

RESEARCH ARTICLE

Epigenomic characterization of latent HIV infection identifies latency regulating transcription factors

Stuart R. Jefferys¹, Samuel D. Burgos², Jackson J. Peterson², Sara R. Selitsky³, Anne-Marie W. Turner⁴, Lindsey I. James⁵, Yi-Hsuan Tsai³, Alisha R. Coffey³, David M. Margolis^{2,4}, Joel Parker¹, Edward P. Browne^{2,4*}

1 Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, **2** Department of Microbiology and Immunology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, **3** Lineberger Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, **4** Department of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, **5** Division of Chemical Biology and Medicinal Chemistry, Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America

* epbrowne@email.unc.edu



OPEN ACCESS

Citation: Jefferys SR, Burgos SD, Peterson JJ, Selitsky SR, Turner A-MW, James LI, et al. (2021) Epigenomic characterization of latent HIV infection identifies latency regulating transcription factors. *PLoS Pathog* 17(2): e1009346. <https://doi.org/10.1371/journal.ppat.1009346>

Editor: Guido Silvestri, Emory University, UNITED STATES

Received: August 7, 2020

Accepted: January 29, 2021

Published: February 26, 2021

Copyright: © 2021 Jefferys et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Underlying data files for ATACseq analysis available at https://webshare.bioinf.unc.edu/public/2020_HivLatency and at https://datadryad.org/stash/share/9dxKTqKu3X80uUg3Ms8H8DaiZmyb_ZTr60j5Gg2Za6c.

Funding: The authors gratefully acknowledge the following funding sources: NIH 1R01AI143381-01A1 (to EPB), 1R61DA047023-01 (to LIJ), and the UNC Center for AIDS research (CFAR) developmental award (to EPB). Research reported

Abstract

Transcriptional silencing of HIV in CD4 T cells generates a reservoir of latently infected cells that can reseed infection after interruption of therapy. As such, these cells represent the principal barrier to curing HIV infection, but little is known about their characteristics. To further our understanding of the molecular mechanisms of latency, we characterized a primary cell model of HIV latency in which infected cells adopt heterogeneous transcriptional fates. In this model, we observed that latency is a stable, heritable state that is transmitted through cell division. Using Assay of Transposon-Accessible Chromatin sequencing (ATACseq) we found that latently infected cells exhibit greatly reduced proviral accessibility, indicating the presence of chromatin-based structural barriers to viral gene expression. By quantifying the activity of host cell transcription factors, we observe elevated activity of Forkhead and Kruppel-like factor transcription factors (TFs), and reduced activity of AP-1, RUNX and GATA TFs in latently infected cells. Interestingly, latency reversing agents with different mechanisms of action caused distinct patterns of chromatin reopening across the provirus. We observe that binding sites for the chromatin insulator CTCF are highly enriched in the differentially open chromatin of infected CD4 T cells. Furthermore, depletion of CTCF inhibited HIV latency, identifying this factor as playing a key role in the initiation or enforcement of latency. These data indicate that HIV latency develops preferentially in cells with a distinct pattern of TF activity that promotes a closed proviral structure and inhibits viral gene expression. Furthermore, these findings identify CTCF as a novel regulator of HIV latency.

in this publication was supported, in part, by 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 the National Institute of Mental Health (NIMH) of the National Institutes of Health, grant number 1UM1AI126619 (to DMM). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

HIV is able to persist during antiviral therapy by entering a state of viral latency, in which viral gene expression is greatly reduced. These latently infected cells can re-seed infection if therapy is interrupted, and thus represent a major obstacle to an HIV cure. Identifying the mechanisms that lead to this state will help to identify strategies to block or eliminate HIV latency, leading to a cure for infection. By observing HIV gene expression in infected CD4 T cells, we isolated cells in which HIV has entered latency and identified characteristics that distinguish them from cells with active viral replication. We found that latently infected cells have elevated activity of specific transcription factors including Forkhead TFs and Kruppel-like factors. We also identify CTCF, a protein responsible for mediating insulation of genome domains from each other, as being required for the establishment of HIV latency. Developing agents to target these factors may lead to new strategies to eliminate the HIV reservoir.

Introduction

HIV infection continues to be a major global health problem, with 37 million infected individuals and approximately one million deaths per year [1]. HIV infection can be potently controlled by combination antiretroviral therapy (ART), with treatment suppressing viral loads to undetectable levels and allowing HIV-infected patients to live lives of roughly normal lifespan. Nevertheless, persistent health complications in treated patients, as well as side effects of ART, make the development of a cure for HIV a high priority [2–5]. Moreover, rapid rebound of viremia after treatment interruption demonstrates the persistence of HIV-infected cells, even after long-term ART. Indeed, longitudinal studies to estimate the half-life of this reservoir indicate that greater than 72 years of treatment would be necessary for elimination of all infected cells [6,7]. The mechanisms of HIV persistence are not completely understood, but likely involve several factors. First, HIV is able to establish a latent infection, characterized by the presence of a transcriptionally silent provirus, in a subset of host cells [8,9]. Sporadic reactivation of these cells may occur continuously, and leads to viral rebound if treatment is interrupted [10]. The precise location of this latently-infected reservoir is debated, but is likely distributed across a number of CD4 T cell subsets, including long-lived memory cells, thereby explaining the long persistence of infection [11–13]. Second, HIV-infected cells can undergo homeostatic expansion *in vivo* during ART, further replenishing the pool of infected cells [14–16]. Understanding the molecular mechanisms of how latency is established and maintained will be critically important to developing strategies to prevent or eliminate the latent reservoir. Certain cell types, such as resting memory CD4 T cells, are a suboptimal environment for HIV transcription, due to limiting availability of transcription factors required for HIV gene expression, including NF- κ B, AP-1 and P-TEFb [17–20]. Stochastic variation in the levels of the viral Tat protein during infection may also contribute to latency [21]. Furthermore, covalent modification of provirus-associated histones by histone-modifying enzymes such as histone deacetylases (HDACs) or histone methyl transferases (HMTs), such as the PRC2 or HUSH complexes, can have a repressive effect on viral transcription [22–26], and their role in the maintenance of latency is strongly supported by evidence of latency reversal *in vivo* [27].

We have previously established, using a primary CD4 T cell model of HIV latency, that latency can occur in diverse host cell environments, but occurs preferentially in cells that express markers of quiescent central memory T cells (T_{cm}), and that exhibit high proliferative potential [28]. These results indicate that the establishment of latency is influenced by the

intrinsic biological program of the host cell. However, the mechanistic details of how specific host cell environments or phenotypes impact the initiation or maintenance of latency are unknown. Gene activity can be regulated by changes in the structure and accessibility of chromatin. These changes can be mediated by chromatin remodeling complexes that are recruited to sites of active transcription by transcription factors (TFs) and result in the removal or repositioning of nucleosomes near transcription start sites (TSS) [29]. Furthermore, enzymes that add or remove covalent modifications to histones tails can create a histone “code” that affects the structure of the chromatin and creates docking sites for additional regulators [30]. Changes in chromatin structure and accessibility can then allow greater access to the promoter by core transcriptional regulators including RNA polymerase, thereby facilitating transcription. To further investigate our observations, we sought to characterize primary CD4 cells in which latency has become established by defining chromatin-based characteristics of these cells. The results revealed an association of the latency phenotype with a distinct pattern of chromatin accessibility and reduced accessibility of the HIV genome. Furthermore, we identify a set of cellular transcription factors with differentially accessible binding sites in latently infected cells, which may play a role in influencing the course of HIV transcriptional silencing and the entry into or maintenance of the latent state. In particular, we investigate and confirm the role of CTCF during HIV latency, implicating this protein as novel latency-regulating factor.

Results

HIV-infected cells enter a stable, heritable state of latency in a cell culture model

We have previously established a cell culture model of HIV latency (Fig 1A). In this model, primary CD4 T cells are activated through TCR stimulation, then infected with a reporter HIV strain that encodes a destabilized eGFP gene (herein referred to as HIV-GFP) [28,31]. Actively infected (GFP+) cells are purified by flow sorting at 2dpi, and then cultured for up to 12 weeks. During this time period, viral gene expression progressively diminishes, and a subset of the cells become GFP- (latently infected), while the remaining cells exhibit variegated levels of viral gene expression. To examine the stability of the viral gene expression phenotype, we resorted this mixed culture at 12wpi into GFP+ (actively infected) and GFP- (latently infected) (Fig 1B) and cultured these cells independently for six additional days. Notably, the infected cells retained their viral gene expression level, that is GFP+ cells remained GFP+, and GFP- cells remained GFP-, for this time period indicating that the level of viral gene expression is a relatively stable property of the infected cells (Fig 1C). We also examined the proviral copy number in the GFP- cells by quantitative PCR, and observed an average of almost exactly one copy per cell (1.05), indicating that these cells retained proviral DNA. To look at the effect of cellular stimulation on viral gene expression in latently infected cells, we isolated the latently infected (GFP-) cells by flow sorting, and then restimulated them through their TCR for 3 days, monitoring GFP expression over time (Fig 1D). At the time of stimulus removal, the latently infected cells had potentially upregulated GFP expression and a large fraction of the cells (90%) had become GFP+. Interestingly, after removal of the TCR stimulus, the majority of reactivated cells then rapidly lost viral gene expression and returned to latency (GFP-) within 14 days. A residual population of around 20% of the reactivated cells retained a low level of GFP expression after this single round of reactivation. To further investigate this finding, we stimulated these cells three additional times over a total period of 48 days (Fig 1E). Similar to the initial restimulation, after each subsequent stimulation the latently infected cells rapidly upregulated GFP expression, then returned to latency with highly consistent kinetics after stimulus removal. Specifically, the cells returned to a largely GFP- phenotype by 7d post

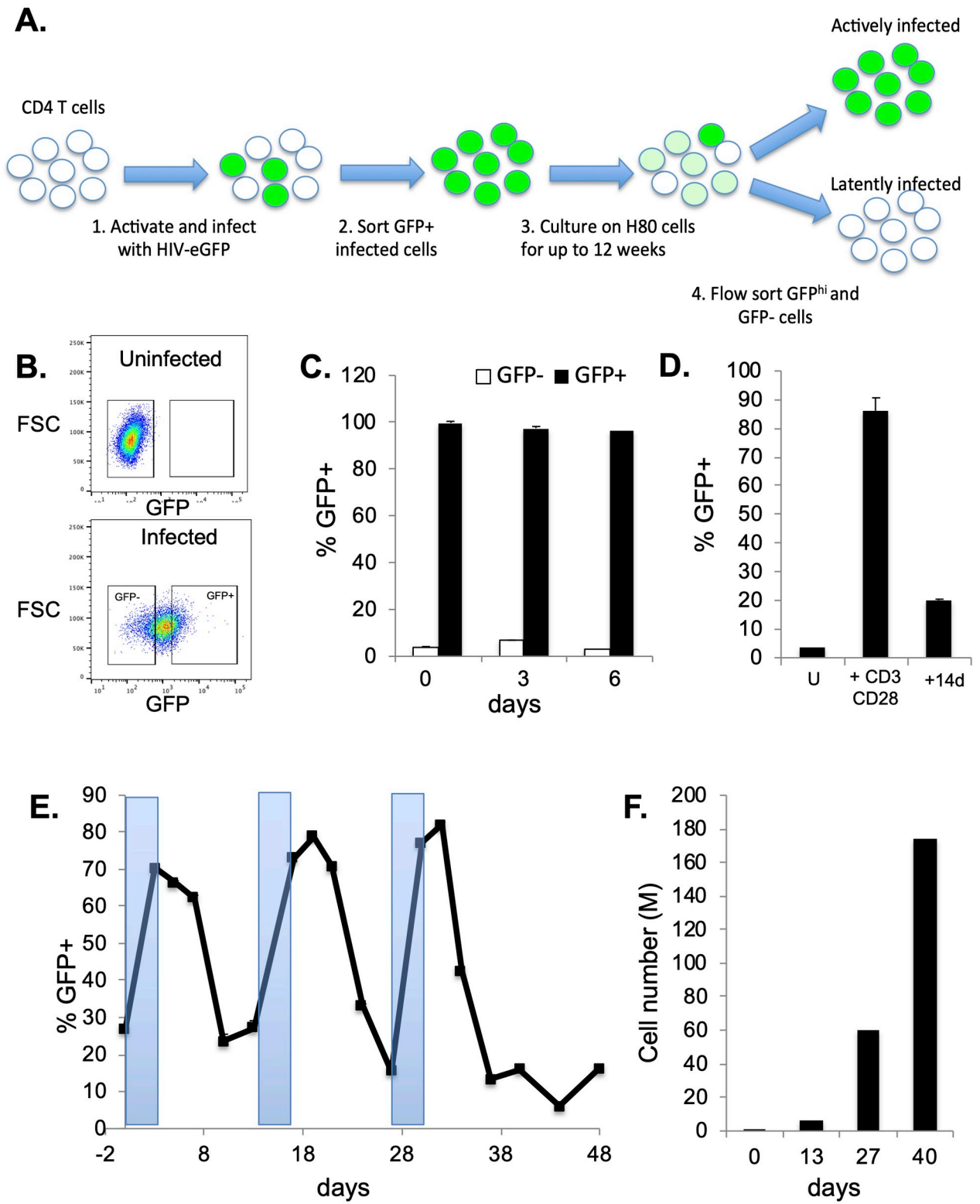


Fig 1. HIV infected CD4 T cells enter a stable, heritable state of viral latency in vitro. A. Schematic overview of primary CD4 T cell HIV latency model. B. Flow cytometry plot showing a representative gating of actively infected (GFP+) and latently infected (GFP-) cells. C. GFP+ and GFP- cells were flow sorted from the HIV-GFP infected population and cultured for 6 days post sorting. Viral gene expression was measured at days 0, 3, and 6 post sorting, and the percentage of cells in the GFP+ gate was measured. Each bar represents the average of three independent replicates. D. Sorted infected GFP- cells at 12wpi were reactivated using α CD3/CD28 beads. At 24h, unstimulated cells (U) and stimulated cells (+ α CD3/CD28) were analyzed by flow cytometry and the percentage of GFP+ cells analyzed. The reactivated cells were then removed from the activating beads and cultured for an additional 14 days (+14d), then analyzed again by flow cytometry. E. Sorted GFP- cells from an infected population were serially stimulated through their TCR (+ α CD3/CD28) and GFP expression was monitored over time by flow cytometry. Shaded areas indicate times during which α CD3/CD28 beads were added to the culture to stimulate the cells. Each datapoint represents the average of two independent replicates. F. Cell numbers were counted at selected timepoints during the serial stimulation shown in E and total cell numbers are shown in millions of cells (M).

<https://doi.org/10.1371/journal.ppat.1009346.g001>

stimulation. Notably, during this period of the time the cells also underwent substantial cell division and cell numbers increased by \sim 170 fold (Fig 1F). This finding suggests that latency is a stable, intrinsic property of infected cells that is rapidly re-established after viral reactivation, and can be transmitted to daughter cells during cell division. This conclusion is consistent with previous studies showing rapid reestablishment of latency after stimulation of infected cells with histone deacetylase inhibitors [32].

Latently infected cells exhibit differentially accessible chromatin

The stable, heritable character of HIV latency in this model system lead us to hypothesize that epigenomic or chromatin-based changes in infected cells were associated with reversible but recurring viral silencing. To examine whether the establishment of latency in our model system was associated with specific changes in chromatin accessibility, both within the HIV genome and within the host cell genome, we performed Assay of Transposon-Accessible Chromatin sequencing (ATACseq) on infected cells. ATACseq uses a hyperactive Tn5 transposon to probe and identify regions of open chromatin [33,34]. To compare cells with latent infection to those with active viral transcription, we sorted HIV-GFP infected cells at 12 weeks after infection based on viral gene expression into GFP+ or GFP- populations, representing cells from the upper and lower 25% of GFP expression intensity respectively, and generated ATACseq libraries from the sorted cells. To ensure biological reproducibility, we analyzed infected cells derived from two independent donors. These data were then aligned to a combined reference genome including the human genome and HIV. Quality control analyses showed nucleosomal laddering of fragment sizes and enrichment of reads at transcription start sites (TSS), indicating high quality data (Fig 2A and 2B). Replicate analysis of cells derived from the same donor clustered closely together by principal component analysis (PCA), indicating high reproducibility (S1 Fig).

Next, we compared accessibility of chromatin from the GFP+ and GFP- infected cells in greater detail by calculating differential accessibility for each of the accessibility peaks present in the samples. These data showed that GFP- and GFP+ cells have distinct chromatin accessibility patterns, with several thousand differentially open peaks between latently infected and actively infected cells (Fig 2C). Specifically, using a false discovery rate (FDR) cutoff value of 0.05, we observed 5021 peaks that were significantly more open in GFP+ cells, and 4068 peaks that were more open in GFP- cells, out of 277992 overall peaks detected across all samples (Fig 2C). These data thus indicate that viral latency in this system is correlated with a specific pattern of overall genomic accessibility.

Notably, a high proportion of differentially accessible peaks (72%) were associated with genes, indicating that the data likely reflect dynamic mechanisms regulating gene expression. To further validate the ATACseq data, we also analyzed matched RNAseq profiles from these infected cells. Importantly, we observed that ATACseq peaks were highly enriched in genes that displayed differential transcriptional levels between the samples ($P = 8.17 \times 10^{-10}$),

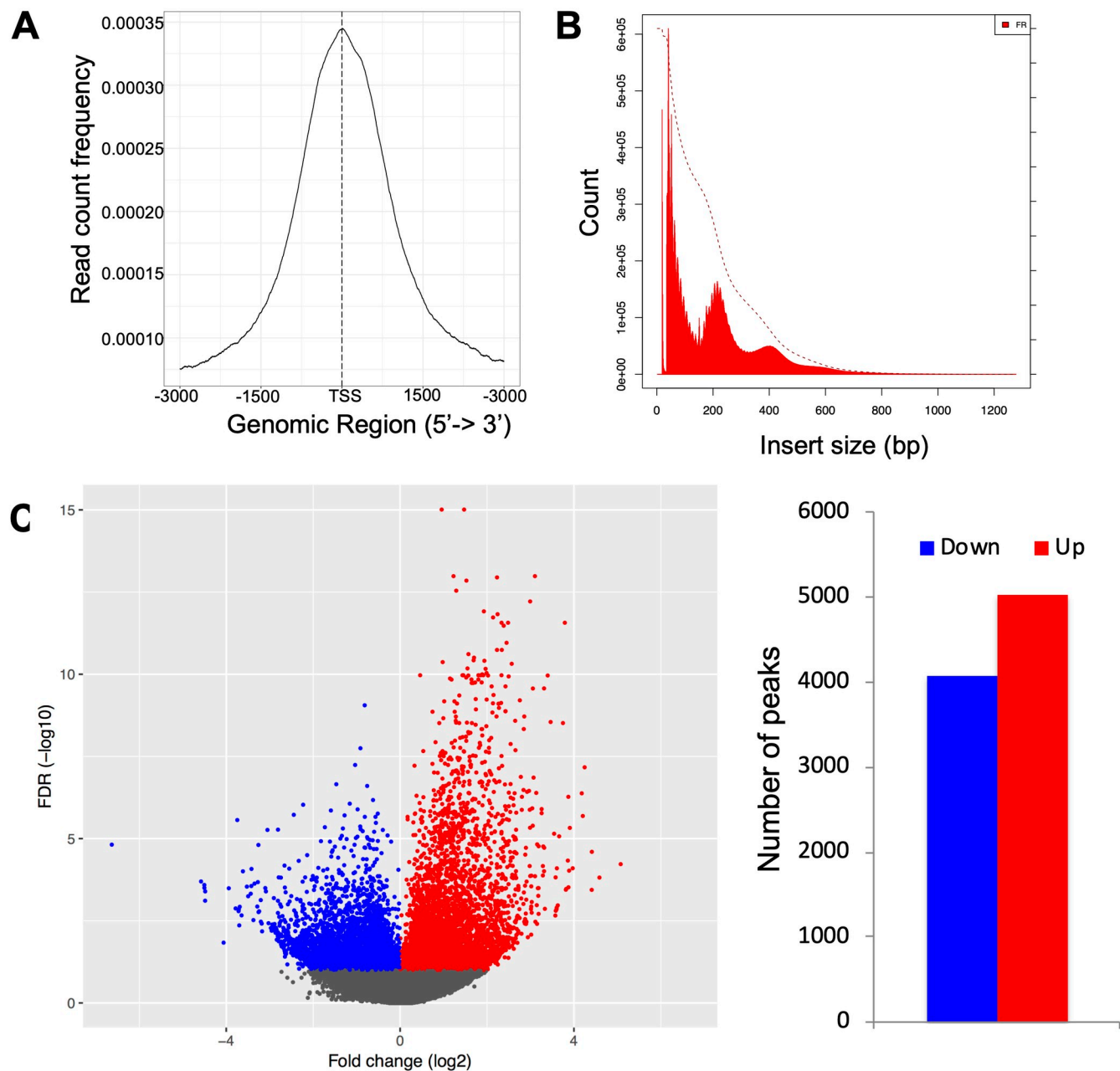


Fig 2. Latently infected cells are epigenomically distinct from actively infected cells. HIV-GFP infected cells at 12 wpi were flow sorted into GFP- and GFP+ populations and analyzed by ATACseq. **A.** Alignment of the reads to the human genome demonstrates enrichment of reads near transcriptional start sites (TSS). **B.** Size distribution of library fragment sizes showing nucleosomal laddering. **C.** Differentially accessible regions of the cellular genome were identified and fold change versus false discovery rate (FDR) displayed as a volcano plot. Peaks that are significantly (FDR < 0.1) more open in GFP+ cells (“up”) are indicated in red, while peaks that are significantly more open in latently infected cells (“down”) are indicated in blue. The data represents an aggregate of infected cells from two independent donors.

<https://doi.org/10.1371/journal.ppat.1009346.g002>

indicating that changes in the chromatin accessibility were associated with transcriptional regulation (S1 Table). Direct comparison of the magnitude of individual peak fold changes to fold changes in transcript levels show low correlation ($r = 0.123$) for genes with significant

changes to both transcript level and peak accessibility, likely reflecting the non-linear relationship between changes in chromatin accessibility and transcription (S2 Fig). 471 genes exhibited both a significant change to transcript level and a change in accessibility (S2 Table). This list included several genes with known roles in HIV latency and transcription (eg RUNX1 and RELB), but also many genes that have not previously been associated with HIV latency. We also examined the phenotypic identity of these cells by comparing the transcriptional profiles of the infected cells to reference datasets of primary immune cell types [35], and observed that these cells most closely resembled resting memory CD4 T cells, the natural host cell type for latent HIV (S3 Fig).

The HIV genome exhibits reduced accessibility in latently infected cells

Next we aligned the ATACseq reads to the HIV genome to determine whether HIV latency is associated with differential accessibility of the provirus. For both GFP+ and GFP- infected cells the HIV genome exhibits major accessibility peaks surrounding the 5' transcription start site (TSS) at nucleotide (nt) 455, and within the 3' LTR (Fig 3A). Downstream of the TSS peak, after nt 800, the virus genome becomes significantly less accessible, consistent with the hypothesis that chromatin structural barriers proximal to TSS impede processive viral transcription [36,37]. We also observed an overall increasing gradient of openness across the 3' half of virus in actively infected cells, leading to the large peak in the 3'LTR region. (Fig 3A). Overall we observed significantly reduced accessibility of the HIV provirus in the latently infected cells relative to the actively infected cells (2.45 fold, $FDR = 2.87 \times 10^{-13}$). To identify specific regions within the HIV genome with differential accessibility, we subdivided the HIV genome into 73bp bins and individually calculated the fold change and false discovery rate for each bin (Fig 3B). This analysis showed significantly reduced accessibility across the viral genome in latency, but a particularly large difference towards the 3' end of the provirus. This finding is consistent with the hypothesis that physical barriers limiting the accessibility of the provirus to factors that regulate transcription likely contribute to latency.

Latency reversal is associated with changes in accessibility in the host cell and viral genomes

We next examined the impact of latency reversing agents (LRAs) on latently infected cells. We stimulated HIV-GFP infected cells at 12wpi with three different LRAs, each with a different mechanism of action—vorinostat (histone deacetylase inhibitor), AZD5582 (SMAC mimetic), and prostratin (PKC agonist) for 24 hr. We have previously examined the transcriptomic impact of these three LRAs in primary CD4 T cells [38]. At 24h, we measured the response of viral gene expression to the LRAs by flow cytometry. Vorinostat and AZD5582 caused modest increases (~5–10%) in the percent GFP+ cells in the culture, indicating weak latency reversal. By contrast, prostratin caused a more substantial increase in viral gene expression (~25%) (Fig 4A).

We then performed ATACseq on the whole population of LRA stimulated cells, as well as on cells that had been exposed to vehicle (DMSO 0.1%) only. We first examined the impact of LRA stimulation on accessibility of the HIV genome. Notably, all three LRAs caused an increase in fraction of reads that mapped to the HIV genome, although trend was most pronounced for prostratin (Fig 4B and 4C). Specifically, vorinostat and AZD5582 caused 1.47 fold ($FDR = 1.49 \times 10^{-2}$) and 1.57 fold ($FDR = 8.81 \times 10^{-2}$) increases in proviral accessibility respectively, while prostratin caused a 2.22 fold increase, ($FDR = 1.87 \times 10^{-8}$). This observation suggests that latency reversal is associated with reopening of the closed latent provirus, and that the potency of LRAs is related to the magnitude of proviral reopening that they induce. To

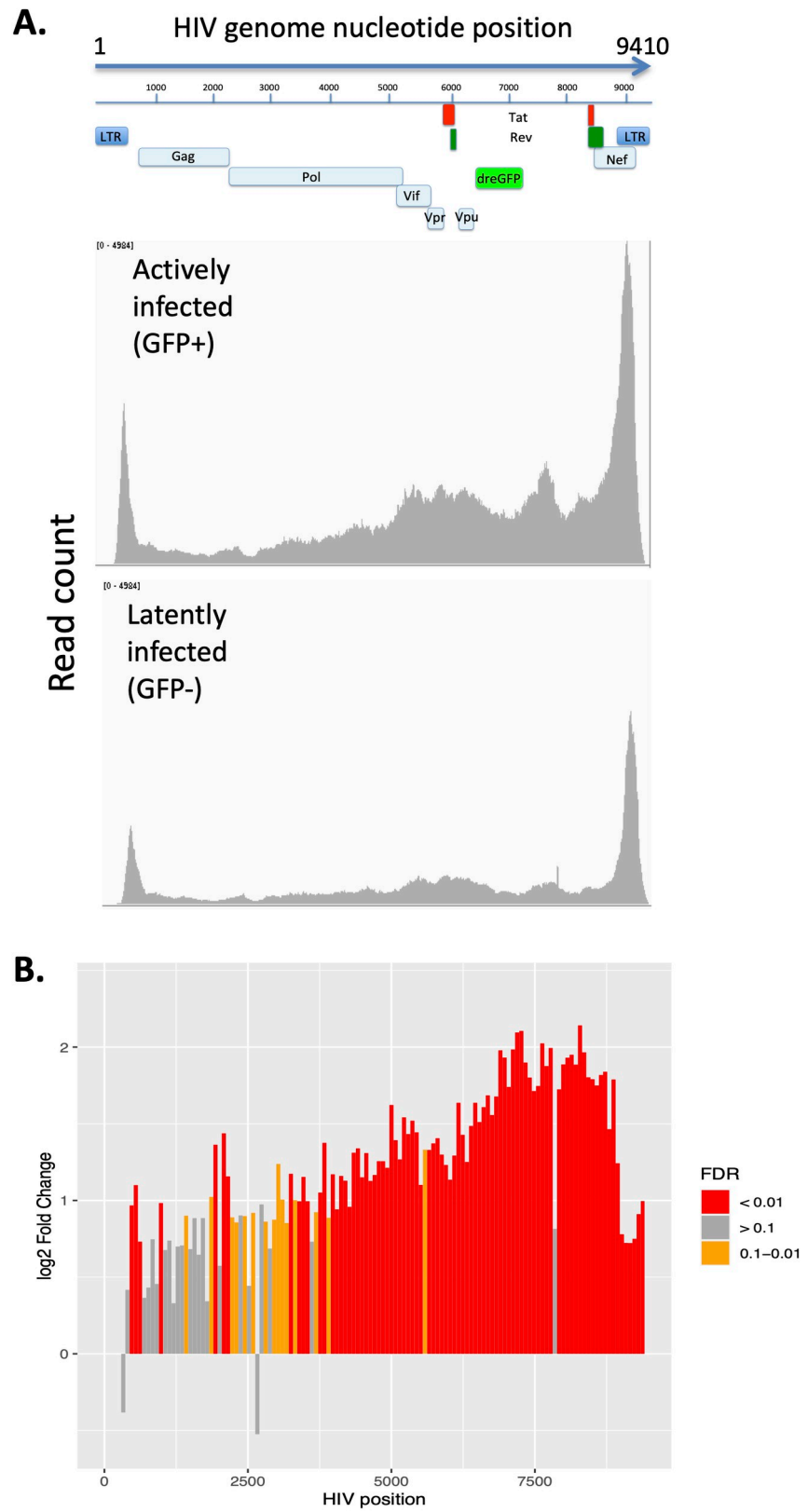


Fig 3. The HIV provirus has reduced accessibility in latently infected cells. **A.** HIV-GFP infected cells at 12wpi were flow sorted into GFP- and GFP+ populations and analyzed by ATACseq. ATACseq reads from the cells were aligned to an HIV reference genome to examine accessibility of the provirus. The intensity of read coverage was graphed across the HIV genome for GFP+ (upper panel) and GFP- (lower panel) cells. Graph shows an aggregate analysis of two donors. **B.** The change in accessibility across the HIV genome during active infection vs latency is shown by determining the fold change between GFP- and GFP+ cells within 73bp bins tiled across HIV. Positive fold change indicates greater accessibility in GFP+ cells. The false discovery rate (FDR) for each 73bp bin is illustrated by the color of each bar: red = <0.01, orange = 0.1 to 0.01, grey = >0.1.

<https://doi.org/10.1371/journal.ppat.1009346.g003>

examine changes in viral accessibility after LRA stimulation in more detail, we subdivided the HIV genome into 73bp bins and individually calculated the fold change and false discovery rate for each bin (Fig 4D). Interestingly, each LRA induced a distinct pattern of changes to proviral accessibility. Prostratin and AZD5582 caused significantly increased accessibility mainly towards the 5' end of the virus, while vorinostat-induced reopening occurred mainly towards the 3' end of the virus. (Fig 4D). The biological basis for the differential behavior of the LRAs in this analysis is unknown, but suggests the existence of previously unappreciated mechanisms of interaction between each agent and the provirus.

By alignment of data from the LRA stimulated cells to the human genome, we were able to identify regions of cellular chromatin for which accessibility was modulated by each agent.

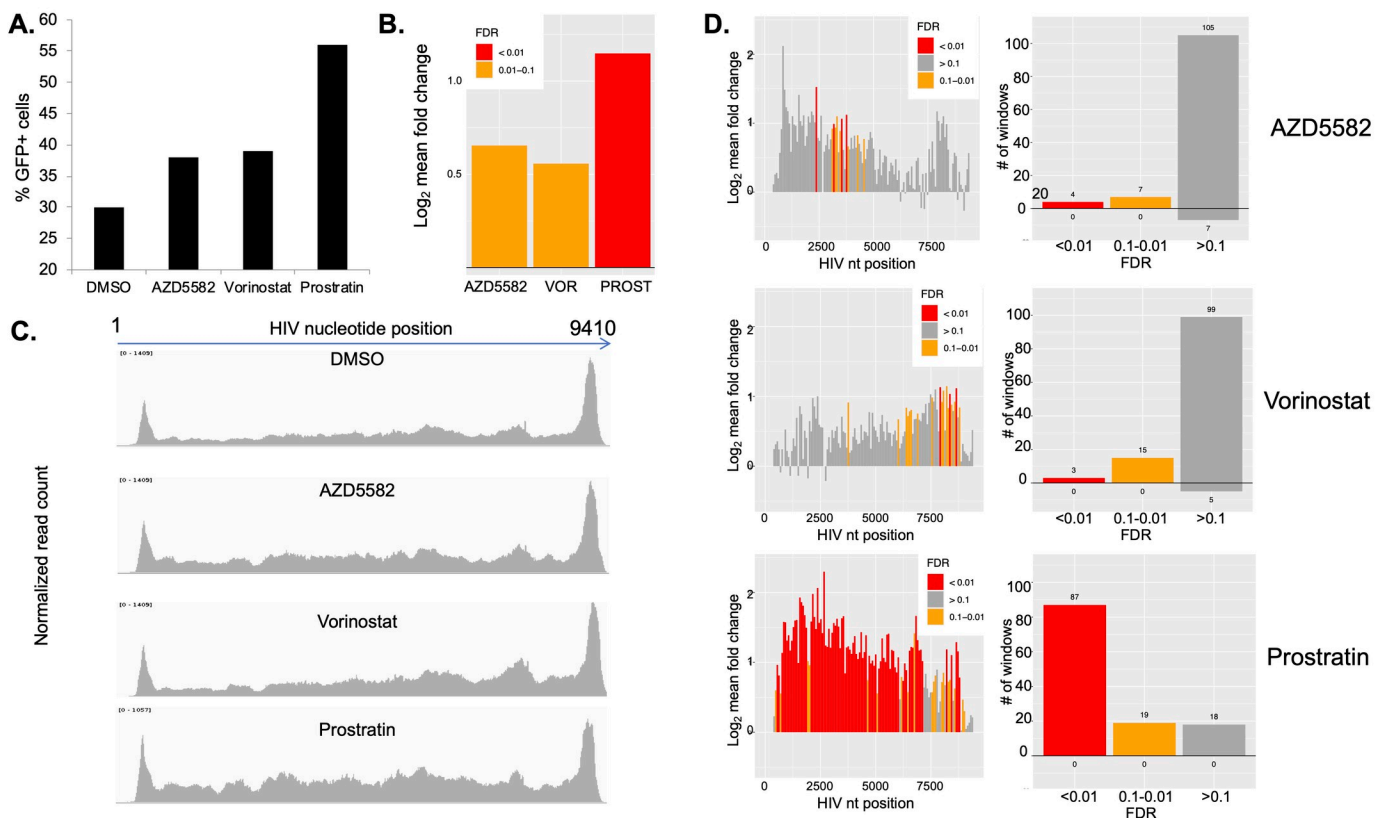


Fig 4. LRAs promote re-opening of the HIV provirus. HIV-GFP infected CD4 cells at 12wpi were stimulated with three different latency reversing agents—vorinostat (250nM), AZD5582 (250nM), prostratin (250nM) or DMSO (0.1%) vehicle for 24hrs. **A.** GFP expression in the culture was analyzed by flow cytometry and the percent GFP+ was plotted. **B.** Stimulated cells were then analyzed by ATACseq. Reads were aligned to the HIV genome and the fold change of reads mapping to HIV in each of the LRA-stimulated conditions are shown. VOR = vorinostat, PROST = prostratin. **C.** The read coverage across the HIV genome for the four conditions (DMSO control and the LRA conditions) are shown. **D.** Fold change in accessibility after LRA stimulation is shown by examination of 73bp bins across the genome (Left panels). Individual bars are color coded by the false discover rate (FDR): red = <0.01, orange = 0.1 to 0.01, grey = >0.1. The total number of differentially accessible 73bp bins across the genome in each FDR category are shown in the right panel. Data represent an aggregate of four independent replicate experiments.

<https://doi.org/10.1371/journal.ppat.1009346.g004>

PCA analysis of accessibility peak data from the four conditions showed that replicate samples stimulated with a given LRA tended to cluster together (S4 Fig). Prostratin and AZD5582 stimulated cells formed distinct clusters separated from each other and vorinostat by PC2. Vorinostat and control vehicle (0.1% DMSO) stimulated cells overlapped in the plot indicating a more limited impact of vorinostat on the accessibility of the host cells. Prostratin had the most dramatic effect on the cells, consistent with its known role as a potent modulator of cellular transcription. Analysis of individual peak dynamics by volcano plot also revealed that each LRA induced a distinct pattern of changes to the accessible chromatin of infected cells (Fig 5). Prostratin had the most pronounced impact and caused numerous accessibility peaks to open and to close. Specifically, 33723 peaks were more open, while 22458 peaks more closed after 24h prostratin stimulation. By contrast, vorinostat caused much more modest changes to host cell chromatin accessibility, with 2326 peaks more open and 3323 peaks more closed. Interestingly, the effects of AZD5582 on chromatin accessibility were highly asymmetric, with 11362 peaks more open and 4945 peaks more closed. This observation is consistent with our previous observation of the asymmetric effect of this compound on transcript levels [38]. Similar to our earlier observations, we found that LRA-induced changes in ATACseq peak accessibility were highly enriched in genes that displayed differential RNA levels after LRA stimulation (S1 Table and S2 Fig). Overall these data indicate that the ATACseq data accurately reflects mechanisms that regulate the transcriptional response to LRAs.

Differential accessibility of specific transcription factor binding sites is associated with latency and latency reversal

To further investigate the regulation of HIV latency, we next examined the enrichment of specific transcription factor (TF) binding sites in the differentially accessible chromatin peaks identified by analysis of the ATACseq data. Since active transcription factors typically promote local opening of chromatin surrounding their binding sites, this information can be used to infer the activity of specific TFs in cells of interest. To investigate this, we analyzed the set of variable peaks using a motif enrichment method, HOMER [39]. This method analyzes defined DNA regions for statistical enrichment of known TF binding site sequences, relative to a background reference sequence. First, we compared the TF site enrichment in the DNA sequences of differential accessible peaks between actively infected (GFP+) and latently infected (GFP-) cells, using the non-differentially accessible peaks as a background baseline (FDR>0.2). From this, we derived a list of transcription factors with binding sites that were enriched in the peaks that were differentially open between latently infected and actively infected cells (Tables 1 and S3 and S4). We identified 406 potential TF binding sequences that were preferentially enriched ($P < 0.05$) in GFP+ cells and 252 sequences that were enriched in GFP- cells. Notably, many of the most highly enriched motifs were variants of consensus binding sites for members of large TF families. Since the binding sites of the individual members in these families are often very similar, precise disambiguation of which specific TFs are active in these cells is difficult to achieve from this data alone. Nevertheless, from examination of the 20 most enriched binding site motifs for each condition, several clear trends were noticeable from the data. Actively infected (GFP+) cells displayed strongly elevated activity for three TF families—AP-1 related TFs, Runt-related (RUNX) TFs, and GATA TFs. Latently infected (GFP-) cells, by contrast, displayed elevated activity of Kruppel-like TFs (KLFs), and Forkhead (FOX) TFs. Notably, several members of these TF families have known roles in regulating HIV gene expression indicating that the results are biologically valid. In particular, AP-1 has well characterized binding sites within the HIV promoter, and has been demonstrated to promote HIV transcription [40–42]. These results are also consistent with our previous scRNAseq findings in this

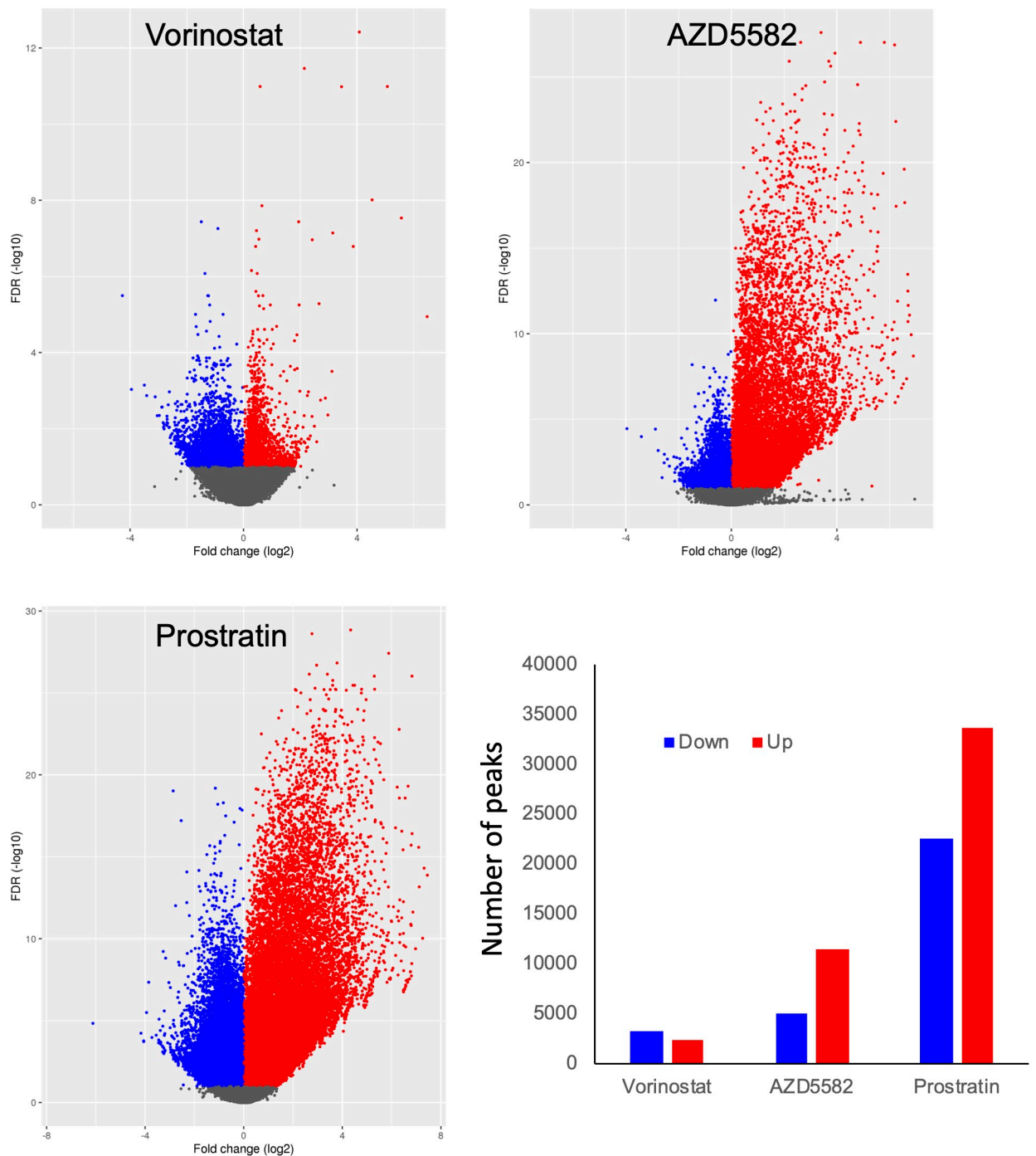


Fig 5. LRAs induce changes to chromatin accessibility in infected cells. ATACseq reads from HIV-GFP infected cells at 12wpi that had been stimulated with vorinostat (250nM), AZD5582 (250nM), prostratin (250nM) or DMSO vehicle for 24hrs were aligned to the human genome. Differentially accessible regions were identified, and fold changes vs FDR plotted as a volcano plot. Data represent an aggregate analysis of four independent replicate experiments. Fold changes versus FDR are displayed as a volcano plot. Peaks that are significantly (FDR < 0.1) more open after LRA stimulation are indicated in red (“up”), while peaks that are significantly more closed are indicated in blue (“down”).

<https://doi.org/10.1371/journal.ppat.1009346.g005>

Table 1. TF motifs enriched in accessible chromatin of GFP+ vs GFP- cells.

Motifs enriched in actively infected cells (GFP+)					
Motif	TF	TF family	P-value	% of Target Sequences with Motif	% of Background Sequences with Motif
DATGASTCAT	BATF	bZIP	1e-449	37.09%	12.04%
DATGASTCATHN	Atf3	bZIP	1e-446	37.52%	12.38%
VTGACTCATC	AP-1	bZIP	1e-443	39.24%	13.54%
NNATGASTCATH	Fra1	bZIP	1e-441	34.27%	10.51%
RATGASTCAT	JunB	bZIP	1e-439	34.21%	10.50%
AAACCACARM	RUNX1	Runt	1e-436	48.42%	20.08%
GGATGACTCATC	Fra2	bZIP	1e-428	31.71%	9.25%
NATGASTCABNN	Fosl2	bZIP	1e-395	25.97%	6.69%
NWAACCACADNN	RUNX2	Runt	1e-375	41.60%	16.72%
GCTGTGGTTW	RUNX-AML	Runt	1e-356	38.68%	15.16%
SAAACCACAG	RUNX	Runt	1e-356	38.06%	14.74%
GATGASTCATCN	Jun-AP1	bZIP	1e-349	21.31%	5.02%
AGATAAGANN	TRPS1	ZF	1.00E-300	53.36%	28.31%
AGATAASR	GATA3	ZF	1.00E-287	46.24%	22.87%
AVCAGCTG	SCL	bHLH	1.00E-248	73.25%	49.88%
YCTTATCTBN	Gata6	ZF	1.00E-234	33.61%	14.99%
NBWGATAAGR	Gata4	ZF	1.00E-234	35.53%	16.39%
RGCCATTAAC	Nanog	Homebox	1.00E-232	67.59%	44.70%
TGCTGAGTCA	Bach2	bZIP	1.00E-213	15.03%	3.92%
TAATCCCN	Pitx1	Homebox	1.00E-213	66.85%	44.94%
Motifs enriched in latently infected cells (GFP-)					
Motif	TF	TF family	P-value	% of Target Sequences with Motif	% of Background Sequences with Motif
NYTGTTTACHN	FOXP1	Forkhead	1.00E-32	12.02%	6.79%
NRGCCCCRCCCHBNN	KLF3	ZF	1.00E-32	11.11%	6.13%
DGTAAACA	Foxo3	Forkhead	1.00E-30	17.55%	11.36%
BSNTGTTTACWYWGN	Foxa3	Forkhead	1.00E-30	9.57%	5.11%
GTGGGCCCCA	ZNF692	ZF	1.00E-29	4.87%	1.94%
DGGGYGKGGC	KLF5	ZF	1.00E-29	19.70%	13.31%
NVWTGTTTAC	FOKK1	Forkhead	1.00E-27	21.44%	14.99%
WWTRTAAACAVG	FoxL2	Forkhead	1.00E-27	18.89%	12.86%
WWATRTAAACAN	Foxf1	Forkhead	1.00E-26	19.92%	13.79%
SCHTGTTTACAT	FOXK2	Forkhead	1.00E-24	14.59%	9.54%
CYTGTTTACWYW	Foxa2	Forkhead	1.00E-23	17.21%	11.84%
MKGGGYGTGGCC	KLF6	ZF	1.00E-22	16.20%	11.03%
AGGTGHCAGACA	Smad2	T-box	1.00E-20	5.58%	2.83%
AGGVNCTTTGT	SOX9	HMG	1.00E-19	16.57%	11.67%
CCCCTCCCCAC	Zfp281	ZF	1.00E-19	4.70%	2.28%
YCTTTGTTCC	Sox4	HMG	1.00E-18	15.79%	11.13%
RGKGGCGKGGC	KLF14	ZF	1.00E-18	22.05%	16.63%
WAAGTAAACA	FOXA1	Forkhead	1.00E-18	22.76%	17.29%
YGGCCCCGCC	Sp2	ZF	1.00E-18	19.68%	14.57%
WAAGTAAACA	FOXA1	Forkhead	1.00E-18	19.82%	14.70%

HOMER motif enrichment analysis of latently infected (GFP-) cells versus actively infected (GFP+) cells, using the total set of open peak sequences in CD4 T cells as a reference baseline. The 20 most enriched sequences in each population relative to the other, and the putative TFs that bind these sequences are displayed.

<https://doi.org/10.1371/journal.ppat.1009346.t001>

Table 2. TF motifs with increased enrichment after LRA stimulation.

AZD5582			Vorinostat			Prostratin		
Motif	TF	P-value	Motif	TF	P-value	Motif	TF	P-value
WGGGGATTCCC	NFκB-p65	1e-3254	RSCACTYRAG	Nkx2.1	1.00E-158	NNATGASTCATH	Fra1	1e-8945
GGAAATTCCC	NFκB-p65	1e-1402	RGCCATTAAC	Nanog	1.00E-158	DATGASTCATHN	Atf3	1e-8905
GGGGGAATCCCC	NFκB-p50	1e-899	TRAGGTCA	THRb	1.00E-156	DATGASTCAT	BATF	1e-8844
TAATCCCN	Pitx1	1e-543	TWGTCTGV	Smad3	1.00E-154	RATGASTCAT	JunB	1e-8731
TWGTCTGV	Smad3	1e-389	TTGAMCTTG	RARa	1.00E-151	GGATGACTCATC	Fra2	1e-8418
AAACCACARM	RUNX1	1e-368	CVGTSCCTCC	Znf263	1.00E-146	VTGACTCATC	AP-1	1e-8322
HTTTCCASG	Rbpj1	1e-366	RGKGGGCGKGGC	KLF14	1.00E-145	NATGASTCABNN	Fosl2	1e-7274
AVCAGCTG	SCL	1e-358	GCTAATCC	CRX	1.00E-144	GATGASTCATCN	Jun-AP1	1e-6346
RGCCATTAAC	Nanog	1e-354	CCAGGAACAG	AR	1.00E-144	TGCTGAGTCA	Bach2	1e-2687
VGCCATAAAA	Hoxd11	1e-352	GGCVGTTR	MYB	1.00E-143	RGCCATTAAC	Nanog	1e-1043

The top 10 most significantly increased TF motifs in accessible chromatin after stimulation of infected with LRAs are shown. Also shown are putative TF binding partners for those motifs and the P-value of significance for enrichment. Data are derived from four independent experiments.

<https://doi.org/10.1371/journal.ppat.1009346.t002>

model, which found elevated and FOXP1 expression in latently infected cells, and elevated GATA3 expression in actively infected cells [28]. To further validate these observations, we also compared our ATACseq data to existing ChIPseq datasets from the Jurkat T cell line [43] for three selected transcription factors (RUNX1, GATA3, MYB) identified by our motif enrichment analysis as exhibiting increased binding site accessibility in actively infected cells. Notably, we observed that peaks with increased accessibility in actively infected (GFP+) cells were significantly enriched with ChIPseq-confirmed binding sites for these TFs, relative to unchanged peaks (S5 Fig and S8 Table).

LRAs induce opening and closing of TF binding sites

To examine whether LRA stimulation triggers changes in the accessibility of bindings sites for specific TFs, we also performed HOMER analysis on the differentially open peaks after LRA stimulation. Each LRA induced a distinct set of TF binding sites to open after stimulation (Tables 2 and S5–S7). For AZD5582, the most strongly enriched TF binding site was NFκB, consistent with this compound's known ability to trigger non canonical NFκB activation [38]. Other AZD5582-induced TF binding sites included PITX1, SMAD3 and RUNX1. For vorinostat, the most highly induced TF binding site was NKX2.1, followed Nanog, THRb and SMAD3. Prostratin induced the opening of binding sites for numerous TFs, most potently for members of the AP-1 family (BATF, FRA1, FRA2, JUNB), suggesting that activation of AP-1 could be a key mechanisms of latency reversal for prostratin and other PKC agonists. We also observed that the three LRAs induced increased accessibility in regions that were enriched for ChIPseq-validated binding sites for the transcription factors we had previously examined (GATA3, RUNX1 and MYB) (S8 Table and S5 Fig). Specifically, AZD5582 increased accessibility of GATA3, RUNX1 and MYB ChipSeq sites, prostratin increased accessibility for GATA3 and RUNX1 sites, while vorinostat increased accessibility for MYB sites (S8 Table).

CTCF regulates HIV latency

We also examined enrichment of TF binding sites present in the set of differentially open peaks of GFP+ and GFP- cells through a different approach—using the overall genome as a background reference baseline (Table 3). Although many of the TFs binding site motifs identified by this approach were also identified by our previous analysis, several notable differences

Table 3. TF motifs enriched in differentially open peaks compared to overall genome.

Motif	Transcription factor (TF)	TF family	P-value	% of Target Sequences with Motif	% of Background Sequences with Motif
AYAGTGCCMYCTRGTGGCCA	CTCF	ZF	1e-323	2.54%	0.21%
CNNBRGCGCCCCCTGSTGGC	CTCF	ZF	1.00E-256	2.80%	0.38%
ACAGGAAGTG	ERG	ETS	1.00E-227	8.24%	3.23%
ACAGGAAGTG	ETS1	ETS	1.00E-223	6.05%	1.98%
DATGASTCAT	BATF	bZIP	1.00E-217	4.34%	1.12%
RATGASTCAT	JunB	bZIP	1.00E-216	3.95%	0.94%
NRYTTCCGGH	Fli1	ETS	1.00E-215	6.32%	2.18%
NNATGASTCATH	Fra1	bZIP	1.00E-211	3.84%	0.92%
AACCGGAAGT	ETV1	ETS	1.00E-202	7.09%	2.72%
GGATGACTCATC	Fra2	ETS	1.00E-196	3.40%	0.77%
VTGACTCATC	AP-1	bZIP	1.00E-196	4.55%	1.32%
DATGASTCATHN	Atf3	bZIP	1.00E-195	4.24%	1.18%
ACCGGAAGTG	ETV4	ETS	1.00E-194	6.29%	2.30%
NATGASTCABNN	Fosl2	bZIP	1.00E-183	2.71%	0.53%
NNAYTTCCTGHN	Etv2	ETS	1.00E-169	5.01%	1.73%
AAACCACARM	RUNX1	Runt	1.00E-165	5.83%	2.24%
RACCGGAAGT	GABPA	ETS	1.00E-164	4.79%	1.63%
GATGASTCATCN	Jun-AP1	bZIP	1.00E-163	2.24%	0.40%
NWAACCACADNN	RUNX2	Runt	1.00E-155	4.74%	1.67%
SAAACCACAG	RUNX1	Runt	1.00E-155	4.44%	1.50%

HOMER motif enrichment analysis of all differentially open/closed peaks between GFP+ and GFP- infected cells, using the whole human genome as a reference baseline. The 20 most enriched sequences and the putative TFs that bind these sequences are displayed.

<https://doi.org/10.1371/journal.ppat.1009346.t003>

were apparent. Interestingly, the transcription factor with the highest degree of enrichment in the differentially open peaks between GFP+ and GFP- cells relative to the overall genome was the insulator protein CTCF. An overall set of 366 CTCF binding sites were differentially accessible between these populations, with 306 CTCF binding sites more accessible in GFP+ cells, and 60 sites more accessible in GFP- cells (Fig 6A). Furthermore, CTCF binding sites were also highly represented within differentially accessible peaks after LRA stimulation (Fig 6B). We also compared our data to ChIPseq datasets for CTCF [43] and confirmed that the differentially accessible peaks were highly enriched for ChIPseq-confirmed CTCF binding sites (S5 Fig and S8 Table). These data suggest that widespread changes in accessibility for CTCF binding sites are associated with HIV latency and latency reversal. These changes occur in both directions, although more CTCF sites were accessible during active infection than in latency. CTCF is a large, ubiquitously expressed regulator of gene expression with an 11 zinc finger central domain and N and C termini that interact with numerous protein partners [44]. The human genome contains 15,000–40,000 CTCF binding sites [44,45], and CTCF homodimerization drives the formation of large topologically-associated DNA domains (TADs) by mediating chromatin loop structures [46]. These loops serve to insulate domains of transcriptional regulation by disconnecting genes from nearby enhancer elements, and by restricting lateral spread of domains of both inhibitory and activating histone modifications [47,48]. Interestingly, CTCF is known to play an important role in regulating transcriptional latency for a number of other viruses, including HSV, KSHV and HPV [49,50]. However, CTCF is not known to have a role in regulation of HIV latency. To investigate the possible role of CTCF in HIV latency we examined the impact of CTCF depletion on viral gene expression in HIV latency models. First, we transduced N6 cells, a latently infected T cell line model that contains

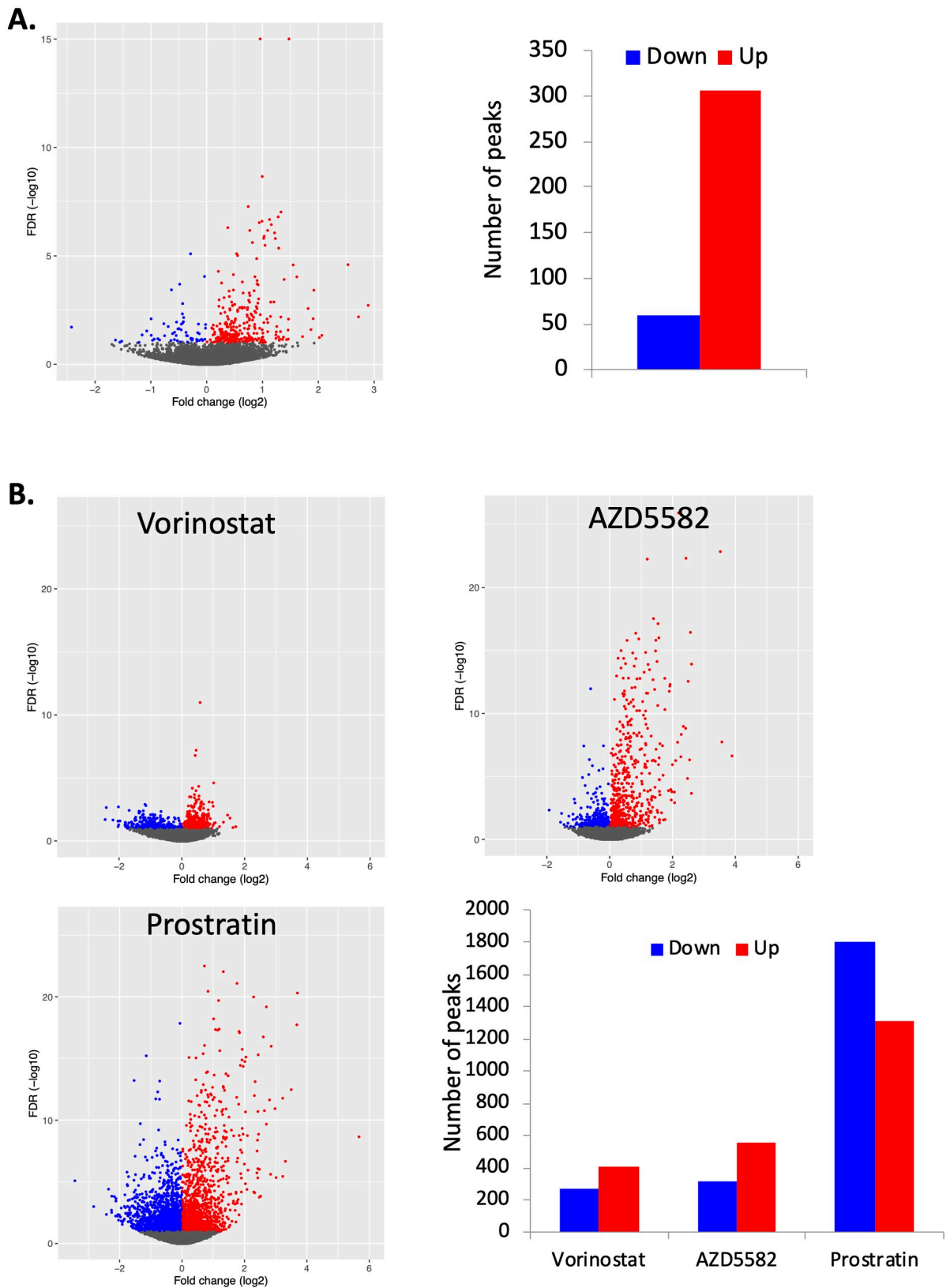


Fig 6. Differential opening of CTCF sites in latency vs active infection, and during latency reversal. **A.** Differentially accessible regions of the cellular genome containing consensus CTCF binding sites were identified by comparison of actively infected (GFP+) and latently infected (GFP-) cells, and fold changes versus FDR values are displayed as a volcano plot. Red datapoints indicate regions that are more accessible in actively infected (GFP+) cells (“up”), while blue datapoints represent regions that are more open in latently infected (GFP-) cells (“down”). **B.** Host cell chromatin peaks containing CTCF binding sites are displayed for the three LRA drug treatment conditions. Peaks with significantly different accessibility (FDR <0.1) after stimulation of HIV infected cells are highlighted in red (upregulated by LRA) and blue (downregulated by LRA). The data represents an aggregate of infected cells from four independent replicate experiments.

<https://doi.org/10.1371/journal.ppat.1009346.g006>

a transcriptionally silent HIV reporter virus that encodes murine heat shock antigen (HSA) [28], with lentiviruses that express an shRNA targeting CTCF or a scrambled control shRNA. Western blot analysis of CTCF confirmed specific reduction in CTCF expression in the cells transduced with the CTCF-targeting shRNA (Fig 7A). At two weeks post transduction, we measured viral reactivation by HSA expression on the surface of the transduced cells using flow cytometry. Notably, we observed potent re-expression of HSA in the CTCF-depleted cells, indicating that CTCF expression is required to maintain repression of HIV transcription in this cell line (Fig 7B and 7C). We next examined the impact of CTCF depletion in our primary CD4 T cell latency model. We activated CD4 T cells through TCR stimulation for 48h, then infected with HIV-GFP. At 3dpi, we purified infected (GFP+) cells, then nucleofected the infected cells with Cas9/sgRNA ribonucleoprotein complexes targeting the viral protein Tat, CTCF or with a non-targeting control (Fig 7D). Quantitative PCR one week after nucleofection confirmed roughly 50% reduction of CTCF expression in the targeted cells (Fig 7D). The infected cells were then cultured for four weeks and the percentage of latently infected (GFP-) cells over time measured by flow cytometry (Figs 7F and S6). Tat targeted cells, as expected, demonstrated a more rapid decline in viral gene expression, confirming the sensitivity of this system to perturbations in regulators of viral transcription (Fig 7E). Notably, CTCF depleted cells exhibited a reduced level of HIV latency beginning at 2wpi, and this effect became more prominent at 3wpi and 4wpi. Overall these results demonstrate a role for CTCF in the establishment or maintenance of HIV latency and highlight a previously unknown mechanism of transcriptional repression for HIV.

Discussion

The presence of a latently infected reservoir is recognized as a primary barrier to curing HIV infection [51]. As such, the mechanisms by which latent infection is established and maintained are of critical importance. By characterizing these mechanisms we may identify strategies to either induce viral gene expression leading to infected cell clearance (“shock and kill”), or induce permanent silencing of the proviruses (“block and lock”) [52,53]. Viral gene expression is regulated by host cell transcription factors and epigenetic mechanisms. Dynamic changes and cell-to-cell variation in these factors likely leads to specific cellular environments that favor or disfavor latency. However, viral transcription can also be influenced by stochastic fluctuations in the Tat transcription factors early in infection [21,54]. Nevertheless, the ability of LRAs targeting host cell factors to induce viral reactivation in both animal models and in human clinical studies demonstrates the value of targeting host cell pathways to manipulate latency [22,38,55]. Thus, latency initiation and reversal both likely result from a combination of stochastic and deterministic processes intrinsic to both the virus and the host cell. These processes yield a diverse population of persistently infected cells, with intermittent viral expression and re-entry into quiescence, undergoing both proliferation and death.

The rare nature of latently infected cells and the lack of markers for their purification make direct observation of cellular events in infected cells from patients extremely difficult. In this regard, model systems have been invaluable in the discovery and validation of new targets. In

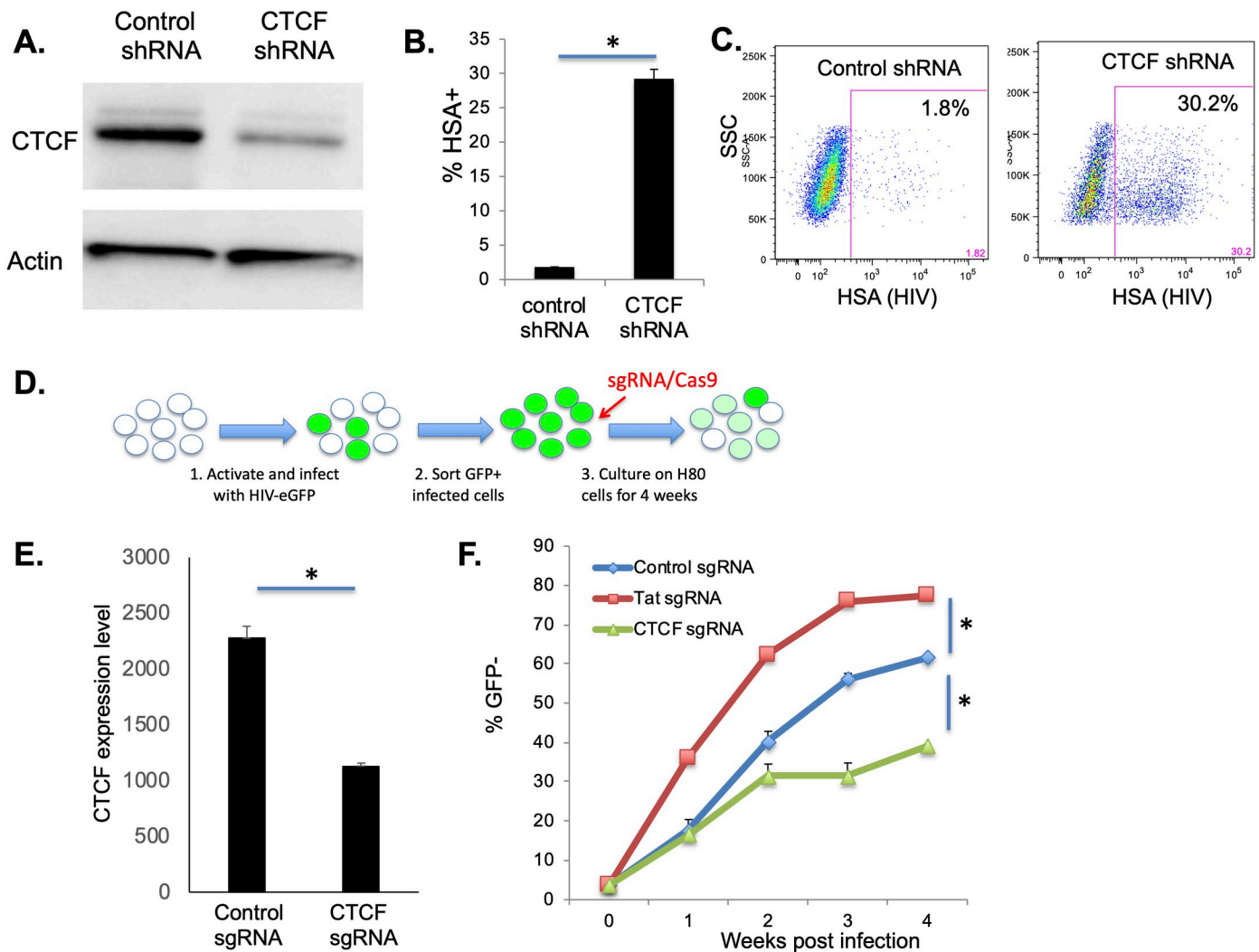


Fig 7. CTCF depletion in latently infected CD4 T cells inhibits HIV latency. N6 cells were transduced with lentiviruses that express mCherry and an shRNA targeting CTCF or a control scrambled shRNA. **A.** Transduced N6 cells were analyzed by western blot to examine CTCF expression as well as a loading control (β -Actin). **B.** Transduced N6 cells were analyzed at 2 weeks post infection by flow cytometry, and the percentage GFP+ cells within the mCherry+ gate determined. Data represent the average of two technical replicates from one of two representative experiments. **C.** A representative FACS plot of HSA expression in shRNA transduced N6 cells. **D.** Primary CD4 T cells at 2 days post activation were infected with HIV-GFP, then nucleofected with Cas9 complexed with a control sgRNA, a Tat sgRNA or a CTCF sgRNA. **E.** At one week post infection, the impact on CTCF mRNA levels, relative to beta-actin was quantified by Taqman qPCR. CTCF expression is displayed in arbitrary units and represents the average of four replicates. **F.** The infected/nucleofected cells were analyzed for GFP expression over four weeks of infection and the percent of latently infected (GFP-) cells was determined. Each datapoint is the average of two independent nucleofection reactions. Asterisks represent significant differences ($P < 0.05$, Student's T test).

<https://doi.org/10.1371/journal.ppat.1009346.g007>

this study, we have discovered, using a primary cell model system, that viral latency is a stable, heritable phenotype that is rapidly re-established after stimulation and transmitted to daughter cells during cell division. These findings expand the understanding of epigenetic mechanisms likely contribute to the maintenance of latency through cell division, an observation that has significant implications for the maintenance of the HIV reservoir *in vivo* through clonal expansion. Epigenetic memory of HIV latency through cell division and T cell development are likely to explain the presence of expanded latently infected clones within diverse T cell subset identities present in infected individuals on therapy. The rapid and repetitive re-

establishment of latency after stimulation also has implications for clearance strategies against reactivated latently infected cells *in vivo*. These results suggest that a temporally limited window of antigen presentation will likely occur after LRA dosing before latency is reestablished and antigen expression wanes. An important caveat to consider is that our model system involves several steps of *in vitro* manipulation which could influence the phenotype or behavior of these cells. These steps include activation with TCR-stimulating beads, and extensive co-culture with H80 cells to permit long term CD4 T cell survival in culture. Our model system also involves the use of a replication defective reporter strain of HIV, which may behave differently from clinical isolates of HIV. Nevertheless, a comparison of the transcriptional profiles of our model cells indicated that these cells closely resemble resting memory CD4 T cells, the natural host for latent HIV.

By analysis of infected cells with ATACseq, we have also discovered that HIV latency is preferentially established in cells that adopt a specific global chromatin structure after activation and return to a resting state. This finding suggests that latency is intricately connected to cell-intrinsic chromatin dynamics, and that these dynamics could represent an important target for preventing latency establishment, reversing latency once established, and preventing a return to latency following latency reversal. The specific mechanisms by which different patterns of overall chromatin accessibility are regulated in these infected cells as latency is established are unclear, and will require further investigation. We have also demonstrated that accessibility of the viral genome is significantly reduced in latently infected cells relative to actively infected cells. This finding confirms previous results in cell line models showing the appearance of restrictive chromatin structures associated with entry of HIV into latency [56]. The closed nature of the provirus likely impedes recruitment of transcription factors essential for initiating viral transcription, and processive RNA polymerase transcription, thereby contributing to latency. Notably, three LRAs with distinct mechanisms of action caused significant re-opening of the provirus, indicating that re-opening of HIV is likely a key feature of latency reversal by LRAs.

An important determinant of proviral chromatin state and expression that cannot be determined from our ATAC-seq data is the influence of the viral integration site. Specifically, the transcription state of neighboring genes and the activity of nearby enhancers will likely modulate the chromatin state and expression of the provirus [57]. HIV frequently integrates near clusters of endogenous enhancer elements [58], and studies in Jurkat cells have shown a clear influence of local endogenous enhancers on HIV expression [57]. While we were able to recover some ATAC-seq fragments that spanned virus-host cell junctions, revealing individual integration sites, these fragments were rare, and the information too sparse to draw significant conclusions about the influence of viral integration site on the proviral chromatin state. Additional experimentation using integrated single cell ATAC-seq and integration site analysis may be useful for resolving this question. New methods, for example, that permit parallel analysis of lentiviral integration sites and host cell chromatin accessibility in single or bulk populations of T cells will help to clarify the relationship between proviral expression, integration site and local chromatin structure [59].

It is also important to note that during clinical infection, infected cells with high viral gene expression are rapidly killed due to a combined effect of viral cytopathic effect and immune pressure. The surviving cells are thus selected to exhibit low residual viral gene expression. Recent studies of individual infected cells demonstrate that a minor population of cells from patients receiving ART express viral transcripts [60], and that these cells can undergo clonal expansion to sustain the reservoir [15,16,60]. HIV infection also frequently results in defective proviruses containing inactivating deletions, and these mutant proviruses constitute the majority of HIV DNA in people receiving ART [61]. For single round reporter viruses such

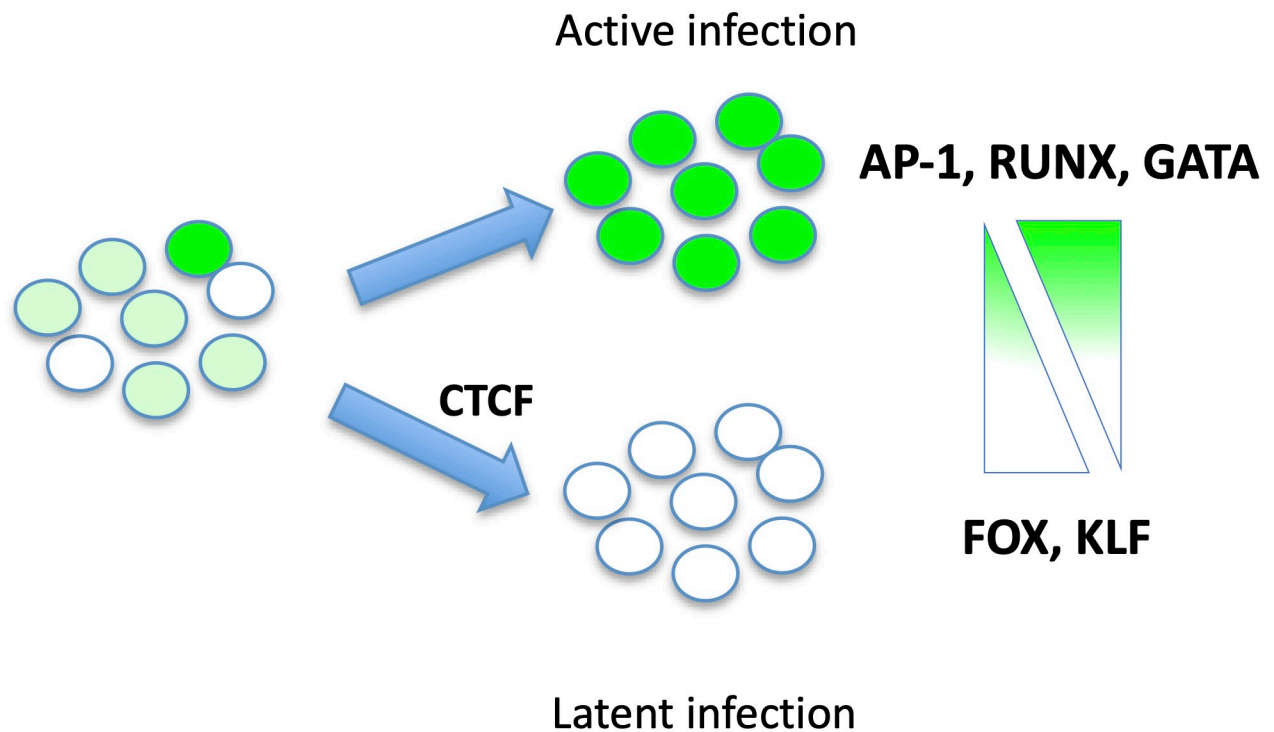


Fig 8. Schematic model of transcription factor activity and HIV latency. HIV-GFP infected cells exhibit variegated levels of transcriptional silencing. Comparison of actively infected (green) and latently infected (white) CD4 T cells reveals latency associated TF activity. Latently infected cells have elevated Forkhead (FOX) and Kruppel-like factor (KLF) TF activity, and reduced AP-1, RUNX and GATA TF activity relative to actively infected cells. The chromatin insulator CTCF promotes HIV latency.

<https://doi.org/10.1371/journal.ppat.1009346.g008>

the one used for this study, we expect the frequency of mutated proviruses to be low, particularly since we initially sort for infected cells with active viral gene expression. Furthermore, our bioinformatic pipeline only mapped intact viral DNA fragments and did not consider fragments with deletions. Future studies of the epigenomic characteristics of provirus may derive insight from also considering defective viruses.

This study also provides new insight into the mechanisms of latency by identifying a set of host cell TFs whose binding site accessibility correlates with viral gene expression. Specifically, by comparing the differentially open regions of chromatin in the latent or actively infected cells to the overall set of open regions across infected CD4 T cells, active viral expression was correlated with elevated accessibility of binding sites for AP-1, GATA, and RUNX TFs, while latency was associated with elevated accessibility of binding sites for Forkhead (FOX) and Kruppel-like factor (KLF) TFs (Fig 8). Importantly, this list contains several TFs that have well-established roles in regulating HIV transcription, indicating the biological relevance of the findings. AP-1, for example has long been identified as a regulator of HIV gene expression [40–42]. AP-1 is a family of heterodimeric basic leucine zipper TFs that recognize a consensus binding site (TGA G/C TCA), and are potently activated by stimuli that promote HIV gene expression, such as phorbol esters. There are several AP-1 binding sites in the HIV LTR, and these have been shown to be important for HIV transcription and replication, suggesting that the correlation between AP-1 activity and viral gene expression in our model is likely through

direct activity of AP-1 at the HIV LTR. Interestingly one AP-1 site has recently been shown to regulate latent infection [40]. GATA TFs have also previously been identified as regulators of HIV transcription. GATA3, in particular, has been shown to bind the HIV LTR and promote HIV gene expression [62]. Notably, we have previously reported that elevated GATA3 transcript levels correlates with active HIV transcription in this model system [28]. From the list of TFs with enriched activity in latently infected (GFP-) cells, FOXP1 was also previously reported by our group as exhibiting elevated expression in latently infected CD4 T cells [28]. This TF is associated with a quiescent central memory T cell (Tcm) phenotype, consistent with the hypothesis that cells that adopt a quiescent Tcm phenotype are more prone to latent infection [63]. Interestingly, a related Forkhead TF, FOXO1, has also recently been reported to promote HIV latency [64,65]. In addition to these known HIV-regulating TFs, several TFs with no known role in HIV transcription were identified by our study, and defining their role in latency will likely lead to the discovery of new mechanisms of transcriptional repression for HIV. However, it is also important to note that changes in TF binding site accessibility do not necessarily reflect differential TF activity, and additional analyses, such as ChIPseq, may be needed to directly confirm differential binding in latently infected cells.

We also examined enrichment of TF motifs in the differentially open chromatin of infected cells relative to the overall human genome. Interestingly, by this comparison, the differentially open chromatin regions of infected CD4 T cells were highly enriched with binding sites for CTCF, indicating that possible dynamic changes to chromatin looping may distinguish latently infected cells from actively infected cells. Furthermore, by selective knockdown of CTCF in a T cell line model of latency (N6) and in a primary cell latency model, we demonstrated that CTCF expression is required for maintaining HIV silencing in HIV infected T cells. These results thus identify CTCF as a novel regulator of HIV latency. Notably, previous studies have shown that CTCF plays a key role in regulating latency for herpesviruses [49,50]. In these cases, dynamic CTCF binding to cognate sites within the viral genomes directly regulates switching between lytic and latent gene expression programs. It is unknown if CTCF binds directly to the HIV genome. As such, the precise mechanism by which CTCF regulates HIV latency is unknown and will require further investigation. CTCF has several known molecular functions that could potentially affect HIV gene expression. In particular, CTCF binding to DNA mediates chromatin looping and the formation of topologically associated domains (TADs), and serves to insulate genes from adjacent enhancer elements and zones of histone modifications [46]. It is conceivable that, in many infected cells, HIV gene expression is impacted by insulation of the HIV promoter from nearby enhancer elements and that disruption of the insulator function of CTCF allows longer-range promoter/enhancer interactions to occur that promote viral reactivation. Alternatively, many latent proviruses may be trapped in TADs that are enriched for repressive histone modifications such as H3K9me3 and H3K27me3, and removal of the CTCF boundary allows replacement of these heterochromatic histone modifications with activating marks. CTCF can also regulate sub-nuclear localization of chromatin regions and mediate interactions with the nuclear lamina [66]. As such, it is also possible that CTCFs ability to regulate latency is determined by spatially restricting access of the provirus to activating transcription factors within the nucleus.

It is also unclear what might regulate differential binding activity of CTCF within our latency model system. CTCF is ubiquitously expressed in mammalian cells, and CTCF transcripts were not differentially expressed between the latently infected (GFP-) and actively infected (GFP+) cells. It is possible that CTCF binding is regulated by expression or differential binding of an interacting partner. CTCF contains three major structural domains—a central DNA binding domain with 11 zinc fingers, and long N and C terminal domains. All three domains interact with an array of binding partners, including Y-box binding protein 1

(YBX1), Chd4 and Snf2h [44]. CTCF itself can be regulated by phosphorylation, sumoylation and ADP ribosylation [67]. CpG methylation of CTCF target sites can also regulate CTCF binding [68]. It is also possible, given CTCFs fundamental role in transcriptional regulation, that CTCFs role in repressing HIV transcription is indirect, and mediated through an as yet unidentified factor whose expression or activity is regulated by CTCF. The precise mechanism behind CTCFs role in latency will need to be clearly defined. No therapeutic strategies for targeting CTCF currently exist, but the results we describe in this manuscript suggest that inhibition of CTCF activity or a pathway regulated by CTCF could serve as a strategy for impacting HIV latency clinically.

Materials and methods

Viruses and cell culture

Stocks of HIV-GFP were generated by co-transfection of 293T cells with the pNL4-3- Δ 6-dreGFP plasmid (generous gift of Robert Siliciano) and the packaging plasmids PAX2-GagPol and MD2-VSVG, using Mirus LT1 reagent. At 2 days post transfection virus was harvested from the supernatant and clarified by low speed centrifugation, followed by filtration through a 45uM filter. shRNA expressing lentivirus plasmids targeting CTCF were purchased from Genecopeia (HSH060478-LVRU6MP) and viral particles generated as for HIV-GFP. Latently infected N6 cells are a clone of Jurkat cells that contain a full length clone of the NL4-3 strain of HIV in which the Nef open reading frame has been replaced by an HSA-P2A-Luciferase cassette (generous gift from David Irlbeck, Viiv Healthcare).

Primary cell latency model

Primary CD4 T cells were isolated from fresh whole blood (purchased from Gulf Coast Regional Blood Center) by Ficoll isolation of peripheral blood mononuclear cells (PBMCs), then magnetic negative selection of CD4 T cells using a CD4 enrichment kit (Stem Cell). Total primary CD4 T cells were activated using anti-CD3/CD28 beads (Thermo Fisher) at a ratio of one bead per cell for 2 days, then infected with HIV-GFP viral supernatant by spinoculation for 2hrs at 600g with 4ug/mL polybrene. The infected cells were then resuspended in fresh RPMI with IL-2 (Peprotech) at 100U/mL and incubated for two days before actively infected (GFP+) cells were sorted using a FACSAria flow sorter (Becton Dickson). The purified infected cells (GFP+) were then maintained at 1-2M/mL with fresh media and IL-2 added every 2-3 days for 2 weeks. From 2wpi, cells were co-cultured with H80 cells (provided by D Bigner, Duke University) in the presence of 20U/mL IL-2 for up to 12wpi. During co-culture the cells were fed with fresh media and IL-2 every 2-3 days, and moved to a fresh flask of H80 feeder cells once a week. For latency reversing agent stimulation, AZD5582, vorinostat and prostratin were generous gifts from David Irlbeck (Viiv Healthcare), and were reconstituted in DMSO at 10mM, before dilution into media to working concentrations of 250nM.

ATACseq library construction and sequencing

Infected cultures of CD4 T cells were stained with the viability dye Zombie Violet (ZV, Biolegend) and 50,000 live (ZV-) cells were sorted into 5mL polypropylene tubes using a FACSAria flowsorter (Becton Dickson) per replicate. The sorted cells were then washed with phosphate buffered saline (PBS) and resuspended in 50ul TD buffer (Illumina) with 0.01% digitonin and 2.5ul TDE Tagmentation enzyme (Illumina). The cells were then agitated in a thermomixer at 37C, 300rpm for 30mins to allow transposon-mediated tagmentation to occur. Tagmented DNA was then purified a Minelute purification kit (Qiagen). DNA fragments were PCR

amplified using dual barcoded primers and re-purified using a Minelute kit followed by a 1.5x final cleanup step with Ampure SPRI beads (Agencourt). Libraries were quantified by Qubit DNA assay (Invitrogen) and size distribution of the library was determined using a High Sensitivity DNA BioAnalyzer chip (Agilent). Samples were sequenced using an Illumina HiSeq2500 to a depth of 50-150M reads with 150bp paired end sequencing.

ATAC-seq bioinformatic analysis pipeline

Sample fastq files were combined by read end and validated with FastQC v0.11.8. Combined fastq read end pair files were aligned with BWA_MEM v0.7.17 [69] against the GRCh38 genome with an added HIV chromosome sequence. BWA is a fast and accurate aligner with which we have significant experience. It generates high-quality alignments while still allowing for soft-clipping and alignment across HIV (non-clonal) integration sites (with the -Y flag). Aligned bam files were sorted and optical duplicates were flagged with Picard Tools v2.21.1 [70]. Picard tools and SAMtools v1.9 [70] were used to verify the quality of the alignments. Filtering prior to read counting was done by SAMtools to remove duplicate, unmapped, and unpaired reads. HIV-aligned reads were duplicated into separate bam files and merged by group with SAMtools for visualizing in IGV [71]. Read counting was done using csaw v1.16.1 [72], with parameters set to count reads with map quality greater than or equal to 10 and inferred fragment length less than or equal to 2000. Read ends were counted separately and combined into bins of 73 bp windows (~1/2 nucleosome width) tiled across the entire genome. Reads were also counted into 5000 bp bins, generating a second count matrix at lower resolution. The csaw function filterWindows() was used to calculate the log₂ signal to noise of all 73 bp windows relative to the median of the large windows counts, scaled by relative window sizes. Low signal windows less than or equal to log₂(2.5) over background were dropped. Per sample TMM normalization factors were calculated using the csaw normFactors() function, retaining potential global changes in accessibility. The normalized noise-filtered count matrices were analyzed by PCA using the prcomp() function of R 3.5 [73]. Review of the PCA results indicated the differential accessibility model for GFP+ vs GFP- data should include donor as a cofactor. For latency reversing agents treated cells, four replicate experiments were carried out using cells from a single donor and a simple treatment-control model for each. Differential accessibility was calculated using edgeR v3.23.4 [74], with global dispersion factors from estimateDisp(). Models were fit with glmQLFit() using robust = TRUE, and the significance calculated using glmQLFTest(). Finally, successive windows were merged into clusters across valleys of up to one bin width using mergeWindows() and combineTests() from csaw. The generated clusters from each modeled comparison were plotted as volcano plots visualizing the FDRs and the mean fold change (calculated as the mean fold change of all the cluster's windows). Clusters were assigned category labels if they overlapped a matching HG38 annotated region, e.g. a "promoter" or "enhancer", as provided by annotatr v1.8.0.

Transcription factor binding site enrichment for each modeled comparison was performed using findMotifsGenome.pl from HOMER v4.10 [75]. Input bed files were generated from the csaw cluster results, separately for the significantly up and significantly down clusters (FDR < 0.1), and for the undifferentiated regions (FDR > 0.2). Using the included default set of transcription factors, HOMER was run to identify known TF binding sites enriched in the significant bed files relative to the undifferentiated background. Separate runs were performed examining the middle 50, the middle 200, and all bases of the significantly up and significantly down cluster regions. Results from the top hits were manually examined for relevant biological associations with HIV. Additionally, HOMER was used to compare the significantly different regions against the genome as a whole. A bed file was generated containing all clusters and

analyzed with findMotifsGenome.pl using the -find parameter to locate all CTCF binding sites. Clusters overlapping CTCF sites were extracted and their differential accessibilities were plotted as volcano plots.

RNAseq analysis

RNA was extracted from infected CD4 T cells using an RNEasy plus kit (Qiagen), and eluted into RNase free water. RNA was quantified by Qubit assay (ThermoFisher) and integrity checked using a BioAnalyzer RNA nano kit (Agilent). RNA was then used for construction of expression libraries using a SMART-seq low input RNA kit (Takara). Libraries were quantified by Qubit and sequenced on an Illumina Hiseq2500 sequencer. Alignment to the Hg38 reference genome was carried out with Star v 2.4.2a [76]. Aligned bam files were sorted and optical duplicates were flagged with Picard Tools v1.86 [70]. Picard tools and SAMtools v1.9 [67] were used to verify the quality of the alignments. Filtering prior to read counting was done by SAMtools to remove duplicate, unmapped, and unpaired reads. Gene transcript counts were obtained from the filtered aligned reads using Salmon v 0.6 [77] and differential gene expression results were quantified using DESeq2 v 1.22.2 [78].

Comparative RNAseq-ATACseq analysis

For each data set, genes with differential expression results were split into two groups, those with significant and large differences (adjusted p-value < 0.1; abs fold change > 1.5x) differences versus all others. Independently, genes with overlapping differential accessibility clusters were split into two groups, those with associated clusters whose differential accessibility is significant (FDR < 0.1), large (abs fold change > 1.5x), and are changing in the same direction (up or down) versus those with either no significant and large difference in accessibility or those with mixed directionality. For genes with both differential expression and differential accessibility data, R 4.0 [73] was used to: (1) perform a Fisher's exact test ("fisher.test" function) for evaluating the significance of the relationship between expression and accessibility changes, (2) to obtain the expected distributions of paired categories ("chi.test" function) for evaluating the directionality of the relationship, and (3) to obtain the linear correlation ("cor" function) between the (log₂) of the gene expression fold change and the (log₂) of the largest fold change of any associated cluster.

Comparative ChIPseq-ATACseq analysis

ChIPseq data from a human CD4 T cell line [43] in the form of reads per million per 50bp bins were downloaded and converted from hg19 to hg38 using rtracklayer V1.48.0 [79] for the RUNX1 (GSM1975912), GATA3 (GSM1975913), MYB (GSM1519635), and CTCF (GSM1689152) transcription factors. Each cluster was assigned a score for each transcription factor based on all chip read bins overlapped = log₂ (1 + sum of bin per million counts). Evaluation of the difference in ChIP score means between the significantly more open clusters (FDR < 0.1 and fold change > 1.5x) and non-significantly different clusters (FDR > 0.2, any fold change) was done using the t-test function of R 4.0 [73] to perform a Welch Two Sample t-test.

Antibodies/Western blots

To extract protein, cells were washed in phosphate-buffered saline (PBS) then lysed in RIPA buffer (Sigma) supplemented with Complete protease inhibitors (Roche), before centrifugation to remove cell debris. Protein concentrations were then quantified using Bradford assay

reagent (BioRad), and 40ug of protein per sample was run on a 4–12% Tris Glycine polyacrylamide gel (Invitrogen). The gel was then transferred to a PVDF membrane. The membranes were then incubated in Tris-buffered saline with 0.1% Tween 20 (TBST) with a primary detection antibody for 2-hrs, before washing with TBST. Primary antibodies used were anti-CTCF (clone D31H2, Cell Signaling) and beta-actin (ab49900, Abcam). Primary antibody staining was followed by incubation with an HRP-linked anti-rabbit antibody. Bands were imaged by incubation of the membrane with ECL chemiluminescence reagents (ThermoFisher) and imaging on a ChemiDoc MP imager (BioRad). For flow cytometry for murine Heat Shock Antigen (HSA), cells were washed in PBS before staining with anti-HSA-APC (Biolegend) at 1:200 dilution for 30mins at 4C, before washing in PBS and analysis using a Fortessa flow cytometer (Becton Dickson).

Quantitative PCR

Quantitative PCR for CTCF was performed using Taqman probes/primers (Thermo Scientific). Specific primer probe sets used were Hs00902016_m1 for CTCF, and Hs01060665_g1 for beta-actin. Reactions were run using primer/probes at 1:20 dilution and Taqman FastVirus one-step reaction mix (Thermo Scientific), and amplified using an Applied Biosystems QS3 cyclor. Expression level was determined using cycle threshold relative to beta-actin and represented in arbitrary units.

CRISPR/Cas9 nucleofection

CRISPR RNAs (crRNAs) targeting CTCF were pre-designed by IDT or designed with the BROAD institute GPP sgRNA designer. crRNAs and trans-activating RNAs (tracrRNA; IDT) were annealed in a thermocycler at a 1:1 ratio according to manufacturer's instructions and stored at -20 C until use. Nucleofection experiments in primary CD4 T cells were performed six days post-infection using methods previously described for CRISPR nucleofection of either resting or activated human T cells [80,81]. CRISPR-Cas9 ribonucleoprotein (RNP) complexes were generated by mixing crRNA:tracrRNA complexes with ALT-R *S.p.* Cas9 nuclease V2 at a 3:1 molar ratio for 10 minutes at room-temperature. Infected CD4 T cells were washed with PBS and 3 million cells per condition were resuspended in 20uL buffer P2 (Lonza) with 4μM IDT electroporation enhancer. CRISPR-Cas9 RNP nucleofection was performed with the EH100 nucleofection protocol on a 4D Nucleofector device (Lonza) [80,81]. Immediately after nucleofection cells were resuspended in fresh pre-warmed media containing 100U/mL recombinant human IL-2. sgRNA targeting sequences were: Non targeting: CGGAGGC-TAAGCGTCGCAA, Tat: CCTTAGGCATCTCCTATGGC, CTCF: GAGCAAACCTGCGTTATACAG.

Supporting information

S1 Fig. Principal component analysis of actively infected and latently infected cells ATAC-seq data. Principal component analysis was performed on ATACseq data from actively infected (GFP+, blue datapoints) and latently infected (GFP-, red datapoints) CD4 T cells. For Donor 1 (circles), two independent replicate experiments are shown. Donor 2 data represented by triangles.
(TIF)

S2 Fig. Fold change comparison—transcript level vs gene-associated ATACseq peaks. Fold changes for both RNA expression and ATACseq accessibility were calculated for each cellular gene and plotted. Each datapoint represents an individual gene. Data points are color coded by

statistical significance—purple = genes with significant change to RNA transcript only ($P < 0.1$, fold change > 1.5), green = genes with significant change to ATACseq peak only ($P < 0.1$, fold change > 1.5), red = genes with significant change to both RNA transcript and ATACseq peak. Correlation coefficients for “red” genes between ATACseq peak change and RNA transcript change are indicated in the lower right corner of each panel.

(TIF)

S3 Fig. CIBERSORT analysis of RNAseq profiles from HIV infected cells. HIV latency model cells were profiled by RNAseq and compared to reference transcriptomes using the CIBERSORT machine learning method [35]. Enrichment of each experimental sample with hallmark signatures from immune cell reference profiles were calculated and plotted as a heat-map.

(TIF)

S4 Fig. Principal component analysis of infected cells after LRA stimulation ATACseq data. Principal component analysis was performed on ATACseq data from HIV-infected cells after 24h stimulation with three different LRAs or vehicle (DMSO). For all conditions, four independent replicate experiments are shown except for prostratin, for which three replicates were used. For all LRAs 250nM concentration was used.

(TIF)

S5 Fig. Comparison of ATACseq data to ChIPseq. Differentially accessible DNA regions between actively infected (GFP+) cells and latently infected (GFP-) cells were compared to ChIPseq datasets for RUNX1, GATA3, MYB, and CTCF [35]. For each accessible peak, a ChIPseq signal score was calculated for the TFs indicated. Box plots of the summed peak cluster ChIPseq scores are shown for peaks that exhibited significantly increased accessibility in GFP+ cells (pink) versus those that showed no significant change (blue). P values for differences between the samples was calculated by Welch two sample t-test (S8 Table; *** denotes $P < 0.0001$).

(TIF)

S6 Fig. Flow cytometry of viral gene expression after CRISPR mediated CTCF knockdown. Underlying flow cytometry and gating for Fig 7F is shown.

(TIF)

S1 Table. Enrichment of ATACseq peaks in differentially expressed genes. For each experimental comparison, the number of cellular genes associated with a significant change for its RNA transcript and/or an ATACseq peak (fold change > 1.5 , $P_{adj} < 0.1$), were calculated. Statistical enrichment of ATACseq peak changes in genes with changing transcript levels was calculated using Fisher’s exact test (lower right panel).

(DOC)

S2 Table. Genes with altered transcript and ATACseq peak levels. Genes exhibiting a significant change to both transcript level and accessibility in actively infected cells (GFP+) vs latently infected cells (GFP-) are shown (fold change > 1.5 , $P_{adj} < 0.1$). Increased RNA levels or accessibility in actively infected cells (GFP+) are designated as changing in the “up” direction.

(DOC)

S3 Table. TF motif enrichment in actively infected cells. HOMER TF motif enrichment analysis of differentially open peaks between actively infected cells (GFP+) and latently infected (GFP-) cells. Top 100 enriched motifs in peaks preferentially open in actively infected cells are shown. Target Sequences represent genomic sequences that have significantly elevated

accessibility in actively infected cells. Background sequences represent all open chromatin regions in CD4 T cells.

(DOC)

S4 Table. TF motif enrichment in latently infected cells. HOMER TF motif enrichment analysis of differentially open peaks between actively infected cells (GFP+) and latently infected (GFP-) cells. Top 100 enriched motifs in peaks preferentially open in latently infected cells are shown. Target Sequences represent genomic sequences that have significantly elevated accessibility in latently infected cells. Background Sequences represent all open chromatin regions in CD4 T cells.

(DOC)

S5 Table. HOMER motif analysis of AZD5582 stimulated latently infected CD4 T cells.

The top 50 most highly enriched TF motifs in the set of chromatin peaks that are more open after 24h AZD5582 (250nM) stimulation are shown. Target sequences represent significantly (FDR <0.1) more open chromatin regions after AZD5582 stimulation. Background sequences represent all open chromatin regions in CD4 T cells.

(DOC)

S6 Table. HOMER motif analysis of vorinostat stimulated latently infected CD4 T cells.

The top 50 most highly enriched TF motifs in the set of chromatin peaks that are more open after 24h vorinostat (250nM) stimulation are shown. Target sequences represent significantly (FDR <0.1) more open chromatin regions after vorinostat stimulation. Background sequences represent all open chromatin regions in CD4 T cells.

(DOC)

S7 Table. HOMER motif analysis of prostratin stimulated latently infected CD4 T cells.

The top 50 most highly enriched TF motifs in the set of chromatin peaks that are more open after 24h prostratin stimulation (250nM) are shown. Target sequences represent significantly (FDR <0.1) more open chromatin regions after prostratin stimulation. Background sequences represent all open chromatin regions in CD4 T cells.

(DOCX)

S8 Table. Enrichment of ATACseq peaks for ChIPseq validated transcription factor binding sites. Change in the mean of the ChIPseq TF score for significantly more accessible peaks (FC > 1.5 and FDR < 0.1, sig) vs non-significantly different (unchanged) peaks (FDR > 0.2 and any FC, nonSig). Calculated using Welch's Two Sample t-test. Box plots of distributions are shown in [S5 Fig](#). AZD = AZD5582, PROST = prostratin, VOR = vorinostat.

(DOCX)

Acknowledgments

The authors acknowledge assistance from the UNC flow cytometry core facility and the UNC Center for AIDS Research.

Author Contributions

Conceptualization: Edward P. Browne.

Data curation: Stuart R. Jefferys, Edward P. Browne.

Formal analysis: Stuart R. Jefferys, Sara R. Selitsky, Yi-Hsuan Tsai, Alisha R. Coffey, Joel Parker, Edward P. Browne.

Funding acquisition: Lindsey I. James, David M. Margolis, Edward P. Browne.

Investigation: Stuart R. Jefferys, Samuel D. Burgos, Jackson J. Peterson, Anne-Marie W. Turner, Edward P. Browne.

Writing – original draft: Stuart R. Jefferys, Joel Parker, Edward P. Browne.

Writing – review & editing: Stuart R. Jefferys, David M. Margolis, Joel Parker, Edward P. Browne.

References

1. Yoshimura K. Current status of HIV/AIDS in the ART era. *J Infect Chemother Off J Jpn Soc Chemother.* 2017; 23: 12–16. <https://doi.org/10.1016/j.jiac.2016.10.002> PMID: 27825722
2. Lederman MM, Calabrese L, Funderburg NT, Clagett B, Medvik K, Bonilla H, et al. Immunologic Failure Despite Suppressive Antiretroviral Therapy Is Related to Activation and Turnover of Memory CD4 Cells. *J Infect Dis.* 2011; 204: 1217–1226. <https://doi.org/10.1093/infdis/jir507> PMID: 21917895
3. Lederman MM, Funderburg NT, Sekaly RP, Klatt NR, Hunt PW. Residual Immune Dysregulation Syndrome in Treated HIV infection. *Adv Immunol.* 2013; 119: 51–83. <https://doi.org/10.1016/B978-0-12-407707-2.00002-3> PMID: 23886064
4. Lichtfuss GF, Hoy J, Rajasuriar R, Kramski M, Crowe SM, Lewin SR. Biomarkers of immune dysfunction following combination antiretroviral therapy for HIV infection. *Biomark Med.* 2011; 5: 171–186. <https://doi.org/10.2217/bmm.11.15> PMID: 21473720
5. Serrano-Villar S, Sainz T, Lee SA, Hunt PW, Sinclair E, Shacklett BL, et al. HIV-Infected Individuals with Low CD4/CD8 Ratio despite Effective Antiretroviral Therapy Exhibit Altered T Cell Subsets, Heightened CD8+ T Cell Activation, and Increased Risk of Non-AIDS Morbidity and Mortality. *PLoS Pathog.* 2014; 10. <https://doi.org/10.1371/journal.ppat.1004078> PMID: 24831517
6. Crooks AM, Bateson R, Cope AB, Dahl NP, Griggs MK, Kuruc JD, et al. Precise Quantitation of the Latent HIV-1 Reservoir: Implications for Eradication Strategies. *J Infect Dis.* 2015; jiv218. <https://doi.org/10.1093/infdis/jiv218> PMID: 25877550
7. Siliciano JD, Kajdas J, Finzi D, Quinn TC, Chadwick K, Margolick JB, et al. Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nat Med.* 2003; 9: 727–728. <https://doi.org/10.1038/nm880> PMID: 12754504
8. Chun T-W, Stuyver L, Mizell SB, Ehler LA, Mican JAM, Baseler M, et al. Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proc Natl Acad Sci.* 1997; 94: 13193–13197. <https://doi.org/10.1073/pnas.94.24.13193> PMID: 9371822
9. Finzi D, Hermankova M, Pierson T, Carruth LM, Buck C, Chaisson RE, et al. Identification of a Reservoir for HIV-1 in Patients on Highly Active Antiretroviral Therapy. *Science.* 1997; 278: 1295–1300. <https://doi.org/10.1126/science.278.5341.1295> PMID: 9360927
10. Pinkevych M, Cromer D, Tolstrup M, Grimm AJ, Cooper DA, Lewin SR, et al. HIV Reactivation from Latency after Treatment Interruption Occurs on Average Every 5–8 Days—Implications for HIV Remission. *PLoS Pathog.* 2015; 11. <https://doi.org/10.1371/journal.ppat.1005000> PMID: 26133551
11. Baxter AE, Niessl J, Fromentin R, Richard J, Porichis F, Charlebois R, et al. Single-Cell Characterization of Viral Translation-Competent Reservoirs in HIV-Infected Individuals. *Cell Host Microbe.* 2016 [cited 23 Aug 2016]. <https://doi.org/10.1016/j.chom.2016.07.015> PMID: 27545045
12. Lee GQ, Lichterfeld M. Diversity of HIV-1 reservoirs in CD4 T cell subsets. *Curr Opin HIV AIDS.* 2016; 11: 383–387. <https://doi.org/10.1097/COH.0000000000000281> PMID: 27023286
13. Soriano-Sarabia N, Bateson RE, Dahl NP, Crooks AM, Kuruc JD, Margolis DM, et al. Quantitation of Replication-Competent HIV-1 in Populations of Resting CD4+ T Cells. *J Virol.* 2014; 88: 14070–14077. <https://doi.org/10.1128/JVI.01900-14> PMID: 25253353
14. Chomont N, El-Far M, Ancuta P, Trautmann L, Procopio FA, Yassine-Diab B, et al. HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nat Med.* 2009; 15: 893–900. <https://doi.org/10.1038/nm.1972> PMID: 19543283
15. Maldarelli F, Wu X, Su L, Simonetti FR, Shao W, Hill S, et al. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science.* 2014; 345: 179–183. <https://doi.org/10.1126/science.1254194> PMID: 24968937
16. Reeves DB, Duke ER, Wagner TA, Palmer SE, Spivak AM, Schiffer JT. A majority of HIV persistence during antiretroviral therapy is due to infected cell proliferation. *Nat Commun.* 2018;9. <https://doi.org/10.1038/s41467-017-01881-x> PMID: 29339724

17. Dahabieh MS, Ooms M, Brumme C, Taylor J, Harrigan PR, Simon V, et al. Direct non-productive HIV-1 infection in a T-cell line is driven by cellular activation state and NFκB. *Retrovirology*. 2014; 11: 17. <https://doi.org/10.1186/1742-4690-11-17> PMID: 24502247
18. Kim YK, Mbonye U, Hokello J, Karn J. T-cell receptor signaling enhances transcriptional elongation from latent HIV proviruses by activating P-TEFb through an ERK-dependent pathway. *J Mol Biol*. 2011; 410: 896–916. <https://doi.org/10.1016/j.jmb.2011.03.054> PMID: 21763495
19. Nabel G, Baltimore D. An inducible transcription factor activates expression of human immunodeficiency virus in T cells. *Nature*. 1987; 326: 711–713. <https://doi.org/10.1038/326711a0> PMID: 3031512
20. Tyagi M, Pearson RJ, Karn J. Establishment of HIV Latency in Primary CD4+ Cells Is due to Epigenetic Transcriptional Silencing and P-TEFb Restriction. *J Virol*. 2010; 84: 6425–6437. <https://doi.org/10.1128/JVI.01519-09> PMID: 20410271
21. Razoooky BS, Pai A, Aull K, Rouzine IM, Weinberger LS. A Hardwired HIV Latency Program. *Cell*. 2015; 160: 990–1001. <https://doi.org/10.1016/j.cell.2015.02.009> PMID: 25723172
22. Archin N, Liberty A, Kashuba A, Choudhary S, Kuruc J, Crooks A, et al. Administration of vorinostat disrupts HIV-1 latency in patients on antiretroviral therapy. *Nature*. 2012; 487: 482–485. <https://doi.org/10.1038/nature11286> PMID: 22837004
23. Barton KM, Archin NM, Keedy KS, Espeseth AS, Zhang Y, Gale J, et al. Selective HDAC Inhibition for the Disruption of Latent HIV-1 Infection. *PLoS ONE*. 2014; 9. <https://doi.org/10.1371/journal.pone.0102684> PMID: 25136952
24. Chougui G, Margottin-Goguet F. HUSH, a Link Between Intrinsic Immunity and HIV Latency. *Front Microbiol*. 2019; 10. <https://doi.org/10.3389/fmicb.2019.00224> PMID: 30809215
25. Friedman J, Cho W-K, Chu CK, Keedy KS, Archin NM, Margolis DM, et al. Epigenetic Silencing of HIV-1 by the Histone H3 Lysine 27 Methyltransferase Enhancer of Zeste 2 ∇ . *J Virol*. 2011; 85: 9078–9089. <https://doi.org/10.1128/JVI.00836-11> PMID: 21715480
26. Tripathy MK, McManamy MEM, Burch BD, Archin NM, Margolis DM. 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*. 2015; 89: 8392–8405. <https://doi.org/10.1128/JVI.00572-15> PMID: 26041287
27. Mzingwane ML, Tiemessen CT. Mechanisms of HIV persistence in HIV reservoirs. *Rev Med Virol*. 2017; 27. <https://doi.org/10.1002/rmv.1924> PMID: 28128885
28. Bradley T, Ferrari G, Haynes BF, Margolis DM, Browne EP. Single-Cell Analysis of Quiescent HIV Infection Reveals Host Transcriptional Profiles that Regulate Proviral Latency. *Cell Rep*. 2018; 25: 107–117.e3. <https://doi.org/10.1016/j.celrep.2018.09.020> PMID: 30282021
29. He G, Ylisastigui L, Margolis DM. The regulation of HIV-1 gene expression: the emerging role of chromatin. *DNA Cell Biol*. 2002; 21: 697–705. <https://doi.org/10.1089/104454902760599672> PMID: 12443539
30. Strahl BD, Allis CD. The language of covalent histone modifications. *Nature*. 2000; 403: 41–45. <https://doi.org/10.1038/47412> PMID: 10638745
31. Yang H-C, Xing S, Shan L, O'Connell K, Dinoso J, Shen A, et al. Small-molecule screening using a human primary cell model of HIV latency identifies compounds that reverse latency without cellular activation. *J Clin Invest*. 2009; 119: 3473–3486. <https://doi.org/10.1172/JCI39199> PMID: 19805909
32. Shan L, Xing S, Yang H-C, Zhang H, Margolick JB, Siliciano RF. Unique characteristics of histone deacetylase inhibitors in reactivation of latent HIV-1 in Bcl-2-transduced primary resting CD4+ T cells. *J Antimicrob Chemother*. 2014; 69: 28–33. <https://doi.org/10.1093/jac/dkt338> PMID: 23999005
33. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol*. 2015; 109: 21.29.1–9. <https://doi.org/10.1002/0471142727.mb2129s109> PMID: 25559105
34. Buenrostro JD, Wu B, Littenberger UM, Ruff D, Gonzales ML, Snyder MP, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*. 2015; 523: 486–490. <https://doi.org/10.1038/nature14590> PMID: 26083756
35. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015; 12: 453–457. <https://doi.org/10.1038/nmeth.3337> PMID: 25822800
36. Battivelli E, Dahabieh MS, Abdel-Mohsen M, Svensson JP, Silva ITD, Cohn LB, et al. Chromatin Functional States Correlate with HIV Latency Reversal in Infected Primary CD4+ T Cells. *bioRxiv*. 2018; 242958. <https://doi.org/10.7554/eLife.34655> PMID: 29714165
37. Rafati H, Parra M, Hakre S, Moshkin Y, Verdin E, Mahmoudi T. Repressive LTR Nucleosome Positioning by the BAF Complex Is Required for HIV Latency. *PLoS Biol*. 2011; 9. <https://doi.org/10.1371/journal.pbio.1001206> PMID: 22140357

38. Nixon CC, Mavigner M, Sampey GC, Brooks AD, Spagnuolo RA, Irlbeck DM, et al. Systemic HIV and SIV latency reversal via non-canonical NF- κ B signalling in vivo. *Nature*. 2020; 578: 160–165. <https://doi.org/10.1038/s41586-020-1951-3> PMID: 31969707
39. Boeva V. Analysis of Genomic Sequence Motifs for Deciphering Transcription Factor Binding and Transcriptional Regulation in Eukaryotic Cells. *Front Genet*. 2016; 7. <https://doi.org/10.3389/fgene.2016.00024> PMID: 26941778
40. Duverger A, Wolschendorf F, Zhang M, Wagner F, Hatcher B, Jones J, et al. An AP-1 Binding Site in the Enhancer/Core Element of the HIV-1 Promoter Controls the Ability of HIV-1 To Establish Latent Infection. *J Virol*. 2013; 87: 2264–2277. <https://doi.org/10.1128/JVI.01594-12> PMID: 23236059
41. Roebuck KA, Gu DS, Kagnoff MF. Activating protein-1 cooperates with phorbol ester activation signals to increase HIV-1 expression. *AIDS Lond Engl*. 1996; 10: 819–826. <https://doi.org/10.1097/00002030-199607000-00004> PMID: 8828738
42. Yang X, Chen Y, Gabuzda D. ERK MAP kinase links cytokine signals to activation of latent HIV-1 infection by stimulating a cooperative interaction of AP-1 and NF- κ B. *J Biol Chem*. 1999; 274: 27981–27988. <https://doi.org/10.1074/jbc.274.39.27981> PMID: 10488148
43. Hnisz D, Weintraub AS, Day DS, Valton A-L, Bak RO, Li CH, et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science*. 2016; 351: 1454–1458. <https://doi.org/10.1126/science.aad9024> PMID: 26940867
44. Ghirlando R, Felsenfeld G. CTCF: making the right connections. *Genes Dev*. 2016; 30: 881–891. <https://doi.org/10.1101/gad.277863.116> PMID: 27083996
45. Holwerda SJB, de Laat W. CTCF: the protein, the binding partners, the binding sites and their chromatin loops. *Philos Trans R Soc B Biol Sci*. 2013; 368. <https://doi.org/10.1098/rstb.2012.0369> PMID: 23650640
46. Chang L-H, Ghosh S, Noordermeer D. TADs and Their Borders: Free Movement or Building a Wall? *J Mol Biol*. 2020; 432: 643–652. <https://doi.org/10.1016/j.jmb.2019.11.025> PMID: 31887284
47. Hansen AS, Hsieh T-HS, Cattoglio C, Pustova I, Saldaña-Meyer R, Reinberg D, et al. Distinct Classes of Chromatin Loops Revealed by Deletion of an RNA-Binding Region in CTCF. *Mol Cell*. 2019; 76: 395–411.e13. <https://doi.org/10.1016/j.molcel.2019.07.039> PMID: 31522987
48. Saldaña-Meyer R, Rodriguez-Hernaez J, Escobar T, Nishana M, Jácome-López K, Nora EP, et al. RNA Interactions Are Essential for CTCF-Mediated Genome Organization. *Mol Cell*. 2019; 76: 412–422.e5. <https://doi.org/10.1016/j.molcel.2019.08.015> PMID: 31522988
49. Chau CM, Zhang X-Y, McMahon SB, Lieberman PM. Regulation of Epstein-Barr Virus Latency Type by the Chromatin Boundary Factor CTCF. *J Virol*. 2006; 80: 5723–5732. <https://doi.org/10.1128/JVI.00025-06> PMID: 16731911
50. Washington SD, Edenfield SI, Lieux C, Watson ZL, Taasan SM, Dhummakupt A, et al. Depletion of the Insulator Protein CTCF Results in Herpes Simplex Virus 1 Reactivation In Vivo. *J Virol*. 2018; 92. <https://doi.org/10.1128/JVI.00173-18> PMID: 29514910
51. Margolis DM. How Might We Cure HIV? *Curr Infect Dis Rep*. 2014; 16: 392. <https://doi.org/10.1007/s11908-014-0392-2> PMID: 24562540
52. Mousseau G, Valente ST. Didehydro-Cortistatin A: a new player in HIV-therapy? *Expert Rev Anti Infect Ther*. 2015; 0: null. <https://doi.org/10.1586/14787210.2016.1122525> PMID: 26581953
53. Siliciano RF, Greene WC. HIV Latency. *Cold Spring Harb Perspect Med*. 2011; 1: a007096. <https://doi.org/10.1101/cshperspect.a007096> PMID: 22229121
54. Weinberger LS, Burnett JC, Toettcher JE, Arkin AP, Schaffer DV. Stochastic Gene Expression in a Lentiviral Positive-Feedback Loop: HIV-1 Tat Fluctuations Drive Phenotypic Diversity. *Cell*. 2005; 122: 169–182. <https://doi.org/10.1016/j.cell.2005.06.006> PMID: 16051143
55. Søggaard OS, Graversen ME, Leth S, Olesen R, Brinkmann CR, Nissen SK, et al. The Depsipeptide Romidepsin Reverses HIV-1 Latency In Vivo. *PLoS Pathog*. 2015; 11: e1005142. <https://doi.org/10.1371/journal.ppat.1005142> PMID: 26379282
56. Pearson R, Kim YK, Hokello J, Lassen K, Friedman J, Tyagi M, et al. Epigenetic Silencing of Human Immunodeficiency Virus (HIV) Transcription by Formation of Restrictive Chromatin Structures at the Viral Long Terminal Repeat Drives the Progressive Entry of HIV into Latency. *J Virol*. 2008; 82: 12291–12303. <https://doi.org/10.1128/JVI.01383-08> PMID: 18829756
57. Vansant G, Chen H-C, Zorita E, Trejbalová K, Miklík D, Filion G, et al. The chromatin landscape at the HIV-1 provirus integration site determines viral expression. *Nucleic Acids Res*. 2020; 48: 7801–7817. <https://doi.org/10.1093/nar/gkaa536> PMID: 32597987
58. Lucic B, Chen H-C, Kuzman M, Zorita E, Wegner J, Minneker V, et al. Spatially clustered loci with multiple enhancers are frequent targets of HIV-1 integration. *Nat Commun*. 2019; 10: 4059. <https://doi.org/10.1038/s41467-019-12046-3> PMID: 31492853

59. Wang W, Fasolino M, Cattau B, Goldman N, Kong W, Frederick MA, et al. Joint profiling of chromatin accessibility and CAR-T integration site analysis at population and single-cell levels. *Proc Natl Acad Sci*. 2020; 117: 5442–5452. <https://doi.org/10.1073/pnas.1919259117> PMID: 32094195
60. Wiegand A, Spindler J, Hong FF, Shao W, Cyktor JC, Cillo AR, et al. Single-cell analysis of HIV-1 transcriptional activity reveals expression of proviruses in expanded clones during ART. *Proc Natl Acad Sci*. 2017; 114: E3659–E3668. <https://doi.org/10.1073/pnas.1617961114> PMID: 28416661
61. Ho Y-C, Shan L, Hosmane NN, Wang J, Laskey SB, Rosenbloom DIS, et al. Replication-Competent Noninduced Proviruses in the Latent Reservoir Increase Barrier to HIV-1 Cure. *Cell*. 2013; 155: 540–551. <https://doi.org/10.1016/j.cell.2013.09.020> PMID: 24243014
62. Yang Z, Engel JD. Human T cell transcription factor GATA-3 stimulates HIV-1 expression. *Nucleic Acids Res*. 1993; 21: 2831–2836. <https://doi.org/10.1093/nar/21.12.2831> PMID: 8332492
63. Garaud S, Roufousse F, De Silva P, Gu-Trantien C, Lodewyckx J-N, Duvillier H, et al. FOXP1 is a regulator of quiescence in healthy human CD4+ T cells and is constitutively repressed in T cells from patients with lymphoproliferative disorders. *Eur J Immunol*. 2017; 47: 168–179. <https://doi.org/10.1002/eji.201646373> PMID: 27861791
64. Roux A, Leroy H, Muylder BD, Bracq L, Oussous S, Dusanter-Fourt I, et al. FOXO1 transcription factor plays a key role in T cell—HIV-1 interaction. *PLOS Pathog*. 2019; 15: e1007669. <https://doi.org/10.1371/journal.ppat.1007669> PMID: 31042779
65. Vallejo-Gracia A, Chen IP, Perrone R, Besnard E, Boehm D, Battivelli E, et al. FOXO1 promotes HIV latency by suppressing ER stress in T cells. *Nat Microbiol*. 2020. <https://doi.org/10.1038/s41564-020-0742-9> PMID: 32541947
66. Fiorito E, Sharma Y, Gilfillan S, Wang S, Singh SK, Satheesh SV, et al. CTCF modulates Estrogen Receptor function through specific chromatin and nuclear matrix interactions. *Nucleic Acids Res*. 2016; 44: 10588–10602. <https://doi.org/10.1093/nar/gkw785> PMID: 27638884
67. Del Rosario BC, Kriz AJ, Del Rosario AM, Anselmo A, Fry CJ, White FM, et al. Exploration of CTCF post-translation modifications uncovers Serine-224 phosphorylation by PLK1 at pericentric regions during the G2/M transition. Tyler JK, Engel N, editors. *eLife*. 2019; 8: e42341. <https://doi.org/10.7554/eLife.42341> PMID: 30676316
68. Wang H, Maurano MT, Qu H, Varley KE, Gertz J, Pauli F, et al. Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res*. 2012; 22: 1680–1688. <https://doi.org/10.1101/gr.136101.111> PMID: 22955980
69. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv13033997 Q-Bio*. 2013 [cited 10 Jul 2020]. Available: <http://arxiv.org/abs/1303.3997>
70. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinforma Oxf Engl*. 2009; 25: 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352> PMID: 19505943
71. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013; 14: 178–192. <https://doi.org/10.1093/bib/bbs017> PMID: 22517427
72. Lun ATL, Smyth GK. csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Res*. 2016; 44: e45. <https://doi.org/10.1093/nar/gkv1191> PMID: 26578583
73. R: The R Project for Statistical Computing. [cited 13 Nov 2020]. Available: <https://www.r-project.org/>
74. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res*. 2012; 40: 4288–4297. <https://doi.org/10.1093/nar/gks042> PMID: 22287627
75. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010; 38: 576–589. <https://doi.org/10.1016/j.molcel.2010.05.004> PMID: 20513432
76. Dobin A, Gingeras TR. Mapping RNA-seq Reads with STAR. *Curr Protoc Bioinforma*. 2015; 51: 11.14.1–11.14.19. <https://doi.org/10.1002/0471250953.bi1114s51> PMID: 26334920
77. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017; 14: 417–419. <https://doi.org/10.1038/nmeth.4197> PMID: 28263959
78. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014; 15: 550. <https://doi.org/10.1186/s13059-014-0550-8> PMID: 25516281
79. Lawrence M, Gentleman R, Carey V. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics*. 2009; 25: 1841–1842. <https://doi.org/10.1093/bioinformatics/btp328> PMID: 19468054

80. Oh SA, Seki A, Rutz S. Ribonucleoprotein Transfection for CRISPR/Cas9-Mediated Gene Knockout in Primary T Cells. *Curr Protoc Immunol*. 2019; 124: e69. <https://doi.org/10.1002/cpim.69> PMID: [30334617](https://pubmed.ncbi.nlm.nih.gov/30334617/)
81. Seki A, Rutz S. Optimized RNP transfection for highly efficient CRISPR/Cas9-mediated gene knockout in primary T cells. *J Exp Med*. 2018; 215: 985–997. <https://doi.org/10.1084/jem.20171626> PMID: [29436394](https://pubmed.ncbi.nlm.nih.gov/29436394/)