIVIERA Consortium (2017). Ultrasensitive HIV-1 p24 assay detects single infected cells and differences in reservoir induction by latency reversal agents. J. Virol. *91*, e02296-16.

Pearson, K. (1901). {On lines and planes of closest fit to systems of points in space}. Philos. Mag. *2*, 559–572.

Rato, S., Golumbeanu, M., Telenti, A., and Ciuffi, A. (2017). Exploring viral infection using single-cell sequencing. Virus Res. *239*, 55–68.

Sahu, G.K., Lee, K., Ji, J., Braciale, V., Baron, S., and Cloyd, M.W. (2006). A novel in vitro system to generate and study latently HIV-infected long-lived normal CD4+ T-lymphocytes. Virology *355*, 127–137.

Scrucca, L., Fop, M., Murphy, T.B., and Raftery, A.E. (2016). mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. R J. *8*, 289–317.

Shan, L., Deng, K., Gao, H., Xing, S., Capoferri, A.A., Durand, C.M., Rabi, S.A., Laird, G.M., Kim, M., Hosmane, N.N., et al. (2017). Transcriptional reprogramming during effector-to-memory transition renders CD4(+) T cells permissive for latent HIV-1 infection. Immunity *47*, 766–775.

Spina, C.A., Anderson, J., Archin, N.M., Bosque, A., Chan, J., Famiglietti, M., Greene, W.C., Kashuba, A., Lewin, S.R., Margolis, D.M., et al. (2013). 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.

Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N.T., Roth, A., Bork, P., et al. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. *45* (*D1*), D362–D368.

Tyagi, M., Pearson, R.J., and Karn, J. (2010). Establishment of HIV latency in primary CD4+ cells is due to epigenetic transcriptional silencing and P-TEFb restriction. J. Virol. *84*, 6425–6437.

van der Sluis, R.M., Jeeninga, R.E., and Berkhout, B. (2013). Establishment and molecular mechanisms of HIV-1 latency in T cells. Curr. Opin. Virol. *3*, 700–706.

Whitney, J.B., Hill, A.L., Sanisetty, S., Penaloza-MacMaster, P., Liu, J., Shetty, M., Parenteau, L., Cabral, C., Shields, J., Blackmore, S., et al. (2014). Rapid seeding of the viral reservoir prior to SIV viraemia in rhesus monkeys. Nature *512*, 74–77.

Yucha, R.W., Hobbs, K.S., Hanhauser, E., Hogan, L.E., Nieves, W., Ozen, M.O., Inci, F., York, V., Gibson, E.A., Thanh, C., et al. (2017). High-throughput characterization of HIV-1 reservoir reactivation using a single-cell-in-droplet PCR assay. EBioMedicine *20*, 217–229.