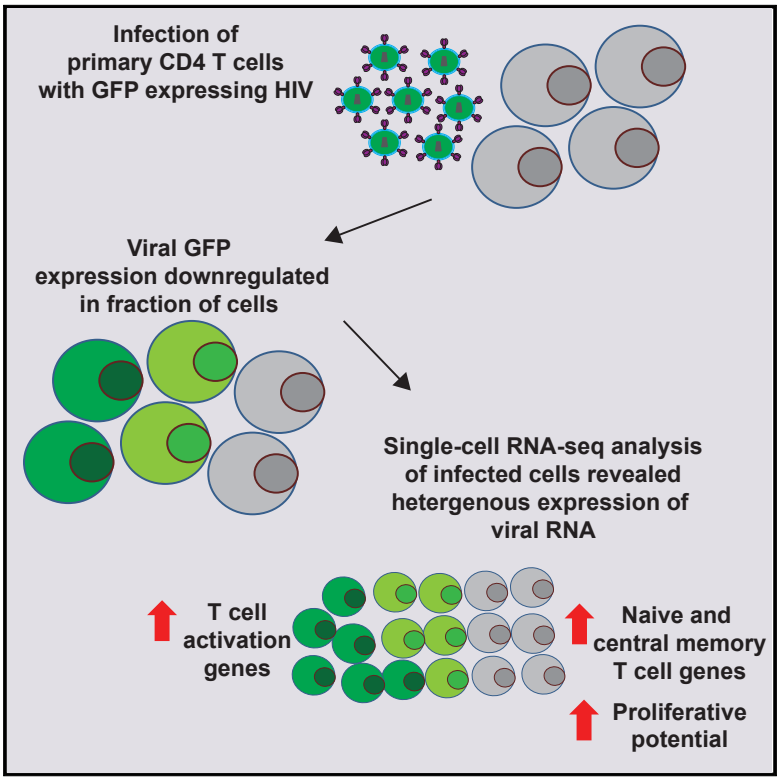
Article

Cell Reports

Single-Cell Analysis of Quiescent HIV Infection Reveals Host Transcriptional Profiles that Regulate Proviral Latency

Graphical Abstract



Authors

Todd Bradley, Guido Ferrari, Barton F. Haynes, David M. Margolis, Edward P. Browne

Correspondence

epbrowne@email.unc.edu

In Brief

Bradley et al. use single-cell RNA-seq to analyze cellular gene expression in a primary cell model of HIV latency. They identify cellular genes that are differentially expressed in cells in which HIV becomes transcriptionally downregulated. These results suggest a relationship between HIV latency and the transcriptional environment of the host cell.

Highlights

- Analyzed primary cell model of latency with scRNA-seq
- Latently infected cells express different genes than cells with viral transcription
- HIV is preferentially downregulated in cells with a naive or central memory phenotype
- HIV is preferentially downregulated in cells with higher proliferative potential

Bradley et al., 2018, Cell Reports *25*, 107–117
October 2, 2018
https://doi.org/10.1016/j.celrep.2018.09.020

CellPress

Cell Reports
Article

Single-Cell Analysis of Quiescent HIV Infection Reveals Host Transcriptional Profiles that Regulate Proviral Latency

Todd Bradley,[1] Guido Ferrari,[1] Barton F. Haynes,[1] David M. Margolis,[2] and Edward P. Browne[2,3,*]
[1]Duke University, Human Vaccine Institute, Duke University School of Medicine, Durham, NC 27710, USA
[2]UNC HIV Cure Center and Department of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[3]Lead Contact
*Correspondence: epbrowne@email.unc.edu
https://doi.org/10.1016/j.celrep.2018.09.020

SUMMARY

A detailed understanding of the mechanisms that establish or maintain the latent reservoir of HIV will guide approaches to eliminate persistent infection. We used a cell line and primary cell models of HIV latency to investigate viral RNA (vRNA) expression and the role of the host transcriptome using single-cell approaches. Single-cell vRNA quantitation identified distinct populations of cells expressing various levels of vRNA, including completely silent populations. Strikingly, single-cell RNA-seq of latently infected primary cells demonstrated that HIV downregulation occurred in diverse transcriptomic environments but was significantly associated with expression of a specific set of cellular genes. In particular, latency was more frequent in cells expressing a transcriptional signature that included markers of naive and central memory T cells. These data reveal that expression of HIV proviruses within the latent reservoir are influenced by the host cell transcriptional program. Therapeutic modulation of these programs may reverse or enforce HIV latency.

INTRODUCTION

Transcriptional downregulation of HIV in latently infected cells results in persistent HIV infection despite antiretroviral therapy. The mechanisms that enforce latency are complex and only partially understood but involve a repressive chromatin state that is regulated by diverse histone modifications (Turner and Margolis, 2017). Such latently infected cells appear to persist in infected patients for decades, their frequency little changed by years of antiretroviral therapy (ART) (Crooks et al., 2015; Siliciano et al., 2003). Furthermore, latently infected $CD4^+$ T cells can contribute to the rebound of infection upon cessation of antiviral therapy and thus are a principal barrier to curing HIV infection (Chun et al., 2010; Davey et al., 1999; Finzi et al., 1997). Intensive efforts currently focus on strategies to eliminate HIV-infected reservoir cells, and a major strategy to achieve this goal involves treatment with pharmacological latency-reversing agents (LRAs) to upregulate HIV expression in infected cells, so that these cells can be recognized and cleared by the host immune system (Archin et al., 2012; Søgaard et al., 2015; Zhu et al., 2012). However, LRAs thus far are ineffective at reactivating HIV in a majority of latently infected cells (Ho et al., 2013). Multiple mechanisms of regulation of proviral expression may limit the response to LRAs acting through a single mechanism, in addition to the heterogeneous phenotypic nature of latently infected cells themselves. Several host factors, such as CDK9 and CyclinT1, are known to regulate HIV transcription in various models of latency (Tyagi et al., 2010), and epigenetic features play a central role in antagonizing or augmenting the role of viral transactivation (He et al., 2002; Turner and Margolis, 2017; Van Lint et al., 2013, Friedman et al., 2011; Tripathy et al., 2015). It has also been demonstrated in model systems that establishment of latency can be driven by stochastic fluctuations in the viral transcription factor Tat (Razooky et al., 2015; Weinberger et al., 2005). Determination of how the host cell transcriptome affects viral expression and latency may identify targets to enhance or reverse latency.

One limitation of previous studies examining HIV latency was using cells considered as bulk populations and not as single cells, missing critical insights. The latently infected $CD4^+$ T cell reservoir is inherently diverse, with each provirus exhibiting a potentially unique combination of the effects of integration site, epigenetic modifications, and infected cell phenotype. For example, $CD4^+$ T cells, the major host cell for HIV infection, can exist as several different developmental stages, categorized as naive (Tn), central (Tcm) and effector memory (Tem), and effector cells. Each of these subtypes have distinct transcriptional and epigenetic programs that could affect the activity of the integrated HIV promoter (Durek et al., 2016). Additionally, biological noise and stochastic fluctuations in transcriptional activity could play an important role in either establishment or reversal of latency (Dar et al., 2014).

Given the multitude of factors that can affect the establishment and maintenance of latent infection, the application of methods that permit the analysis of single cells will be required to fully characterize latency and the mechanisms that determine HIV latency. Recent technological breakthroughs in analysis of individual cells by single-cell RNA sequencing (scRNA-seq) now permit detailed characterization of heterogeneous behaviors of individual cells (Hashimshony et al., 2016; Macosko et al., 2015; Picelli et al., 2013; Villani and Shekhar, 2017). These methods have provided insights into biological systems and

revealed surprising diversity in cultures of cells previously assumed to be uniform (Buettner et al., 2015; Shalek et al., 2014; Villani and Shekhar, 2017). We hypothesized that within latently infected populations, there exist subpopulations with differing patterns of viral and host gene transcription at rest and after host cell activation. Moreover, we hypothesized that this transcriptional diversity within latently infected cell populations is influenced by an identifiable set of genes. To investigate these hypotheses, we applied two single-cell assays to models of latent HIV infection. Analysis of vRNA at the single-cell level revealed the existence of diverse levels of vRNA expression both at rest and after LRA stimulation, and scRNA-seq of latently infected primary cells indicated that HIV downregulation occurs in diverse environments but was significantly associated with expression of a specific set of host cell genes. This latency associated signature suggests that downregulation of HIV in primary cells is regulated by the underlying transcriptional program of the infected cells. These insights illustrate an important role for the host cell environment in HIV latency and will guide the development of therapies that can achieve optimal reactivation of the latent reservoir.

RESULTS

Heterogeneous Viral RNA Induction by LRAs and a Threshold for Viral Protein Expression

To measure viral RNA (vRNA) in individual latently infected $CD4^+$ T cells, we used an approach that combined flow sorting of single cells into 96-well PCR plates followed by real-time qPCR for unspliced HIV RNA. To determine the utility of this approach, we first analyzed vRNA levels in N6 cells, a Jurkat-derived $CD4^+$ T cell line that is latently infected with HIV. This cell line contains a full-length integrated copy of the NL4-3 strain of HIV, with the *nef* open reading frame replaced by coding sequence for the murine Heat-shock antigen (HSA) reporter. Flow cytometry for surface expression of the HSA reporter protein encoded by this viral clone allowed us to identify HIV reactivation in N6 cells. We stimulated N6 cells with three different LRAs—the histone deacetylase inhibitor vorinostat (3 μM), the protein kinase C (PKC) agonist prostratin (3 μM), and tumor necrosis factor-alpha (TNFα; 100ng/mL)—and assayed vRNA levels in 144 cells for each condition at 24 hr (Figure S1A). In unstimulated cells, the majority of cells had undetectable vRNA levels, but a subpopulation (19%) expressed low levels of vRNA, detectable but typically below the lower limit of quantification. Upon stimulation, cells treated with either of the three LRAs exhibited a strong upregulation of vRNA and the appearance of detectable HSA protein expression. A wide range of vRNA levels from 1 to 3,442 copies per cell was detected across the population, with coefficients of variation ranging from 93% for vorinostat to 164% for TNFα. vRNA levels did not exceed 3,442 copies per cell, suggesting a uniform restriction of vRNA across the clonal cell line. Notably, expression of virally encoded HSA became pronounced only when vRNA levels were above 500 copies per cell, suggesting that a threshold of vRNA is required for translation of a detectable quantity of HSA in these cells. For all LRAs, the percentage of $vRNA^+$ cells exceeded the percent of HSA^+ cells, indicating that protein-based viral reporters significantly underestimate the fraction of responding cells (Figure S1B). Thus, these data demonstrate that stimulation of latently infected cells with LRAs induces a broad spectrum of diverse vRNA/antigen responses and that a significant population of $vRNA^+$ cells can be detected that are not producing detectable viral antigen. We also examined the response of N6 cells to vorinostat at different times after stimulation (Figure S2A) and at different concentrations (Figure S2B). These data demonstrated significant variation in the kinetics and thresholds of reactivation for vRNA expression among individual cells.

Establishment of a Primary $CD4^+$ T Cell Model for HIV Latency

To further examine the behavior of latently infected cells, we established a primary $CD4^+$ T cell model of HIV latency. This model is similar to models from other laboratories (Kim et al., 2014; Mohammadi et al., 2014; Sahu et al., 2006; Tyagi et al., 2010) and involves infecting activated $CD4^+$ T cells with a GFP-expressing HIV strain (Yang et al., 2009) and sorting to obtain a pure infected population, followed by long-term (8–12 weeks) co-culture with H80 cells (Figure 1A). During this period of culture, we observed highly variegated transcriptional downregulation of HIV gene expression within the infected population. Some cells exhibited downregulation of GFP expression to undetectable levels (GFP^-), while others maintained intermediate or high levels of expression (Figures 1B and 1C). To test the ability of the latently infected cells to reactivate viral gene expression in response to T cell receptor (TCR) engagement, we purified GFP^- cells from the infected cell culture and stimulated them with anti-CD3/CD28 beads. These cells re-expressed GFP to nearly 90% by 3 days, indicating that they were indeed latently infected and that initial loss of GFP expression was not due to deletion of the provirus or outgrowth of a contaminating uninfected population (Figure 1D).

Single-Cell Analysis of vRNA in Latently Infected Primary $CD4^+$ T Cells

To investigate the diversity of vRNA expression in latently infected primary cells, we analyzed sorted single cells from the infected population at 12 weeks post-infection (wpi) without stimulation (Figure 2). For comparison, we also examined productively infected cells at 2 days post-infection (dpi). Using linear regression, we found that vRNA levels in primary $CD4^+$ T cells were linearly correlated with GFP protein expression for both time points ($R^2 = 0.179$, $p = 1.71 \times 10^{-7}$ for 12 wpi; $R^2 = 0.475$, $p = 1.35 \times 10^{-21}$ for 2 dpi). This contrasts with the apparent threshold of vRNA required for viral protein expression in N6 Jurkat cells (Figure S1). The 2 dpi population expressed significantly higher overall levels of vRNA than cells at 12 wpi ($p < 0.00001$, Mann-Whitney test), indicating that transcriptional downregulation accounted for part of the reduction in GFP protein expression over time, but interestingly, a subset of 12 wpi cells expressed vRNA to a similar level to that seen at 2 dpi. At 12 wpi, 12% of cells exhibited undetectable vRNA, indicating transcriptional latency. Overall, this analysis shows that after long periods in culture, most infected cells downregulate viral gene expression at both the RNA and protein level, but the extent of this downregulation varied greatly between individual cells

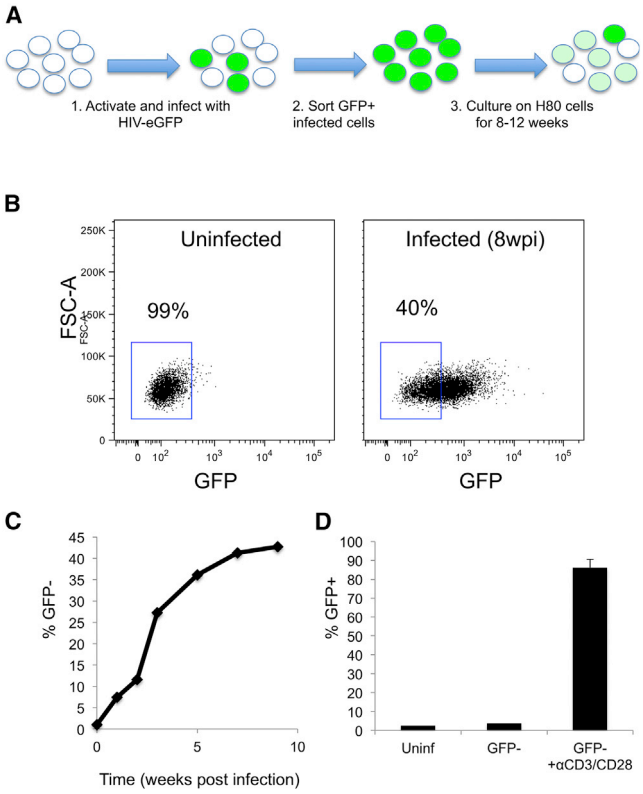


Figure 1. Primary CD4⁺ T Cell Model of HIV Latency

(A) Graphical depiction of primary cell model of latency. Activated cells were infected with pNL4-3-Δ6-dreGFP, a GFP-expressing HIV clone, and infected (GFP^+) cells were isolated by flow sorting.

(B) After 8 weeks of co-culturing the sorted GFP^+ population with H80 cells, cells displayed heterogeneous levels of virally encoded GFP expression, with 40% now being GFP^-.

(C) The percentage GFP^- cells over time was determined by flow cytometry. Data shown are a representative sample from one of three separate donors.

(D) Latently infected cells were isolated by flow sorting of the lowest 15% GFP-expressing cells from the infected cell culture. These cells were then stimulated with anti-CD3/CD28 beads for 3 days. GFP expression was measured by flow cytometry before (GFP^-) and after (GFP^- + αCD3/CD28) stimulation and compared with uninfected (uninf) cells. Error bars represent SDs of two technical replicates.

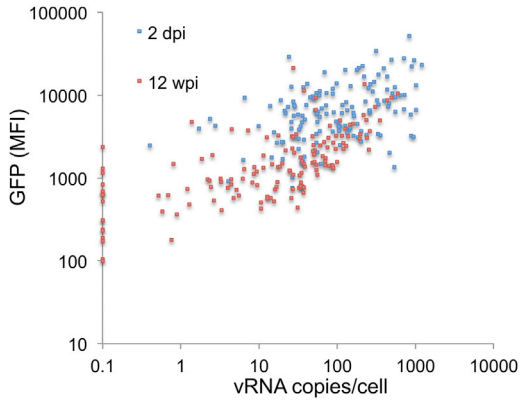


Figure 2. Single-Cell Viral RNA and Antigen Expression in the Primary Cell Latency Model Is Heterogeneous

Single infected primary $CD4^+$ T cells at 2 dpi (blue) and 12 wpi (red) were flow sorted and analyzed for vRNA expression by single-cell qPCR (sc-qPCR). GFP mean fluorescence intensity (MFI) for each cell was plotted against vRNA copies per cell. Each dot represents data from an individual cell.

and that an HIV-infected culture contains cells with diverse transcription levels.

Differential Proviral Transcription in Cells with Different Host Cell Gene Expression Patterns

We next sought to determine the relationship between host cell gene expression patterns and expression of vRNA by performing single-cell RNA-seq (scRNA-seq). We first compared the single-cell transcriptomes of 5,666 uninfected and 3,565 infected $CD4^+$ T cell populations from two donors. Reduction of the dimensionality of the transcriptomes of single cells to two dimensions by t-distributed stochastic neighbor embedding (tSNE) created a single-cell map of the clusters of cells. Overall, the infected cells clustered near the uninfected cells, indicating overall similar transcriptome profiles (Figure 3A). However, the cells from the two donors clustered distinctly, indicating differences in transcriptomic profiles between donors (Figure 3A). The inter-donor differences could represent true biological differences but may also incorporate some batch specific phenomena, as the donors were cultured and processed for scRNA-seq at different times. As such, we focused our attention on analyzing differences between infected and uninfected cells and between infected cells within the donors. Unsupervised graph-based clustering identified seven clusters on the basis of gene expression (Figure 3B). For the most part, these clusters represent a continuous gradient of T cell gene expression rather than discrete clusters of cell subsets. This result is consistent with previous scRNA-seq analysis of human T cells (Villani and Shekhar, 2017). Nevertheless, the clusters exhibited significant differences in gene expression. For example, clusters 0, 5, and 3 had higher levels of expression for *TNFSF10* and *ID2*, while clusters 4, 1, and 2 had higher expression of *CCR7* and *SELL* (Figure 3C). Cluster 6 had upregulation of genes that indicated feeder cell contamination and were excluded from further analysis. Among 21,258 total genes detected in all samples, we identified 12 upregulated and 6 downregulated transcripts ($p \leq 0.001$, likelihood-ratio test) when comparing infected and uninfected cells (Figure 3D; Table S1). As a marker of viral expression, *GFP* was the most differentially expressed transcript, followed by *IFITM1* and *MIF* (Figure 3D; Table S1). *IFITM1* has been shown to be upregulated upon HIV infection and even identified as a candidate marker of latently infected cells (Raposo et al., 2017), and *MIF* has been shown to be elevated in plasma from HIV-infected individuals and may play a role in viral replication (Regis et al., 2010). These results suggested that after 12 weeks of culture, infected and uninfected $CD4^+$ T cells have overall similar transcriptomes, but there are a small number of transcripts that are differentially expressed as a result of infection.

Next, we focused our analysis on differences within the infected cell population. We included the scRNA-seq from an additional 641 cells from a third donor, interrogating in total the

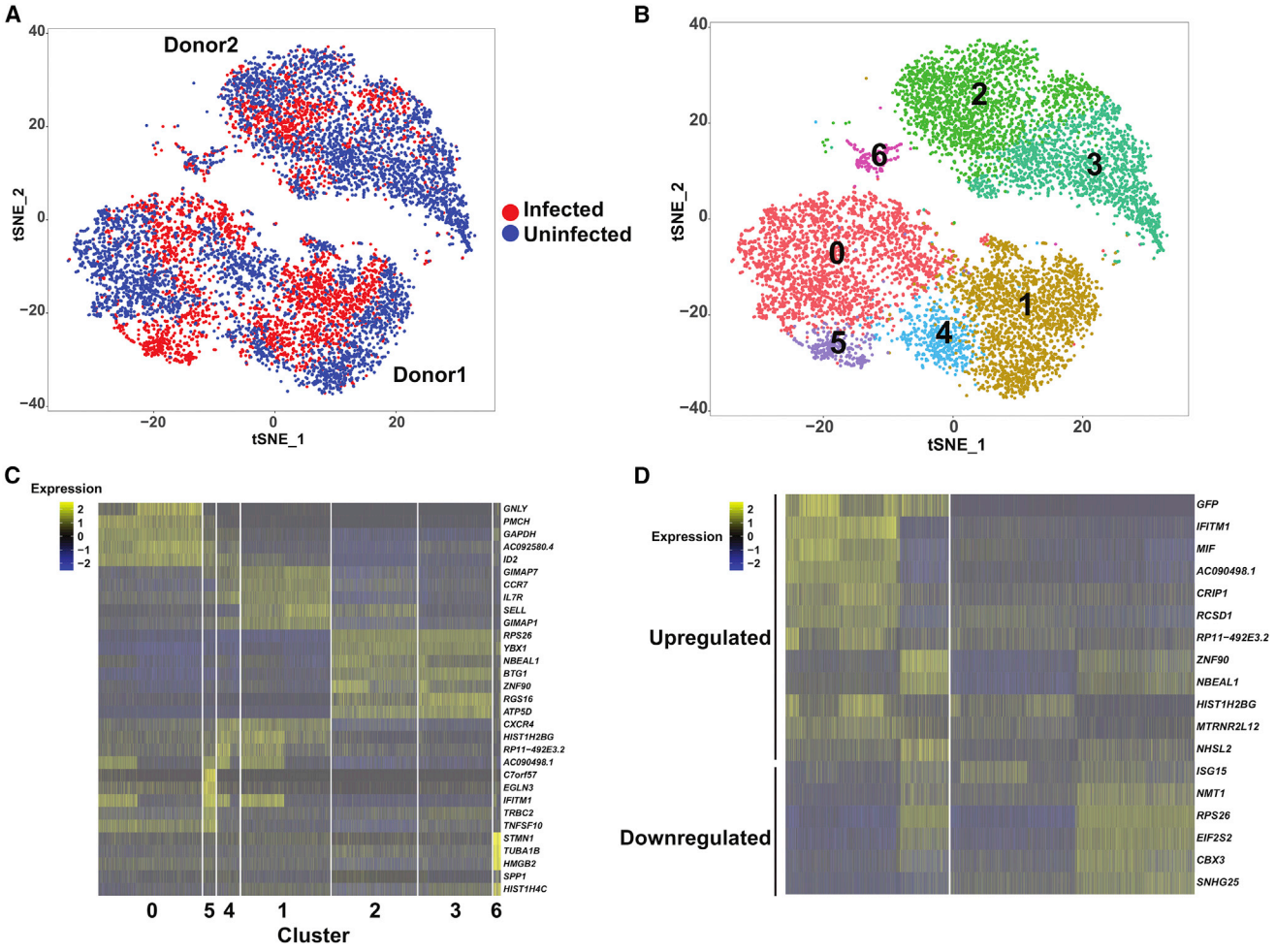


Figure 3. Single-Cell Transcriptome Analysis of Infected and Uninfected CD4⁺ T Cells

(A) Two-dimensional plot from unsupervised clustering by t-distributed stochastic neighbor embedding (tSNE) of the single-cell transcriptomes of 5,666 uninfected (blue) and 3,565 infected (red) $CD4^+$ T cell populations from two donors. Individual dots represent single cells. Uninfected cells were cultured in parallel with the infected cells under identical conditions.

(B) Two-dimensional tSNE plot with identified distinct clusters determined by graph-based clustering.

(C) Heatmap of the top five transcripts differentially expressed between each cluster (likelihood ratio test; $p \leq 0.05$). Each pixel column is the expression of an individual cell. Transcripts are on the rows with the normalized expression (*Z* scores) colored by the legend (yellow, more upregulated; blue, more downregulated).

(D) Heatmap of transcripts differentially expressed between infected and uninfected cells (likelihood ratio test; $p \leq 0.05$). Each pixel column is the expression of an individual cell. Transcripts are on the rows with the normalized expression (*Z* scores) colored by the legend (yellow, more upregulated; blue, more downregulated).

transcriptomes of 4,206 infected cells from three donors. Again, the three donors clustered distinctly, indicating differences in overall transcriptomes among the donors (Figure 4A). We identified 2,435 cells with detectable *GFP* transcript expression representing 68.5%, 28.9%, and 48.4% of the infected cells in the three donors, respectively (Figure 4C). The cells with no detectable *GFP* expression are candidate latently infected cells. Consistent with our previous observation, viral gene expression, as measured by *GFP* RNA levels, was heterogeneous within the infected cell population. Importantly, cells with undetectable viral transcription were observable in all the clusters, indicating that latency can be established in diverse cellular environments. However, significant differences in viral gene expression were found between some clusters (Figure S3). Specifically, cells from two of the three donors (donors 1 and 3) contained clusters that were significantly enriched for high levels of *GFP* expression, as well as clusters that were enriched for GFP^- cells (Figures 4B and S3). The other donor (donor 2) also exhibited apparent differences in viral gene expression between clusters, but these differences were not statistically significant, possibly because of the smaller number of cells recovered for this donor. To investigate this observation further, we then determined the differences in gene expression between cells actively transcribing vRNA (GFP^+) and cells with no vRNA expression (GFP^-). We found 33 upregulated and 13 downregulated transcripts in GFP^+ cells compared with GFP^- cells ($p \leq 0.001$,

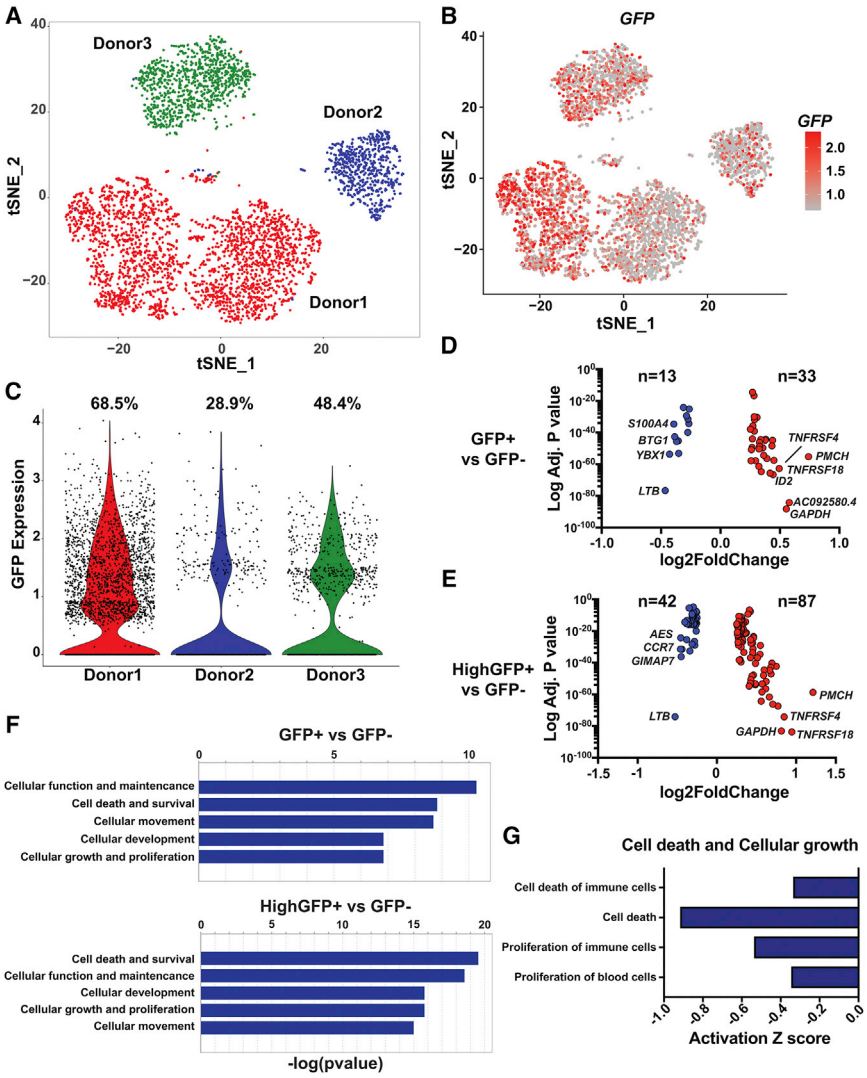


Figure 4. Differential Gene Expression in Infected Cells with Different Levels of Viral Transcription

(A) Two-dimensional plot from unsupervised clustering by t-distributed stochastic neighbor embedding (tSNE) of the single-cell transcriptomes of 4,206 infected CD4$^+$ T cell populations from three donors. Individual dots represent single cells. Cells were taken from infected cultures at 12 wpi and consist of a mixture of GFP$^+$ and GFP$^-$ cells.

(B) Normalized transcript expression level of GFP on the tSNE plot representation. Legend indicates normalized transcript values.

(C) Violin plots detailing the expression levels of GFP in each of the three donors infected cells.

(D and E) Volcano plots of genes significantly differentially expressed genes in GFP$^+$ cells (D) or high-GFP$^+$ cells ($\geq$2 GFP expression) (E) with adjusted log p value and $\log_2$ fold change graphed. Only significant genes are shown. Upregulated transcripts in red and downregulated in blue.

(F) Ingenuity Pathway Analysis (IPA; QIAGEN) of predicted functional annotations (diseases and functions) for gene lists comparing all GFP$^+$ or the high-GFP$^+$ and GFP$^-$ cells.

(G) Predicted activation or inhibition of canonical pathways. *Z* score indicates activation versus inhibition of indicated functions.

likelihood-ratio test; Figure 4D; Table S2). We also determined the differences in expression in cells with high vRNA expression (GFP^{hi}; top 10 percentile *GFP* expression; normalized expression $\geq$ 2) compared with GFP$^-$ cells and found 87 upregulated and 42 downregulated transcripts ($p \leq 0.001$, likelihood-ratio test; Figure 4E; Table S3). Genes with higher expression in the GFP^{hi} subset included *IL2RA* (CD25), *TNFRSF4* (OX40L), and *TNFRSF18* (GITR), consistent with previous observations that activated T cells permit higher levels of HIV transcription (Williams and Greene, 2007). In contrast, cells with undetectable HIV gene expression exhibited higher expression of *CCR7*, *CXCR4*, *SELL* (CD62L), and the cytokine receptor CD127. Visualization of the expression patterns for a number of these genes in individual donors also suggested an enrichment of latently infected cells in clusters with higher *CCR7*, *CD27*, and *SELL* expression and a greater frequency of cells with higher viral transcription in clusters with expression of *IL2RA*, *HLA-DRA*, and *CD38* (Figure 5). Thus, although HIV transcriptional downregulation can occur in cells with diverse phenotypes, it occurs with a higher frequency in cells that express a specific set of cellular genes.

To investigate the temporal dynamics of the observed gene changes, we performed scRNA-seq on 7,300 infected cells from an additional donor after 6 weeks of culture and identified 2,599 GFP$^+$-expressing cells (Figure S4A). As in our analysis of cells at 12 wpi, we observed HIV downregulation in all transcriptomic clusters but also found that some clusters were significantly enriched for high HIV gene expression or for downregulation (Figure S4B). We found seven or nine upregulated and three or two downregulated genes when comparing the GFP$^+$ ($p \leq 0.001$, likelihood-ratio test; Table S4) or high GFP$^+$ (top 10 percentile of *GFP* expression; Table S5) with GFP$^-$ cells, respectively (Figure S4C). Interestingly, even after only 6 weeks of culture, we found that cells with undetectable vRNA expression as measured by *GFP* expression had higher expression of several of the same genes we observed at 12 wpi, including TCF7 and CD27 (Figures S4C and S4D).

HIV Transcription Is Preferentially Downregulated in Cells with Greater Proliferative Potential

To investigate whether the set of genes that differentiate transcriptionally silent HIV infection (GFP$^-$) from transcriptionally active infection (GFP$^+$) represented differential activity of specific biological pathways, we performed Ingenuity Pathway Analysis (Jiménez-Marín et al., 2009) and identified significant enrichment

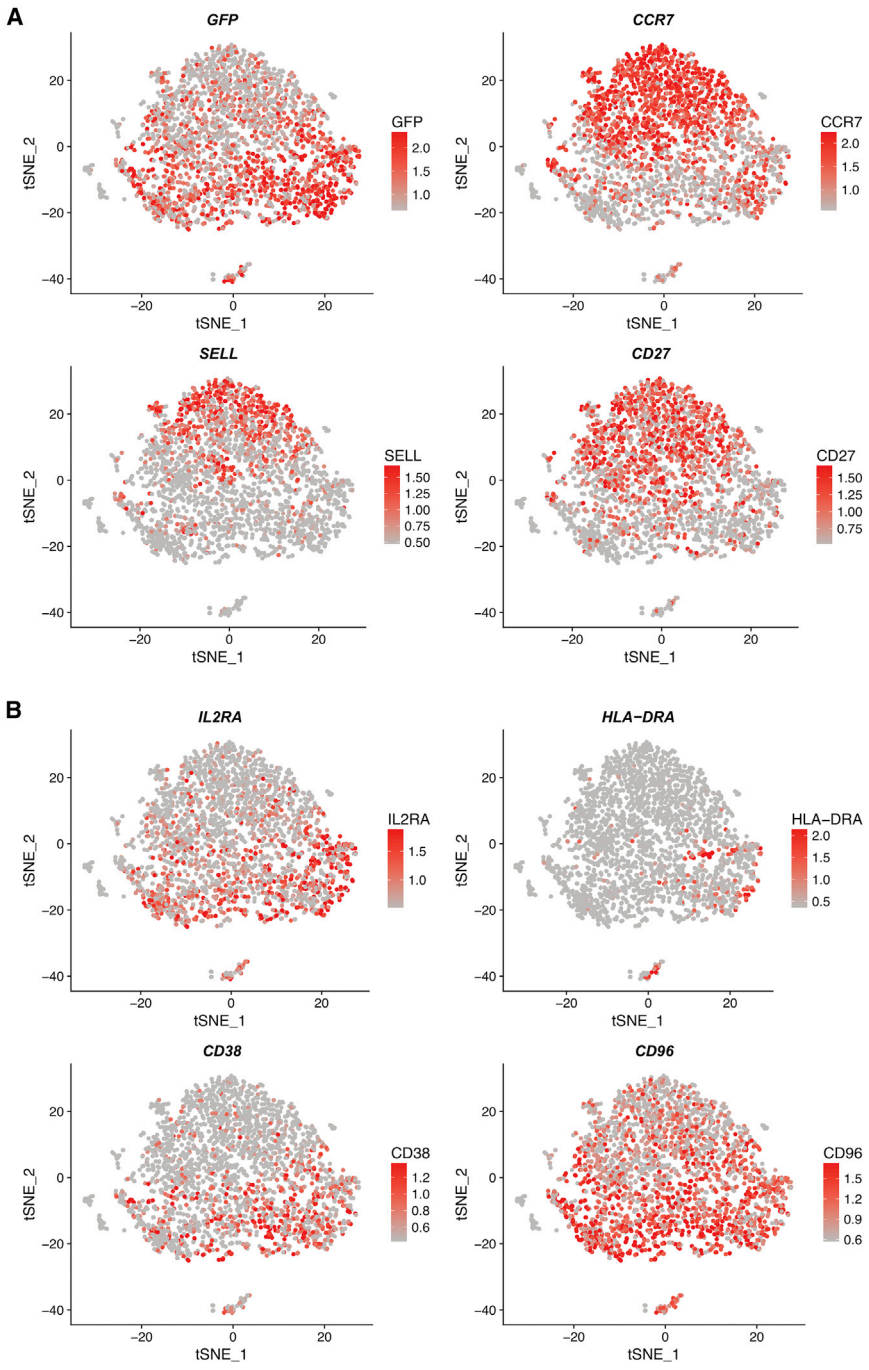


Figure 5. Expression Patterns of Host Cell Genes that Associate with HIV Transcription Level

(A and B) Normalized expression pattern for selected genes whose expression negatively (A) or positively (B) associated with viral gene expression (GFP) plotted on the tSNE plot representation. Each dot represents and individual cell. Legends indicate normalized transcript values. Data are derived from one of three donors.

of genes involved in cell death and survival and cellular proliferation when comparing all GFP$^+$ or GFP^{hi} cells to GFP$^-$ cells (Figure 4F). We then determined the predicted activation status of the genes in the cell death and survival and cellular growth and proliferation categories in the high-GFP$^+$ cells and found that transcripts were changed in a way that predicted inhibition of cell death and proliferation compared with GFP$^-$ cells (Figure 4G). A recent study also found activation of cellular survival and OX40 pathways during HIV infection (Kuo et al., 2018). On the basis of this analysis, we hypothesized that HIV transcriptional downregulation might be correlated with differential proliferative potential of the infected cells. To test this hypothesis, we sorted infected cells at 8 wpi into two populations on the basis of GFP expression level (GFP$^-$, GFP+) (Figure 6, top). We then stimulated each population with anti-CD3/CD28 beads in the presence of IL-2 100 U/mL for 3 days, then allowed the culture to expand further in the absence of stimulation. After 14 days of expansion, we measured expansion by counting viable cells. Notably, GFP$^-$ cells exhibited the greater fold expansion of the two cultures (Figure 6, bottom). These data are consistent with a model in which HIV transcriptional downregulation occurs preferentially in a specific subset of cells with greater proliferative potential and suggest that HIV transcriptional downregulation in primary CD4$^+$ T cells is associated with intrinsic biological properties of host cells.

Viral Downregulation Is Associated with T Cell Subset Identity

The observation that several transcripts that define specific T cell subsets (CCR7, SELL, CD27) are differentially expressed within infected cells with the lowest level of HIV gene expression suggested a relationship between specific CD4$^+$ T cell subsets and transcriptional downregulation of the HIV promoter. CD4$^+$ T cells emerge from the thymus as naive T cells (Tn), then, through a combination of antigenic simulation and cytokine cues, develop linearly into central memory T cells (Tcm), transitional memory T cells (Ttm), and finally effector memory T cells (Tem). These subsets are defined by differential expression of a set of surface markers including CD45RO and CCR7. We examined expression of CCR7 and CD45RO on the surface of infected cells by flow cytometry after 8 weeks of co-culture with H80 cells. This staining confirmed that the infected cell population consisted of a mixture of cells with

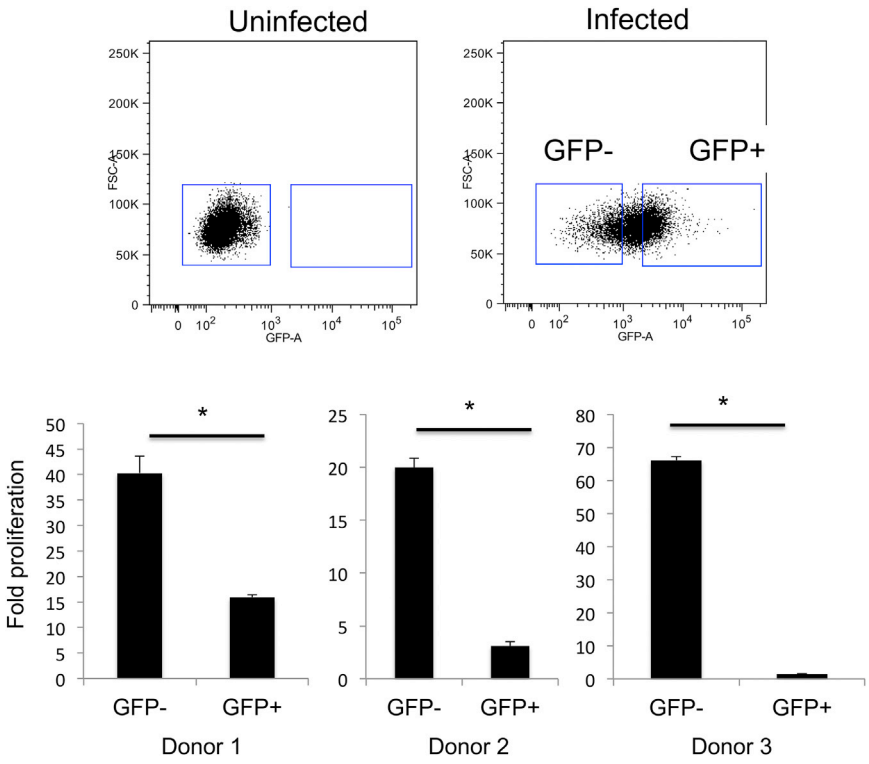


Figure 6. HIV Is Preferentially Downregulated in Cells with Higher Proliferative Capacity

An infected culture of primary $CD4^+$ T cells at 8 wpi was sorted into two populations on the basis of the level of GFP expression: GFP^- and GFP^+ (top right). Equivalent numbers of sorted cells were then stimulated with αCD3/αCD28 beads for 3 days and then expanded in 100 U/mL IL-2 for 2 weeks. At the end of this period, the fold expansion for each culture was calculated by counting live cells using a hemocytometer and trypan blue exclusion (bottom). Data shown are from three independent donors. Error bars indicate SDs of two technical replicates. *$p < 0.05$ (Student's t test).

"naive" ($CCR7^+$ $CD45RO^-$) phenotype, cells with a Tcm phenotype ($CCR7^+$ $CD45RO^+$), and cells with a Tem/Ttm ($CCR7^-$ $CD45RO^+$) phenotype (Figure 7, left). Notably, the distribution of cells within these populations differs from other reports using a similar model (Tyagi et al., 2010; Yang et al., 2009). This difference is likely explained by methodological differences in the preparation of the cells: in one of the previous studies (Tyagi et al., 2010), the infected cells were reactivated and expanded post-sorting, while in our experiments, the infected cells were not re-stimulated after the initial activation. To determine the level of viral gene expression in each subset, we examined the percentage of GFP^- (latent) cells in each subset (Figure 7, right). All three populations exhibited a mixture of GFP^- and GFP^+ cells, but, consistent with the scRNA-seq, we observed that there was a gradient of HIV downregulation across the populations, with $CD45RO^-$ $CCR7^+$ cells having the highest percentage of GFP^- cells and $CD45RO^+$ $CCR7^-$ cells exhibiting the lowest. These results demonstrated that in this latency model, HIV transcription was preferentially downregulated in cells at the Tn/Tcm end of the developmental spectrum. Nevertheless, we also observe heterogeneous viral expression levels within each subset, consistent with the hypothesis that latency is the output of a convergence of factors at the level of each individual cell. In this model, we would expect a diverse population of latently infected cells, the most frequent phenotype of which would reflect the most common convergence of these influences: (1) a cell that had recently returned to the central memory pool from a highly activated effector population prone to the initial stages of HIV infection and (2) a provirus whose expression had become restricted by epigenetic marks.

DISCUSSION

In this study we used single-cell approaches to determine the expression of viral and host cell genes in individual latently infected cells. Our approach provides a view of the latently infected cellular subsets, revealing that (1) latency occurs in diverse cellular environments, (2) transcriptional downregulation is significantly associated with a distinct host cell transcriptional signature, and (3) transcriptional downregulation occurs preferentially in cells that possess greater proliferation potential than cells in which viral transcription remains high.

Over the past few years, single-cell methods have become widely used and have provided powerful new insights into heterogeneity in biological systems that were previously assumed to be homogeneous (Linnarsson and Teichmann, 2016; Shalek et al., 2014; Villani and Shekhar, 2017). These methods are particularly useful to the study of virus-host interactions, in which multiple layers of complexity arise from variation in both host cells and infecting viruses (Ciuffi et al., 2016). The HIV latent reservoir, for example, consists of thousands of genetically distinct viral genomes, including those encoding immune escape variants, that persist as transcriptionally silent proviruses integrated into a range of genomic locations. Furthermore, expression of these proviruses is regulated by a wide range of dynamic chromatin modifications and host cell transcription factors that can be divergently expressed and differentially active in different $CD4^+$ T cell subpopulations. The latent HIV reservoir thus represents not a single, uniform target but a diverse mixture of several subpopulations of infected cells, potentially with fundamentally different characteristics. The development of strategies to reactivate or eliminate these cells may need to account for this diversity within the latent reservoir. For example, targeting this reservoir with LRAs may require separate strategies for different subtypes of cells rather than a "one size fits all" approach. Consistent with this notion, upon a single round of induction, individual LRAs typically disrupt latency in only a fraction of replication competent proviruses (Ho et al., 2013). As such there is an

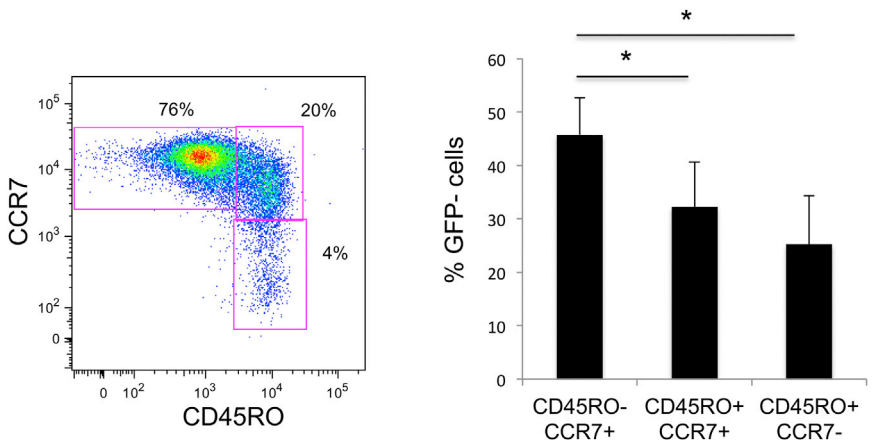


Figure 7. HIV Is Preferentially Downregulated in Cells Expressing Tn and Tcm Markers

Infected cells at 8 wpi were stained for surface markers CD4, CD45RO, and CCR7 to identify different $CD4^+$ T cell subsets. $CD45RO^-$ $CCR7^+$ cells represent naive T cells (Tn), $CD45RO^+$ $CCR7^+$ cells represent central memory cells (Tcm), and $CD45RO^+$ $CCR7^-$ cells represent a mix of transitional memory T cells (Ttm) and effector memory T cells (Tem). The fraction of GFP^- cells in each gate was then calculated and plotted (right). Data shown on the right represent the average of four independent donors. Error bars indicate SDs of biological replicates. *$p < 0.05$ (Student's t test).

urgent need to study latency in the context of approaches that can observe and characterize individual latently infected cells.

Prior studies have also applied single-cell approaches to the study of HIV latency, although with significantly different methodologies. Wiegand et al. (2017) used cell-associated HIV RNA/DNA quantitation combined with single HIV genome sequencing, both at limiting dilution, to detect diversity of viral transcription from individual proviruses within samples from patients on antiretroviral therapy. Similar to our results, they found that a minor subset of latently infected cells express detectable RNA in the absence of stimulation and that vRNA levels differ greatly from infected cell to infected cell. Likewise, Bui et al. (2015) found that the burst size of virus release from single infected cells varies over a wide range. Flow cytometry using probes for viral transcripts has also yielded insights into diversity of virus and host gene expression in individual cells during infection (Baxter et al., 2016.; Bolton et al., 2017; Martrus et al., 2016). Two recent publications have also examined HIV latency using scRNA-seq of a primary cell model (Golumbeanu et al., 2018) and from sorted patient samples (Cohn et al., 2018), yielding important insights. In particular, Golumbeanu et al. (2018) observed two clusters of cells in a primary cell model system that exhibited distinct transcriptional phenotypes, associated with different levels of viral gene expression. Furthermore, these investigators identified a transcriptional signature that distinguished these clusters. This signature contains some of the same genes we observe as being associated with HIV gene expressing in our study, such as IL-32, GAPDH, HLA-E, and CD96, but also many different genes (Golumbeanu et al., 2018).

Using a single-cell vRNA assay, we found that viral transcription levels are variable among individual latently infected cells. This observation was true for both a Jurkat-derived cell line (N6) and latently infected primary cells. In both model systems, a sizable subpopulation of latently infected cells transcribed low levels of vRNA in the absence of stimulation. Furthermore, the data demonstrate that antigen-based assays for latency reversal significantly underestimate the fraction of responding cells after stimulation with LRAs. N6 cells exhibited an apparent threshold of vRNA expression before virally encoded antigen became detectable, while for primary cells, this relationship was more complex. The reason for this difference is unclear but could be related to the greater levels of underlying heterogeneity in primary cells.

In our primary cell latency model, we observed only minor differences between infected and uninfected cells using scRNA-seq, suggesting that viral reprogramming of infected cells during latency is limited. This finding is consistent with a previous population analysis of the transcriptome of infected cells using a similar model (Mohammadi et al., 2014). However, the virus used in these studies lacks expression of most viral proteins in order to limit cytopathic effect and thus may not reflect the full impact of an intact replication competent virus on the host cell. In contrast, we observed a significant association between activity of the HIV promoter and the host cell transcriptome within the infected cell population. Thus, our data argue that viral transcription during latency is influenced by the underlying environment of the infected host cells. In particular, we observed a clear preference for HIV downregulation in cells expressing markers of Tn and Tcm subsets, while effector and activated cells were associated with higher levels of viral gene expression. These results thus suggest a role for T cell subset identity and intracellular environment in regulating the outcome of infection. Consistent with this notion, it has recently been demonstrated that HIV preferentially enters latency if infection occurs during a period of global cellular transcriptional downregulation as cells return to rest from activation (Shan et al., 2017). It is unlikely that the preferential downregulation we observe in Tn and Tcm in our primary cell model is related to differential mutagenesis of the provirus, because almost 90% of sorted latent (GFP^-) primary cells in our model re-expressed GFP upon TCR stimulation. Our finding of preferential HIV downregulation in Tn and Tcm cells is also consistent with outgrowth studies from suppressed patents (Soriano-Sarabia et al., 2014).

Several other studies have also noted significant differences between T cells subsets with respect to the establishment, maintenance, or reversal of latency, confirming that this is an area that will require further investigation. Tsunetsugu-Yokota et al. (2016) reported using an *ex vivo* infection model that latent proviruses in cells with a naive phenotype are difficult to reactivate compared with proviruses within the memory cell compartment, suggesting a distinct mechanism of latency in these cells. Consistent with our findings, Grau-Expósito et al. (2017) used a single-cell

flow-fluorescence *in situ* hybridization (FISH) assay of patient derived samples to identify effector memory T cells as the subset with the highest level of proviral transcription. Another recent study used flow cytometry to characterize cells expressing vRNA and Gag protein in both viremic and antiretroviral therapy-treated cells and found that proviruses in Tcm and Tem cells exhibited distinct patterns of reactivation with LRAs (Baxter et al., 2016).

Our finding that HIV is preferentially downregulated in cells with greater proliferation potential is also consistent with this hypothesis, and this phenomenon may contribute to persistence of the reservoir. It was also recently reported that HIV-infected cells exhibit long-term cell survival through expression of BIRC5 (Kuo et al., 2018). We did not observe upregulation of *BIRC5* in our model system, although we did observe a related protein, *BIRC3*, being upregulated in cells with active HIV transcription. It is not yet clear whether viral gene expression contributes to increased infected cell survival in our system or whether, conversely, the viral infection trajectory is guided by pre-existing cellular phenotypic diversity.

The molecular basis for differential viral downregulation across $CD4^+$ T cell subtypes is unclear. As T cells develop along a linear developmental trajectory from Tn to Tcm, and then to Ttm and Tem cells, they undergo progressive epigenetic remodeling, characterized by de-repression of cellular genes and loss of histone methylation islands that regulate expression of these genes (Durek et al., 2016). Thus, the pattern of HIV downregulation we observe mirrors the overall epigenetic program of these cells, suggesting a link between the two. In addition to epigenetic differences, the activity of other important mechanisms of HIV transcriptional regulation, such as Tat or transcriptional elongation factors (Kim et al., 2011; Razooky et al., 2015; Weinberger et al., 2005; Yukl et al., 2018), may differ in distinct T cell subsets. Nevertheless, with each subset, we see a mixture of GFP^+ and GFP^- cells, indicating that subset identity may result in a given provirus being vulnerable to viral downregulation, but other factors, such as integration site or stochasticity, must also contribute to the process of entry into latency (Chen et al., 2017; Weinberger et al., 2005).

Overall these findings indicate that the latent reservoir is a complex and diverse population of cells but that expression of specific host cell genes likely plays a role in the preferential downregulation of the HIV provirus in some T cell subtypes. The functional roles of the individual genes whose expression is enriched in the different transcriptional clusters in HIV gene expression should be investigated. Further studies to understand the mechanisms of latency may lead to more effective therapeutic approaches to clear persistent proviral infection.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - N6 Jurkat cell model
  - Primary cell model of HIV latency
- METHOD DETAILS
  - Generation of HIV stocks
  - Single cell vRNA (sc-vRNA) assay
  - scRNA-seq
  - Antibodies/flow cytometry
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and five tables and can be found with this article online at https://doi.org/10.1016/j.celrep.2018.09.020.

ACKNOWLEDGMENTS

Research reported in this publication was supported, in part, by the Collaboratory of AIDS Researchers for Eradication (CARE), a Martin Delaney Collaboratory of the National Institute of Allergy and Infectious Diseases (NIAID), National Institute of Neurological Disorders and Stroke (NINDS), National Institute on Drug Abuse (NIDA), and National Institute of Mental Health (NIMH) of the NIH (grant number 1UM1AI126619-01). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. We thank Nancy Fisher and the University of North Carolina (UNC) Flow Cytometry Core Facility, supported in part by Cancer Center Core Support Grant P30 CA016086 to the UNC Lineberger Comprehensive Cancer Center and in part by North Carolina Biotech Center Institutional Support Grant 2012-IDG-1006. This work was also supported by the UNC Center for AIDS Research (P30 AI50410). We acknowledge Connor Hart and Bhavna Hora for technical assistance and the Duke Human Vaccine Institute Viral Genetic Analysis Core.

AUTHOR CONTRIBUTIONS

E.P.B., G.F., and B.F.H. conceived the study. T.B. and E.P.B. performed the experiments. T.B., E.P.B., and D.M.M. analyzed the data. T.B., E.P.B., and D.M.M. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 11, 2018
Revised: June 26, 2018
Accepted: September 7, 2018
Published: October 2, 2018

REFERENCES

Archin, N.M., Liberty, A.L., Kashuba, A.D., Choudhary, S.K., Kuruc, J.D., Crooks, A.M., Parker, D.C., Anderson, E.M., Kearney, M.F., Strain, M.C., et al. (2012). Administration of vorinostat disrupts HIV-1 latency in patients on antiretroviral therapy. Nature *487*, 482–485.

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.

Bolton, D.L., McGinnis, K., Finak, G., Chattopadhyay, P., Gottardo, R., and Roederer, M. (2017). Combined single-cell quantitation of host and SIV genes and proteins ex vivo reveals host-pathogen interactions in individual cells. PLoS Pathog. *13*, e1006445.

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.

Bui, J.K., Mellors, J.W., and Cillo, A.R. (2015). HIV-1 virion production from single inducible proviruses following T-cell activation ex vivo. J. Virol. *90*, 1673–1676.

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., Justement, J.S., Murray, D., Hallahan, C.W., Maenza, J., Collier, A.C., Sheth, P.M., Kaul, R., Ostrowski, M., Moir, S., et al. (2010). Rebound of plasma viremia following cessation of antiretroviral therapy despite profoundly low levels of HIV reservoir: implications for eradication. AIDS *24*, 2803–2808.

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

Cohn, L.B., da Silva, I.T., Valieris, R., Huang, A.S., Lorenzi, J.C.C., Cohen, Y.Z., Pai, J.A., Butler, A.L., Caskey, M., Jankovic, M., and Nussenzweig, M.C. (2018). Clonal CD4$^+$ T cells in the HIV-1 latent reservoir display a distinct gene profile upon reactivation. Nat. Med. *24*, 604–609.

Crooks, A.M., Bateson, R., Cope, A.B., Dahl, N.P., Griggs, M.K., Kuruc, J.D., Gay, C.L., Eron, J.J., Margolis, D.M., Bosch, R.J., and Archin, N.M. (2015). Precise quantitation of the latent HIV-1 reservoir: implications for eradication strategies. J. Infect. Dis. *212*, 1361–1365.

Dar, R.D., Hosmane, N.N., Arkin, M.R., Siliciano, R.F., and Weinberger, L.S. (2014). Screening for noise in gene expression identifies drug synergies. Science *344*, 1392–1396.

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. U S A *96*, 15109–15114.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Durek, P., Nordström, K., Gasparoni, G., Salhab, A., Kressler, C., de Almeida, M., Bassler, K., Ulas, T., Schmidt, F., Xiong, J., et al.; DEEP Consortium (2016). Epigenomic profiling of human CD4$^+$ T cells supports a linear differentiation model and highlights molecular regulators of memory development. Immunity *45*, 1148–1161.

Finzi, D., Hermankova, M., Pierson, T., Carruth, L.M., Buck, C., Chaisson, R.E., Quinn, T.C., Chadwick, K., Margolick, J., Brookmeyer, R., et al. (1997). Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. Science *278*, 1295–1300.

Friedman, J., Cho, W.-K., Chu, C.K., Keedy, K.S., Archin, N.M., Margolis, D.M., and Karn, J. (2011). Epigenetic Silencing of HIV-1 by the Histone H3 Lysine 27 Methyltransferase Enhancer of Zeste 2 ▽. J. Virol *85*, 9078–9089.

Golumbeanu, M., Cristinelli, S., Rato, S., Munoz, M., Cavassini, M., Beerenwinkel, N., and Ciuffi, A. (2018). Single-cell RNA-seq reveals transcriptional heterogeneity in latent and reactivated HIV-infected cells. Cell Rep. *23*, 942–950.

Grau-Expósito, J., Serra-Peinado, C., Miguel, L., Navarro, J., Curran, A., Burgos, J., Ocaña, I., Ribera, E., Torrella, A., Planas, B., et al. (2017). A novel single-cell FISH-flow assay identifies effector memory CD4$^+$ T cells as a major niche for HIV-1 transcription in HIV-infected patients. MBio *8*, 8.

Hashimshony, T., Senderovich, N., Avital, G., Klochendler, A., de Leeuw, Y., Anavy, L., Gennert, D., Li, S., Livak, K.J., Rozenblatt-Rosen, O., et al. (2016). CEL-Seq2: sensitive highly-multiplexed single-cell RNA-seq. Genome Biol. *17*, 77.

He, G., Ylisastigui, L., and Margolis, D.M. (2002). The regulation of HIV-1 gene expression: the emerging role of chromatin. DNA Cell Biol. *21*, 697–705.

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

Jiménez-Marín, Á., Collado-Romero, M., Ramirez-Boo, M., Arce, C., and Garrido, J.J. (2009). Biological pathway analysis by ArrayUnlock and Ingenuity Pathway Analysis. BMC Proc. *3*, S6.

Kim, Y.K., Mbonye, U., Hokello, J., and Karn, J. (2011). T-cell receptor signaling enhances transcriptional elongation from latent HIV proviruses by activating P-TEFb through an ERK-dependent pathway. J. Mol. Biol. *410*, 896–916.

Kim, M., Hosmane, N.N., Bullen, C.K., Capoferri, A., Yang, H.-C., Siliciano, J.D., and Siliciano, R.F. (2014). A primary cell model of HIV-1 latency that uses activation through the T cell receptor and return to quiescence to establish latent infection. Nat. Protoc. *9*, 2755–2770.

Kuo, H.-H., Ahmad, R., Lee, G.Q., Gao, C., Chen, H.-R., Ouyang, Z., Szucs, M.J., Kim, D., Tsibris, A., Chun, T.-W., et al. (2018). Anti-apoptotic protein BIRC5 maintains survival of HIV-1-infected CD4+ T cells. Immunity *48*, 1183–1194.

Linnarsson, S., and Teichmann, S.A. (2016). Single-cell genomics: coming of age. Genome Biol. *17*, 97.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell *161*, 1202–1214.

Martrus, G., Niehrs, A., Cornelis, R., Rechtien, A., García-Beltran, W., Lütgehetmann, M., Hoffmann, C., and Altfeld, M. (2016). Kinetics of HIV-1 latency reversal quantified on the single cell level using a novel flow-based technique. J. Virol. *90*, 9018–9028.

McDavid, A., Finak, G., Chattopadyay, P.K., Dominguez, M., Lamoreaux, L., Ma, S.S., Roederer, M., and Gottardo, R. (2013). Data exploration, quality control and testing in single-cell qPCR-based gene expression experiments. Bioinformatics *29*, 461–467.

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.

Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat. Methods *10*, 1096–1098.

Raposo, R.A.S., de Mulder Rougvie, M., Paquin-Proulx, D., Brailey, P.M., Cabido, V.D., Zdinak, P.M., Thomas, A.S., Huang, S.H., Beckerle, G.A., Jones, R.B., and Nixon, D.F. (2017). IFITM1 targets HIV-1 latently infected cells for antibody-dependent cytolysis. JCI Insight *2*, e85811.

Razooky, B.S., Pai, A., Aull, K., Rouzine, I.M., and Weinberger, L.S. (2015). A hardwired HIV latency program. Cell *160*, 990–1001.

Regis, E.G., Barreto-de-Souza, V., Morgado, M.G., Bozza, M.T., Leng, L., Bucala, R., and Bou-Habib, D.C. (2010). Elevated levels of macrophage migration inhibitory factor (MIF) in the plasma of HIV-1-infected patients and in HIV-1-infected cell cultures: a relevant role on viral replication. Virology *399*, 31–38.

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.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. *33*, 495–502.

Shalek, A.K., Satija, R., Shuga, J., Trombetta, J.J., Gennert, D., Lu, D., Chen, P., Gertner, R.S., Gaublomme, J.T., Yosef, N., et al. (2014). Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature *510*, 363–369.

Shan, L., Deng, K., Gao, H., Xing, S., Capoferri, A.A., Durand, C.M., Rabi, S.A., Laird, G.M., Kim, M., Hosmane, N.N., et al. (2017). Transcriptional reprogramming during effector-to-memory transition renders CD4$^+$ T cells permissive for latent HIV-1 infection. Immunity *47*, 766–775.e3.

Siliciano, J.D., Kajdas, J., Finzi, D., Quinn, T.C., Chadwick, K., Margolick, J.B., Kovacs, C., Gange, S.J., and Siliciano, R.F. (2003). 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.

Søgaard, O.S., Graversen, M.E., Leth, S., Olesen, R., Brinkmann, C.R., Nissen, S.K., Kjaer, A.S., Schleimann, M.H., Denton, P.W., Hey-Cunningham, W.J., et al. (2015). The depsipeptide romidepsin reverses HIV-1 latency in vivo. PLoS Pathog. *11*, e1005142.

Soriano-Sarabia, N., Bateson, R.E., Dahl, N.P., Crooks, A.M., Kuruc, J.D., Margolis, D.M., and Archin, N.M. (2014). Quantitation of replication-competent HIV-1 in populations of resting CD4+ T cells. J. Virol. *88*, 14070–14077.

Tripathy, M.K., McManamy, M.E.M., Burch, B.D., Archin, N.M., and Margolis, D.M. (2015). H3K27 Demethylation at the Proviral Promoter Sensitizes Latent HIV to the Effects of Vorinostat in Ex Vivo Cultures of Resting CD4+ T Cells. J. Virol *89*, 8392–8405.

Tsunetsugu-Yokota, Y., Kobayahi-Ishihara, M., Wada, Y., Terahara, K., Takeyama, H., Kawana-Tachikawa, A., Tokunaga, K., Yamagishi, M., Martinez, J.P., and Meyerhans, A. (2016). Homeostatically maintained resting naive $CD4^+$ T cells resist latent HIV reactivation. Front. Microbiol. *7*, 1944.

Turner, A.W., and Margolis, D.M. (2017). Chromatin regulation and the histone code in HIV latency. Yale J. Biol. Med. *90*, 229–243.

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 Lint, C., Bouchat, S., and Marcello, A. (2013). HIV-1 transcription and latency: an update. Retrovirology *10*, 67.

Villani, A.-C., and Shekhar, K. (2017). Single-cell RNA sequencing of human T cells. Methods Mol. Biol. *1514*, 203–239.

Weinberger, L.S., Burnett, J.C., Toettcher, J.E., Arkin, A.P., and Schaffer, D.V. (2005). Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. Cell *122*, 169–182.

Wiegand, A., Spindler, J., Hong, F.F., Shao, W., Cyktor, J.C., Cillo, A.R., Halvas, E.K., Coffin, J.M., Mellors, J.W., and Kearney, M.F. (2017). Single-cell analysis of HIV-1 transcriptional activity reveals expression of proviruses in expanded clones during ART. Proc. Natl. Acad. Sci. U S A *114*, E3659–E3668.

Williams, S.A., and Greene, W.C. (2007). Regulation of HIV-1 latency by T-cell activation. Cytokine *39*, 63–74.

Yang, H.-C., Xing, S., Shan, L., O'Connell, K., Dinoso, J., Shen, A., Zhou, Y., Shrum, C.K., Han, Y., Liu, J.O., et al. (2009). Small-molecule screening using a human primary cell model of HIV latency identifies compounds that reverse latency without cellular activation. J. Clin. Invest. *119*, 3473–3486.

Yukl, S.A., Kaiser, P., Kim, P., Telwatte, S., Joshi, S.K., Vu, M., Lampiris, H., and Wong, J.K. (2018). HIV latency in isolated patient CD4+ T cells may be due to blocks in HIV transcriptional elongation, completion, and splicing. Sci. Transl. Med. *10*, eaap9927.

Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. *8*, 14049.

Zhu, J., Gaiha, G.D., John, S.P., Pertel, T., Chin, C.R., Gao, G., Qu, H., Walker, B.D., Elledge, S.J., and Brass, A.L. (2012). Reactivation of latent HIV-1 by inhibition of BRD4. Cell Rep. *2*, 807–816.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
CD24-PE	Biolegend	Cat# 138504; RRID: AB_10578416
CCR7-APC	Biolegend	Cat# 353213; RRID: AB_10915474
CD45RO-BV786	Biolegend	Cat# 304227; RRID: AB_10899580
CD4-BV605	Biolegend	Cat# 300555; RRID: AB_2564390
Bacterial and Virus Strains		
pNL43-Δ6-dreGFP	Dr Robert Siliciano	N/A
pcDNA3-gp160 (envelope CXCR4)	Dr Ron Swanstrom	N/A
psPAX2	Addgene	12260
Biological Samples		
Human Blood	Gulf Coast	N/A
Tissue Culture Media		
RPMI 1640	GIBCO	11875-093
DMEM	GIBCO	11995065
Chemicals, Peptides, and Recombinant Proteins		
Mirus LT1	Mirus Bio	Mir 2300
RNA SPRI beads	Agencourt	A63987
Zombie Violet live dead stain	Biolegend	423113
TCL buffer	QIAGEN	1031576
Polybrene	Sigma-Aldrich	107689-10G
Critical Commercial Assays		
ChromiumTM Single Cell 3′ Library and Gel Bead Kit v2, 4 rxns	10X Genomics	120267
NextSeq500 Hi-Output kit v2	NextSeq500 Hi-Output kit v2	NextSeq500 Hi-Output kit v2
Taqman fastvirus 1 step RT-PCR	ThermoFisher	4444432
CD4 Isolation Kit	Stem Cell	17952
T cell activation beads	ThermoFisher	1131D
Library quantification kit	KAPA	KK4824
Deposited Data		
scRNaseq raw data 1	Biosample	SAMN08685499
scRNaseq raw data 2	Biosample	SAMN08685500
scRNaseq raw data 3	Biosample	SAMN08685501
scRNaseq raw data 4	Biosample	SAMN08685502
Experimental Models: Cell Lines		
HEK293T	ATCC	CRL-3216
N6	David Irlbeck	N/A
H80	Darrell Bigner	N/A
Oligonucleotides		
HIV-GAG F ATCAAGCAGCCATGCAAATGTT	This study	N/A
HIV-GAG R CTGAAGGGTACTAGTAGTTCCTGCTATGTC	This study	N/A
HIV-GAG probe FAM/ZEN-ACCATCAATGAGGAAGCTGCAGAATGGGA-IBFQ	This study	N/A
BAC F TCACCCACACTGTGCCCATCTACGA	This study	N/A
BAC R CAGCGGAACCGCTCATTGCCAATGG	This study	N/A
BAC Probe HEX-ATGCCCTCCCCCATGCCATCCTGCGT-IBFQ	This study	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software		
STAR	Dobin et al., 2013	N/A
Ingenuity Pathway Analysis	QIAGEN	N/A
CellRanger	Zheng et al., 2017	N/A
Seurat	Satija et al., 2015	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Edward P. Browne (epbrowne@email.unc.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

N6 Jurkat cell model

N6 cells were a generous gift from David Irlbeck (Glaxo SmithKline, Chapel Hill, NC). This cell line was derived from the human Jurkat T cell line and was made using HIV-1 engineered to express a luciferase reporter in place of the HIV-1 nef gene with an additional mouse heat stable antigen CD24 (HSA) reporter located just downstream of the luciferase open reading frame, separated by a T2A element (NLCH-Luci-HSA). NLCH, the parent molecular infectious clone was kindly provided by the laboratory of Dr. Ron Swanstrom (UNC-Chapel Hill, Chapel Hill, NC) and is a modification of HIV-1 NL4-3 (GenBank U26942) where flanking sequences were removed. To generate latently infected clones, NLCH-Luci-HSA infected Jurkat cells expressing low levels of HSA were selected, limit diluted, and individual clones, including clone N6, were expanded in culture. This cell line was maintained in RPMI media with 10% Fetal calf serum (FCS) and penicillin/streptomycin. 500nM Efavirenz was added to the media to prevent spontaneous virus outgrowth.

Primary cell model of HIV latency

For generation of infected primary model cells, human blood was purchased from Gulf Coast Regional Blood Center (Houston, TX). Total CD4 T cells were then isolated from human peripheral blood mononuclear cells (PBMCs) by negative selection using Easysep total CD4 T cell isolation kit (Stem Cell, Vancouver, BC). Purity was determined by staining with anti-CD4-FITC and flow cytometry and was typically ~98%–99%. For infection, 20 million CD4 T cells were activated by mixing with anti-CD3/CD28 beads (Thermo Fisher) at one bead per cell for 2 days with 100U/mL IL-2. At 2d, the beads were magnetically removed, and the cells infected with pNL43-Δ6-dreGFP virus by centrifugation at 600 g for 2h at room temperature, in the presence of 4ug/mL polybrene. At 2dpi, cells were resuspended in staining buffer, and infected (GFP+) cells were isolated using a FACSAria flow sorter (Becton Dickson). The recovered GFP+ cells were then co-cultured with the human H80 cell line in RPMI media (provided by Darrell Bigner, Duke University) and 20 U/mL IL-2 for up to 12 weeks. Media for the co-culture was replaced every 2-3 days. Infected cells were moved to flasks with fresh H80 cells every two weeks. All donors were anonymous. This study was reviewed by the UNC-Chapel Hill institutional review board (IRB) and was deemed to not constitute human research.

METHOD DETAILS

Generation of HIV stocks

HIV stocks were produced by transient transfection of the human 293T cell line with pNL43-Δ6-dreGFP plasmid, as well as the packaging plasmids PAX2 and gp160 envelope, using Mirus LT1 tranfection reagent (Mirus Bio, Madison, WI). pNL43-Δ6-dreGFP was kindly provided by Robert Siliciano (Johns Hopkins). This virus contains premature stop codons in all viral genes except tat and rev, and contains a destabilized eGFP gene in the envelope open reading frame (Yang et al., 2009). The gp160 expression plasmid was derived from the CXCR4-tropic NL4-3 strain and was kindly provided by Ronald Swanstrom (UNC Chapel Hill). Tissue culture media was replaced at 24 h. At 48 h. supernatant was harvested and spun at 2000 rpm for 5 mins to remove cell debris before filtering through a 0.45μM filter (Millipore, Burlington MA). Aliquots of virus were frozen at −80°C.

Single cell vRNA (sc-vRNA) assay

Cells were stained at 1:1000 in phosphate buffered saline (PBS) with the live/dead dye Zombie Violet (ZV; Biolegend, San Diego CA) for 20 mins. Cells were then washed and resuspended in staining buffer (PBS with 2% fetal calf serum and 1mM EDTA). Next, single cells were sorted into 96-well PCR plates with 10 μl of TCL buffer (QIAGEN, Hilden, Germany) containing 1% Beta-mercaptoethanol using an Aria flow sorter (Becton Dickinson, Franklin Lakes, NJ). Doublets were identified and excluded by forward scatter (FSC)/side scatter

(SSC) gating, and dead cells were excluded by gating on ZV- cells. Index sorting was used to record fluorophore/GFP intensity for each cell. After sorting plates were briefly spun and frozen at −80°C. To extract RNA, plates were thawed and 22 μl of RNA-Clean XP beads (Beckman Coulter, Brea, CA) added and incubated for 10 mins. Beads were then isolated and washed in 80% ethanol, then eluted into 10 μL of 1x PCR master mix using Fastvirus (Thermo, Waltham, MA) and primer sets for HIV Gag (GAG-F: ATCAAGCAGC CATGCAAATGTT, GAG-R: CTGAAGGGTACTAGTAGTTCCTGCTATGTC, GAG-Probe: FAM/ZEN-ACCATCAATGAGGAAGCTGCA GAATGGGA-IBFQ) and Beta-actin (BAC-F: TCACCCACACTGTGCCCATCTACGA, BAC-R: CAGCGGAACCGCTCATTGCCAATGG, BAC-Probe: HEX-ATGCCCTCCCCCATGCCATCCTGCGT-IBFQ). This plate was then run on a QS3 (Applied Biosystems, Foster City, CA) real time thermocycler with a 5 minute reverse transcription step at 50°C, followed by 40 cycles of 94°C (3 s.), 60°C (30 s.). RNA copy number was determined by comparison to a standard curve of synthesized Gag gblock purchased from Integrated DNA Technologies (Coralville IA). Wells that failed to amplify Beta actin were excluded from analysis. The lower limit of quantification (LLOQ) for this assay was seven copies of Gag RNA. Coefficient of variation for technical replicates in the standard curve ranged from 10%–15% for 5000 copies to 40%–45% for 7 copies. Analysis of uninfected cells indicated a false positive amplification rate of approximately 1%, with the majority of false positive signals falling below the LLOQ. Sensitivity and precision of this assay was also benchmarked by analysis of vRNA standards (purchased from Bio-synthesis, Lewisville TX) that were spiked into assay wells with lysis buffer (Figure S5).

scRNA-seq

scRNA-seq was performed as described (Zheng et al., 2017). Briefly, cellular suspensions were loaded on a GemCode Single-Cell instrument (10X Genomics, Pleasanton, CA) to generate single-cell beads in emulsion. scRNA-seq libraries were then prepared using GemCode Single Cell 3′ Gel bead and library kit (10X Genomics). Single-cell barcoded cDNA libraries were quantified by quantitative PCR (Kappa Biosystems, Wilmington, MA) and sequenced on an Illumina NextSeq 500 (San Diego CA). Read lengths were 26 bp for read 1, 8 bp i7 index, and 98 bp read 2. Cells were sequenced to greater than 50,000 reads per cell. The Cell Ranger Single Cell Software Suite was used to perform sample de-multiplexing, barcode processing and single-cell 3′ gene counting (Zheng et al., 2017). Reads were aligned to human genome release Hg38 with GFP nucleotide sequence added as an additional gene. Graph based cell clustering, dimensionality reduction and data visualization were analyzed by the Seurat R package (Satija et al., 2015). Cells that exhibited high transcript counts indicative of cells with abnormal expression or multiple cells within single droplet (> 7500 total transcripts), > 0.1% mitochondrial transcripts (cellular stress), or transcripts characterized by the H80 feeder cells were excluded from analysis. scRNA-seq data shown in the main text are from three donors at 12 wpi. An additional donor was analyzed at 6 wpi and these data are shown in the supplemental information.

Antibodies/flow cytometry

For flow cytometry, cells were stained in 100 μl staining buffer (PBS with 2% FCS and 1 mM EDTA). Staining was carried out for 30 mins. on ice, before washing with PBS and resuspension in staining buffer. Flow cytometry was carried out on a BD Fortessa (Becton Dickson). Antibodies/fluorophores used for staining were anti-CD24-PE, anti-CD4-BV605, anti-CD45RO-BV785, anti-CCR7-APC. All antibodies were purchased from Biolegend.

QUANTIFICATION AND STATISTICAL ANALYSIS

Differentially expressed transcripts were determined in the Seurat R package utilizing the Likelihood-ratio test for single cell gene expression statistical test (McDavid et al., 2013) and resulting p values were adjusted for multiple comparisons using Bonferroni correction within Seurat. Significantly changed genes were identified as an adjusted p value ≤ 0.05. Enriched gene pathway analysis and predicted pathway activation were determined using Ingenuity Pathway Analysis (QIAGEN) software. For Figure 2, R2 values were calculated by linear regression, with p values determined for a null hypothesis of the slope being zero. For data in Figures 6 and 7, analysis was performed with two tailed Students t test, with significance indicated by $p > 0.05$.

DATA AND SOFTWARE AVAILABILITY

scRNaseq data has been deposited in the Biosample database. Accession numbers are: SAMN08685499, SAMN08685500, SAMN08685501, SAMN08685502.