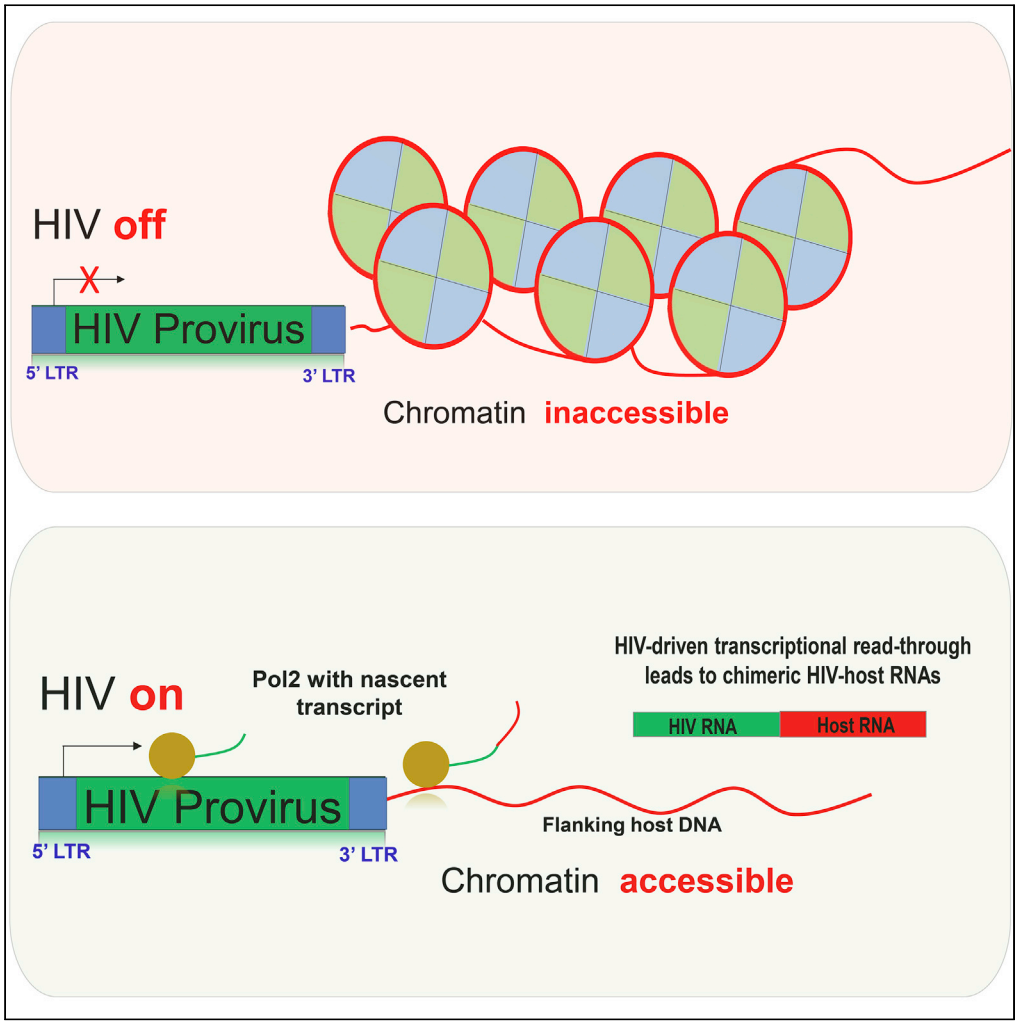
iScience

CellPress
OPEN ACCESS

Article

Activation of HIV-1 proviruses increases downstream chromatin accessibility



Raven Shah, Christian M. Gallardo, Yoonhee H. Jung, ..., Bruce E. Torbett, Philip R. Tedbury, Stefan G. Sarafianos

stefanos.sarafianos@emory.edu

Highlights

Integrative approach to profile chromatin and transcription at HIV-1 proviruses

Activation of HIV-1 transcription can alter downstream chromatin accessibility

Major chromatin reorganization is not needed for HIV integration and transcription

Long-read sequencing enables accurate detection of chimeric virus-host RNAs

Shah et al., iScience 25, 105490
December 22, 2022 © 2022
https://doi.org/10.1016/j.isci.2022.105490

iScience

CellPress
OPEN ACCESS

Article

Activation of HIV-1 proviruses increases downstream chromatin accessibility

Raven Shah,[1,2] Christian M. Gallardo,[3] Yoonhee H. Jung,[4] Ben Clock,[5] Jesse R. Dixon,[5] William M. McFadden,[1,2] Kinjal Majumder,[6] David J. Pintel,[7] Victor G. Corces,[4] Bruce E. Torbett,[3,8] Philip R. Tedbury,[1,2] and Stefan G. Sarafianos[1,2,9,*]

SUMMARY

It is unclear how the activation of HIV-1 transcription affects chromatin structure. We interrogated chromatin organization both genome-wide and nearby HIV-1 integration sites using Hi-C and ATAC-seq. In conjunction, we analyzed the transcription of the HIV-1 genome and neighboring genes. We found that long-range chromatin contacts did not differ significantly between uninfected cells and those harboring an integrated HIV-1 genome, whether the HIV-1 genome was actively transcribed or inactive. Instead, the activation of HIV-1 transcription changes chromatin accessibility immediately downstream of the provirus, demonstrating that HIV-1 can alter local cellular chromatin structure. Finally, we examined HIV-1 and neighboring host gene transcripts with long-read sequencing and found populations of chimeric RNAs both virus-to-host and host-to-virus. Thus, multiomics profiling revealed that the activation of HIV-1 transcription led to local changes in chromatin organization and altered the expression of neighboring host genes.

INTRODUCTION

HIV-1 is a retrovirus that targets immune cells including T cells,[1,2] macrophages,[3] and dendritic cells,[4] establishing permanent infection by integrating a double-stranded complementary DNA (cDNA) copy of its RNA genome into the targeted cellular chromatin.[5] Prior to integration, HIV nuclear import is likely driven by interactions with capsid (CA) core[6–8] and host factors that facilitate nuclear entry of the pre-integration complex (PIC). The PIC, containing HIV-1 integrase and the viral cDNA genome, interacts with nuclear factors such as the cellular cofactor lens epithelium-derived growth factor (LEDGF/p75), which help direct the PIC to intronic sites in genes with accessible chromatin for integration.[5,9,10] Post-integration, the virus exploits cellular transcriptional machinery to express viral mRNAs required for viral replication.

The transcription of integrated HIV-1 genomes (provirus) is regulated by viral and cellular factors.[11–16] These include- but are not limited to-the chromatin environment into which the provirus has integrated,[17] transcriptional orientation of the provirus relative to its proximal host genes,[18] and occupancy of the HIV-1 5′ LTR promoter by host transcription factors.[19] Since chromatin environment, transcriptional orientation, and accessibility are regulated by nuclear topology, the association between integration site selection and higher-order nuclear organization are of particular interest.[10,20,21] Consistent with this possibility, HIV-1 integration sites are enriched in the vicinity of cellular super-enhancers (SE), intricate chromatin structures that drive expression of critical regulatory gene networks.[22] Additionally, integration site analysis and live-cell imaging showed that HIV-1 cDNA traffics to nuclear speckles, ultimately integrating into nuclear structures dubbed Speckle-Associated Domains (SPADs).[9] SPADs and SE networks are generally found in highly structured genomic compartments with elevated transcriptional activity.[23] While there appears to be a direct relationship between HIV-1 integration preference and hierarchical nuclear compartmentalization, we do not have a clear understanding of how these higher-order chromatin structures affect HIV-1 transcription and to what extent the HIV-1 provirus alters the nuclear environment and transcriptome, globally or in the vicinity of the integration site.

Chromosome Conformation Capture (3C)-based methods have facilitated the study of the spatial organization of the genome. Based on the chemical cross-linking of distally located cellular sites that are closely associated in 3D space, these technologies have enabled us to dissect the structure of the nucleome. Among its

[1]Laboratory of Biochemical Pharmacology, Department of Pediatrics, Emory University School of Medicine, Atlanta, GA 30329, USA

[2]Children's Healthcare of Atlanta, Atlanta, GA 30329, USA

[3]Center for Immunity and Immunotherapies, Seattle Children's Research Institute, Seattle, WA 98101, USA

[4]Department of Biology, Emory University, Atlanta, GA 30329, USA

[5]Salk Institute for Biological Studies, La Jolla, CA 92037, USA

[6]Institute for Molecular Virology and McArdle Laboratory for Cancer Research, University of Wisconsin-Madison, Madison, WI 53705, USA

[7]Department of Molecular Microbiology and Immunology, Christopher S. Bond Life Sciences Center, University of Missouri School of Medicine, Columbia, MO 65211, USA

[8]Department of Pediatrics, University of Washington School of Medicine, Seattle, WA 98101, USA

[9]Lead contact

*Correspondence: stefanos.sarafianos@emory.edu

https://doi.org/10.1016/j.isci.2022.105490

many iterations, the 3C-based method, Hi-C, enables the interrogation of the global contact network of all cellular genomic sites, revealing the intricate hierarchal organization of nuclear chromatin.[24–30] Simultaneously, Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) has facilitated the genome-wide identification of open chromatin sites.[31–33] Overlaying Hi-C and ATAC-seq have revealed that the genome is segregated into two self-associating compartments (compartment A and B) with differential chromatin accessibility states. Compartment A is mainly composed of euchromatin and localized toward the nuclear interior, while compartment B is mainly composed of heterochromatin[34–36] and enriched toward the nuclear periphery.[37,38] Euchromatin is highly accessible to nuclear factors and packed with histones marked with active epigenetic modifications, promoting RNA polymerase occupancy and transcriptional activity, whereas, heterochromatin contains densely packed nucleosomes with repressive epigenetic modifications and is often transcriptionally silent.[39] The discovery of A and B compartments provided an early example of the role that the spatial position of a gene can play in the regulation of its expression.

A further level of organization is provided by Topologically Associating Domains (TADs), structural genomic units formed by the coalescence of multiple loops over a distinct section of the linear chromosome, that vary in size from ~40 kb to 1 Mb[28,40,41] and are characterized by sharp boundaries that isolate them from the adjacent TAD (32-33). In doing so, *cis*-regulatory sequences that are far apart on the chromosome may be brought close together as a part of a TAD.[42] Protein complexes at TAD boundaries insulate these domains from one another, permitting the formation of discrete epigenetic environments and providing an additional level of transcriptional control. Interphase chromatin is highly dynamic, and as the transcriptional state of TAD changes, chromatin contacts within the domain are altered, remodeling chromatin architecture.[36] The dynamic formation and dissolution of TADs is mediated by the action of CCCTC-binding factor (CTCF) in coordination with the cohesin complex.[33,41,43,44] This dynamic nature of chromatin reorganization plays an important role in gene regulation in cells. As an example, T cell activation induces large-scale nuclear reorganization, repositioning SE networks proximate to the nuclear pore and dramatically altering cellular transcriptional patterns and cellular function.[22,23]

Viral pathogens that enter the host cell's nucleus also exploit the regulatory mechanisms that maintain chromatin states and genome topology. Integration into heterochromatin mediated by knock-down of the Human Silencing Hub (HUSH) complex has been shown to repress retroviral transcription, demonstrating position-effect variegation of viral gene expression.[45–47] Viruses have been shown to also interact with cellular chromatin structure to regulate viral transcription, as well alter local cellular transcriptional patterns via virus-induced chromatin remodeling. The Epstein-Barr virus (EBV) genome preferentially associates with repressive nuclear compartments during latency and with active compartments during viral transcriptional activation, with its own transcription regulated alongside that of the host.[48] Conversely, Human T-cell leukemia virus type 1 (HTLV-1) integration directly restructures host chromatin upon integration. The HTLV-1 genome contains a CTCF binding site that drives the formation of chromatin loops and dysregulation of local host gene transcription.[49–51] Similarly, bovine leukemia virus (BLV) was shown to dysregulate host chromatin looping via the formation of virus-host chromatin loops, which influenced proviral transcriptional status and latency.[15] Through applications of a viral chromosome conformation capture assay to measure virus-host chromatin interactions, the parvovirus minute virus of mice (MVM) genome was shown to closely interact with host chromatin conformation. MVM DNA is found at previously identified TADs that are packaged in Type A chromatin and accrue DNA damage early in S phase, facilitating the establishment of MVM replication centers in their vicinity.[52–55] Some of these cellular sites of DNA damage are fragile sites where DNA breaks are generated by transcription-replication conflict and have previously been shown to serve as sites for the localization of Human Papillomavirus.[56] Although the MVM genome is bound by CTCF, this *cis*-element is involved in the processing of virally generated mRNA transcripts, consistent with newly uncovered role of CTCF in RNA processing.[52,57,58] White spot syndrome virus (WSSV) globally influences cellular chromatin structure by decreasing the compaction of chromatin through activity of viral protein 9 (VP9).[59] During influenza A virus (IAV) infection, viral NS1 protein induces cellular readthrough transcription to displace cohesin from CTCF sites on the host genome, thereby converting Type B compartments into Type A compartments.[60] Therefore, although proper host chromatin organization is essential for proper gene regulation, it is vulnerable to exploitation and manipulation by a broad spectrum of viral pathogens through diverse mechanisms.

Host chromatin influences HIV-1 transcription via epigenetic marks and integration site interactions (*trans*-acting factors and long-range chromatin interactions). Studies suggest that the chromatin functional state

(active vs. repressed) and the location of the provirus within the nucleus are important for HIV-1 gene regulation.[22,61–63] An important recent study also demonstrated that chromatin could restructure proximate to transcriptionally active HIV-1 proviruses,[64] and another previous study showed that the integrated HIV-1 genome forms a transcription-dependent gene loop structure between the 5′ LTR promoter and 3′ LTR poly(A) signal, further suggesting chromatin state and structure affect proviral transcription.[65] Many important questions remain regarding how HIV-1 directly influences structure and function of host chromatin. Unlike HTLV-1, the HIV-1 genome does not appear to contain a CTCF binding-site[50]; and, although HIV-1 integration can perturb the transcription of the cellular gene into which the provirus is integrated,[18,66,67] the impact of HIV-1 on host chromatin organization has not been directly investigated.

In this work, we have examined chromatin interactions in uninfected T cells, T cells harboring an inactive HIV-1 provirus, and T cells harboring an actively transcribing HIV-1 provirus; we have considered changes in long-range chromatin interactions globally and changes local to the HIV-1 integration site. Additionally, we interrogated chromatin accessibility around sites of HIV-1 integration via ATAC-seq. By pairing these high-resolution chromatin mapping methods with RNA-seq, we evaluated how HIV-1 affects chromatin structure and host transcription, demonstrating HIV-induced chromatin remodeling that is highly localized to neighboring regions downstream of the viral genome.

RESULTS

Description of inducible J-Lat cellular models for multiomics profiling studies

The methods used to study chromatin structure cannot be reliably applied to single cells and consequently would be difficult or impossible to apply, in combination with integration site mapping and transcription analysis, to a heterogeneous mixture of cells, such as would be produced by *de novo* infection of cells; therefore, in our studies, we required clonal cell lines with known proviral integration sites. Additionally, to compare the effect of HIV-1 integration into the transcription of the integrated genome, we required a cell system where HIV-1 transcription can be regulated. For these reasons, we selected well-established J-Lat cell lines. Clonal J-Lat models are derived from Jurkat T cells latently infected with an HIV-1 provirus carrying a long terminal repeat (LTR)-driven GFP reporter; HIV-1 transcription can be activated by treatment with stimulating agents such as TNF-α or PMA/ionomycin.[68,69] Following stimulation, J-Lat cells can be sorted into HIV-inactive (GFP-) and -active (GFP+) populations for downstream studies (cell-sorting scheme in Figure S1). To probe how HIV-1 transcriptional orientation ((5′ → 3′) or (3′ ←5′)) relative to the flanking host gene and how local nuclear environment impacts HIV-1 transcription, we used two J-Lat cell lines, both containing proviruses in canonical integration sites (intron within transcriptionally active gene)[20,62]: in the J-Lat 10.6 model, the provirus is integrated in the opposite orientation relative to the host gene (*SEC16A*); in the J-Lat 8.4 model, the provirus and the host gene (*FUBP1*) are in the same transcriptional orientation (Figures 1A and 1B).[68,70] Assessment of occupancy of active epigenetic markers (H3K27ac, Pol2, BRD4) and presence of higher-order chromatin structures (SPADs and super-enhancers) near these respective J-Lat integration sites was performed utilizing publicly available NGS datasets of Jurkat T cells (ATAC- and ChIP-seq). In addition, we overlaid predicted Jurkat super-enhancer sites and SPADs found in K562 cells (See materials availability and STAR Methods). Our analysis found that there is a high enrichment of active epigenetic markers, putative SE networks, and SPADs within 500 kilobases (kb) of the J-Lat 10.6 provirus. These findings suggested that within this model, HIV-1 has integrated into a very transcriptionally active environment, yet HIV-1 remains transcriptionally repressed. Markers associated with active chromatin are less abundant within 500 kb of the J-Lat 8.4 provirus, suggesting that this provirus may be in a less transcriptionally active environment (Figure 1C). In the case of these two models, local proviral epigenetic landscape was a good predictor of HIV-1 transcriptional competency, as the J-Lat 10.6 cells not only activated more frequently following TNF-α stimulation but also demonstrated significantly higher levels of transcription (>10-fold) and GFP output compared to the J-Lat 8.4 model (Figure 1D). These models provide identical HIV-1 proviruses in two distinct chromatin environments for our downstream studies.

Hi-C profiling to map the chromatin contact network of HIV-infected cells

To evaluate the impact of proviral transcription on higher-order chromatin structure, we performed Hi-C assays in wild-type (WT) Jurkat and J-Lat model T-cell lines. First, we characterized the native local 3D genome folding in the Jurkat cell line (Figure 2A and 2B). We obtained ~680 million contacts in the WT Jurkat cells, sufficient for analysis of the 3D chromatin folding at ~5 kb resolution. We then examined the chromatin interaction profiles in the vicinity of the two proviral integration site-harboring genes in

A

Jurkat T cell derived J Lat clones

J-Lat clone	Chromosome[a]	Position (bp)[b]	Host Gene[c]	Host Gene Orientation[d]	HIV Orientation[e]
10.6	9	136,468,579	SEC16A	-	+
8.4	1	77,946,384	FUBP1	-	-

[a] Chromosome where HIV provirus integrated
[b] First human genome nucleotide following 3'-end LTR
[c] Cellular gene where HIV integrated
[d-e] Transcription direction of the host or HIV gene

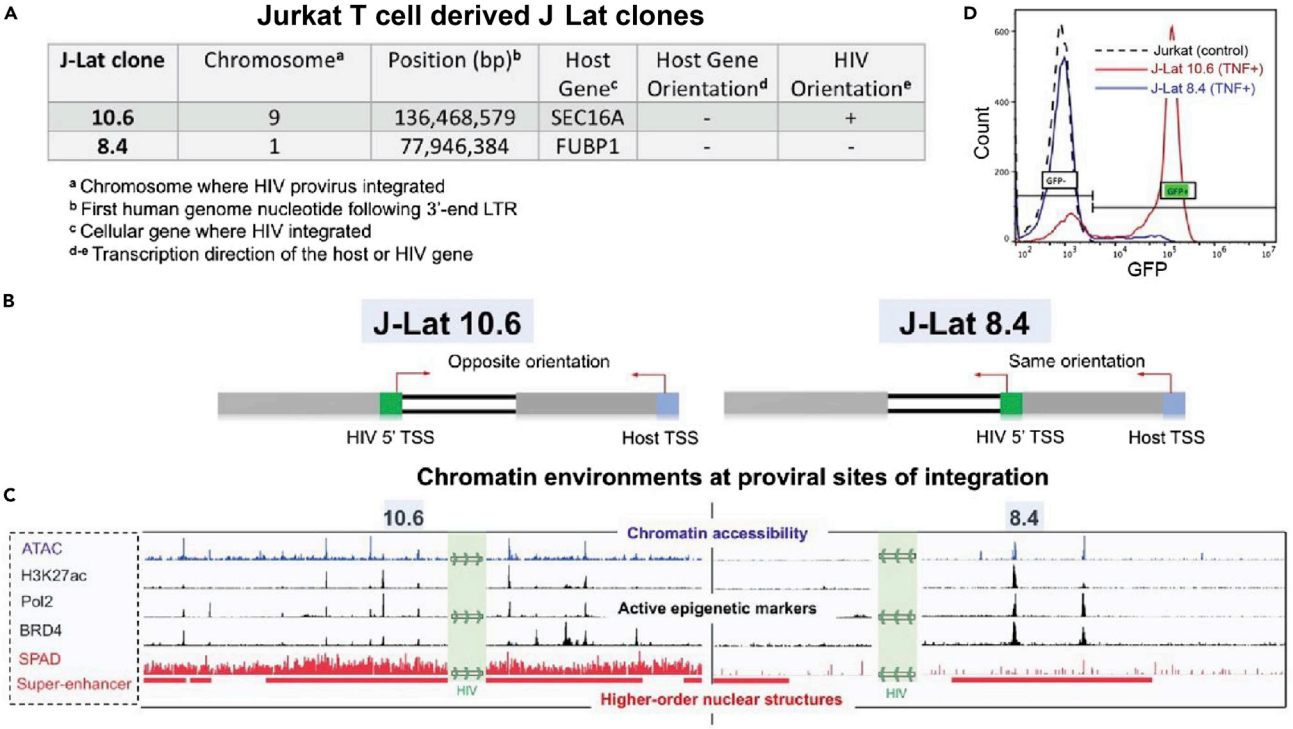


Figure 1. Variability in chromatin environments at HIV-1 sites of integration

(A) Table adapted from Symons et al. 2018,[70] displaying HIV-1 site of integration in J-Lat 10.6 and 8.4 models.

(B) Schematic depicting the orientation of HIV-1 provirus and respective host gene in J-Lat 10.6 and 8.4 models.

(C) Epigenomic profile of chromatin environment in WT Jurkat cells within 500 kb of HIV-1 integration sites in J-Lat 10.6 and J-Lat 8.4 models. ATAC-seq (blue; Jurkat T cells), ChIP-seq against H3K27ac (Jurkat T cells), Pol2 (Jurkat T cells), and BRD4 (black tracks; Jurkat T cells), and speckle-associated domains (SPADs) in K562 cells and predicted super-enhancers in Jurkat T cells (both red tracks). Accession numbers of these sequencing datasets can be found in materials availability and STAR Methods. HIV-1 marker (green) denoting the exact sites of integration is not to scale. Sequencing tracks are visualized using Integrated Genomics Viewer (IGV).[71]

(D) Flow cytometric analysis of J-Lat activation potentials. Cells are treated with TNF-α. HIV-active cells are GFP+. Jurkat T cells (no reporter provirus; black) serve as a gating control.

J-Lat 10.6 and 8.4 cell lines (*SEC16A* and *FUBP1*, respectively). In both cases, in WT Jurkat T cells (absence of virus), these genes are located within TADs. There are chromatin looping events several kilobases (kb) from HIV-1 integration sites. In the J-Lat 10.6 cells, this is a loop surrounding the *NOTCH1* gene (>10 kb from HIV-1), while in the 8.4 cells, this is a loop surrounding the *PTGFR* gene (>500 kb from HIV-1).

Next, to examine the impact of proviral integration and transcription on 3D chromatin folding in the nucleus overall and in the vicinity of the provirus, we performed Hi-C in the J-Lat 10.6 and 8.4 cell lines under untreated conditions or following the activation of HIV-1 transcription with TNF-α (Figure 2C). For the J-Lat 10.6 cell line, we obtained 823 and 641 million contacts in the untreated and TNF-α treated conditions, respectively. J-Lat cells are sorted post-treatment to isolate HIV-inactive and -active cell populations. In Figure S2, we show that J-Lat cells that are treated with TNF-α, but do not activate, still undergo canonical TNF-signaling with an increase in expression of the NF-κB subunits. All cell populations, therefore, appear to have responded to TNF-stimulus, albeit not all exhibit HIV-1 activation. In the J-Lat 8.4 cell line, we obtained 578 and 549 million contacts for the untreated and treated conditions, respectively. Mapped reads and quality control statistics of all Hi-C libraries for this study are shown in Figure S3. Examining the patterns of chromosome folding at the sites of integration we observe only modest differences in chromatin structure in the context of integration or proviral activation (Figure 2D). In both the 10.6 and 8.4 cell lines, there were no major changes in chromatin interaction frequency at either the sites of integration, or in the downstream loops proximal to each gene. We do observe subtle differences at specific loci. For example, the loop downstream of the *SEC16A* integration site on chromosome 9 in the J-Lat 10.6 cell line shows modestly reduced chromatin interaction frequency, though the differences are not significant (Figure 2E).

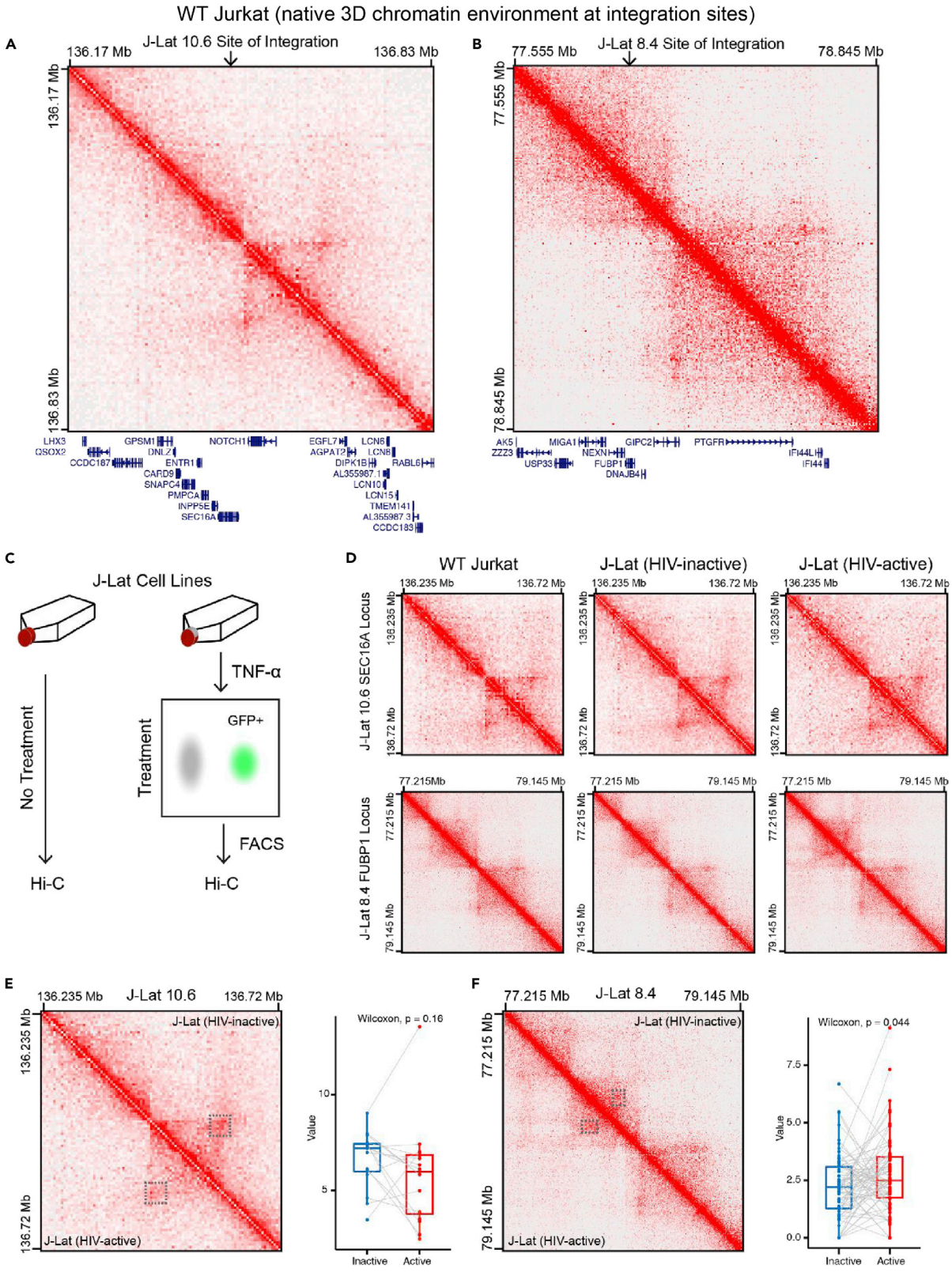

WT Jurkat (native 3D chromatin environment at integration sites)
A
136.17 Mb
J-Lat 10.6 Site of Integration
136.83 Mb
136.17 Mb
136.83 Mb
LHX3
QSOX2
GPSM1
DNLZ
CCDC187
ENTR1
CARD9
SNAPC4
PMPCA
INPPSE
SEC16A
NOTCH1
EGFL7
AGPAT2
LCN6
LCN8
DIPK1B
AL355987.1
LCN10
LCN15
TMEM141
AL355987.3
CCDC183
RABL6
B
77.555 Mb
J-Lat 8.4 Site of Integration
78.845 Mb
77.555 Mb
78.845 Mb
AK5
ZZZ3
USP33
MIGA1
NEXN
FUBP1
DNAJB4
GIPC2
PTGFR
IFI44L
IFI44
C
J-Lat Cell Lines
TNF-α
No Treatment
Treatment
GFP+
FACS
Hi-C
Hi-C
D
WT Jurkat
J-Lat (HIV-inactive)
J-Lat (HIV-active)
J-Lat 10.6 SEC16A Locus
J-Lat 8.4 FUBP1 Locus
136.235 Mb
136.72 Mb
77.215Mb
79.145 Mb
E
136.235 Mb
J-Lat 10.6
136.72 Mb
J-Lat (HIV-inactive)
J-Lat (HIV-active)
Wilcoxon, p = 0.16
Value
Inactive
Active
F
77.215 Mb
J-Lat 8.4
79.145 Mb
J-Lat (HIV-inactive)
J-Lat (HIV-active)
Wilcoxon, p = 0.044
Value
Inactive
Active

Figure 2. Minimal changes in integration site higher-order chromatin structure upon HIV-1 transcriptional activation

(A) Hi-C interaction map in wild-type Jurkat T-cells of the region on chr9 where the integration site is located in the J-Lat 10.6 cell line. The arrow indicates the site of integration, within the *SEC16A* gene. Below the heatmap is the location of genes at the locus, including *SEC16A*.

(B) Similar as in panel A but depicting the region on chr1 where the integration site is located in the J-Lat 8.4 model within the *FUBP1* gene.

(C) Schematic showing the experimental design of Hi-C in the context of HIV-1 transcriptional induction. TNF-α is added for 24 h to the J-Lat model lines. GFP + cells that have successfully activated and are then sorted and processed for Hi-C. Untreated J-Lat cells are used as controls.

(D) Hi-C heat maps of the integration sites in WT Jurkat cells (left column), untreated J-Lat lines (middle column), and HIV-active (right column), for the J-Lat 10.6 sample (top row) and the J-Lat 8.4 sample (bottom row).

(E) Comparison of the *SEC16A* locus in the J-Lat 10.6 cell line. The heatmap shows the untreated conditions in the upper right-hand half of the heatmap, and the activated population in the lower left-hand. Dashed lines mark a loop downstream of the *SEC16A* locus that surrounds the *NOTCH1* gene. The plot on the right shows the quantification of the observed/expected Hi-C interaction frequencies for the HIV-inactive (blue) and -activated (yellow) conditions of pixels within the dashed line. The loop generally is weakened upon HIV activation, but the results do not reach a threshold for statistical significance ($p = 0.16$, Wilcoxon).

(F) Comparison of the *FUBP1* locus in the J-Lat 8.4 cell line. The heatmap shows the untreated conditions in the upper right-hand half of the heatmap, and the activated cells in the lower left-hand. Dashed lines mark a region of increased interactions between the *FUBP1* gene and a region upstream. The plot on the right shows the quantification of the observed/expected Hi-C interaction frequencies for the HIV-inactive (blue) and -active (yellow) conditions of pixels within the dashed line ($p = 0.0.44$, Wilcoxon).

In addition, at the *FUBP1* site on chromosome 1, the *FUBP1* gene does show increased interactions with an upstream region near the *MIGA1* gene (Figure 2F), but these differences are modest and there are no other notable changes in chromatin interaction frequency at the *FUBP1* locus. Reconstruction of compartments A or B via eigenvector analysis revealed the provirus in both J-Lat models in nuclear Compartment A in both inactive and active transcriptional states; the lack of change in the nuclear compartment is consistent with minimal changes in the chromatin interaction network around HIV-1 integration sites. In summary, 3D genome organization was preserved following HIV-1 integration or proviral transcriptional activation. In addition to the conservation of the 3D chromatin organization around HIV-1 proviruses, we do not observe any significant global changes in the chromatin architecture of HIV-infected cells relative to the uninfected WT Jurkat T cells.

Profiling chromatin accessibility and transcription at HIV-1 integration sites reveal a unique chromatin signature associated with active proviruses

To understand how HIV-1 integration and transcription affect the functional state of chromatin around integration sites, we applied ATAC-seq to our WT Jurkat and J-Lat T cell models pre- and post-activation of transcription. ATAC-seq reads displayed in Figure 3 were filtered for shorter reads (1-115 bp), as performed previously,[33] to improve the resolution of ATAC reads at HIV-1 integration sites. The majority of these short reads map to gene transcriptional start sites. Read distribution is shown in Figure S3. ATAC-seq read density is associated with open chromatin, as such, active promoters typically generate distinct peaks, along with corresponding ChIP-seq read density peaks for epigenetic markers of active chromatin, such as H3K27ac, BRD4, and Pol2; accordingly, markers of active transcription were found associated with the two host promoters (*INPP5E* and *SEC16A*) either side of the HIV-1 provirus in the J-Lat 10.6 model (Figure 3A). Following the activation of proviral transcription in the J-Lat 10.6 model, we saw both the expected increased chromatin openness (ATAC-seq) at the 5′ LTR and an overall increase in chromatin accessibility extending through the HIV-1 genome and over ten kilobase pairs into the host sequence downstream of the 3′ LTR; the most prominent changes in chromatin openness were in the 3′ LTR and approximately one kilobase downstream of the 3′ LTR (Figures 3A-3C). In the J-Lat 8.4 model, the transcription-induced changes in chromatin openness were less dramatic. There was little to no change in ATAC-seq read density observed within the proviral chromatin, even at the 5′ LTR; however, we observed a modest increase in chromatin accessibility downstream of the 3′ LTR (Figures 3C and S6). In control Jurkat T cells, lacking an HIV-1 provirus, no equivalent increase in chromatin accessibility was observed at the locus corresponding to the region downstream of the 3′ LTR in J-Lat 10.6 or 8.4 models following treatment with TNF-α (Figures 3B and S6; zoomed out view of J-Lat 8.4 chromatin (ATAC-seq) and transcriptional environment), suggesting that phenotype is associated with HIV-1 transcription. Quantification of normalized reads at ATAC peaks at nearby host gene promoters, the HIV-1 provirus, and at the junction between HIV-1 and the host genome is shown in Figures 3D and 3E, further demonstrating that the most prominent change in chromatin accessibility following HIV-1 activation is localized downstream of the HIV-1 genome. This increased chromatin accessibility was found downstream of the HIV-1 genome in both J-Lat models but is not typically associated with the transcription of host genes. Chromatin accessibility of constitutively active house-keeping genes such as GAPDH reveals the most accessible chromatin concentrated at the 5′ gene promoter and not the 3′ end of the gene (Figure S7). The increase in chromatin accessibility

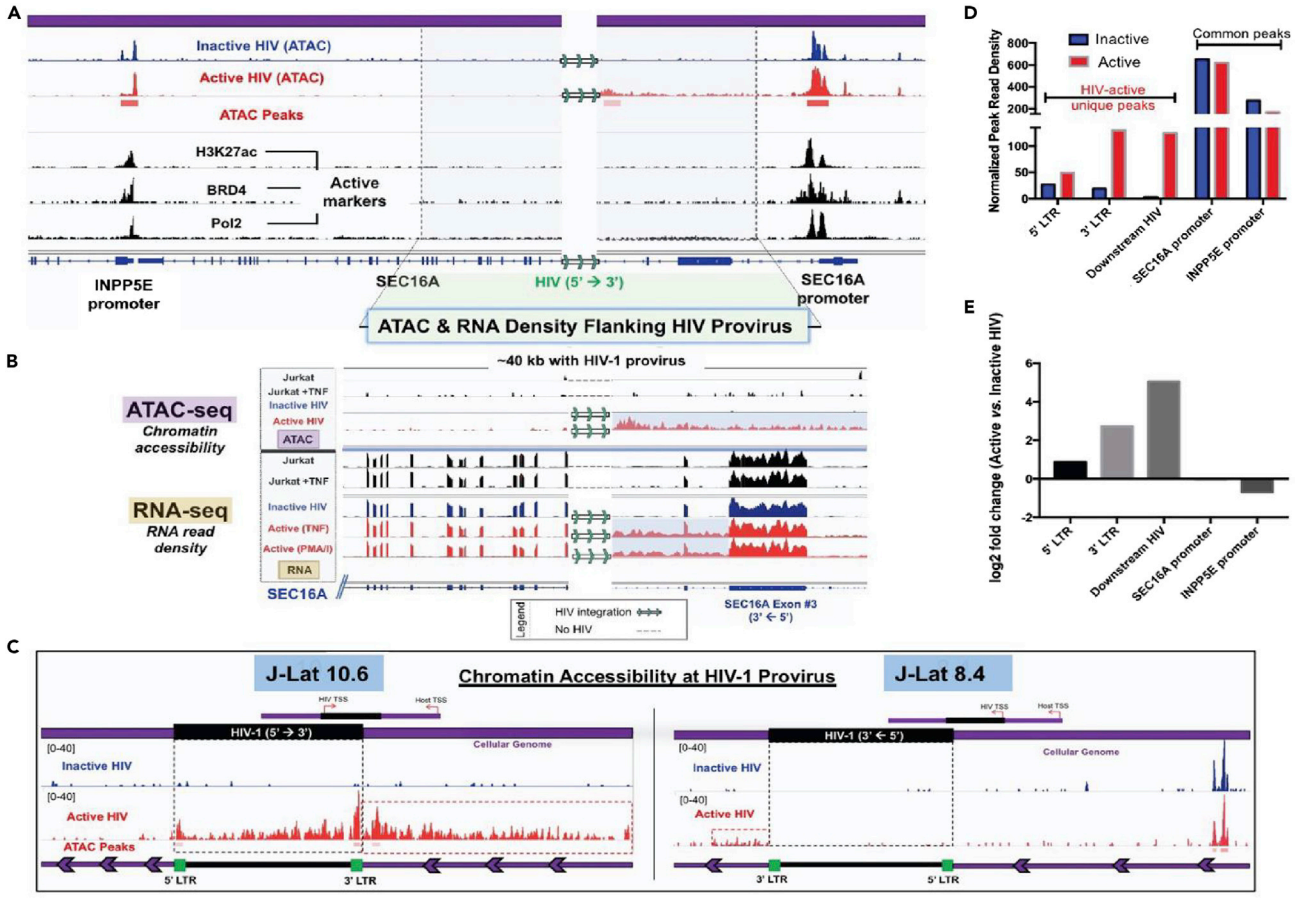


Figure 3. ATAC- & RNA-seq read density at HIV-1 integration sites to profile local HIV-host chromatin environment and transcription

(A) Representative ATAC-seq tracks of inactive (blue) and active (red) populations of J-Lat 10.6 cells. ChIP-seq of Jurkat T cells (black) against active enhancer and promoter marks (H3K27ac, Pol2, and BRD4) are overlaid to demarcate close host promoters (*INPP5E* & *SEC16A*).

(B) Close-up of the highlighted region in panel (A), visualizing ATAC-seq and RNA-seq (Illumina) read density of the flanking host genome (reads aligned to a human reference genome (hg38)). Sequencing tracks to the left are upstream of HIV-1, flanking the 5′ LTR, and tracks to the right are downstream. The highlighted regions in ATAC and RNA density demarcate regions of increased read pileup or density directly downstream of the activated provirus. The sequencing tracks in black are of uninfected Jurkat T cells +/− TNF-α to visualize what the native chromatin and transcriptional state is in the absence of the provirus. The host gene, *SEC16A*, has a 3′ ← 5′ orientation, whereas HIV-1 has the opposite 5′ → 3′ transcriptional orientation.

(C) ATAC read density at the proviral site of integration in the J-Lat 10.6 and 8.4 models. Reads were mapped to a custom reference genome that includes the HIV-1 proviral genome. The red dotted boxes highlight changes in ATAC density downstream from the HIV 3′ LTR. HIV-1 and host transcriptional start sites (TSS) are shown to bring attention to the transcriptional orientation of HIV-1 and the respective host genes in both J-Lat models.

(D and E) Quantification of differential ATAC read density at provirus, flanking cellular genome, and proximal host promoters (*SEC16A* and *INPP5E* 5′ promoters) in the J-Lat 10.6 cells (inactive vs. active states). Differential peak analysis is performed from a merged dataset from two biological replicates for each condition. Common peaks across the two biological replicates were selected and used to normalize the dataset using MAnorm.[72] Normalized peaks from the merged datasets were used for differential peak analysis. All ATAC- and RNA-seq experiments were performed with two biological replicates. ATAC-seq replicates for J-Lat 10.6 cells are shown in Figure S5.

downstream of actively transcribing HIV-1 chromatin appears to be a distinctive feature and not observed with highly and or inducibly expressed cellular genes.

We next applied RNA-seq to measure the transcription of viral and host genes before and after the activation of HIV-1 transcription. Comparing activated and inactive cells in both the J-Lat 10.6 and 8.4 models, we observed an increase in read density downstream of the 3′ LTR in both models following HIV-1 activation (Figure 3B in J-Lat 10.6 and S6 in J-Lat 8.4). The strandedness or polarity of the host RNA fragments corresponding to this flanking region are in the same transcriptional orientation as the respective HIV-1 genome (both 3′ ← 5′), consistent with prior reports that HIV-driven transcription can run into the cellular

genome.[18,66,67] In Jurkat T cells, we found that there was no difference in gene expression +/− TNF-α at the *SEC16A* and *FUBP1* loci. In addition, there is no change in chromatin accessibility within the regulatory regions of nearby host genes, further suggesting changes in chromatin organization are localized to the provirus and the region directly downstream of the 3′ LTR following HIV-1 transcriptional activation. Based on the comparison of J-Lat 10.6 and 8.4 cells to WT Jurkat T cells, and of the provirus to host genes either at nearby loci, exhibiting similar expression levels or that are induced by TNF-α treatment, run-on transcription is more commonly a feature of HIV-1 transcription than of host gene transcription. We observe similar levels of HIV-driven read-through in J-Lat cells treated with either TNF-α or PMA/I (Figure 3B). In Figure S8, we show the read pile-up from RNA-seq at TNF-responsive genes and genes that are highly expressed in our J-Lat models. In these representative sequencing tracks, transcriptional read-through is not detectable or thus does not occur or occurs at very low frequency. We performed similar analysis quantifying read density downstream of the 3′ UTR of the top quartile of highly expressed genes in our RNA-seq analysis and did not find evidence of read-through, suggesting that the transcription read-through present at HIV-1 integration sites is a signature of HIV-1.

Proviral transcriptional read-through observed regardless of the orientation of HIV-1 relative to the flanking host gene

To visualize transcriptional read-through at HIV-host intergenic boundaries, we applied stranded single-molecule RNA FISH (smRNA FISH), monitoring the expression of positive (+) and antisense (−) HIV-1 RNA[73,74] with or without activation. Imaging inactive J-Lat 10.6 cells revealed antisense HIV-1 (−) RNA (Figures 4A and S9). The HIV-1 (−) RNA was exclusively nuclear, suggesting it cannot be translated into HIV-1 antisense protein (ASP).[75–77] Following the activation of the J-Lat 10.6 cells, we observed HIV-1 (+) RNA as expected, and the loss of most of the HIV-1 (−) RNA staining. As the DNA probes can anneal to complementary RNA and DNA, we confirmed the staining was predominantly RNA by RNase treatment. This removed the entire (+) RNA signal and left just a single HIV-1 DNA focus in the cell, presumed to be the integrated provirus (Figure 4A). We found similar results in J-Lat 8.4, albeit the HIV-1 (+) RNA was present prior to activation (Figures 4A and S9). Note, however, that although there was HIV-1 (+) RNA in the cytoplasm of J-Lat 8.4 cells, there was no detectable GFP or HIV-1 protein expression, suggesting that the RNA was not capable of being translated and did not reflect authentic HIV-1 genomic RNA or mRNA.[78] Since we found HIV-1 (−) RNA in the inactive J-Lat 10.6 cells, where the provirus and *SEC16A* are in opposite orientations, and found HIV-1 (+) RNA in the inactive J-Lat 8.4 cells, where the provirus and *FUBP1* are in the same orientation, we suspect the HIV-1 RNA staining represents chimeric RNAs that are the by-product of host-driven transcription, originating at the *SEC16A* or *FUBP1* promoter (in J-Lat 10.6 and J-Lat 8.4, respectively) continuing through the HIV-1 genome. A similar observation of host-driven transcriptional read-through has been reported previously, where "transcriptional interference" was proposed to suppress proviral transcription via the convergence of Pol2 originating at the host promoter with Pol2 originating at the HIV-1 5′ LTR.[18] The authors used semi-quantitative RT-qPCR assays to quantify differential transcriptional read-through products in these studies using clonal J-Lat systems.[18]

To confirm the presence of the virus-host chimeric RNAs suggested by smRNA FISH imaging, we used long-read Nanopore sequencing to identify hybrid transcripts and changes in viral and host RNA splicing.[79,80] In the inactive J-Lat 8.4 cells, we found host-driven chimeric transcripts consistent with our smFISH data (Figure 4B). Approximately 1-5% of the total HIV-1 transcripts sequenced from the activated J-Lat 10.6 and 8.4 models were chimeric RNAs (Figure 4B). A schematic of transcriptional read-through occurring at HIV-host gene boundaries is depicted in Figure 4C. In both cases of HIV- or host-driven transcriptional read-through, we can capture these complex virus-host hybrid RNAs. Long-read Nanopore sequencing enabled the identification of the transcriptional start and end sites of the chimeric transcripts in the 8.4 model for both transcriptionally active and inactive cells. In the inactive 8.4 cells, the chimeric RNAs were originating from the flanking host promoter (*FUBP1*). In active J-Lat 8.4 cells, we saw that transcription originated at the HIV-1 5′ LTR and terminated at the 3′ UTR and poly-adenylation (poly(A)) site of the *FUBP1* gene (Figure 5). We found that in both J-Lat models, activation of proviral transcription did not affect the overall abundance of local host gene transcripts (Figure 6). However, isoform analysis comparing *SEC16A* and *FUBP1* transcripts in the presence and absence of HIV-1 integration identified a subset of transcripts that appear to have prematurely terminated when both the provirus and host gene are in the same orientation; the HIV-1 provirus is located in an intron of *FUBP1* and is usually removed when the intron is spliced out of the pre-mRNA, however, in some instances transcription appears to have ended when the

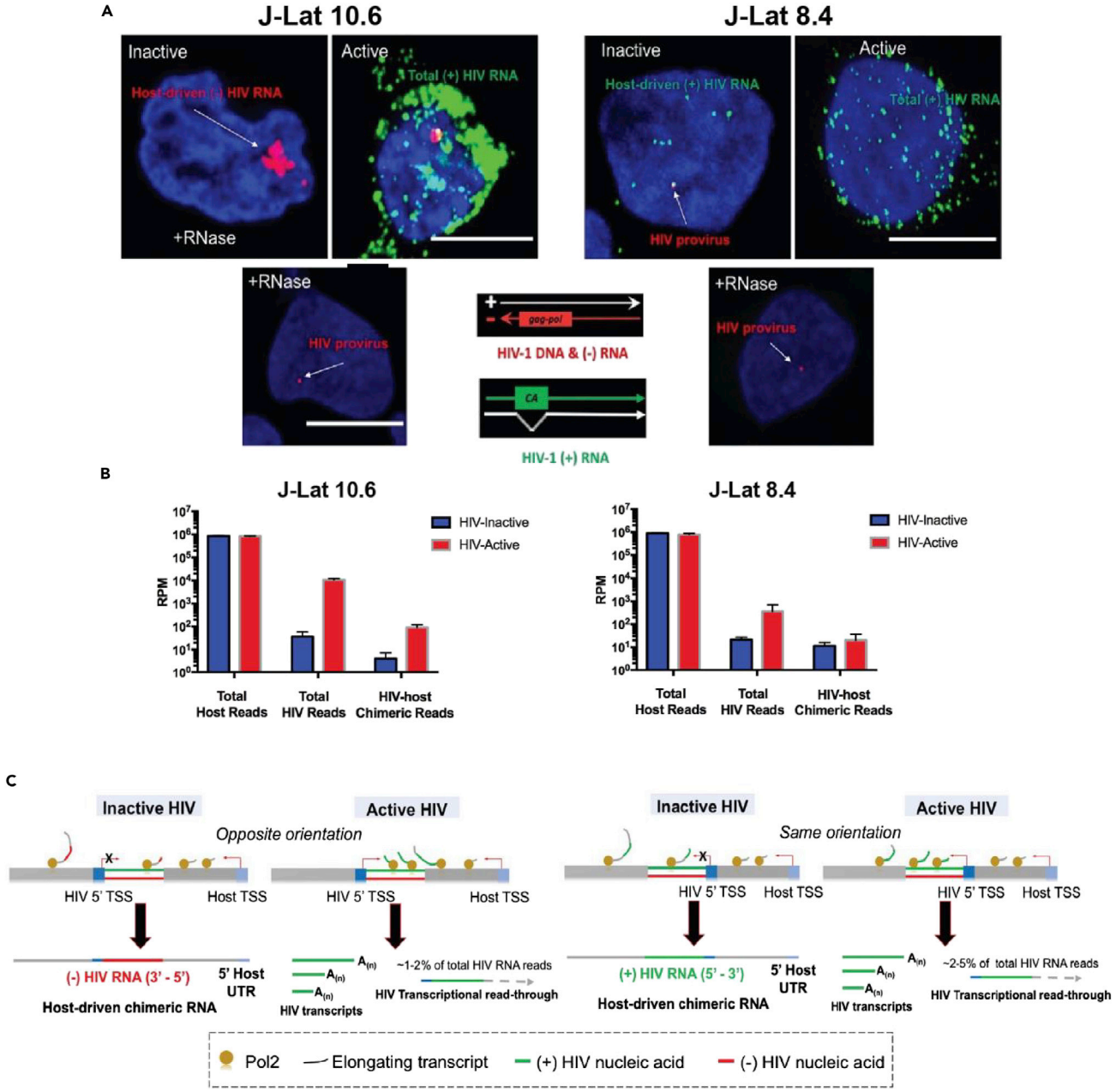


Figure 4. smFISH to visualize transcriptional read-through at HIV-host genomic boundaries

(A) Visualization of HIV-1 (+) RNA, and (–) RNA via single-molecule FISH (smFISH). The oligonucleotide hybridization probe shown in red targets template HIV-1 cDNA (3′ ← 5′) and antisense (–) HIV-1 transcripts. The probe depicted in green is designed to target positive-sense (+) HIV-1 RNA. Both HIV-1 transcriptional patterns for HIV-inactive and -active cells is shown for J-Lat 10.6 and 8.4 models. RNase-treatment control. Scale bars represent 10 μm.

(B) Normalized read counts of total HIV and chimeric HIV-host RNAs using long-read Nanopore sequencing. Normalized read counts were averaged across four biological replicates/condition (HIV-inactive and -active) for both the J-Lat 10.6 and 8.4 cells. Chimeric RNAs were defined by identifying "multi-mappers," reads that align to both hg38 and custom HIV-1 reference genome.

(C) Schematic of transcriptional read-through at HIV-host junctions. The red nascent transcript in the J-Lat 10.6 inactive cells indicates the expression of HIV-1 (–) RNA driven by the flanking host promoter. Similarly, in the inactive 8.4 cells, HIV-1 (+) RNA is expressed with transcription originating at the host promoter and Pol2 running into the HIV-1 genome. In the active J-Lat cells, (+) RNA is expressed, and both canonical poly(A) transcripts and chimeric HIV-host RNAs are generated.

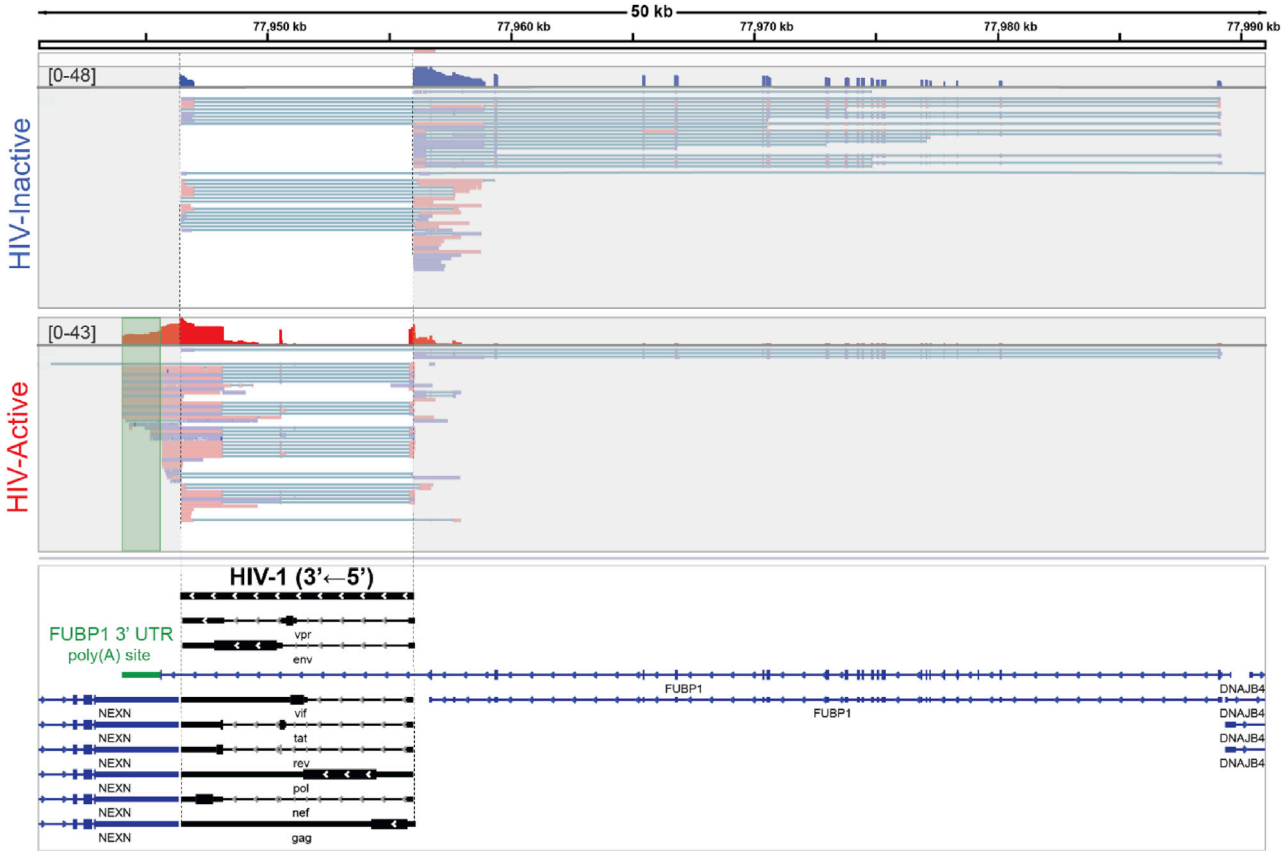


Figure 5. Long-read Nanopore RNA-sequencing of J-Lat 8.4 cells enables identification and characterization of long chimeric HIV-host RNA isoforms

Chimeric HIV-host RNA reads were filtered from the total RNA reads and displayed above. Chimeric read pile-up from HIV-inactive (blue) and -active (red) cells are displayed. Under the respective histograms, individual mapped reads are shown. In the HIV-active cell population (red), HIV-driven transcriptional read-through is highlighted by the green box (*FUPB1* 3′ UTR/poly(A) site). HIV-1 and cellular transcript variants are shown below the mapped sequencing reads (HIV-1 = black, *FUBP1* = blue). The HIV-1 and *FUBP1* sequences are in the 3′ ← 5′ transcriptional orientation.

RNA Pol2 complex encountered termination sites within the HIV-1 genome, preventing the completion of the *FUBP1* transcript (Figure 5, S6, and S11B; all in J-Lat 8.4 model).[83,84] Similarly, intergenic splicing appears to have generated chimeric transcripts where an *FUBP1* exon splice donor combined with an acceptor sequence in the HIV-1 genome and again transcription prematurely terminated within the HIV genome (Figure 5), such transcripts are consistent with previous reports.[66] This effect is also apparent using Illumina-based RNA-seq, where expression of full-length *FUBP1* in our J-Lat 8.4 is significantly attenuated in both HIV-inactive and -active cells relative to uninfected Jurkat T cells (Figure S6). There is little to no detectable *FUBP1* exon expression downstream of the HIV-1 provirus in the J-Lat 8.4 model relative to uninfected Jurkat T cells, demonstrating that HIV-1 can function as a "gene trap," interfering with expression of a cellular gene if HIV-1 and the host gene into which it integrated are in the same orientation. Accordingly, this effect was not observed in the J-Lat 10.6 model (Figures 3B and S11A), where HIV-1 and the host gene (*SEC16A*) are in the opposite orientation. *CDK9* or cyclin-dependent kinase 9 levels, a factor critical for RNA Pol2 transcriptional initiation, elongation, and termination,[81] are also similar across conditions, suggesting *CDK9* is not a limiting factor for proviral transcriptional activation in these models. While our analyses support the notion of HIV- and host-driven transcriptional read-through at the HIV-host boundary, as previously reported,[18,66] our findings indicate that HIV-1 integration and proviral activation only modestly alter local host RNA splicing patterns and transcriptional output (Figures S11 and 6, respectively).

DISCUSSION

The HIV-1 genome integrates into that of its host and, from that point on, transcription of viral RNA is carried out by host transcriptional machinery. The nucleus is a heterogeneous environment, and it is

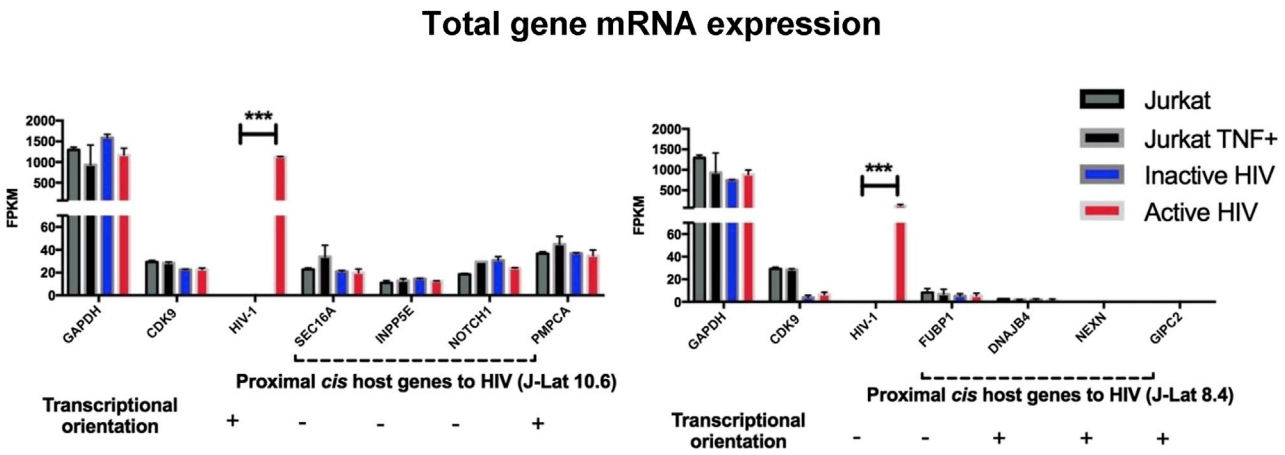


Figure 6. Total mRNA reads for genes proximate to HIV-1 integration sites

mRNA reads for HIV-1 and cellular genes (GAPDH and genes within 100 kb of J-Lat 10.6 (*SEC16A*, *INPP5E*, *NOTCH1*, *PMPCA*) or 8.4 (*FUBP1*, *DNAJB4*, *NEXN*, *GIPC2*) integration sites). *CDK9* or cyclin-dependent kinase 9 is critical for Pol2 transcription, initiation, elongation, and termination.[81] Post-read alignment, assembled mRNA transcripts were selected vs. total reads by cross-referencing to hg38 (human) and custom HIV-1 annotation files (See STAR Methods) to generate read count matrices for protein-coding transcripts and lncRNAs. Jurkat T cells +/− TNF (no HIV-1 integration; black) were also assessed to collectively evaluate the role HIV-1 integration, TNF stimulation, and HIV-1 transcriptional state has on the local cellular RNA landscape. Differential analysis of normalized mRNA read-counts (FPKM) was performed using DESeq2.[82] *** = p value $\leq$ 0.001.

increasingly understood that the architecture and organization of chromatin is a key factor in the regulation of host gene expression. Thus, chromatin organization is a significant factor when attempting to understand both regulation of HIV-1 transcription at different sites in the genome and the potential influence of the HIV-1 provirus on the chromatin and genes in the vicinity of the integration site; however, few studies to date have addressed these relationships. In our work, we have investigated the nuclear environment around two distinct proviral integration sites, characterizing the epigenome and transcriptome at these sites in the presence and absence of the provirus, and addressing the impact of viral transcription on global and local chromatin structure.

Integrative profiling of chromatin organization and RNA transcription at HIV-1 integration sites allowed us to determine how HIV-1 can influence the cellular chromatin environment and identify signatures associated with transcriptionally active HIV-1 proviruses and the changes induced in host chromatin by the virus. Previous studies have demonstrated that another retrovirus, HTLV-1, can alter cellular chromatin structure following integration via induction of CTCF-mediated looping, disrupting local transcriptional patterns.[49–51] The HTLV-1 genome contains a CTCF binding site that mediates long-range chromatin interactions with the host genome, leading to significant chromatin remodeling. By contrast, HIV-1 does not contain an obvious CTCF binding site, and our Hi-C analysis found that, unlike HTLV-1, HIV-1 does not induce large-scale genome reorganization either globally or local to the integration site (Figure 2). These findings suggested that major chromatin restructuring is not associated with HIV-1 integration. Interestingly, analysis of chromatin accessibility revealed that the activation of HIV-1 transcription increased chromatin accessibility downstream of the HIV-1 genome, extending 1-10 kb into the cellular chromatin in both J-Lat models. Since this increased chromatin accessibility was not observed in the absence of HIV-1 integration (WT Jurkat T cells +/− TNF-α), with highly expressed or induced cellular genes, our results suggest that this is an HIV-induced effect. However, we cannot conclude whether the effector is an HIV-1 protein, RNA, or genome motif. Additionally, although the chromatin opening was more pronounced in the J-Lat 10.6 model than J-Lat 8.4, with only two models to compare, we cannot determine whether the differences in chromatin opening are a consequence of the level of proviral transcription or the local chromatin environment. Finally, as both cell lines used in this study were selected for the properties of being first, latent, and second, reactivatable, we cannot exclude the possibility that the observed chromatin opening is directly related to these selected properties. Nevertheless, this observation is contrary to the conventional understanding of the relationship between chromatin accessibility and transcription (active genes typically exhibit open chromatin at the promoter but not through the body of the gene or downstream of it[85–88]), the main questions remaining relate to mechanism, and whether this chromatin opening is induced by a majority of integrated proviruses, or a subset latent proviruses that are amenable to reactivation.

Extensive studies of HIV-1 gene regulation have identified nuclear factors that can affect proviral transcription. The HIV-1 provirus can be silenced by repressive epigenetic marks at the Nuc-0 and Nuc-1 nucleosomes flanking the HIV-1 5′ LTR, reducing the accessibility of the proviral promoter to host transcription factors.[89] Furthermore, repressive chromatin marks such as H3K9me3, H3K27me3, and CpG methylation are also enriched at latent or transcriptionally inactive proviruses.[13,63,90–94] In contrast, active proviruses are associated with open chromatin and active epigenetic markers at the 5′ LTR.[13,95–97] The relationship between chromatin functional state, HIV-1 integration preference, and HIV-1 transcriptional competency have been established[9,22,62,63,97]; in this study, we expand our understanding by revealing how integrated HIV-1 can restructure the local nuclear environment.

In addition to addressing chromatin structure, we further characterized chimeric HIV-host RNAs and the transcriptional landscape in the vicinity of integrated HIV-1 in the J-Lat 8.4 and 10.6 models. Data from both short- (Illumina) and long- (Nanopore) read RNA-seq methodologies indicated that HIV-1 integration has minimal impact on the local transcription landscape; however, retroviral integration can promote expression of abnormal isoforms of the specific gene into which HIV is integrated. Although we did not see significant perturbations to local cellular transcription in the presence of integrated HIV-1 proviruses, whether active or inactive, it is conceivable that HIV-1 can contribute to aberrant transcription of host genes, depending on the site of integration (e.g., HIV-1 can block the formation of functional RNAs or promote expression of an abnormal gene product).

Expression of chimeric HIV-host RNAs[66,98,99] has a notable resemblance to stress-induced transcriptional readthrough[100–102] and chromatin remodeling following lytic Herpes Simplex Virus 1 (HSV-1) infection.[103,104] HIV-1 transcription is highly efficient following assembly of the Tat/TAR transactivation axis,[13] a *trans*-acting regulatory mechanism critical for proviral transcriptional elongation. Activation of processive viral transcription enables HIV-1 to dominate the local transcriptional environment. Similarly, HSV-1, a double-stranded DNA virus, hijacks the transcription machinery of the host cell during lytic infection. Despite not integrating into the host chromatin, HSV-1 transcriptional activation leads to significant dysregulation of host transcription by inducing genome-wide transcriptional read-through via the disruption of transcriptional termination at the 3′ UTR of cellular genes.[100–102] These transcripts, termed "Downstream-of-Genes" or DoGs, are also induced by cellular stress[105] and are not translated, evident by their absence within polysome fractions.[103] Further studies have demonstrated that transcriptional read-through of host genes during lytic HSV-1 infection also coincides with an increase in chromatin accessibility downstream of the 3′ UTR of the disrupted host gene. These changes in chromatin accessibility and transcriptional patterns are analogous to our findings following the activation of HIV-1 transcription, however, while HSV-1-induced effects are observed in a subset of stress-responsive genes, the HIV-1-induced effects are far more specific, localized to the site of HIV-1 integration and not dependent on the local gene having a propensity toward increased chromatin accessibility or transcription readthrough. While cellular transcription may occasionally fail to terminate, basal levels of transcriptional read-through are found well below the 1-5% of HIV-1 read-through products that we report in this study and that have been reported previously,[100–102] suggesting that the propensity of HIV-1 to open chromatin and generate transcription run-through products may mimic that of host stress response genes.

Recent studies have demonstrated an association between cellular stress responses and latency reversal (activation of HIV transcription),[106,107] with stress-inducing conditions increasing HIV-1 reactivation frequency. Other studies have implicated the integration of HIV-1 near specific host genes in the formation highly proliferative clones.[108–110] Combined with our findings, these studies collectively suggest that there is a significant interplay between virus and host, mediated through cellular stress pathways and chromatin structure. The capacity of HIV-1 to behave analogs to a stress-responsive element that may integrate into, and influence the function of, diverse host genes may be an important component in understanding the role of the HIV-1 integration site in both virus and host outcomes.

Limitation of study

Our work investigates the long-standing question of the impact of HIV-1 integration on the host cellular genome, cellular 3D chromatin organization, and cellular gene expression using an integrative approach. In this study, we used two well-defined HIV-1 latency models: J-Lat 10.6 and 8.4. These two models enabled us to assess how HIV-1 orientation relative to the host gene affects cellular chromatin state and transcription. However, this study was limited to only two cellular models. While the use of clonally expanded

models enabled us to interrogate defined HIV-1 integration sites at high-resolution using multiple NGS-based approaches, future studies will aim to expand upon this work using primary cell models following bulk *de novo* infection. A limiting factor using primary cellular models is that NGS-based studies will need to be multiplexed with single-cell integration site identification. If individual integration sites can be mapped, single-cell sequencing may not provide the resolution to observe subtle changes in local transcriptional patterns and chromatin state at retroviral integration sites. Therefore, we used clonal models and bulk sequencing for these studies, which enabled us to identify subtle molecular changes at the HIV-1 genome (increase in chromatin accessibility downstream active HIV-1).

Another limitation of the study is the resolution of Hi-C data in the context of investigating changes in chromatin structure in the presence of HIV-1 integration. We generated high-quality Hi-C libraries that were sequenced at sufficient depth (> billion reads/library) to enable us to conclude that major chromatin reorganization is not a prerequisite for HIV-1 transcriptional activation. However, we do not have the resolution to confidently assess direct HIV-host chromatin interactions using Hi-C, as the HIV-1 genome is small relative to host chromatin domains. Measuring HIV-host chromatin interactions may be important with the consideration that Jurkat T cells are known to be hypotetraploid,[111] which may obscure assessing the impact individual HIV-1 integration events have on their local chromatin structure using the Jurkat-derived J-Lat models. However, overall, very little to no alteration in chromatin organization and chromatin looping around the sites of integration in both J-Lat models makes us confident that changes in direct HIV-host chromatin interactions are also minimal following HIV-1 transcriptional activation in the 10.6 and 8.4 models.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Cell culture of T cell lines
  - Activation of clonal J-Lat cells
  - Flow cytometry and sorting of J-Lat cell lines/flow cytometric analysis
  - In situ Hi-C library preparation & analysis
  - Assay for transposase-accessible chromatin using sequencing (ATAC-seq)
  - ATAC-seq data analysis
  - Illumina-based RNA-seq library preparation
  - Illumina-based RNA-seq analysis
  - Nanopore RNA-sequencing library preparation and sequencing/base-calling
  - Long-read nanopore sequencing analysis
  - Quantification and productivity analysis of host cell isoforms at integration site
  - Analysis of publicly available datasets (ChIP-seq, TSA-seq, & SEduper)
  - Single-molecule DNA & RNA FISH
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.105490.

ACKNOWLEDGMENTS

This study was supported in part by NIH funding U54 AI150472 (SGS, BET, and JRD). We acknowledge the Emory Flow Cytometry Core (EFCC), one of the Emory Integrated Core Facilities (EICF), subsidized by the Emory University School of Medicine. The following reagents were obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: J-Lat 10.6 and 8.4 cells from Dr. Eric Verdin, and Jurkat (E6-1) Cells, ARP-177, contributed by ATCC from Dr. Arthur Weiss.

AUTHOR CONTRIBUTIONS

Conceptualization, R.S., P.R.T., and S.G.S.; investigation, R.S., C.M.G., Y.H.J., B.C., J.R.D., and W.M.F.; data curation, R.S., C.M.G., and J.R.D.; writing – original draft, R.S., P.R.T, S.G.S.; writing – review & editing, R.S., C.M.D., B.E.T., P.R.T., and S.G.S.; funding acquisition, B.E.T. and S.G.S.; supervision, B.R.T., P.R.T., and S.G.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in science.
One or more of the authors of this paper self-identifies as a member of the LGBTQ+ community.

Received: April 25, 2022
Revised: September 15, 2022
Accepted: October 31, 2022
Published: December 22, 2022

REFERENCES

1. Chun, T.W., Engel, D., Mizell, S.B., Hallahan, C.W., Fischette, M., Park, S., Davey, R.T., Jr., Dybul, M., Kovacs, J.A., Metcalf, J.A., et al. (1999). Effect of interleukin-2 on the pool of latently infected, resting CD4+ T cells in HIV-1-infected patients receiving highly active anti-retroviral therapy. Nat. Med. *5*, 651–655.

2. Brenchley, J.M., Hill, B.J., Ambrozak, D.R., Price, D.A., Guenaga, F.J., Casazza, J.P., Kuruppu, J., Yazdani, J., Migueles, S.A., Connors, M., et al. (2004). T-cell subsets that harbor human immunodeficiency virus (HIV) in vivo: implications for HIV pathogenesis. J. Virol. *78*, 1160–1168.

3. Sonza, S., Mutimer, H.P., Oelrichs, R., Jardine, D., Harvey, K., Dunne, A., Purcell, D.F., Birch, C., and Crowe, S.M. (2001). Monocytes harbour replication-competent, non-latent HIV-1 in patients on highly active antiretroviral therapy. AIDS *15*, 17–22.

4. Wu, L., and KewalRamani, V.N. (2006). Dendritic-cell interactions with HIV: infection and viral dissemination. Nat. Rev. Immunol. *6*, 859–868.

5. Craigie, R., and Bushman, F.D. (2014). Host factors in retroviral integration and the selection of integration target sites. Microbiol. Spectr. *2*.

6. Burdick, R.C., Li, C., Munshi, M., Rawson, J.M.O., Nagashima, K., Hu, W.S., and Pathak, V.K. (2020). HIV-1 uncoats in the nucleus near sites of integration. Proc. Natl. Acad. Sci. USA *117*, 5486–5493.

7. Li, C., Burdick, R.C., Nagashima, K., Hu, W.S., and Pathak, V.K. (2021). HIV-1 cores retain their integrity until minutes before uncoating in the nucleus. Proc. Natl. Acad. Sci. USA *118*. e2019467118.

8. Zila, V., Margiotta, E., Turoňová, B., Müller, T.G., Zimmerli, C.E., Mattei, S., Allegretti, M., Börner, K., Rada, J., Müller, B., et al. (2021). Cone-shaped HIV-1 capsids are transported through intact nuclear pores. Cell *184*, 1032–1046.e18.

9. Francis, A.C., Marin, M., Singh, P.K., Achuthan, V., Prellberg, M.J., Palermino-Rowland, K., Lan, S., Tedbury, P.R., Sarafianos, S.G., Engelman, A.N., and Melikyan, G.B. (2020). HIV-1 replication complexes accumulate in nuclear speckles and integrate into speckle-associated genomic domains. Nat. Commun. *11*, 6165.

10. Achuthan, V., Perreira, J.M., Ahn, J.J., Brass, A.L., and Engelman, A.N. (2019). Capsid-CPSF6 interaction: master regulator of nuclear HIV-1 positioning and integration. J. Life Sci. *1*, 39–45.

11. Jadlowsky, J.K., Wong, J.Y., Graham, A.C., Dobrowolski, C., Devor, R.L., Adams, M.D., Fujinaga, K., and Karn, J. (2014). Negative elongation factor is required for the maintenance of proviral latency but does not induce promoter-proximal pausing of RNA polymerase II on the HIV long terminal repeat. Mol. Cell Biol. *34*, 1911–1928.

12. Mann, D.A., Mikaélian, I., Zemmel, R.W., Green, S.M., Lowe, A.D., Kimura, T., Singh, M., Butler, P.J., Gait, M.J., and Karn, J. (1994). A molecular rheostat. Co-operative rev binding to stem I of the rev-response element modulates human immunodeficiency virus type-1 late gene expression. J. Mol. Biol. *241*, 193–207.

13. Mbonye, U., and Karn, J. (2014). Transcriptional control of HIV latency: cellular signaling pathways, epigenetics, happenstance and the hope for a cure. Virology *454–455*, 328–339.

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

15. Bellefroid, M., Rodari, A., Galais, M., Krijger, P.H.L., Tjalsma, S.J.D., Nestola, L., Plant, E., Vos, E.S.M., Cristinelli, S., Van Driessche, B., et al. (2022). Role of the cellular factor CTCF in the regulation of bovine leukemia virus latency and three-dimensional chromatin organization. Nucleic Acids Res. *50*, 3190–3202.

16. Zhang, Q., Wang, S., Li, W., Yau, E., Hui, H., Singh, P.K., Achuthan, V., Young Karris, M.A., Engelman, A.N., and Rana, T.M. (2022). Genome-wide CRISPR/Cas9 transcriptional activation screen identifies a histone acetyltransferase inhibitor complex as a regulator of HIV-1 integration. Nucleic Acids Res. *50*, 6687–6701.

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

18. Lenasi, T., Contreras, X., and Peterlin, B.M. (2008). Transcriptional interference antagonizes proviral gene expression to promote HIV latency. Cell Host Microbe *4*, 123–133.

19. Dahabieh, M.S., Battivelli, E., and Verdin, E. (2015). Understanding HIV latency: the road to an HIV cure. Annu. Rev. Med. *66*, 407–421.

20. Schröder, A.R.W., Shinn, P., Chen, H., Berry, C., Ecker, J.R., and Bushman, F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. Cell *110*, 521–529.

21. Bushman, F.D., Fujiwara, T., and Craigie, R. (1990). Retroviral DNA integration directed by HIV integration protein in vitro. Science *249*, 1555–1558.

22. Lucic, B., Chen, H.C., Kuzman, M., Zorita, E., Wegner, J., Minneker, V., Wang, W., Fronza, R., Laufs, S., Schmidt, M., et al. (2021). Spatially clustered loci with multiple enhancers are frequent targets of HIV-1 integration. Nat. Commun. *12*, 6326.

23. Chen, Y., Zhang, Y., Wang, Y., Zhang, L., Brinkman, E.K., Adam, S.A., Goldman, R., van Steensel, B., Ma, J., and Belmont, A.S. (2018). Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. J. Cell Biol. *217*, 4025–4048.

24. Towbin, B.D., Gonzalez-Sandoval, A., and Gasser, S.M. (2013). Mechanisms of heterochromatin subnuclear localization. Trends Biochem. Sci. *38*, 356–363.

25. Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W.M., Solovei, I., Brugman, W., Gräf, S., Flicek, P., Kerkhoven, R.M., van Lohuizen, M., et al. (2010). Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Mol. Cell *38*, 603–613.

26. Andrey, G., and Mundlos, S. (2017). The three-dimensional genome: regulating gene expression during pluripotency and development. Development *144*, 3646–3658.

27. Marsman, J., and Horsfield, J.A. (2012). Long distance relationships: enhancer-promoter communication and dynamic gene transcription. Biochim. Biophys. Acta *1819*, 1217–1227.

28. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

29. Dixon, J.R., Xu, J., Dileep, V., Zhan, Y., Song, F., Le, V.T., Yardımcı, G.G., Chakraborty, A., Bann, D.V., Wang, Y., et al. (2018). Integrative detection and analysis of structural variation in cancer genomes. Nat. Genet. *50*, 1388–1398.

30. Lee, D.S., Luo, C., Zhou, J., Chandran, S., Rivkin, A., Bartlett, A., Nery, J.R., Fitzpatrick, C., O'Connor, C., Dixon, J.R., and Ecker, J.R. (2019). Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. Nat. Methods *16*, 999–1006.

31. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

32. Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., et al. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods *14*, 959–962.

33. Jung, Y.H., Kremsky, I., Gold, H.B., Rowley, M.J., Punyawai, K., Buonanotte, A., Lyu, X., Bixler, B.J., Chan, A.W.S., and Corces, V.G. (2019). Maintenance of CTCF- and transcription factor-mediated interactions from the gametes to the early mouse embryo. Mol. Cell *75*, 154–171.e5.

34. Hansen, A.S., Cattoglio, C., Darzacq, X., and Tjian, R. (2018). Recent evidence that TADs and chromatin loops are dynamic structures. Nucleus *9*, 20–32.

35. Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

36. Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289–293.

37. Rae, M.M., and Franke, W.W. (1972). The interphase distribution of satellite DNA-containing heterochromatin in mouse nuclei. Chromosoma *39*, 443–456.

38. Schermelleh, L., Carlton, P.M., Haase, S., Shao, L., Winoto, L., Kner, P., Burke, B., Cardoso, M.C., Agard, D.A., Gustafsson, M.G.L., et al. (2008). Subdiffraction multicolor imaging of the nuclear periphery with 3D structured illumination microscopy. Science *320*, 1332–1336.

39. Huisinga, K.L., Brower-Toland, B., and Elgin, S.C.R. (2006). The contradictory definitions of heterochromatin: transcription and silencing. Chromosoma *115*, 110–122.

40. Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

41. Rowley, M.J., and Corces, V.G. (2018). Organizational principles of 3D genome architecture. Nat. Rev. Genet. *19*, 789–800.

42. Kleinjan, D.A., and van Heyningen, V. (2005). Long-range control of gene expression: emerging mechanisms and disruption in disease. Am. J. Hum. Genet. *76*, 8–32.

43. Xiang, J.F., and Corces, V.G. (2021). Regulation of 3D chromatin organization by CTCF. Curr. Opin. Genet. Dev. *67*, 33–40.

44. Phillips, J.E., and Corces, V.G. (2009). CTCF: master weaver of the genome. Cell *137*, 1194–1211.

45. Tchasovnikarova, I.A., Timms, R.T., Matheson, N.J., Wals, K., Antrobus, R., Göttgens, B., Dougan, G., Dawson, M.A., and Lehner, P.J. (2015). GENE SILENCING. Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. Science *348*, 1481–1485.

46. Chougui, G., Munir-Matloob, S., Matkovic, R., Martin, M.M., Morel, M., Lahouassa, H., Leduc, M., Ramirez, B.C., Etienne, L., and Margottin-Goguet, F. (2018). HIV-2/SIV viral protein X counteracts HUSH repressor complex. Nat. Microbiol. *3*, 891–897.

47. Yurkovetskiy, L., Guney, M.H., Kim, K., Goh, S.L., McCauley, S., Dauphin, A., Diehl, W.E., and Luban, J. (2018). Primate immunodeficiency virus proteins Vpx and Vpr counteract transcriptional repression of proviruses by the HUSH complex. Nat. Microbiol. *3*, 1354–1361.

48. Moquin, S.A., Thomas, S., Whalen, S., Warburton, A., Fernandez, S.G., McBride, A.A., Pollard, K.S., and Miranda, J.L. (2018). The epstein-barr virus episome maneuvers between nuclear chromatin compartments during reactivation. J. Virol. *92*. e01413-17.

49. Cook, L., Melamed, A., Yaguchi, H., and Bangham, C.R. (2017). The impact of HTLV-1 on the cellular genome. Curr. Opin. Virol. *26*, 125–131.

50. Melamed, A., Yaguchi, H., Miura, M., Witkover, A., Fitzgerald, T.W., Birney, E., and Bangham, C.R. (2018). The human leukemia virus HTLV-1 alters the structure and transcription of host chromatin in cis. Elife *7*, e36245.

51. Satou, Y., Miyazato, P., Ishihara, K., Yaguchi, H., Melamed, A., Miura, M., Fukuda, A., Nosaka, K., Watanabe, T., Rowan, A.G., et al. (2016). The retrovirus HTLV-1 inserts an ectopic CTCF-binding site into the human genome. Proc. Natl. Acad. Sci. USA *113*, 3054–3059.

52. Boftsi, M., Majumder, K., Burger, L.R., and Pintel, D.J. (2020). Binding of CCCTC-binding factor (CTCF) to the minute virus of mice genome is important for proper processing of viral P4-generated pre-mRNAs. Viruses *12*, 1368.

53. Majumder, K., Boftsi, M., and Pintel, D.J. (2019). Viral chromosome conformation capture (V3C) assays for identifying trans-interaction sites between lytic viruses and the cellular genome. Bio. Protoc. *9*, e3198.

54. Majumder, K., Boftsi, M., Whittle, F.B., Wang, J., Fuller, M.S., Joshi, T., and Pintel, D.J. (2020). The NS1 protein of the parvovirus MVM Aids in the localization of the viral genome to cellular sites of DNA damage. PLoS Pathog. *16*. e1009002.

55. Majumder, K., Wang, J., Boftsi, M., Fuller, M.S., Rede, J.E., Joshi, T., and Pintel, D.J. (2018). Parvovirus minute virus of mice interacts with sites of cellular DNA damage to establish and amplify its lytic infection. Elife *7*, e37750.

56. Jang, M.K., Shen, K., and McBride, A.A. (2014). Papillomavirus genomes associate with BRD4 to replicate at fragile sites in the host genome. PLoS Pathog. *10*. e1004117.

57. Saldaña-Meyer, R., González-Buendía, E., Guerrero, G., Narendra, V., Bonasio, R., Recillas-Targa, F., and Reinberg, D. (2014). CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript. Genes Dev. *28*, 723–734.

58. Kaneko, S., Bonasio, R., Saldaña-Meyer, R., Yoshida, T., Son, J., Nishino, K., Umezawa, A., and Reinberg, D. (2014). Interactions between JARID2 and noncoding RNAs

regulate PRC2 recruitment to chromatin. Mol. Cell *53*, 290–300.

59. Tan, X., Ravasio, A., Ong, H.T., Wu, J., and Hew, C.L. (2020). White spot syndrome viral protein VP9 alters the cellular higher-order chromatin structure. FASEB Bioadv. *2*, 264–279.

60. Heinz, S., Texari, L., Hayes, M.G.B., Urbanowski, M., Chang, M.W., Givarkes, N., Rialdi, A., White, K.M., Albrecht, R.A., Pache, L., et al. (2018). Transcription elongation can affect genome 3D structure. Cell *174*, 1522–1536.e22.

61. Dieudonné, M., Maiuri, P., Biancotto, C., Knezevich, A., Kula, A., Lusic, M., and Marcello, A. (2009). Transcriptional competence of the integrated HIV-1 provirus at the nuclear periphery. EMBO J. *28*, 2231–2243.

62. Marini, B., Kertesz-Farkas, A., Ali, H., Lucic, B., Lisek, K., Manganaro, L., Pongor, S., Luzzati, R., Recchia, A., Mavilio, F., et al. (2015). Nuclear architecture dictates HIV-1 integration site selection. Nature *521*, 227–231.

63. Vansant, G., Chen, H.C., Zorita, E., Trejbalová, K., Miklík, D., Filion, G., and Debyser, Z. (2020). The chromatin landscape at the HIV-1 provirus integration site determines viral expression. Nucleic Acids Res. *48*, 7801–7817.

64. Einkauf, K.B., Osborn, M.R., Gao, C., Sun, W., Sun, X., Lian, X., Parsons, E.M., Gladkov, G.T., Seiger, K.W., Blackmer, J.E., et al. (2022). Parallel analysis of transcription, integration, and sequence of single HIV-1 proviruses. Cell *185*, 266–282.e15.

65. Perkins, K.J., Lusic, M., Mitar, I., Giacca, M., and Proudfoot, N.J. (2008). Transcription-dependent gene looping of the HIV-1 provirus is dictated by recognition of pre-mRNA processing signals. Mol. Cell *29*, 56–68.

66. Liu, R., Yeh, Y.H.J., Varabyou, A., Collora, J.A., Sherrill-Mix, S., Talbot, C.C., Jr., Mehta, S., Albrecht, K., Hao, H., Zhang, H., et al. (2020). Single-cell transcriptional landscapes reveal HIV-1-driven aberrant host gene transcription as a potential therapeutic target. Sci. Transl. Med. *12*, eaaz0802.

67. Han, Y., Lin, Y.B., An, W., Xu, J., Yang, H.C., O'Connell, K., Dordai, D., Boeke, J.D., Siliciano, J.D., and Siliciano, R.F. (2008). Orientation-dependent regulation of integrated HIV-1 expression by host gene transcriptional readthrough. Cell Host Microbe *4*, 134–146.

68. Jordan, A., Bisgrove, D., and Verdin, E. (2003). HIV reproducibly establishes a latent infection after acute infection of T cells in vitro. EMBO J. *22*, 1868–1877.

69. Spina, C.A., Anderson, J., Archin, N.M., Bosque, A., Chan, J., Famiglietti, M., Greene, W.C., Kashuba, A., Lewin, S.R., Margolis, D.M., et al. (2013). An in-depth comparison of latent HIV-1 reactivation in multiple cell model systems and resting CD4+ T cells from aviremic patients. PLoS Pathog. *9*, e1003834.

70. Symons, J., Cameron, P.U., and Lewin, S.R. (2018). HIV integration sites and implications for maintenance of the reservoir. Curr. Opin. HIV AIDS *13*, 152–159.

71. Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. *14*, 178–192.

72. Shao, Z., Zhang, Y., Yuan, G.C., Orkin, S.H., and Waxman, D.J. (2012). MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets. Genome Biol. *13*, R16.

73. Shah, R., Lan, S., Puray-Chavez, M.N., Liu, D., Tedbury, P.R., and Sarafianos, S.G. (2020). Single-cell multiplexed fluorescence imaging to visualize viral nucleic acids and proteins and monitor HIV, HTLV, HBV, HCV, zika virus, and influenza infection. J. Vis. Exp. https://doi.org/10.3791/61843.

74. Puray-Chavez, M., Tedbury, P.R., Huber, A.D., Ukah, O.B., Yapo, V., Liu, D., Ji, J., Wolf, J.J., Engelman, A.N., and Sarafianos, S.G. (2017). Multiplex single-cell visualization of nucleic acids and protein during HIV infection. Nat. Commun. *8*, 1882.

75. Gholizadeh, Z., Iqbal, M.S., Li, R., and Romerio, F. (2021). The HIV-1 antisense gene ASP: the new kid on the block. Vaccines *9*, 513.

76. Ma, G., Yasunaga, J.I., Shimura, K., Takemoto, K., Watanabe, M., Amano, M., Nakata, H., Liu, B., Zuo, X., and Matsuoka, M. (2021). Human retroviral antisense mRNAs are retained in the nuclei of infected cells for viral persistence. Proc. Natl. Acad. Sci. USA *118*. e2014783118.

77. Mancarella, A., Procopio, F.A., Achsel, T., De Crignis, E., Foley, B.T., Corradin, G., Bagni, C., Pantaleo, G., and Graziosi, C. (2019). Detection of antisense protein (ASP) RNA transcripts in individuals infected with human immunodeficiency virus type 1 (HIV-1). J. Gen. Virol. *100*, 863–876.

78. Fernandez, G., and Zeichner, S.L. (2010). Cell line-dependent variability in HIV activation employing DNMT inhibitors. Virol. J. *7*, 266.

79. Depledge, D.P., Srinivas, K.P., Sadaoka, T., Bready, D., Mori, Y., Placantonakis, D.G., Mohr, I., and Wilson, A.C. (2019). Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. Nat. Commun. *10*, 754.

80. Gallardo, C.M., Wang, S., Montiel-Garcia, D.J., Little, S.J., Smith, D.M., Routh, A.L., and Torbett, B.E. (2021). MrHAMER yields highly accurate single molecule viral sequences enabling analysis of intra-host evolution. Nucleic Acids Res. *49*, e70.

81. Bacon, C.W., and D'Orso, I. (2019). CDK9: a signaling hub for transcriptional control. Transcription *10*, 57–75.

82. Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

83. Kao, S.Y., Calman, A.F., Luciw, P.A., and Peterlin, B.M. (1987). Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. Nature *330*, 489–493.

84. Roebuck, K.A., and Saifuddin, M. (1999). Regulation of HIV-1 transcription. Gene Expr. *8*, 67–84.

85. Miyamoto, K., Nguyen, K.T., Allen, G.E., Jullien, J., Kumar, D., Otani, T., Bradshaw, C.R., Livesey, F.J., Kellis, M., and Gurdon, J.B. (2018). Chromatin accessibility impacts transcriptional reprogramming in oocytes. Cell Rep. *24*, 304–311.

86. Teif, V.B., Vainshtein, Y., Caudron-Herger, M., Mallm, J.P., Marth, C., Höfer, T., and Rippe, K. (2012). Genome-wide nucleosome positioning during embryonic stem cell development. Nat. Struct. Mol. Biol. *19*, 1185–1192.

87. Lu, F., Liu, Y., Inoue, A., Suzuki, T., Zhao, K., and Zhang, Y. (2016). Establishing chromatin regulatory landscape during mouse preimplantation development. Cell *165*, 1375–1388.

88. Wu, J., Huang, B., Chen, H., Yin, Q., Liu, Y., Xiang, Y., Zhang, B., Liu, B., Wang, Q., Xia, W., et al. (2016). The landscape of accessible chromatin in mammalian preimplantation embryos. Nature *534*, 652–657.

89. Verdin, E., Paras, P., Jr., and Van Lint, C. (1993). Chromatin disruption in the promoter of human immunodeficiency virus type 1 during transcriptional activation. EMBO J. *12*, 3249–3259.

90. du Chéné, I., Basyuk, E., Lin, Y.L., Triboulet, R., Knezevich, A., Chable-Bessia, C., Mettling, C., Baillat, V., Reynes, J., Corbeau, P., et al. (2007). Suv39H1 and HP1gamma are responsible for chromatin-mediated HIV-1 transcriptional silencing and post-integration latency. EMBO J *26*, 424–435.

91. Kauder, S.E., Bosque, A., Lindqvist, A., Planelles, V., and Verdin, E. (2009). Epigenetic regulation of HIV-1 latency by cytosine methylation. PLoS Pathog. *5*, e1000495.

92. Blazkova, J., Trejbalova, K., Gondois-Rey, F., Halfon, P., Philibert, P., Guiguen, A., Verdin, E., Olive, D., Van Lint, C., Hejnar, J., and Hirsch, I. (2009). CpG methylation controls reactivation of HIV from latency. PLoS Pathog. *5*, e1000554.

93. Friedman, J., Cho, W.K., Chu, C.K., Keedy, K.S., Archin, N.M., Margolis, D.M., and Karn, J. (2011). Epigenetic silencing of HIV-1 by the histone H3 lysine 27 methyltransferase enhancer of Zeste 2. J. Virol. *85*, 9078–9089.

94. He, G., and Margolis, D.M. (2002). Counterregulation of chromatin deacetylation and histone deacetylase occupancy at the integrated promoter of human immunodeficiency virus type 1 (HIV-1) by the HIV-1 repressor YY1 and HIV-1 activator Tat. Mol. Cell Biol. *22*, 2965–2973.

95. Miller-Jensen, K., Dey, S.S., Pham, N., Foley, J.E., Arkin, A.P., and Schaffer, D.V. (2012). Chromatin accessibility at the HIV LTR promoter sets a threshold for NF-kappaB mediated viral gene expression. Integr. Biol. *4*, 661–671.

96. Battivelli, E., and Verdin, E. (2018). HIVGKO: a Tool to assess HIV-1 latency reversal agents in human primary CD4(+) T cells. Bio. Protoc. *8*, e3050.

97. Battivelli, E., Dahabieh, M.S., Abdel-Mohsen, M., Svensson, J.P., Tojal Da Silva, I., Cohn, L.B., Gramatica, A., Deeks, S., Greene, W.C., Pillai, S.K., and Verdin, E. (2018). Distinct chromatin functional states correlate with HIV latency reactivation in infected primary CD4(+) T cells. Elife *7*, e34655.

98. Cesana, D., Santoni de Sio, F.R., Rudilosso, L., Gallina, P., Calabria, A., Beretta, S., Merelli, I., Bruzzesi, E., Passerini, L., Nozza, S., et al. (2017). HIV-1-mediated insertional activation of STAT5B and BACH2 trigger viral reservoir in T regulatory cells. Nat. Commun. *8*, 498.

99. Christian, M.L., Dapp, M.J., Scharffenberger, S.C., Jones, H., Song, C., Frenkel, L.M., Krumm, A., Mullins, J.I., and Rawlings, D.J. (2022). CRISPR/Cas9-Mediated insertion of HIV long terminal repeat within BACH2 promotes expansion of T regulatory-like cells. J. Immunol. *208*, 1700–1710.

100. Rosa-Mercado, N.A., Zimmer, J.T., Apostolidi, M., Rinehart, J., Simon, M.D., and Steitz, J.A. (2021). Hyperosmotic stress alters the RNA polymerase II interactome and induces readthrough transcription despite widespread transcriptional repression. Mol. Cell *81*, 502–513.e4.

101. Vilborg, A., Sabath, N., Wiesel, Y., Nathans, J., Levy-Adam, F., Yario, T.A., Steitz, J.A., and Shalgi, R. (2017). Comparative analysis reveals genomic features of stress-induced transcriptional readthrough. Proc. Natl. Acad. Sci. USA *114*, E8362–E8371.

102. Vilborg, A., and Steitz, J.A. (2017). Readthrough transcription: how are DoGs made and what do they do? RNA Biol. *14*, 632–636.

103. Rutkowski, A.J., Erhard, F., L'Hernault, A., Bonfert, T., Schilhabel, M., Crump, C., Rosenstiel, P., Efstathiou, S., Zimmer, R., Friedel, C.C., and Dölken, L. (2015). Widespread disruption of host transcription termination in HSV-1 infection. Nat. Commun. *6*, 7126.

104. Hennig, T., Michalski, M., Rutkowski, A.J., Djakovic, L., Whisnant, A.W., Friedl, M.S., Jha, B.A., Baptista, M.A.P., L'Hernault, A., Erhard, F., et al. (2018). HSV-1-induced disruption of transcription termination resembles a cellular stress response but selectively increases chromatin accessibility downstream of genes. PLoS Pathog. *14*, e1006954.

105. Cardiello, J.F., Goodrich, J.A., and Kugel, J.F. (2018). Heat Shock causes a reversible increase in RNA polymerase II occupancy downstream of mRNA genes, consistent with a global loss in transcriptional termination. Mol. Cell Biol. *38*, 001811-18.

106. Pan, X.Y., Zhao, W., Zeng, X.Y., Lin, J., Li, M.M., Shen, X.T., and Liu, S.W. (2016). Heat Shock factor 1 mediates latent HIV reactivation. Sci. Rep. *6*, 26294.

107. Timmons, A., Fray, E., Kumar, M., Wu, F., Dai, W., Bullen, C.K., Kim, P., Hetzel, C., Yang, C., Beg, S., et al. (2020). HSF1 inhibition attenuates HIV-1 latency reversal mediated by several candidate LRAs in Vitro and Ex Vivo. Proc. Natl. Acad. Sci. USA *117*, 15763–15771.

108. Maldarelli, F., Wu, X., Su, L., Simonetti, F.R., Shao, W., Hill, S., Spindler, J., Ferris, A.L., Mellors, J.W., Kearney, M.F., et al. (2014). HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. Science *345*, 179–183.

109. Mellors, J.W., Guo, S., Naqvi, A., Brandt, L.D., Su, L., Sun, Z., Joseph, K.W., Demirov, D., Halvas, E.K., Butcher, D., et al. (2021). Insertional activation of STAT3 and LCK by HIV-1 proviruses in T cell lymphomas. Sci. Adv. *7*, eabi8795.

110. Patro, S.C., Brandt, L.D., Bale, M.J., Halvas, E.K., Joseph, K.W., Shao, W., Wu, X., Guo, S., Murrell, B., Wiegand, A., et al. (2019). Combined HIV-1 sequence and integration site analysis informs viral dynamics and allows reconstruction of replicating viral ancestors. Proc. Natl. Acad. Sci. USA *116*, 25891–25899.

111. Gioia, L., Siddique, A., Head, S.R., Salomon, D.R., and Su, A.I. (2018). A genome-wide survey of mutations in the Jurkat cell line. BMC Genom. *19*, 334.

112. Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., and Aiden, E.L. (2016). Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. *3*, 99–101.

113. Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. *3*, 95–98.

114. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

115. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.

116. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

117. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. *37*, 907–915.

118. Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. *33*, 290–295.

119. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

120. Liu, T. (2014). Use model-based Analysis of ChIP-Seq (MACS) to analyze short reads generated by sequencing protein-DNA interactions in embryonic stem cells. Methods Mol. Biol. *1150*, 81–95.

121. Khan, A., and Zhang, X. (2016). dbSUPER: a database of super-enhancers in mouse and human genome. Nucleic Acids Res. *44*, D164–D171.

122. Tang, A.D., Soulette, C.M., van Baren, M.J., Hart, K., Hrabeta-Robinson, E., Wu, C.J., and Brooks, A.N. (2020). Full-length transcript characterization of SF3B1 mutation in chronic lymphocytic leukemia reveals downregulation of retained introns. Nat. Commun. *11*, 1438.

123. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100.

124. Weiss, A., Wiskocil, R.L., and Stobo, J.D. (1984). The role of T3 surface molecules in the activation of human T cells: a two-stimulus requirement for IL 2 production reflects events occurring at a pre-translational level. J. Immunol. *133*, 123–128.

125. Ukah, O.B., Puray-Chavez, M., Tedbury, P.R., Herschhorn, A., Sodroski, J.G., and Sarafianos, S.G. (2018). Visualization of HIV-1 RNA transcription from integrated HIV-1 DNA in reactivated latently infected cells. Viruses *10*, 534.

126. Shen, L., Shao, N., Liu, X., and Nestler, E. (2014). ngs.plot: quick mining and visualization of next-generation sequencing data by integrating genomic databases. BMC Genom. *15*, 284.

127. Gallardo, C.M., Nguyen, A.V.T., Routh, A.L., and Torbett, B.E. (2022). Selective ablation of 3′ RNA ends and processive RTs facilitate direct cDNA sequencing of full-length host cell and viral transcripts. Nucleic Acids Res. *50*, e98.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
RPMI 1640 medium	Gibco	11875093
TNF-α	Genscript	Z01001
Phorbol 12-myristate/ionomycin	eBioscience	00-4970-03
poly-*d*-lysine	Gibco	A3890401
4′,6′-diamino-2-phenylinndole (DAPI)	Advanced Cell Diagnostics	320858
Prolong Gold Antifade	Invitrogen	P36930
Critical commercial assays		
2x KAPA HiFi mix	Kapa Biosystems	KK2601
RNeasy Plus mini kit	Qiagen	74134
TrueSeq Stranded Total RNA sample preparation kit	Illumina	20020596
Super-Script IV Reverse Transcriptase	ThermoFisher	18090010
Monarch PCR and DNA Cleanup Kit	New England Biosciences	T1030S
Native Barcoding kit	Oxford Nanopore Technologies	EXP-NBD104, EXP-NBD114
Ligation Sequencing Kit	Oxford Nanopore Technologies	SQK-LSK109
RNAscope Fluorescent Multiplex Assay	Advanced Cell Diagnostics	320850
RNA Clean & Concentrator	Zymo Research	R1013
Deposited data		
Hi-C data	This study	GSE189178
ATAC-seq data	This study	PRJNA764194
RNA-seq data	This study	PRJNA771201
ChIP-seq (Jurkat; H3K27ac)	GEO	GSM2691418
ChIP-seq (Jurkat; BRD4)	GEO	GSM2218755
ChIP-seq (Jurkat; Pol2)	GEO	GSM1850204
TSA-seq (K562 cells; SPADs)	GEO	GSM3111194
Experimental models: Cell lines		
Jurkat E6-1 cells	NIH AIDS Reagent Program catalog	#ARP-177
J-Lat 8.4 full-length cells	NIH AIDS Reagent Program catalog	#9847
J-Lat full-length 10.6 cells	NIH AIDS Reagent Program catalog	#9849
Oligonucleotides		
HIV-1 (+) RNA (HIV-nongagpol-C3)	Advanced Cell Diagnostics	317711-C
HIV-gagpol-C1	Advanced Cell Diagnostics	317701
Software and algorithms		
FlowJo v10.8	FlowJo, LLC	https://www.flowjo.com/solutions/flowjo
Juicer	Durand et al., 2016[112,113]	http://www.aidenlab.org/software.html
Hiccups	Rao et al., 2014[35]	https://www.bioinformatics.babraham.ac.uk/projects/hicup/
Bowtie2	Langmead and Salzberg, 2012[114]	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Samtools	Li et al., 2009[115]	http://samtools.sourceforge.net/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bedtools	Quinlan et al., 2010[116]	https://sourceforge.net/projects/bedtools/
Picard	Broad Institute	https://broadinstitute.github.io/picard/
HISAT2	Kim et al., 2019[117]	http://daehwankimlab.github.io/hisat2/
StringTie	Pertea et al., 2015[118]	https://ccb.jhu.edu/software/stringtie/
DEseq2	Love et al., 2014[119]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
MACS2	Liu et al., 2014[120]	https://pypi.org/project/MACS2/
MAnorm	Shao et al., 2012[72]	http://bioinfo.sibs.ac.cn/zhanglab/MAnorm/MAnorm.htm
R programming language (Version R-4.2.1)	www.r-project.org	http://cran.r-project.org/bin/windows/base
dbSuper	Khan et al., 2016[121]	https://asntech.org/dbsuper/
Guppy Basecaller version 4.2.3+8aca2af	Oxford Nanopore Technologies	https://github.co/nanoporetech/rerio
Re-form Tool web portal used to generate custom reference files (hg38 with J-Lat integration)	https://reform.bio.nyu.edu/	https://reform.bio.nyu.edu/
FLAIR isoform analysis pipeline	Tang et al., 2020[122]	https://github.conm/BrooksLabUCSC/flair
Minimap2	Li, 2018[123]	https://github.con/ih3/minimap2
Integrated Genomics Viewer (IGV)	Thorvaldsdóttir et al., 2013[71]	https://software.broadinstitute.org/software/igv/
Other		
40 µm cell strainer	Corning	431750
Long-read RNA-seq performed using MinION Mk1B using an FLO-MIN106D flow cell	Oxford Nanopore Technology	R.9.4.1
HybEZ oven	Advanced Cell Diagnostics	PN 321710/321720

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Stefan Sarafianos (stefanos.sarafianos@emory.edu).

Materials availability

No new materials or reagents were generated in this study.

Data and code availability

Raw fastq datasets (Illumina and Nanopore) generated as part of this study can be downloaded from NCBI Sequence Read Archive (SRA) under the following study accessions: ATAC-seq (PRJNA764194), RNA-seq (PRJNA771201), and Hi-C (GSE189178). Custom scripts were used to separate ATAC-seq reads into subnucleosomal size ranges. This analysis is included in Figures 3 and S4–S7. These scripts are available upon request. The following publicly mined sequencing datasets were obtained from NCBI SRA: ChIP-seq datasets from Jurkat T cells include H3K27ac (GSM2691418), Pol2 (GSM1850204), and BRD4 (GSM2218755) and TSA-seq dataset of SPADs (GSM3111194) in K562 cells. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The following reagents were obtained though the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: J-Lat 10.6 and 8.4 cells from Dr. Eric Verdin, and Jurkat (E6-1) Cells, ARP-177, contributed by ATCC from Dr. Arthur Weiss. No *in vivo* animal studies or human studies were performed for this work.

METHOD DETAILS

Cell culture of T cell lines

Jurkat E6-1 cells (NIH AIDS Reagent Program catalog #ARP-177) are a human T-cell lymphoma derived cell line.[124] J-Lat full-length 10.6 cells (NIH AIDS Reagent Program catalog #9849) and J-Lat 8.4 full-length cells (NIH AIDS Reagent Program catalog #9847) are Jurkat-derived cells that are latently infected with the packaged retroviral construct HIV-R7/E-/GFP, a full-length HIV-1 minus *env* and *nef*.[68] Cells were cultured in RPMI 1640 medium (Gibco, Waltham, MA, USA), supplemented with 10% heat-inactivated fetal bovine serum (FBS) and 2 mM L-glutamine (Gibco), in a humidified incubator at 37°C with 5% CO_2.

Activation of clonal J-Lat cells

Compounds used for activation in HIV-inducible cell lines included: (1) 10 ng/μL TNF-α (Genscript, Piscataway, NJ, USA) and (2) 40.5 nM phorbol 12-myristate 13-acetate with 670 nM ionomycin (PMA/I) (eBioscience, San Diego, CA, USA). Cells were treated for 24 h prior to collection for NGS-based, flow cytometric, and smFISH assays. Treatment conditions were based on parameters previously described.[125]

Flow cytometry and sorting of J-Lat cell lines/flow cytometric analysis

The BC Cytoflex cytometer was used for data collection. 10^5 Jurkat or J-Lat cells were seeded per well in a 96-well plate format. Cells were treated with TNF-α or PMA/I and the numbers of GFP + cells following treatment were quantified. Gates for GFP + cells were set by assessing the fluorescence profiles of uninfected Jurkat T cells. FlowJo v10.8 software was used for data analysis. Gating strategy is depicted in Figure S1.

Fluorescence-activated cell sorting (FACS) of inactive (GFP-) and active (GFP+) populations of cells was performed using the BD FACS Aria II SORP. Prior to cell sorting, T cells were resuspended in PBS and passed through a 40 μm cell strainer (Corning, 431570). Final cell concentrations were ~5 × 10^6 cells/mL in PBS prior to sorting.

In situ Hi-C library preparation & analysis

Hi-C was performed using the *in situ* method with the MboI enzyme as previously described.[35] Hi-C libraries were generated from either wild-type (WT) Jurkat T-cell lines, untreated clonal J-Lat model cell lines, or sorted GFP + J-Lat cells after stimulation with TNF-α. About two million cells per library were fixed in solution using 1% formaldehyde. For sorted samples, GFP + cells were sorted into PBS and subsequently fixed with 1% formaldehyde. Hi-C libraries were sequenced on Novaseq 6000 S4 flow cells. Reads were aligned to the hg38 genome, filtered, and deduplicated as previously described.[29] After alignment, contact files were generated and processed into hic or cool files using the Juicer[112,113] or cooler pipelines, respectively. Loop calling was performed using hiccups.[35]

Assay for transposase-accessible chromatin using sequencing (ATAC-seq)

ATAC-seq was performed using the OMNI-ATAC protocol.[32] After counting 50,000 cells/library, nuclei were isolated with lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$) containing 0.1% NP40, 0.1% Tween 20, and 0.01% digitonin. The purified Jurkat or J-Lat nuclei were then resuspended in a transposase reaction buffer containing 0.05% digitonin and incubated for 30 min at 37°C. Following incubation, samples were treated with Proteinase K at 55°C for 2 h, and genomic DNA was isolated via phenol:chloroform:isoamyl alcohol and EtOH precipitation. Library amplification was done with 2x KAPA HiFi mix (Kapa Biosystems) and 1.25 μM indexed Illumina primers using the following PCR conditions: 72°C for 5 min; 98°C for 30 s; and 10-11 cycles at 98°C for 10s, 63°C for 30 s, and 72°C for 1 min. Libraries were generated with two biological replicates per condition sampled.

ATAC-seq data analysis

All ATAC-seq libraries were sequenced using an Illumina Hiseq2500 v4 sequencer and 50 bp paired-end format. Paired reads were aligned to the human reference genome hg38 or custom HIV-host genome (See "*Reference alignment of Nanopore-seq datasets*" for description of reference genome assembly) using Bowtie2.[114] Reads were aligned using default parameters except -X 2000 -m 1. PCR duplicates were removed using Picard Tools (https://broadinstitute.github.io/picard/). To adjust for fragment size, we aligned all reads as (+) strands offset by +4 bp and (−) strands offset by −5 bp.[31] The reads corresponding to Tn5 hypersensitive sites and mononucleosomes were separated by filtering for fragments 50-115 bp and

180-247 bp in length, respectively.[33] MACS2 was used for peak calling for Tn5 hypersensitive sites.[120] Quality-control (QC) data was visualized via ggplot2 in the R programming language and through ngs.plot.[126] Differential peaks between HIV-inactive and -active populations was performed using MAnorm.[72]

Illumina-based RNA-seq library preparation

Total RNA was extracted from one million cells/condition using the RNeasy Plus mini kit (Qiagen; 74134). Preparation included an on-column DNase digestion. Strand-specific cDNA libraries were generated using a TrueSeq Stranded Total RNA sample preparation kit (Illumina). Sequencing was performed on an Illumina Novaseq 6000 to obtain ~150 bp paired-end reads. Each library, performed in replicate, had ≥40 million read pairs per sample.

Illumina-based RNA-seq analysis

Reads were mapped to the human reference genome (hg38) or a custom concatenated HIV-host reference genome for the J-Lat 10.6 and 8.4 models. The J-Lat specific sequence that was used for the J-Lat custom reference genomes was generated *de novo* from activated J-Lat 10.6 cells using our Nanopore pipeline discussed below. Reads were mapped using HISAT2.[117] HTseq was used to quantify RNAs based on annotations.[82] StringTie was used for transcript assembly[118] and for quantification of different isoforms. Differential expression analysis was performed using DESeq2[119] with read counts normalized to fragments per kilobase of transcript per million mapped reads (FPKM).[82]

Nanopore RNA-sequencing library preparation and sequencing/base-calling

Total RNA was isolated from cell pellets (~50,000 cells) using the RNeasy Mini Kit (QIAGEN, 74134), according to the manufacturer's instructions, with elution in nuclease-free water. Total RNA was treated with the CASPR reagent at a concentration of 2 mg/mL for 30 min at room temperature in the dark, followed by RNA cleanup with RNA Clean & Concentration (Zymo Research, R1013) and elution in nuclease-free water.[127] Reverse Transcription was carried out with Super-Script IV Reverse Transcriptase (ThermoFisher, 18090010) in 20 μL volume with the following components and final concentrations: 1X Reaction Buffer, 0.5 mM dNTPs, 2U RNAse OUT, 1 μM Oligo-d(T) primer, 5 mM DTT, and 200 U Super-Script IV Reverse Transcriptase. Primer was annealed to CASPR-treated template RNA in the presence of dNTPs by heating to 65°C for 5 min, followed by snap cooling to 4°C for at least 2 min. The remaining components were added after a snap-cool step, then incubated at 50°C for 1.5 h, and heat inactivated at 85°C for 5 min. Second-strand synthesis was carried out in a single pot format using modified Gubler and Hoffman procedure from Invitrogen's A48570 kit by direct addition of second strand buffer, dNTPs, *E. coli* DNA polymerase I, RNAse H, and *E. Coli* DNA ligase to heat-inactivated first strand reaction and incubation at 16°C for 2 h. DNA was then isolated with Monarch PCR and DNA Cleanup Kit (NEB, T1030S), and eluted in 0.1X TE. Quality control and yield of double-stranded cDNA was determined with NanoDrop spectrometer.

Double-stranded cDNA samples were barcoded with the Native Barcoding kit (Oxford Nanopore Technologies, EXP-NBD104, EXP-NBD114), and library prepped using the Ligation Sequencing Kit (ONT, SQK-LSK109). Samples were sequenced with MinION Mk1B using an FLO-MIN106D flow cell (R.9.4.1). Reads were basecalled with Guppy Basecaller version 4.2.3+8aca2af with high accuracy mode.

Long-read nanopore sequencing analysis

For human reference alignment, the hg38 UCSC analysis set of December 2013 human genome (GCA_000001405.15) without alt-scaffolds was used along with its associated annotation file in GTF format. For HIV alignments a custom R7 strain reference sequence was generated *de novo* from reads from activated J-Lat 10.6 cells, an accompanying custom annotation file in GTF format was also generated containing all canonical splice variants. For alignment of host/viral chimeric reads at integration site, the R7 reference sequence and its associated annotation was inserted into the hg38 reference and annotation file using the Re-form Tool web portal (https://reform.bio.nyu.edu/) at the following hg38 positions and R7 orientations according to validated integration sites: chr1 77,946,384 (R7 antisense orientation) for J-Lat 8.4, and chr9 136,468,579 (R7 sense in sense orientation) for J-Lat 10.6 [70].

For extraction of HIV-host chimeric reads and analysis of HIV-host readthrough products, Nanopore reads were mapped to the R7 HIV reference genome using *minimap2*[123] with *map-ont* preset and *–secondary=no* option. R7-mapped reads were extracted from sam output using *samtools view* with following options

-h -F4 and converted to FASTQ using *samtools bam2fq*.[115] R7-mapped reads were aligned to unmodified UCSC hg38 reference using *minimap2* with *splice* preset and *–secondary=no* option. Hg38-mapped reads were extracted and converted into FASTQ as before. The resulting FASTQ file contains HIV/host chimeric reads that mapped to both R7 HIV and hg38 references. For readthrough analysis, HIV/host chimeric reads were mapped to the custom hg38 reference containing R7 HIV sequence using *minimap2* with *splice* preset and *–secondary=no* option. Mapped reads were extracted using *samtools view* with *-F4 -h* options and sorted using *samtools sort*. Sorted bam outputs were indexed with *samtools sort* and visualized using Integrative Genomics Viewer version 2.8.9.

Quantification and productivity analysis of host cell isoforms at integration site

For this workflow, the FLAIR isoform analysis pipeline was used.[122] Reads were mapped to the UCSC hg38 reference using FLAIR *align* module using option *-p*, followed by splice junction correction using the *correct* module with option *-c* and using hg38 gtf annotation file. Resulting bed files from FLAIR *correct* step were parsed to extract isoforms that overlap with *FUBP1* and *SEC16A* genes, the genes containing the HIV-1 integration sites in J-Lat 8.4 and 10.6, respectively. For this purpose, a bed file was generated containing the following boundaries for FUBP1 and SEC16A with care to avoid overlap with any neighboring genes: chr1: 77,949,155–77,979,130 (*FUBP1*), and chr9 136,440,096–136,482,938 (*SEC16A*). The *bedtools intersect* command was used with options *-wa* and *-a* to extract all isoform that intersect with the query regions, with resulting intersect bed file concatenated according to treatment.[116] The FLAIR *collapse* module was used with default settings to create a transcript model for *FUBP1* and *SEC16A* isoforms following treatment with HIV-1 activating agents. These transcript models were plotted with plot_isoform_usage.py script to obtain isoform structures and cross-referenced with hg38 gtf annotation files to predict productivity of each isoform using the predictProductivity.py script. Following isoform collapse, the FLAIR *quantify* module was used with option *–tpm* to obtain transcript per million (TPM) values for each isoform present in each replicate and treatment.

Analysis of publicly available datasets (ChIP-seq, TSA-seq, & SEduper)

For ChIP-seq analysis of publicly mined data (NCBI GEO), reads were mapped using Bowtie2[114] to hg38. PCR duplicates were removed using Picard Tools. MACS2[120] was used to call peaks using default parameters with IgG ChIP-seq data as input control. ChIP-seq datasets from Jurkat T cells include: H3K27ac (GSM2691418), Pol2 (GSM1850204), and BRD4 (GSM2218755). TSA-seq dataset of SPADs (GSM3111194) in K562 cells were processed as stated in Chen et al.[23] Predicted SE networks were obtained from dbSuper.[121]

Single-molecule DNA & RNA FISH

Branched DNA *in situ* hybridization (bDNA FISH) was used for detection of HIV-1 RNA in J-Lat cells using the RNAscope method, with modifications (1). One million cells per condition were harvested and spun down at 1,500 rpm for 5 min. These cells were resuspended in 50 μL of phosphate buffered saline (PBS) and seeded on poly-*d*-lysine (Gibco, cat: A3890401) coated coverslips for 30 min. Cells were then fixed in 4% paraformaldehyde for 30 min at room temperature (RT) and washed three times in PBS. Samples were then incubated in PBS supplemented with 0.1% Tween 20 for 10 min at RT and washed twice in PBS. The manufacturer's protease solution (Pretreat 3; Advanced Cell Diagnostics (ACD)) was diluted 1:5 in PBS and applied to samples. Samples were then incubated in a humidified HybEZ oven (ACD) at 40°C for 15 min. Protease solution was decanted, and samples were washed twice in PBS. A probe that recognizes HIV-1 (+) RNA (HIV-nongagpol-C3; ACD; 317711-C) and HIV-gagpol-C1 (317,701) diluted in hybridization buffer was applied to the sample. Incubation with the probe was performed at 40°C for 2 h in the HybEZ oven. The remaining wash steps and hybridization of preamplifiers, amplifiers, and fluorescent label were performed as performed previously.[73,74] HIV-1 (−) RNA or cDNA was labeled with ATTO 550, and HIV-1 (+) RNA was labeled with Alexa 647. Nuclei were counter-stained with 4′,6′-diamino-2-phenylinndole (DAPI, ACD) for 1 min at RT and washed twice in PBS. Coverslips were mounted on slides using Prolong Gold Antifade (Invitrogen).

Imaging was performed using a Nikon C2 confocal microscope with a 60x oil-immersion objective. The excitation/emission bandpass wavelengths used to detect DAPI, ATTO 550, and Alexa 647 were set to 405/420-480, 550/560-610, and 647/655-705 nm, respectively.

QUANTIFICATION AND STATISTICAL ANALYSIS

Analysis pipelines for Hi-C, ATAC-seq, ChIP-seq, and RNA-seq are presented in detail in the respective Method details section. For Hi-C analysis, single replicate libraries were generated per condition, with each library consisting of over one billion sequencing reads. Quality control analysis of the Hi-C libraries shown in Figure S3 indicates *cis:trans* interaction ratios representative of high quality Hi-C datasets. Quality control of ATAC-seq datasets is shown in Figure S4, where we plot size distribution of ATAC reads. For ATAC-seq, each condition was performed in biological duplicate. Publicly mined ChIP-seq datasets were used to assess protein and histone mark occupancy at defined HIV-1 integration sites. Enriched peaks within ChIP-seq datasets were called using MACS2[120] with default parameters and using the respective ChIP-seq IgG dataset as control. Long-read RNA-seq experiments were performed in biological triplicate per condition and subjected to a custom analysis pipeline outlined in detail in the respective Method details section.

Several libraries/packages of R v4.2.1 were used to generate quality control plots for Hi-C and ATAC-seq shown in Figures S3 and S4, such as 'ggplot2' and 'ngs.plot'. DESeq2[119] was used for data normalization of RNA-seq datasets and quantification of differential splicing (Figure S11). For visualization of read pile-up of the NGS datasets, we used Integrated Genomics Viewer (IGV).[71] Other details on statistics applied for specific analysis are provided in the respective figure legends.