

## RESEARCH ARTICLE

## Intragenic proviral elements support transcription of defective HIV-1 proviruses

Jeffrey Kuniholm<sup>1</sup>, Elise Armstrong<sup>2</sup>, Brandy Bernabe<sup>3\*</sup>, Carolyn Coote<sup>2</sup>, Anna Berenson<sup>4</sup>, Samantha D. Patalano<sup>4</sup>, Alex Olson<sup>2</sup>, Xianbao He<sup>2</sup>, Nina H. Lin<sup>2</sup>, Juan I. Fuxman Bass<sup>4</sup>, Andrew J. Henderson<sup>1,2,3,\*</sup>

**1** Boston University School of Medicine, Department of Microbiology, Boston, Massachusetts, United States of America, **2** Boston University School of Medicine, Department of Medicine, Section of Infectious Diseases; Boston, Massachusetts, United States of America, **3** Boston University School of Medicine Graduate Medical Sciences, Boston, Massachusetts, United States of America, **4** Boston University, Department of Biology, Boston, Massachusetts, United States of America

✉ Current address: Emory University, Laney Graduate School, Atlanta, Georgia, United States of America  
\* [hender@bu.edu](mailto:hender@bu.edu)



## OPEN ACCESS

**Citation:** Kuniholm J, Armstrong E, Bernabe B, Coote C, Berenson A, Patalano SD, et al. (2021) Intragenic proviral elements support transcription of defective HIV-1 proviruses. *PLoS Pathog* 17(12): e1009982. <https://doi.org/10.1371/journal.ppat.1009982>

**Editor:** Ronald Swanstrom, University of North Carolina at Chapel Hill, UNITED STATES

**Received:** September 25, 2021

**Accepted:** December 7, 2021

**Published:** December 28, 2021

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.ppat.1009982>

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

**Data Availability Statement:** Relevant data are within the manuscript and its [supporting information](#) files.

## Abstract

HIV-1 establishes a persistent proviral reservoir by integrating into the genome of infected host cells. Current antiretroviral treatments do not target this persistent population of proviruses which include latently infected cells that upon treatment interruption can be reactivated to contribute to HIV-1 rebound. Deep sequencing of persistent HIV proviruses has revealed that greater than 90% of integrated HIV genomes are defective and unable to produce infectious virions. We hypothesized that intragenic elements in the HIV genome support transcription of aberrant HIV-1 RNAs from defective proviruses that lack long terminal repeats (LTRs). Using an intact provirus detection assay, we observed that resting CD4+ T cells and monocyte-derived macrophages (MDMs) are biased towards generating defective HIV-1 proviruses. Multiplex reverse transcription droplet digital PCR identified *env* and *nef* transcripts which lacked 5' untranslated regions (UTR) in acutely infected CD4+ T cells and MDMs indicating transcripts are generated that do not utilize the promoter within the LTR. 5'UTR-deficient *env* transcripts were also identified in a cohort of people living with HIV (PLWH) on ART, suggesting that these aberrant RNAs are produced *in vivo*. Using 5' rapid amplification of cDNA ends (RACE), we mapped the start site of these transcripts within the *Env* gene. This region bound several cellular transcription factors and functioned as a transcriptional regulatory element that could support transcription and translation of downstream HIV-1 RNAs. These studies provide mechanistic insights into how defective HIV-1 proviruses are persistently expressed to potentially drive inflammation in PLWH.

## Author summary

People living with HIV establish a persistent reservoir which includes latently infected cells that fuel viral rebound upon treatment interruption. However, the majority of HIV-1 genomes in these persistently infected cells are defective. Whether these defective HIV

**Funding:** This work was in part funded by NIH including R01 AI138960, R01DA055488 and DA047032 R61/R33 to A.J.H. and NIH NIGMS R35 GM128625 to J.I.F.B. J.K. was supported by the Immunology Training Program (NIAID T32 AI007309-31A1), and B.B. was supported by BU PREP, R25 GM125511. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

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

genomes are expressed and whether they contribute to HIV associated diseases including accelerated aging, neurodegenerative symptoms, and cardiovascular diseases are still outstanding questions. In this paper, we demonstrate that acute infection of macrophages and resting T cells is biased towards generating defective viruses which are expressed by DNA regulatory elements in the HIV genome. These studies describe an alternative mechanism for chronic expression of HIV genomes.

## Introduction

HIV-1 establishes a persistent infection by integrating into the host genome that can only be fully eradicated by the elimination of infected cells. Although antiretroviral therapy (ART) restricts viral replication and disease progression, it only slowly diminishes the persistent reservoir [1,2]. Furthermore, a subset of persistently infected cells harbor transcriptionally repressed HIV-1 proviruses which, upon reactivation, fuel the rapid rebound of HIV-1 replication and infection upon ART discontinuation. Understanding the mechanisms that establish and maintain persistent and latent HIV-1 infections are required for effective HIV cure strategies.

Persistent HIV-1 infection is also postulated to contribute to inflammation and chronic immune activation associated with systemic diseases, even in people treated with antiretroviral drugs that have undetectable HIV-1 expression [3–10]. People living with HIV-1 (PLWH) receiving ART continue to suffer from inflammatory associated co-morbidities including neurological deficits, frailty, and cardiovascular disease [8]. It has been demonstrated that innate immune sensing of HIV-1 intron containing RNAs in infected macrophages, results in induction of MAVS-dependent interferon type I and pro-inflammatory responses [11,12]. However, whether other mechanisms contribute to this persistent inflammation and immune dysfunction are poorly understood.

Studies examining the proviral landscape using next generation deep sequencing approaches show that the HIV-1 reservoir is diverse and dynamic, changing over the course of time [13–16]. Approximately 90% of HIV-1 proviruses sequenced from peripheral blood samples from antiretroviral treated PLWH have crippling mutations including large deletions of the genome that would prevent the generation of infectious HIV-1 particles [17,18]. A subset of CD4+ T cells harboring defective proviruses generate polypeptides that are presented in the context of MHC I. Cells presenting such peptides can be targeted by CD8+ cytotoxic T lymphocytes (CTLs), thereby shaping the HIV-1 provirus reservoir and potentially redirecting adaptive HIV-1 immunity [19,20]. These reports indicate that subsets of defective proviruses are expressed and influence immune cell function. Importantly, what regulates the expression of defective proviruses and how they contribute to persistent HIV-associated inflammation has not been reported.

We hypothesize that defective provirus transcription is regulated by intragenic transcriptional elements. We demonstrate that acute infection of resting CD4+ T cells and monocyte-derived-macrophages are biased towards generating defective proviruses, and despite these defects still express HIV-1 RNAs and proteins. Importantly, we detect these mRNAs in a subset of PLWH on ART. We also identify the intragenic transcriptional elements that regulate the expression of these aberrant RNAs. These studies provide mechanistic insights into how defective HIV-1 proviruses are persistently expressed to potentially drive inflammation.

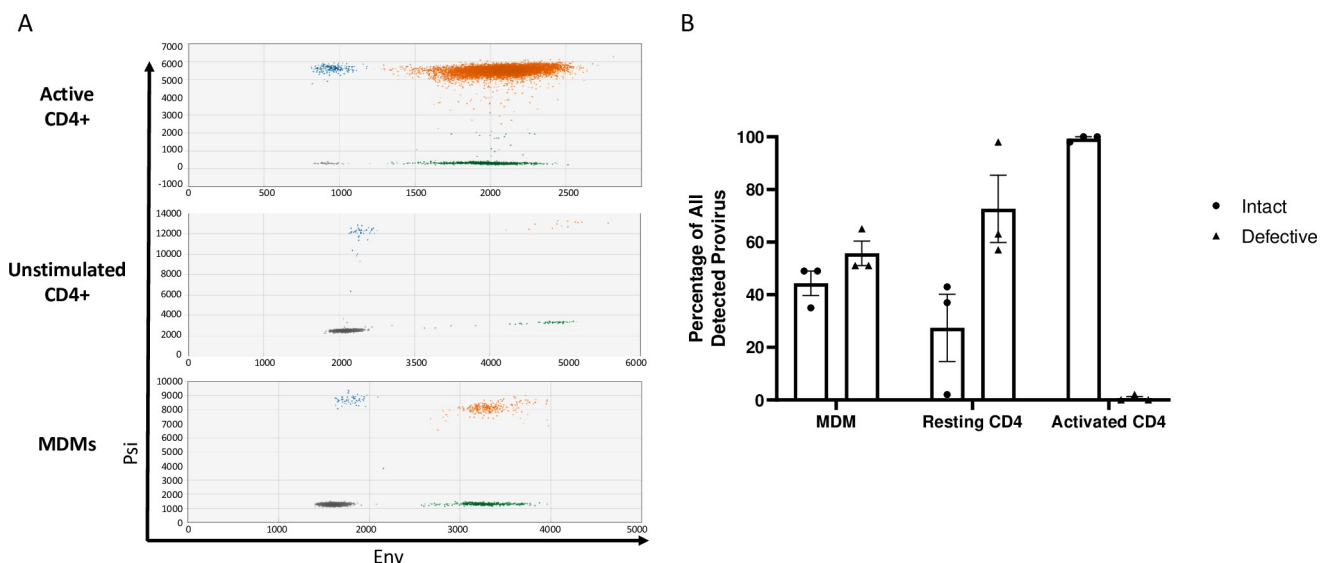
## Results

### Acute infection of resting CD4+ T cells and macrophages generate defective HIV-1 proviruses

We hypothesized that acute HIV-1 infection of resting and quiescent cells biases infection towards defective proviruses. To determine the ratio of intact and defective provirus present in resting CD4+ T cells, activated CD4+ T cells, and macrophages following acute HIV-1 infection, we utilized an intact provirus detection assay (IPDA) [18] (Figs 1A and S1A and S1 Table). CD4+ T cells were either unstimulated or activated with anti-CD3/anti-CD28 beads prior to infection with HIV-1<sub>VSVgNL4-3</sub>, whereas monocyte derived macrophages (MDMs) were infected with R5-tropic HIV-1<sub>NL4-3 BaL</sub>. Infected cells were cultured for at least 48-hours to allow the completion of reverse transcription and integration (S1B Fig). Genomic DNA from infected cells was used as template for IPDA droplet digital PCR (ddPCR). IPDA showed that activated CD4+ T cells had a higher frequency (>90%) of intact HIV provirus when compared to resting CD4+ T cells or MDMs (Fig 1A and 1B) which had on average 25% and 40% intact proviruses, respectively. Approximately 50–60% of proviruses detected in both resting CD4+ T cells and MDMs were defective with the majority of proviruses harboring 5' mutations that prevented detection by the *psi* probe. The non-nucleoside reverse transcriptase inhibitor Efavirenz (EFV) was used as a negative control to assure that IPDA signals were dependent on reverse transcription and were not contaminating DNA (S2 Fig). These results support that defective HIV proviruses are generated in both resting CD4+ T cells and MDMs after an acute infection, and the activation state of CD4+ T cells influences efficient establishment of intact proviruses.

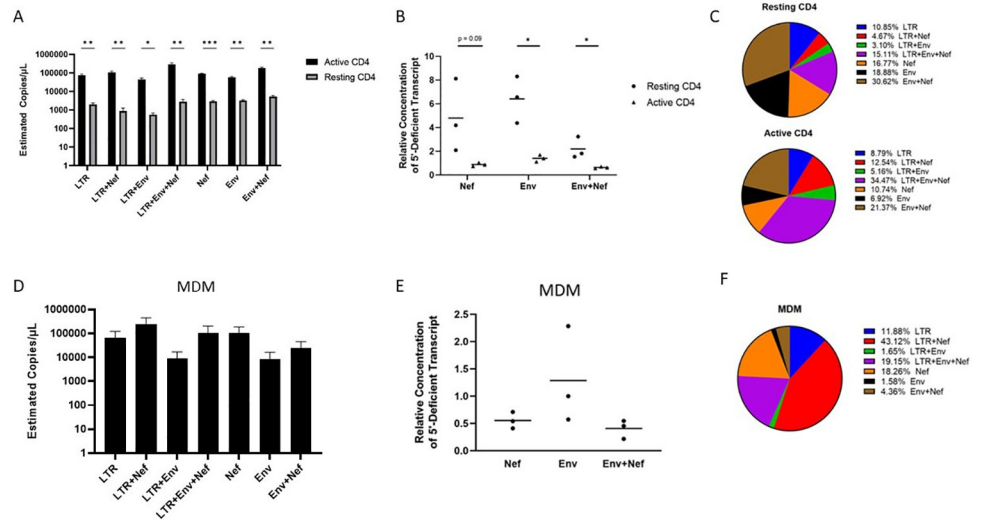
### Defective proviruses generate proviral RNAs that lack 5'untranslated regions

Several studies demonstrated that defective HIV proviruses are transcribed and translated [19,20]. To determine if the provirus populations detected in HIV-1 infected CD4+ T cells and



**Fig 1. Acute HIV-1 infection generates defective proviruses in MDMs and resting CD4+ T cells.** (A) Representative raw IPDA data of (top) activated CD4+ T cells, (middle) unstimulated CD4+ T cells, and (bottom) monocyte-derived macrophages (MDMs) infected with HIV-1. Droplets are color coded based on manual gating of positive and negative probe signals (gray = empty or double mutation/deletion droplets, blue = droplets single positive for *psi*, green = single positive *env* droplets, orange = double positive *psi* and *env* intact proviral droplets). A parallel reaction to detect the host cell gene *RPP30* was used as a correction for DNA shearing. (B) Percentages of intact and defective HIV genomes detected in MDMs and CD4+ T cells following HIV-1 infection. MDMs were infected with HIV-1<sub>NL4-3-BaL</sub> for 48 hours prior to DNA isolation. Resting and anti-CD3/CD-28 bead activated CD4+ T cells were infected with HIV-1<sub>NL4-3-VSVg</sub> for 72 hours prior to DNA isolation. Data represents three independent infections using cells generated from three different donors. Resting and activated CD4+ T cell sample data are participant matched.

<https://doi.org/10.1371/journal.ppat.1009982.g001>



**Fig 2. CD4+ T cells and MDMs express aberrant HIV-1 RNAs.** (A) Multiplex RT-ddPCR was performed on RNA prepared from unstimulated and activated CD4+ T cells infected with HIV-1<sub>NL4-3</sub> by spinoculation for 90 minutes at an MOI of 0.02 and incubated for 72 hours prior to RNA isolation. HIV-1 infection was limited to a single round by addition of 10 μM saquinavir 30 minutes after spinoculation. Activated CD4+ T cells were cultured with anti-CD3/CD28 beads for 72 hours prior to spinoculation. Estimated copies/μL for transcripts were calculated with the QuantaSoft droplet reader software after manual gating of multiplex RT-ddPCR data (See S3 Fig) and dilution factor correction. (B) Ratios of HIV 5'UTR defective transcripts relative to transcripts with intact 5'UTRs for resting and activated HIV infected CD4+ T cells. (C) Frequencies of detected HIV transcripts as a percentage of all transcripts quantified by multiplex RT-ddPCR for resting (top) and CD3/CD28 activated CD4+ T cells. (D) Multiplex RT-ddPCR performed with RNA from HIV-1<sub>NL4.3-BaL</sub> infected MDMs. MDMs were infected with HIV-1<sub>NL4.3-BaL</sub> at an MOI of 0.01 for 4 hours by adding virus directly to adherent cell cultures for 48 hours prior to RNA isolation. (E) Ratios of 5'UTR defective transcripts relative to transcripts with intact 5'UTR for HIV infected MDMs. (F) Frequencies of HIV-1 transcripts as a percentage of all transcripts quantified by multiplex RT-ddPCR for HIV infected MDMs. Data represent three independent experiments using cells from three donors. Resting and activated CD4+ T cell sample data are donor matched. Bars represent standard error of the mean. Significance values were generated using multiple unpaired Two-tailed T tests. \* denotes  $p < 0.05$ ; \*\* denotes  $p < 0.01$ , and \*\*\* denotes  $p < 0.001$ .

<https://doi.org/10.1371/journal.ppat.1009982.g002>

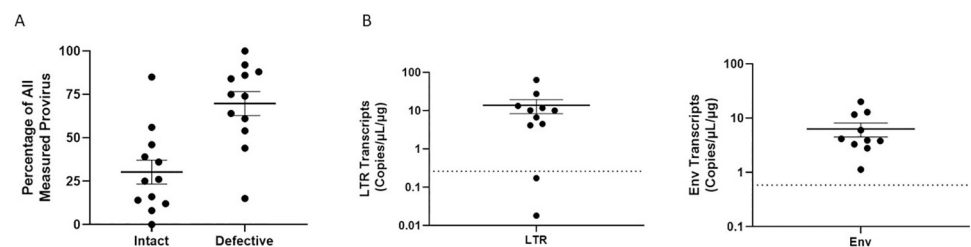
MDMs were transcriptionally competent, we adapted a reverse transcription ddPCR assay (RT-ddPCR) originally described by Yukl et al [21] by multiplexing probes for simultaneous detection of transcripts spanning the R-U5-Gag junction (“Late LTR” probe), distal transcripts (“Nef” probe), and sequences that included both upstream and distal sequences (5’ UTR/Nef double positives, S3 Fig) [22,23]. Multiplex RT-ddPCR performed using infected CD4+ T cell cDNA detected transcripts which included 5’ untranslated region (UTR) sequence (R-U5-gag positive) only and double-positive 5’ UTR-Nef transcripts. We also detected *nef*-positive transcripts which did not include 5’ UTR sequence in HIV infected CD4+ T cells suggesting these transcripts were not generated from the transcriptional start site located in the 5’ LTR (Fig 2A). We further multiplexed the RT-ddPCR by including an *env* primer and probe combination allowing for the detection of 7 different HIV-1 RNAs (S3 Fig). The multiplex RT-ddPCR identified two general sets of transcripts in CD4+ T cells and MDMs; a set of RNAs that contained 5’ UTR sequences and a set of RNAs that were positive for distal RNAs *env* and *nef* but lacked 5’ UTRs. The relative levels of RNAs with and without 5’ R-U5 sequence differed between the infected CD4+ T cell populations with approximately 40% of RNAs in activated CD4+ T cells lacking 5’ UTRs compared to 66% of RNAs lacking 5’ UTRs in resting cells. These frequencies correlated with IPDA data which showed that resting CD4+ T cells harbor more defective proviruses (Fig 2B and 2C). We also observed approximately 25% of RNAs in infected MDMs lacked 5’ UTRs (Fig 2D–2F). Using a poly d(T) primer instead of random

hexamers for the RT reactions resulted in the amplification and detection of 5' UTR deficient transcripts in CD4+ T cells and MDMs, indicating that a subset of these RNAs were polyadenylated (S4 Fig). These data support that aberrant HIV transcripts are produced after acute HIV-1 infection, with their relative expression influenced by cell type and activation state.

Chronically infected individuals on ART demonstrate a gradual decrease in intact HIV proviruses concurrent with selection and expansion of defective proviruses due to sustained immunological pressure and clearance [15,17–19,24]. Notably, defective HIV proviruses are observed to decay at a slower rate than intact provirus clones, presumably due to their relatively low HIV-1 expression levels [13,23–26]. We posit that HIV-1 mRNA lacking 5' UTRs should be detected in people living with HIV (PLWH) on ART. Using DNA isolated from PBMCs of PLWH on ART (S2 Table), we first confirmed the presence of HIV-1 proviruses in chronically infected individuals by IPDA with most proviruses detected in these individuals being defective (Fig 3A). To determine if aberrant RNAs that lack 5'-UTRs were also present in individuals chronically infected with HIV-1, multiplex RT-ddPCR was performed using RNA from PLWH donor PBMCs. Probes specific for R-U5 5' UTR sequences and *env* were used in tandem to determine the presence of one or both probe sites on HIV-1 transcripts expressed in HIV-positive donor PBMCs. Multiplex RT-ddPCR assays successfully detected HIV-1 RNAs with 5' UTR as well as *env* transcripts that lacked 5' UTRs start sites (Fig 3B and 3C). *Env* sequences lacking 5' UTRs were detected with signal above uninfected PBMC background controls in 10/11 RNA samples from HIV-positive PBMC donors (Fig 3B and S2 Table). These data show that aberrant HIV transcripts are detected in a subset of individuals chronically infected with HIV even during ART suggesting these RNAs are persistently expressed in PLWH.

### Defective proviral transcripts are generated from intragenic transcriptional elements

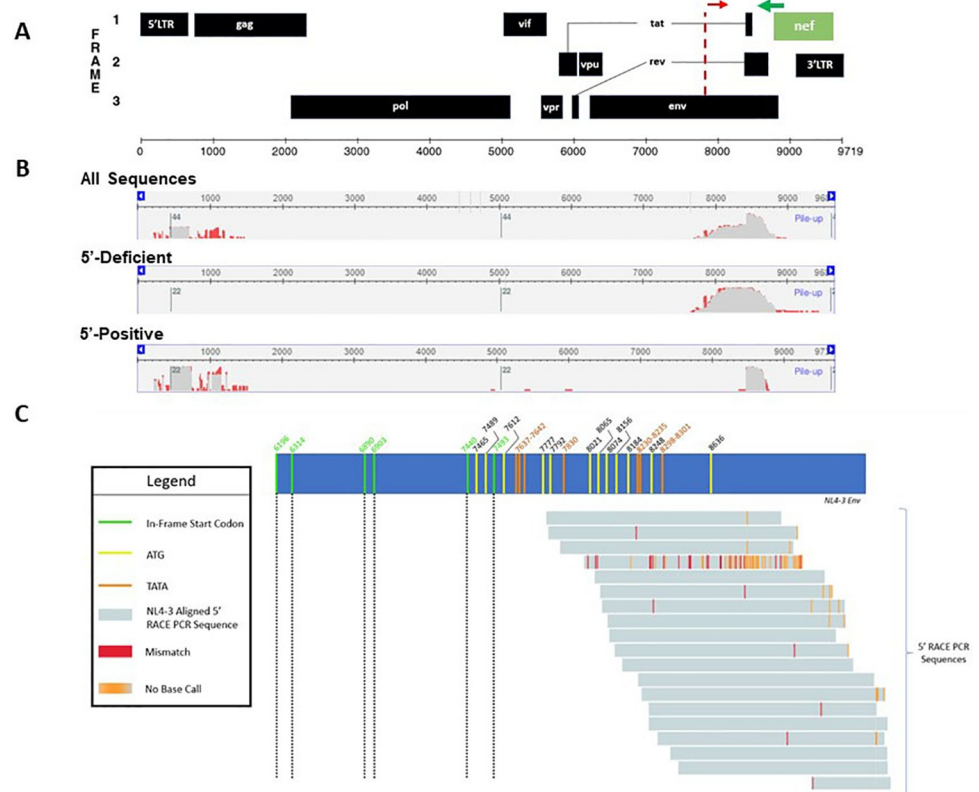
In the context of intact proviruses, the 5' LTR includes the cis-regulatory elements that recruit and coordinate host transcription factors to initiate and regulate HIV-1 transcription. The mechanisms by which HIV transcription is regulated in defective HIV proviruses that lack a functional 5' LTR remains incompletely described. We hypothesized that there are intragenic sequences that support transcription of defective HIV-1 proviruses. Previous reports that support this hypothesis include descriptions of transcription factor binding and DNase I hypersensitive sites downstream of the HIV-1 LTR [27–32].



**Fig 3. Detection of HIV-1 aberrant transcripts in PLWH on ART.** (A) Percentage of intact and defective HIV provirus detected in DNA samples from PBMCs of HIV-positive ART-treated individuals relative to all provirus quantified by IPDA. (B) Multiplex RT-ddPCR was performed to detect intact 5' UTR containing transcripts (LTR) and 5' UTR-deficient *env* transcripts (*Env*) using RNA samples prepared from PBMCs of HIV-positive ART-treated individuals. No RT enzyme control reactions were carried out for each assay and background signal was subtracted from experimental readings. Dots represent estimated copies/uL of single positive transcripts per  $\mu\text{g}$  of RNA. Dotted line represents the mean assay background signal calculated from five HIV-negative PBMC samples. Each point represents an independent participant sample. A total of 11 participant samples were assessed.

<https://doi.org/10.1371/journal.ppat.1009982.g003>

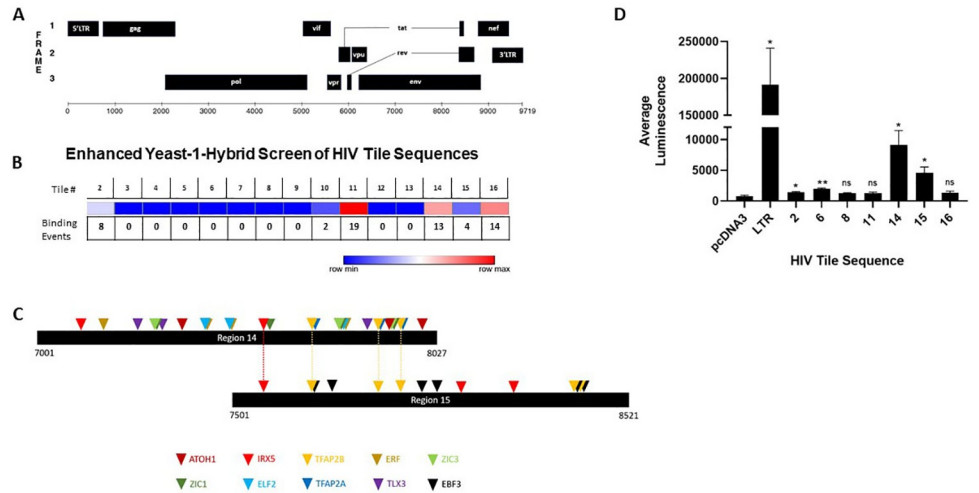




**Fig 4. Sequencing of 5' RACE PCR products identified HIV-1 transcripts with intragenic start sites.** (A) Schematic of HIV-1 genome with putative promoters indicated as dotted red lines as predicted by Promoter 2.0 [33] and ElementNT V2 [34]. The green line represents the 5' RACE gene-specific Nef primer. (B) 5' RACE results analyzed using the NCBI Multiple Sequence Alignment Viewer. (Top) All 5' RACE sequences aligned to NL4-3 sequence (N = 41 sequences). (Middle) 5' RACE sequences which lacked 5' UTR aligned to NL4-3 sequence (N = 19). (Bottom) 5' RACE sequences which contained 5' UTR aligned to NL4-3 sequence (N = 22). (C) Schematic alignment of 5' RACE sequences lacking 5' UTR sequence (gray bars, N = 19) with the NL4-3 env gene (blue bar). Green and black dotted lines indicate in-frame start codons (numbered using GenBank: AF324493.2). Yellow lines represent out of frame start codons. Orange lines (top) represent TATA box sequences. Red lines represent sequence mismatches between 5' RACE reads and NL4-3 env. Orange lines (bottom) indicate that no nucleotide was assigned for that position during the sequencing reaction. Start codons and TATA box sequences were mapped and multiple sequence alignments were performed using Benchling Biology Software (<https://benchling.com>, 2021).

<https://doi.org/10.1371/journal.ppat.1009982.g004>

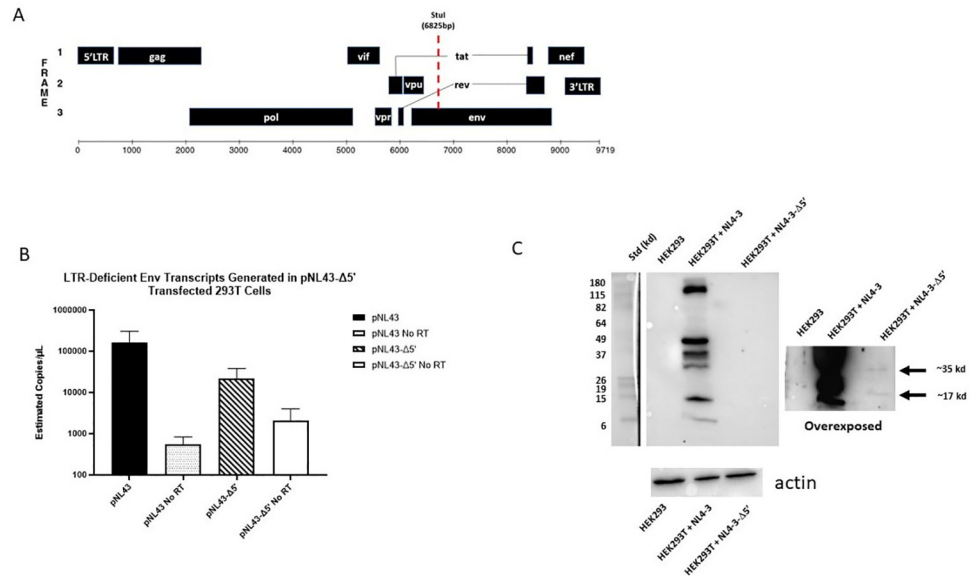
To characterize transcriptional start sites of defective HIV transcripts, we performed 5' RACE PCR using RNA isolated from HIV infected activated CD4+ T cells and MDMs. A universal forward primer and a gene-specific reverse primer for *nef* (S1 Table) were used to amplify the populations of *nef* containing transcripts from infected cells. PCR products were separated by gel electrophoresis and ~1kb RACE PCR products were gel extracted, cloned into a pUC19-based vector, and 41 products were sequenced to map their 5' ends (S3 Table). MDM and CD4+ T cell sequences demonstrated that 53.6% of the transcripts were splice products which initiated at the transcriptional start site within the 5' LTR sequence (Fig 4A and 4B). The remaining 46.3% of sequences initiated transcription within *env*, across a span of approximately 200 nucleotides located between +7760 to +8178 bp downstream of the canonical 5' LTR transcriptional start site (Fig 4C and S3 Table). These data support that transcription initiates from intragenic sites of the HIV genome. Based on the RACE data, we used Promoter 2.0 [33] and ElementNT V2 [34] to predict if promoter characteristics were within the *env* region. These bioinformatic tools predicted a transcriptional start site (7800 bp, score 1.141) and



**Fig 5. Identification of intragenic transcriptional elements.** (A) Schematic of HIV-1 genome. (B) Heat map for transcription factor binding within the HIV-1 provirus. HIV-1 tile sequences (1–16) were used in yeast-1-hybrid screen (See S4 Table). Dark blue represents regions lacking transcription factor binding whereas red indicates regions of transcription factor binding. Binding events represent the number of transcription factors that bound the specific HIV-1 genomic sequence. (C) Location of predicted binding sites for transcription factors identified by yeast-1 hybrid assay that bind baits 14 and 15. Motifs were identified using FIMO with a threshold set at  $p < 0.001$  ([38] and S6 Table). (D) Intragenic sequences enhance transcription. HIV-1 tile sequences were cloned upstream of a luciferase gene in pcDNA3.1 Luciferase and transfected in HEK-293T cells. Luciferase was measured and reported as fold-change in signal over cells transfected with the negative control pcDNA3.1-Luc plasmid. HIV-LTR-Luciferase was used as a positive control. These data include three independent transfections performed in triplicate. Error bars represent standard error of the mean for all values. Two-tailed unpaired T-tests were used to calculate statistical significance between raw luminescence values for HIV-1 tile sequences and negative control pcDNA3.1-Luc plasmid. \* denotes  $p < 0.05$ ; \*\* denotes  $p < 0.01$ , and \*\*\* denotes  $p < 0.001$ .

<https://doi.org/10.1371/journal.ppat.1009982.g005>

TATA box motifs (7483–7488 bp and 7662–7669 bp) within the *env* sequence, adjacent to the start sites identified by 5' RACE (Fig 4B and 4C). Alignment of 5' RACE PCR sequences to the NL4-3 *env* gene revealed a cluster of out-of-frame start codons in this region, suggesting that multiple alternative transcriptional start sites may be operative in the context of defective provirus (Fig 4C). To survey direct binding of human transcription factors within the HIV proviral genome, we performed a functional enhanced yeast one-hybrid screen (eY1H) [35–37]. This method allows for high-throughput screening of host transcription factor binding events to the HIV-1 genome. In brief, HIV bait sequences were generated by tiling the HIV-1<sub>NL4-3</sub> genome into 17 overlapping fragments (S4 Table). Fragments were cloned upstream of the two reporter genes, *LacZ* or *HIS3*. Yeasts were transformed with reporter constructs as well as a library representing 66% of the known human transcription factors fused to the yeast activation domain, *Gal4*. Transcription factor binding hits were identified as yeast colonies that grew in the absence of histidine and were blue when provided X-Gal. The eY1H screen detected 60 transcription factor-DNA interactions, representing 43 unique transcription factors which bound to six regions within the HIV-1 genome (designated 2, 10, 11, 14, 15 and 16) (S5 Table and Fig 5A and 5B). Transcription factor binding hot spots were clustered within and near HIV *env* gene (Fig 5A–5C), whereas, transcription factor binding was not detected in *pol* (Regions 3–9). Identified transcription factors included homeobox, basic helix-loop-helix, and zinc finger families (S5 Table). The distribution of transcription factor binding hotspots near the HIV *env* gene correlated with transcriptional start sites suggested by the 5' RACE results. To investigate if these intragenic sequences that bound host transcription factors function as cis-regulatory elements, reporter vectors were generated with HIV-1 fragments cloned



**Fig 6. Truncated HIV-1 proviral construct expresses RNA and protein.** (A) Schematic of HIV-1 NL4-3 in pNL4-3 expression plasmid and location of *StuI*-restriction digest site used to generate pNL4-3Δ5' (bottom). (B) RT-ddPCR data from HEK-293T cells transfected with either full length pNL4-3 (black) or pNL4-3Δ5' (striped) for 24-hours. RNA samples were subjected to multiplex ddPCR using a probe for the 5'LTR and a probe for env and droplets were quantified as indicated above. Reaction mixtures without RT enzyme (No RT) were used as negative controls for contaminating plasmid DNA. No LTR signal was detected for HEK293T cells transfected with pNL4-3Δ5'. Data include 4 independent pNL4-3 transfections and three pNL4-3 Δ5' transfections. (C) Western blots using cell lysate from HEK293T cells transfected with either full length pNL4-3 plasmid or pNL4-3Δ5'. Lysates were run on 12% SDS-PAGE gel, transferred to a PDVF membrane, and probed with human polyclonal anti-HIV IgG. Two exposures are shown: a short exposure demonstrating HIV-1 proteins (left) and an extended exposure which shows weaker HIV protein bands at approximately 35 kDa and 19 kDa, as indicated by the arrows (right). Non-transfected HEK293T cell lysates were also probed as a negative control. The lower panel is a separate gel from the above lysates probed with a monoclonal against β-actin antibody demonstrating the presence of proteins in the different lysates.

<https://doi.org/10.1371/journal.ppat.1009982.g006>

upstream of the luciferase reporter gene. Reporter plasmids were transiently transfected in HEK293T cells, and firefly luciferase was assessed 24 hours post-transfection (Fig 5D). Most of the intragenic sequences, including Regions 11 and 16, which bound several transcription factors in the eY1H assays, modestly influence transcription of the reporter gene. However, regions 14 and 15, which spanned sequences 7001 to 8521 bp in *env* and bound several transcription factors in the eY1H assays, exhibited 5–10 fold higher luciferase signal compared to pcDNA3 control. These sequences also spanned sites that included putative promoter elements and the RACE transcriptional start sites. The predicted transcription factor binding motifs generated using Finding Individual Motif Occurance (FIMO) software [38] for the sequences that span regions #14 and #15 are shown in S6 Table.

In addition, we generated a viral expression construct, pNL4-3Δ5' in which we deleted approximately 6,000 bp of the 5' end of the provirus, including the 5' LTR, *gag* and *pol* genes, and transfected this construct into HEK293T cells (Fig 6A). Transcription of *env* containing sequences was detected by RT-ddPCR a log-fold higher than the assay background (Fig 6B). Furthermore, 35 kd and 19 kd proteins were detected by immunoblotting when whole cell lysates from pNL4-3Δ5' transfected cells were probed with human HIV anti-serum suggesting that these RNAs are translated (Fig 6C). It should be noted that immunoblots required longer exposure times to visualize the proteins expressed from pNL4-3Δ5', probably reflecting that these intragenic elements are much weaker than the HIV-1 5'-LTR which increased luciferase expression on the order of ~200-fold compared to pcDNA control in transfected HEK293T



cells (Fig 5D). Taken together, these data support that HIV-1 intragenic sequences, including those within the *env* gene, act as cis-regulatory elements which support the transcription of cryptic RNAs from a subset of defective proviruses.

## Discussion

The reservoir of persistent HIV-1 proviruses is dynamic, being shaped over time by immune recognition and clearance of cells expressing HIV-1 proteins [17,18,39], as well as clonal expansion and homeostasis of memory T cell subsets that harbor HIV-1 proviruses [13,14,17]. This model is supported by longitudinal tracking of the persistent HIV-1 proviruses in PLWH on ART and long-term controllers which demonstrated a relatively rapid decay of intact HIV-1 provirus compared to defective proviruses [23–25,40]. These dynamics result in the progressive accumulation of defective HIV-1 proviral genomes which comprise the vast majority of the latent reservoir in chronically infected individuals that are undergoing treatment or are controllers of HIV infection [13–16,23,24,40]. In this study, we provide evidence that a subset of these defective proviruses are expressed during acute HIV-1 infection and provide insights into the mechanisms that control their transcription.

We hypothesized that certain cells may be predisposed to generating defective proviruses upon infection. It has been observed that quiescent and resting cell populations such as macrophages, resting CD4+ cells and memory T cell subsets are difficult to infect and biased towards unproductive or latent infections [41–43]. We compared the fate of HIV-1 infections in macrophages, unstimulated CD4+ T cells, and CD4+ T cells activated through the CD3/CD28 signaling axis. At the time of infection, macrophages and unstimulated CD4+ T cells harbor more defective proviruses compared to activated CD4+ T cells with up to 80% of the infected cells containing 5' deletions or mutations. Our results suggest that upon acute infection, even in the absence of immune selection, a foundation of cells harboring defective viruses is rapidly established [44].

The shaping of the reservoir and persistent proviral genomes has been linked to provirus expression and selection by anti-HIV immune responses. For example, it has been shown that a subset of defective viruses from individuals treated with ART can be expressed and targeted by CD8+ T cells [17,39]. However, it is not clear how these defective proviruses are transcribed, especially if a significant number of proviruses have large upstream deletions or LTR mutations. The existence of intragenic promoters has been suggested for several retroviruses [45–49]. The presence of proviral cis-regulatory elements has also been proposed within the HIV-1 genome, with reports of DNase hypersensitive sites, methylated CpG islands, transcription factor DNA binding sites and modest transcriptional activation potential [27,31,32,50,51]. However, the regulatory mechanisms which control transcription of alternative transcriptional start sites in HIV remain largely undescribed. We hypothesized that intragenic cis-elements exist and drive transcription of a subset of the defective HIV proviral genomes. Taken together, our data generated from yeast-one-hybrid transcription factor binding assays, promoter prediction software, transfections with deletion and reporter plasmids, and 5' RACE sequencing, indicate cis-transcriptional elements and promoters located at approximately +7,800 bp in the HIV-1 proviral genome. This observed transcriptional activity overlaps with previously reported CpG islands [51]. The transcripts generated originate in the *env* gene, span the *tat* splice junction, and include intronic sequence.

Despite minimal detectable virus replication in PLWH undergoing anti-retroviral treatment, immune stimulation and inflammation persist, contributing to co-morbidities including neuroinflammation, cardiovascular disease, and signs of accelerated aging [52–54]. Previous work has indicated that HIV-1 transcripts are produced by defective proviral clones and a

subset of these transcripts can be translated into viral proteins that stimulate CD8+ T cell activity [19,20,39]. Furthermore, cryptic HIV-1 peptides are produced by using alternative reading frames (ARFs) which are distributed throughout the HIV-1 genome. Previous studies have demonstrated that ARFs can be loaded into and presented by MHC-I to activate CD8+ T cells, inducing cytokine release and killing activity [55–57]. Taken together, these data support that defective persistent HIV proviruses influence and perpetuate immune responses against HIV. Our ability to detect HIV-1 protein from proviral sequences that lack intact 5' sequences suggest that defective viruses may provide an alternative mechanism to generate ARFs.

In addition, it has been demonstrated that partially spliced and unspliced HIV-1 RNAs are recognized by MAVS-dependent nucleic acid sensing pathways and mediate IFN-I responses [11,12,58]. The defective mRNA sequences generated from the intragenic elements that we have characterized initiate transcription within *env* and span the intron leading us to speculate that these mRNAs could be proinflammatory. Importantly, we detect transcripts that are initiated from intragenic elements in samples from PLWH on ART. Overall, our study provides potential insight into the mechanisms and proviral elements which regulate transcription of these defective HIV-1 genomes and the potential of cryptic peptides that perpetuate HIV pathogenesis.

## Methods

### Ethics statement

These studies utilize PBMCs from PLWH treated with ART enrolled in a study assessing inflammation and immune dysfunction in people living with HIV and aging which was approved by Boston University IRB (IRB-33095) [10]. Inclusion criteria included individuals between 18–35 and  $\geq 50$  years old who were on antiretroviral treatment for at least 6 months. Subjects with recent active infection (30 days) or on immunosuppressive therapy were excluded from enrollment. Written informed consent was obtained prior to specimen and data collection. See [S2 Table](#) for details. PBMCs were isolated using Ficoll-Hypaque density gradient separation and cryopreserved at  $-150^{\circ}\text{C}$  until use.

### Cells

HEK293T cells (ATCC) were maintained and cultured in Dulbecco's Modified Eagle Medium (DMEM; Invitrogen) supplemented with 100 U/mL penicillin/streptomycin (P/S; Invitrogen), 2 mM L-glutamine (Invitrogen), and 10% fetal bovine serum (FBS) (Gemini Bio-Products). CD4+ T cells were enriched from peripheral blood mononuclear cells (PBMC) from leukapheresis packs (New York Biologics). Negative selection was performed using EasySep Human CD4+ T Cell Enrichment Kit (Stemcell Technologies) after Lymphoprep density gradient (Stemcell Technologies) separation of PBMCs. Resting CD4+ T cells were cultured in Roswell Park Memorial Institute medium (RPMI; Invitrogen) with 100 U/mL P/S, 2 mM of L-glutamine, and 10% FBS. Flow cytometry using CD25 and CD69 markers confirmed resting CD4+ T cell phenotypes [42]. CD4+ T cells were activated with anti-CD3/CD28 Dynabeads (Invitrogen) at a ratio of one bead per cell for 72 h and maintained in RPMI with 100 U/mL IL-2 (AIDS Reagents Program) and 100 ng/mL IL-7 (Miltenyi Biotec). Monocyte-derived-macrophages (MDMs) were generated by positive selection of CD14+ cells from PBMCs using the EasySep Human CD14 Positive Selection Kit II (STEMCELL Technologies, Cat: 17858) and differentiating cells for at least 7 days in RPMI supplemented with 10% Human AB Serum (Millipore Sigma), 100U/mL P/S, 2mM of L-glutamine, and 20 ng/mL M-CSF (BioLegend, Cat: 574802) at  $37^{\circ}\text{C}$  and 5%  $\text{CO}_2$ . 50% fresh media was added 2 days after plating followed by full media changes every 2–3 days to remove non-adherent cells.

## Viruses and infections

Viral stocks of molecular clones HIV-1<sub>NL4-3</sub> and HIV-1<sub>NL4-3 BaL</sub> (AIDS Reagents Program) were generated by transfecting HEK293T cells using PEI transfection reagent (Sigma, Cat: 408727). Pseudotyped HIV-1<sub>VSVg-NL4-3</sub>, HEK293T cells were co-transfected with an expression vector containing VSVg and HIV-1<sub>NL4-3</sub>. HIV-1<sub>NL4-3</sub> virus stocks titers were estimated using CEM-GFP cells obtained from the AIDS Reagent Program [59]. HIV-1<sub>NL4-3 BaL</sub> virus stocks were titrated on primary MDMs and IU/mL were estimated from IPDA results. All infections were performed using MOI < 1 (HIV-1<sub>VSVg-NL4-3</sub> MOI = 0.1, HIV-1<sub>NL4-3</sub> MOI = 0.02, HIV-1<sub>NL4-3 BaL</sub> MOI = 0.01). DNase/MgCl<sub>2</sub> was added to cells at time of infection prior to DNA/RNA isolation to limit contaminating plasmid DNA. CD4+ T cells were infected by spinoculation as previously described [60]. MDMs were infected by adding HIV-1<sub>NL4-3BaL</sub> directly to the cells at 37°C for 4 h, washed twice with PBS and maintained in fresh media. 1μM of the non-nucleoside reverse transcriptase (RT) inhibitor Efavirenz was added prior to infection and post-infection as a negative control.

## DNA/RNA isolation

DNA and RNA were isolated from cells using the AllPrep DNA/RNA Mini Kit (Qiagen, Cat: 80204). Briefly, cells were washed twice with PBS, treated with DNase for 10 min prior to lysis directly in cell culture plate (MDM) or as cell pellets (CD4+ T cells) using Qiagen RLT buffer supplemented with β-Mercaptoethanol. Cell lysates were homogenized with Qias shredder columns (Qiagen, Cat: 79654) following manufacturer's protocol for AllPrep DNA/RNA Mini Kits. RNA isolation columns were treated with DNase for > 10 minutes to remove contaminating DNA.

## Intact Proviral DNA Assay (IPDA)

Probes targeting 5' and 3' regions of the HIV provirus were designed to bind minimally variant locations as described in detail by Bruner et al [18]. Briefly, primers and probes to amplify signal from the psi region and RRE of the env region of the HIV-1 provirus simultaneously were used (S1 Table). Signals were identified using a Bio-Rad QX200 Droplet Digital PCR System and Quantasoft Data Analysis Software.

A hydrolysable FAM probe was designed to bind Psi sequence whereas hydrolysable VIC probe was designed to bind the RRE sequence. Primers flanked probes. An unlabeled "dark probe" specific for a commonly hypermutated env sequence was included in each IPDA reaction to distinguish hypermutated provirus sequences. Proviruses containing hypermutated signals were detected as single-positive droplets containing FAM-only signal.

## Droplet Digital PCR (ddPCR)

Droplet digital PCR Reaction mixes were comprised of: 1X ddPCR Supermix (no dUTP) (Bio-Rad, Cat: 186-3024), 750 nM primers, 250 nM probes. DNA was diluted to ~10–100 ng in 8 μL of nuclease free H<sub>2</sub>O per reaction. A control reaction using probes and primers for human RPP30 gene was performed to calculate shearing index and correct the estimated concentrations of intact and defective HIV-1 proviruses. Droplets were generated using Bio-Rad's QX200 Droplet Generator (Cat: 1864002).

PCR conditions were: 95°C for 10 min enzyme activation, 94°C for 30 sec denaturation, 59°C for 1 min annealing/extension, repeat denaturation and annealing/extension 45 cycles, 98°C for 10 min to inactivate enzyme.

Droplets containing PCR product were assessed on the Bio-Rad QX200 Droplet Digital PCR System. Data were collected using Quantasoft Software set to copy-number variation (CNV) mode. All experiments included a no template (water only) control and/or an uninfected cell sample negative control for background probe signal. RPP30 single probe was used to estimate shearing index and to correct the estimation of defective and intact proviruses. Estimations of intact and defective provirus frequencies were calculated using droplet counts of single- and double-positives determined by the Quantasoft software.

For RT-ddPCR, RNA was reverse transcribed to generate cDNA with RT reaction mixtures of 14  $\mu\text{L}$  of 5 $\times$  SuperScript III buffer, 7  $\mu\text{L}$  of 50 mM MgCl<sub>2</sub>, 3.5  $\mu\text{L}$  of random hexamers (50 ng/ $\mu\text{L}$ ), 1.35  $\mu\text{L}$  of oligo dT15 (500ug/mL), 3.5  $\mu\text{L}$  of 10 mM deoxynucleoside triphosphates (dNTPs), 1.75  $\mu\text{L}$  of RNaseOUT (40 U/ $\mu\text{L}$ ; Invitrogen), 2  $\mu\text{L}$  of SuperScript III RT (200 U/ $\mu\text{L}$ ; Invitrogen), and nuclease-free dH<sub>2</sub>O to bring the final reaction volume to 70  $\mu\text{L}$ . RT was performed at 25°C for 10 min, 50°C for 50 min, followed by an inactivation step at 85°C for 5 min. Negative controls included reactions with no RT enzyme. [S1 Table](#) shows primers and probes for ddPCR. FAM probes were designed and directed to sites throughout expected HIV transcripts. RT-ddPCR probe design was based on those reported by Yukl et al [21]. ddPCR probes were multiplexed to measure the frequency of probe sites simultaneously in ddPCR assays.

## 5' RACE PCR

5'-RACE PCR was performed on total RNA from VSVg-NL43 infected CD4+ T cells and MDMs following the manufacturer's protocol for SMARTer RACE 5'/3' Kit (Takara Bio, Cat: 634858). Random primers were annealed to template RNA using 10X Random Primer Mix, RNA (1 $\mu\text{g}$ ), and water and incubating the mixture at 72°C for 3 min followed by cooling to 42°C for 2 min. SMARTer II A Oligonucleotide (1 $\mu\text{L}$ ) was added to tag the 5'-end of the generated cDNA providing a universal forward primer binding site for subsequent PCRs. The RT reaction included template RNA annealed with random primers, 5X First-Strand Buffer, DTT (100 mM), dNTPs (20 mM), RNase Inhibitor (40 U/ $\mu\text{L}$ ), Reverse Transcriptase (100U) and was performed at 42°C for 90 min and then cooled to 70°C for 10 min. This first-strand reaction product was diluted with 90  $\mu\text{L}$  of Tricine-EDTA Buffer before RACE-PCR. Sequences generated are shown in [S3 Table](#).

For 5' RACE PCR a nef specific reverse primer was designed (*See S1 Table*). The RACE reaction mixture consisted of 5' RACE PCR sample: PCR-Grade H<sub>2</sub>O (15.5  $\mu\text{L}$ ), 2X SeqAmp Buffer (25  $\mu\text{L}$ ), SeqAmp DNA Polymerase (1  $\mu\text{L}$ ), 5' RACE-Ready cDNA (2.5  $\mu\text{L}$ ), 10X Universal Primer Mix (5  $\mu\text{L}$ ), 5' gene-specific Primer (10  $\mu\text{M}$ ). Negative controls included only the universal primer mix or only the gene-specific primers. 5' RACE PCR thermal cycling was: 94°C for 30 sec, 68°C for 30 sec, 72°C for 3 min, repeated for 25 cycles. 5' RACE-PCR products were separated by gel electrophoresis (1% agarose) and major PCR products were excised, pooled, and subcloned into a linearized puC19-based vector (pRACE) using linearized pRACE vector (1 $\mu\text{L}$ ), Gel-purified RACE product (7  $\mu\text{L}$ ), and In-Fusion HD Master Mix (2  $\mu\text{L}$ ). The ligation reaction was incubated for 15 min at 50°C before transforming Stellar Competent Cells (TakaraBio). Colonies were selected for direct bacterial colony sequencing using M13F/R primers which flanked the In-Fusion cloning site of the pRACE vector (Genewiz).

## Enhanced yeast-1-hybrid screen

Enhanced yeast one-hybrid (eY1H) assays were performed as previously described [35–37]. Briefly, the HIV-1 NL4-3 genome was tiled into 17 overlapping fragments of 1000-1200bp in length ([S3 Table](#)). "Bait" HIV DNA were cloned upstream to the yeast selection marker *HIS3*

or the reporter gene *LacZ*. Constructs were transformed and integrated into yeast chromatin. A library of 1,086 human “prey” transcription factors were linked with the yeast Gal4 activation domain. Binding of the transcription factors to the DNA fragment resulted in transcription of reporter genes. Positive interactions were determined as two of the four replicate colonies that survived in the absence of histidine and experienced a color change in the presence of the substrate X-Gal. More than 90% of positive interactions were identified in 4/4 colonies [35–37]. [S5 Table](#) highlights transcription factors that bound HIV-1 tile sequences. Putative binding sites were predicted by FIMO software [38] with a  $p < 0.001$  and these data are shown in [S6 Table](#).

### Transfection with HIV-1 vectors

HIV proviral fragments, designated regions 1–17 ([S3 Table](#)), were cloned in vector pGL4.23 (Addgene, Cat: E8411) upstream of the luciferase reporter gene. pcDNA3.1+ (ThermoFisher, Cat: V79020) was used as a negative control. The pNL4-3- $\Delta 5'$  construct was derived from pNL4-3 (GenBank: AF324493.2) using enzymatic digestion with *StuI* which cuts within the 5' cellular genomic sequence of the NL4-3 plasmid and within the *env* gene at 6,825 bp. Following digestion, pNL4-3- $\Delta 5'$  was generated by blunt end ligation to delete the 5' end of the proviral genome. Successful deletion of 5' end was confirmed by DNA sequencing.

Transfection of HEK293T cells was performed in black-walled, clear-bottom 96-well plates. Cells were plated at a density of 20,000 cells/well 24 h prior to transfection. Cells were transfected at cell confluency of 60–70%. 300 ng of PEI and 50 ng of luciferase reporter constructs in up to 10  $\mu$ l of Opti-MEM were transfected in triplicate wells. 24 hours post transfection, growth media was removed, and cells were lysed with cell culture lysis buffer for 10 min at room temperature. Firefly D-Luciferin was added at a 1:1 volume ratio and the plate was immediately read on BioTek Synergy HT plate reader at 260 nm for 1 sec/well.

### Western blot

Cells transfected with NL-43 or pNL4-3- $\Delta 5'$  were lysed using freshly made lysis buffer (0.5% Igepal CA-630, Sigma; 50 mM Tris, pH 8.0; 150 mM NaCl; and Complete Protease Inhibitor (Millipore Sigma) in Milli-Q H<sub>2</sub>O. Lysates were mixed in Laemmli's SDS-Sample Buffer (Boston BioProducts, Cat: BP-111R) and heated at 95°C for 5 min. Proteins were separated by 12% SDS-polyacrylamide gel electrophoresis (PAGE) and transferred by electroblotting onto polyvinylidene difluoride membrane (Millipore). After blocking 1 h with 5% nonfat dry milk, blots were incubated with primary antibodies (Monoclonal Mouse anti- $\beta$ -Actin, BioRad, Cat: VMA00048 or Polyclonal Human Anti-HIV IgG, NIH-ARP3957, AIDS Reagent Program) overnight at 4°C and probed with horseradish peroxidase (HRP)-conjugated secondary antibodies. Membranes were developed with ECL Prime Western Blotting System (GE Healthcare) and visualized by merging a calorimetric (white light) and chemiluminescent image.

### Statistical analysis

All *in vitro* experiments and technical replicates were conducted at least three times. Experiments with primary cells included cells from at least three different donors. Data are presented as mean values  $\pm$  standard error of the mean. *p* values were calculated based on the two tailed two sample T-test using GraphPad Prism software unless otherwise noted. \* denotes  $p < 0.05$ ; \*\* denotes  $p < 0.01$  and \*\*\* denotes  $p < 0.001$ .



## Supporting information

### S1 Table. Probe and Primer Sequences.

(PDF)

### S2 Table. Participant Details.

(PDF)

### S3 Table. 5' RACE Sequences.

(XLSX)

### S4 Table. HIV Tiling Sequences.

(XLSX)

### S5 Table. Transcription Factor Binding.

(PDF)

### S6 Table. HIV-1 Baits 14+15 Motif Analysis.

(XLSX)

**S1 Fig. IPDA and MDM infection time course.** (A) Schematic of primer (arrows) and probe (stars) binding sites used for IPDA. For details see methods. (B) IPDA time course data for HIV-1 NL4-3-BaL infected MDMs. MDMs differentiated from 3 separate donors were infected as described in Methods and incubated for either 2 or 6 days before DNA isolation. IPDA was used to estimate intact and defective provirus frequencies per  $1 \times 10^6$  cells at each time point. Day 2 and Day 6 intact and defective provirus estimates were not statistically different when analyzed by Two-Sample T Test.

(TIF)

**S2 Fig. Reverse transcriptase inhibitor Efavirenz (EFV) treatment of CD4+ T cells and MDMs prevented intact and defective HIV provirus establishment regardless of activation state.** (A) IPDA data for EFV pretreated CD4+ T cells infected with HIV-1NL4-3. For CD4+ T cells, EFV was added at a concentration of 10  $\mu$ M 30 minutes prior to spinoculation with HIV-1NL4-3. Infection was limited to a single round by addition of the viral protease inhibitor Saquinavir 30 minutes after spinoculation. Resting and activated CD4+ T cells were participant matched. Data are from 3 separate donors. (B) IPDA data for EFV pretreated MDMs infected with HIV-1NL4-3-BaL. For MDMs, EFV was added at a concentration of 10  $\mu$ M 24-hours prior to addition of HIV-1NL4-3-BaL to the cell cultures. Data are from 3 separate donors.

(TIF)

**S3 Fig. Multiplex RT-ddPCR probe locations and representative data.** (A) Schematic of RT-ddPCR probe binding sites aligned to the HIV genome. Traditional HIV transcripts are labeled with population numbers that correlate to the droplet populations shown in (B). Putative cryptic HIV transcripts lacking 5'UTR sequence are labeled as populations 2, 6, and 5. Population 1 represents empty droplets or HIV transcripts which lack all probed sequence sites. (B) Representative data acquired during multiplex RT-ddPCR. Data shown reflects the following multiplexed probe reactions: 1X concentration LTR probe reaction (y-axis, FAM), 0.5X Nef probe reaction (y-axis, FAM), and 1X Env probe reaction (x-axis, VIC). The multiplexed assay detected 8 distinct populations which were manually gated as reflected by different droplet colors and numbers above. Numbers represent droplets harboring the following distinct transcriptional species: (1) Empty droplets, (2) Nef Only Transcripts, (3) LTR Only Transcripts, (4) LTR+Nef Transcripts, (5) Env Only Transcripts, (6) Env+Nef Transcripts, (7) LTR+Env

Transcripts, (8) LTR+Env+Nef Transcripts.  
(TIF)

**S4 Fig. Env+Nef HIV transcripts are polyadenylated.** RT was performed as described in Methods using a poly d(T) primer for reverse transcription of 500ng of RNA from either HIV-1<sub>NL4-3-BaL</sub> infected MDMs or HIV-1<sub>NL4-3</sub> infected CD4+ T cells which were unstimulated or activated for 72 hours using anti-CD3/CD28 beads. No Infection negative control sample is representative of N = 1. All remaining data are representative of three separate infections. CD4 + T cell data are donor matched between unstimulated and activated cells.  
(TIF)

## Acknowledgments

We thank the Providence/Boston CFAR Basic Science and Biobehavioral Cores (P30 AI042853) and Dr. Nelson Lau of the BUSM Genome Science Institute for assistance in drop-let digital PCR protocols. We also thank Allison Burdie who provided technical support in characterizing putative promoters and Drs. Manish Sagar and Suryaram Gummuluru (Boston University School of Medicine) for critical comments regarding the manuscript.

## Author Contributions

**Conceptualization:** Jeffrey Kuniholm, Elise Armstrong, Juan I. Fuxman Bass, Andrew J. Henderson.

**Data curation:** Jeffrey Kuniholm, Elise Armstrong, Brandy Bernabe, Carolyn Coote, Anna Berenson, Samantha D. Patalano, Alex Olson, Xianbao He, Juan I. Fuxman Bass.

**Formal analysis:** Jeffrey Kuniholm, Elise Armstrong, Brandy Bernabe, Carolyn Coote, Anna Berenson, Samantha D. Patalano, Alex Olson, Xianbao He, Nina H. Lin, Juan I. Fuxman Bass, Andrew J. Henderson.

**Funding acquisition:** Juan I. Fuxman Bass, Andrew J. Henderson.

**Investigation:** Jeffrey Kuniholm, Elise Armstrong, Brandy Bernabe, Carolyn Coote, Nina H. Lin, Juan I. Fuxman Bass.

**Methodology:** Jeffrey Kuniholm, Elise Armstrong, Brandy Bernabe, Carolyn Coote, Anna Berenson, Samantha D. Patalano, Alex Olson, Xianbao He, Juan I. Fuxman Bass.

**Resources:** Alex Olson, Nina H. Lin, Juan I. Fuxman Bass, Andrew J. Henderson.

**Supervision:** Juan I. Fuxman Bass, Andrew J. Henderson.

**Validation:** Jeffrey Kuniholm.

**Writing – original draft:** Jeffrey Kuniholm, Elise Armstrong, Andrew J. Henderson.

**Writing – review & editing:** Jeffrey Kuniholm, Alex Olson, Juan I. Fuxman Bass, Andrew J. Henderson.

## References

1. Crooks AM, Bateson R, Cope AB, Dahl NP, Griggs MK, Kuruc JD, et al. Precise Quantitation of the Latent HIV-1 Reservoir: Implications for Eradication Strategies. *J Infect Dis* [Internet]. 2015 Nov 1 [cited 2021 Sep 15]; 212(9):1361–5. Available from: <https://academic.oup.com/jid/article/212/9/1361/1025016>. <https://doi.org/10.1093/infdis/jiv218> PMID: 25877550
2. Siliciano JD, Kajdas J, Finzi D, Quinn TC, Chadwick K, Margolick JB, et al. Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nat Med* 2003 96 [Internet].

- 2003 May 18 [cited 2021 Sep 23]; 9(6):727–8. Available from: <https://www.nature.com/articles/nm880>. <https://doi.org/10.1038/nm880> PMID: 12754504
3. Araínga M, Edagwa B, Mosley RL, Poluektova LY, Gorantla S, Gendelman HE. A mature macrophage is a principal HIV-1 cellular reservoir in humanized mice after treatment with long acting antiretroviral therapy. *Retrovirology* [Internet]. 2017 Mar 9 [cited 2021 Sep 15]; 14(1):17. Available from: /pmc/articles/PMC5345240/. <https://doi.org/10.1186/s12977-017-0344-7> PMID: 28279181
  4. Burdo TH, Weiffenbach A, Woods SP, Letendre S, Ellis RJ, Williams KC. Elevated sCD163 in plasma but not cerebrospinal fluid is a marker of neurocognitive impairment in HIV infection. *AIDS* [Internet]. 2013 Jun 1 [cited 2021 Sep 15]; 27(9):1387–95. Available from: /pmc/articles/PMC3844286/. <https://doi.org/10.1097/QAD.0b013e32836010bd> PMID: 23435298
  5. Rappaport J, Volsky DJ. Role of the Macrophage in HIV-Associated Neurocognitive Disorders and Other Comorbidities in Patients on Effective Antiretroviral Treatment. *J Neurovirology* [Internet]. 2015 Jun 16 [cited 2021 Sep 15]; 21(3):235. Available from: /pmc/articles/PMC4445403/. <https://doi.org/10.1007/s13365-015-0346-y> PMID: 25933548
  6. Hendricks CM, Cordeiro T, Gomes AP, Stevenson M. The Interplay of HIV-1 and Macrophages in Viral Persistence. *Front Microbiol* [Internet]. 2021 Apr 7 [cited 2021 Sep 15]; 12. Available from: /pmc/articles/PMC8058371/ <https://doi.org/10.3389/fmicb.2021.646447> PMID: 33897659
  7. Porcheray F, Samah B, Léone C, Dereuddre-Bosquet N, Gras G. Macrophage activation and human immunodeficiency virus infection: HIV replication directs macrophages towards a pro-inflammatory phenotype while previous activation modulates macrophage susceptibility to infection and viral production. *Virology*. 2006 May 25; 349(1):112–20. <https://doi.org/10.1016/j.virol.2006.02.031> PMID: 16564558
  8. Deeks SG, Lewin SR, Havlir D V. The End of AIDS: HIV Infection as a Chronic Disease. *Lancet* [Internet]. 2013 [cited 2021 Sep 15]; 382(9903):1525. Available from: /pmc/articles/PMC4058441/ [https://doi.org/10.1016/S0140-6736\(13\)61809-7](https://doi.org/10.1016/S0140-6736(13)61809-7) PMID: 24152939
  9. Campbell JH, Hearps AC, Martin GE, Williams KC, Crowe SM. The importance of monocytes and macrophages in HIV pathogenesis, treatment, and cure. *AIDS* [Internet]. 2014 [cited 2021 Sep 15]; 28(15):2175. Available from: /pmc/articles/PMC6331181/ <https://doi.org/10.1097/QAD.0000000000000408> PMID: 25144219
  10. Olson A, Coote C, Snyder-Cappione JE, Lin N, Sagar M. HIV-1 Transcription but Not Intact Provirus Levels are Associated With Systemic Inflammation. *J Infect Dis* [Internet]. 2021 Jun 4 [cited 2021 Sep 23]; 223(11):1934–42. Available from: <https://academic.oup.com/jid/article/223/11/1934/5930820> <https://doi.org/10.1093/infdis/jiaa657> PMID: 33075121
  11. Akiyama H, Miller CM, Ettinger CR, Belkina AC, Snyder-Cappione JE, Gummuluru S. HIV-1 intron-containing RNA expression induces innate immune activation and T cell dysfunction. *Nat Commun* 2018 91 [Internet]. 2018 Aug 27 [cited 2021 Sep 15]; 9(1):1–12. Available from: <https://www.nature.com/articles/s41467-018-05899-7> PMID: 30150664
  12. McCauley SM, Kim K, Nowosielska A, Dauphin A, Yurkovetskiy L, Diehl WE, et al. Intron-containing RNA from the HIV-1 provirus activates type I interferon and inflammatory cytokines. *Nat Commun*. 2018 Dec 1; 9(1):1–10. <https://doi.org/10.1038/s41467-017-02088-w> PMID: 29317637
  13. Pinzone MR, VanBelzen DJ, Weissman S, Bertuccio MP, Cannon L, Venanzi-Rullo E, et al. Longitudinal HIV sequencing reveals reservoir expression leading to decay which is obscured by clonal expansion. *Nat Commun* [Internet]. 2019 Dec 1 [cited 2021 Sep 15]; 10(1). Available from: /pmc/articles/PMC6374386/ <https://doi.org/10.1038/s41467-019-08431-7> PMID: 30760706
  14. Liu R, Simonetti FR, Ho Y-C. The forces driving clonal expansion of the HIV-1 latent reservoir. *Virology* 2020 171 [Internet]. 2020 Jan 7 [cited 2021 Jul 28]; 17(1):1–13. Available from: <https://virologyj.biomedcentral.com/articles/10.1186/s12985-019-1276-8> PMID: 31910871
  15. Anderson EM, Simonetti FR, Gorelick RJ, Hill S, Gouzoulis MA, Bell J, et al. Dynamic Shifts in the HIV Proviral Landscape During Long Term Combination Antiretroviral Therapy: Implications for Persistence and Control of HIV Infections. *Viruses* 2020, Vol 12, Page 136 [Internet]. 2020 Jan 25 [cited 2021 Sep 23]; 12(2):136. Available from: <https://www.mdpi.com/1999-4915/12/2/136/html>
  16. Wang Z, Simonetti FR, Siliciano RF, Laird GM. Measuring replication competent HIV-1: advances and challenges in defining the latent reservoir. *Retrovirology* [Internet]. 2018 Feb 13 [cited 2021 Sep 15]; 15(1):21. Available from: /pmc/articles/PMC5810003/ <https://doi.org/10.1186/s12977-018-0404-7> PMID: 29433524
  17. Bruner KM, Murray AJ, Pollack RA, Soliman MG, Laskey SB, Capoferri AA, et al. Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nat Med*. 2016 Sep 1; 22(9):1043–9. <https://doi.org/10.1038/nm.4156> PMID: 27500724
  18. Bruner KM, Wang Z, Simonetti FR, Bender AM, Kwon KJ, Sengupta S, et al. A novel quantitative approach for measuring the reservoir of latent HIV-1 proviruses. *Nature* [Internet]. 2019 Feb 7 [cited

- 2021 Sep 15]; 566(7742):120. Available from: /pmc/articles/PMC6447073/ <https://doi.org/10.1038/s41586-019-0898-8> PMID: 30700913
19. Pollack RA, Jones RB, Perteau M, Bruner KM, Martin AR, Thomas AS, et al. Defective HIV-1 Proviruses Are Expressed and Can Be Recognized by Cytotoxic T Lymphocytes, which Shape the Proviral Landscape. *Cell Host Microbe*. 2017 Apr 12; 21(4):494–506.e4. <https://doi.org/10.1016/j.chom.2017.03.008> PMID: 28407485
  20. Imamichi H, Smith M, Adelsberger JW, Izumi T, Scrimieri F, Sherman BT, et al. Defective HIV-1 proviruses produce viral proteins. *Proc Natl Acad Sci U S A* [Internet]. 2020 Feb 18 [cited 2021 Sep 15]; 117(7):3704. Available from: /pmc/articles/PMC7035625/ <https://doi.org/10.1073/pnas.1917876117> PMID: 32029589
  21. Yukl SA, Kaiser P, Kim P, Telwatte S, Joshi SK, Vu M, et al. HIV latency in isolated patient CD4+ T cells may be due to blocks in HIV transcriptional elongation, completion, and splicing. *Sci Transl Med* [Internet]. 2018 Feb 28 [cited 2021 Sep 15]; 10(430):9927. Available from: /pmc/articles/PMC5959841/
  22. Levy CN, Hughes SM, Roychoudhury P, Reeves DB, Amstutz C, Zhu H, et al. A highly multiplexed droplet digital PCR assay to measure the intact HIV-1 proviral reservoir. *Cell Reports Med*. 2021 Apr 20; 2(4):100243. <https://doi.org/10.1016/j.xcrm.2021.100243> PMID: 33948574
  23. Liu R, Catalano AA, Ho Y-C. Measuring the size and decay dynamics of the HIV-1 latent reservoir. *Cell Reports Med* [Internet]. 2021 Apr 20 [cited 2021 Sep 15]; 2(4):100249. Available from: /pmc/articles/PMC8080242/ <https://doi.org/10.1016/j.xcrm.2021.100249> PMID: 33948579
  24. Antar AAR, Jenike KM, Jang S, Rigau DN, Reeves DB, Hoh R, et al. Longitudinal study reveals HIV-1–infected CD4+ T cell dynamics during long-term antiretroviral therapy. *J Clin Invest* [Internet]. 2020 Jul 1 [cited 2021 Sep 15]; 130(7):3543. Available from: /pmc/articles/PMC7324206/ <https://doi.org/10.1172/JCI135953> PMID: 32191639
  25. Peluso MJ, Bacchetti P, Ritter KD, Beg S, Lai J, Martin JN, et al. Differential decay of intact and defective proviral DNA in HIV-1–infected individuals on suppressive antiretroviral therapy. *JCI Insight*. 2020 Feb 27; 5(4). <https://doi.org/10.1172/jci.insight.132997> PMID: 32045386
  26. Falcinelli SD, Kilpatrick KW, Read J, Murtagh R, Allard B, Ghofrani S, et al. Longitudinal Dynamics of Intact HIV Proviral DNA and Outgrowth Virus Frequencies in a Cohort of Individuals Receiving Antiretroviral Therapy. *J Infect Dis* [Internet]. 2021 Jul 2 [cited 2021 Sep 23]; 224(1):92–100. Available from: <https://academic.oup.com/jid/article/224/1/92/5995716> <https://doi.org/10.1093/infdis/jiaa718> PMID: 33216132
  27. Cochrane AW, Jones KS, Beidas S, Dillon PJ, Skalka AM, Rosen CA. Identification and characterization of intragenic sequences which repress human immunodeficiency virus structural gene expression. *J Virol*. 1991; <https://doi.org/10.1128/JVI.65.10.5305-5313.1991> PMID: 1895385
  28. Colin J, Libri D, Porrua O. Cryptic Transcription and Early Termination in the Control of Gene Expression. *Genet Res Int*. 2011 Nov 24; 2011:1–10. <https://doi.org/10.4061/2011/653494> PMID: 22567365
  29. Goffin V, Demonté D, Vanhulle C, de Walque S, de Launoit Y, Burny A, et al. Transcription factor binding sites in the pol gene intragenic regulatory region of HIV-1 are important for virus infectivity. *Nucleic Acids Res* [Internet]. 2005 Jul 1 [cited 2021 Sep 15]; 33(13):4285–310. Available from: <https://academic.oup.com/nar/article/33/13/4285/1094465> <https://doi.org/10.1093/nar/gki720> PMID: 16061936
  30. Lint C Van, Ghysdael J, Paras P Jr, Burny A, Verdin E. A transcriptional regulatory element is associated with a nuclease-hypersensitive site in the pol gene of human immunodeficiency virus type 1. *J Virol* [Internet]. 1994 Apr [cited 2021 Sep 15]; 68(4):2632. Available from: /pmc/articles/PMC236741/?report=abstract <https://doi.org/10.1128/JVI.68.4.2632-2648.1994> PMID: 8139041
  31. Verdin E. DNase I-hypersensitive sites are associated with both long terminal repeats and with the intragenic enhancer of integrated human immunodeficiency virus type 1. *J Virol*. 1991; <https://doi.org/10.1128/JVI.65.12.6790-6799.1991> PMID: 1942252
  32. Verdin E, Becker N, Bex F, Droogmans L, Burny A. Identification and characterization of an enhancer in the coding region of the genome of human immunodeficiency virus type 1. *Proc Natl Acad Sci U S A*. 1990; <https://doi.org/10.1073/pnas.87.12.4874> PMID: 2352955
  33. Knudsen S. Promoter2.0: for the recognition of PolII promoter sequences. *Bioinformatics* [Internet]. 1999 May 1 [cited 2021 Sep 23]; 15(5):356–61. Available from: <https://academic.oup.com/bioinformatics/article/15/5/356/312175> <https://doi.org/10.1093/bioinformatics/15.5.356> PMID: 10366655
  34. Sloutskin A, Danino YM, Orenstein Y, Zehavi Y, Doniger T, Shamir R, et al. ElemeNT: a computational tool for detecting core promoter elements. *Transcription* [Internet]. 2015 May 27 [cited 2021 Sep 15]; 6(3):41. Available from: /pmc/articles/PMC4581360/ <https://doi.org/10.1080/21541264.2015.1067286> PMID: 26226151

35. Fuxman Bass JI, Sahni N, Shrestha S, Garcia-Gonzalez A, Mori A, Bhat N, et al. Human Gene-Centered Transcription Factor Networks for Enhancers and Disease Variants. *Cell*. 2015 Apr 23; 161(3):661–73. <https://doi.org/10.1016/j.cell.2015.03.003> PMID: 25910213
36. Pedro KD, Agosto LM, Sewell JA, Eberenz KA, He X, Bass JIF, et al. A functional screen identifies transcriptional networks that regulate HIV-1 and HIV-2. *Proc Natl Acad Sci [Internet]*. 2021 Mar 16 [cited 2021 Aug 18]; 118(11). Available from: <https://www.pnas.org/content/118/11/e2012835118> <https://doi.org/10.1073/pnas.2012835118> PMID: 33836568
37. Santoso CS, Li Z, Lal S, Yuan S, Gan KA, Agosto LM, et al. Comprehensive mapping of the human cytokine gene regulatory network. *Nucleic Acids Res [Internet]*. 2020 Dec 2 [cited 2021 Sep 23]; 48(21):12055–73. Available from: <https://academic.oup.com/nar/article/48/21/12055/5976935> <https://doi.org/10.1093/nar/gkaa1055> PMID: 33179750
38. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. *Bioinformatics [Internet]*. 2011 [cited 2021 Nov 30]; 27(7):1017. Available from: [/pmc/articles/PMC3065696/](https://pmc/articles/PMC3065696/) <https://doi.org/10.1093/bioinformatics/btr064> PMID: 21330290
39. Imamichi H, Dewar RL, Adelsberger JW, Rehm CA, O'Doherty U, Paxinos EE, et al. Defective HIV-1 proviruses produce novel protein-coding RNA species in HIV-infected patients on combination antiretroviral therapy. *Proc Natl Acad Sci U S A [Internet]*. 2016 Aug 2 [cited 2021 Sep 15]; 113(31):8783. Available from: [/pmc/articles/PMC4978246/](https://pmc/articles/PMC4978246/) <https://doi.org/10.1073/pnas.1609057113> PMID: 27432972
40. Jiang C, Lian X, Gao C, Sun X, Einkauf KB, Chevalier JM, et al. A unique viral reservoir landscape in HIV-1 elite controllers. *Nature [Internet]*. 2020 Sep 10 [cited 2021 Sep 15]; 585(7824):261. Available from: [/pmc/articles/PMC7837306/](https://pmc/articles/PMC7837306/) <https://doi.org/10.1038/s41586-020-2651-8> PMID: 32848246
41. Pinzone MR, O'Doherty U. Measuring integrated HIV DNA ex vivo and in vitro provides insights about how reservoirs are formed and maintained. *Retrovirology [Internet]*. 2018 Feb 17 [cited 2021 Sep 15]; 15(1):22. Available from: [/pmc/articles/PMC5816390/](https://pmc/articles/PMC5816390/) <https://doi.org/10.1186/s12977-018-0396-3> PMID: 29452580
42. Agosto LM, Yu JJ, Dai J, Kaletsky R, Monie D, O'Doherty U. HIV-1 integrates into resting CD4+ T cells even at low inoculums as demonstrated with an improved assay for HIV-1 integration. *Virology [Internet]*. 2007 Nov 10 [cited 2021 Sep 23]; 368(1):60. Available from: [/pmc/articles/PMC2140271/](https://pmc/articles/PMC2140271/) <https://doi.org/10.1016/j.virol.2007.06.001> PMID: 17631931
43. Agosto LM, Herring MB, Mothes W, Henderson AJ. HIV-1-Infected CD4+ T Cells Facilitate Latent Infection of Resting CD4+ T Cells through Cell-Cell Contact. *Cell Rep*. 2018 Aug 21; 24(8):2088–100. <https://doi.org/10.1016/j.celrep.2018.07.079> PMID: 30134170
44. Abrahams MR, Joseph SB, Garrett N, Tyers L, Moeser M, Archin N, et al. The replication-competent HIV-1 latent reservoir is primarily established near the time of therapy initiation. *Sci Transl Med*. 2019 Oct 9; 11(513). <https://doi.org/10.1126/scitranslmed.aaw5589> PMID: 31597754
45. Arrigo S, Yun M, Beemon K. cis-acting regulatory elements within gag genes of avian retroviruses. *Mol Cell Biol*. 1987; <https://doi.org/10.1128/mcb.7.1.388-397.1987> PMID: 3031470
46. Campbell M, Eng C, Luciw PA. The simian foamy virus type 1 transcriptional transactivator (Tas) binds and activates an enhancer element in the gag gene. *J Virol*. 1996; <https://doi.org/10.1128/JVI.70.10.6847-6855.1996> PMID: 8794326
47. Löchelt M, Flügel RM, Aboud M. The human foamy virus internal promoter directs the expression of the functional Bel 1 transactivator and Bet protein early after infection. *J Virol*. 1994; <https://doi.org/10.1128/JVI.68.2.638-645.1994> PMID: 8289367
48. Lochelt M, Muranyi W, Flügel RM. Human foamy virus genome possesses an internal, Bel-1-dependent and functional promoter. *Proc Natl Acad Sci U S A*. 1993; <https://doi.org/10.1073/pnas.90.15.7317> PMID: 8394017
49. Reuss FU, Coffin JM. Mouse Mammary Tumor Virus Superantigen Expression in B Cells Is Regulated by a Central Enhancer within the pol Gene. *J Virol*. 1998;
50. Graf M, Bojak A, Deml L, Bieler K, Wolf H, Wagner R. Concerted Action of Multiple cis-Acting Sequences Is Required for Rev Dependence of Late Human Immunodeficiency Virus Type 1 Gene Expression. *J Virol*. 2000; <https://doi.org/10.1128/jvi.74.22.10822-10826.2000> PMID: 11044131
51. Kint S, Trypsteen W, De Spiegelaere W, Malatinkova E, Kinloch-de Loes S, De Meyer T, et al. Underestimated effect of intragenic HIV-1 DNA methylation on viral transcription in infected individuals. *Clin Epigenetics* 2020 121 [Internet]. 2020 Feb 28 [cited 2021 Sep 15]; 12(1):1–11. Available from: <https://link.springer.com/articles/10.1186/s13148-020-00829-1> PMID: 32111236
52. Deeks SG, Tracy R, Douek DC. Systemic Effects of Inflammation on Health during Chronic HIV Infection. *Immunity [Internet]*. 2013 Oct 17 [cited 2021 Jul 28]; 39(4):633. Available from: [/pmc/articles/PMC4012895/](https://pmc/articles/PMC4012895/) <https://doi.org/10.1016/j.immuni.2013.10.001> PMID: 24138880
53. Višković K, Židovec Lepej S, Gorenc A, Grgić I, Lukas D, Zekan Š, et al. Cardiovascular markers of inflammation and serum lipid levels in HIV-infected patients with undetectable viremia. *Sci Reports*



- 2018 81 [Internet]. 2018 Apr 17 [cited 2021 Jul 28]; 8(1):1–8. Available from: <https://www.nature.com/articles/s41598-018-24446-4> <https://doi.org/10.1038/s41598-018-24446-4> PMID: 29666424
54. Angel CJL, Pham EA, Du H, Vallania F, Fram BJ, Perez K, et al. Signatures of immune dysfunction in HIV and HCV infection share features with chronic inflammation in aging and persist after viral reduction or elimination. *Proc Natl Acad Sci* [Internet]. 2021 Apr 6 [cited 2021 Sep 23]; 118(14):2021. Available from: <https://www.pnas.org/content/118/14/e2022928118>
  55. Champiat S, Raposo RAS, Maness NJ, Lehman JL, Purtell SE, Hasenkrug AM, et al. Influence of HAART on Alternative Reading Frame Immune Responses over the Course of HIV-1 Infection. *PLoS One* [Internet]. 2012 Jun 29 [cited 2021 Aug 16]; 7(6). Available from: [/pmc/articles/PMC3387156/https://doi.org/10.1371/journal.pone.0039311](https://doi.org/10.1371/journal.pone.0039311) PMID: 22768072
  56. Bansal A, Carlson J, Yan J, Akinsiku OT, Schaefer M, Sabbaj S, et al. CD8 T cell response and evolutionary pressure to HIV-1 cryptic epitopes derived from antisense transcription. *J Exp Med*. 2010 Jan 18; 207(1):51–9. <https://doi.org/10.1084/jem.20092060> PMID: 20065064
  57. Cardinaud S, Moris A, Février M, Rohrlisch PS, Weiss L, Langlade-Demoyen P, et al. Identification of Cryptic MHC I-restricted Epitopes Encoded by HIV-1 Alternative Reading Frames. *J Exp Med*. 2004 Apr 19; 199(8):1053–63. <https://doi.org/10.1084/jem.20031869> PMID: 15078897
  58. Akiyama H, Miller CM, Ettinger CR, Belkina AC, Snyder-Cappione J, Gummuluru S. HIV-1 Unspliced RNA Expression Induces Innate Immune Activation in Macrophages. [cited 2020 Mar 2]; Available from: <https://doi.org/https%3A/doi.org/10.1101/174508>
  59. Gervaix A, West D, Leoni LM, Richman DD, Wong-Staal F, Corbeil J. A new reporter cell line to monitor HIV infection and drug susceptibility in vitro. *Proc Natl Acad Sci U S A* [Internet]. 1997 Apr 29 [cited 2021 Sep 17]; 94(9):4653. Available from: [/pmc/articles/PMC20779/https://doi.org/10.1073/pnas.94.9.4653](https://doi.org/10.1073/pnas.94.9.4653) PMID: 9114046
  60. Gagne M, Michaels D, Schiralli Lester GM, Gummuluru S, Wong WW, Henderson AJ. Strength of T cell signaling regulates HIV-1 replication and establishment of latency. Swanstrom R, editor. *PLOS Pathog* [Internet]. 2019 May 22 [cited 2020 Jan 2]; 15(5):e1007802. Available from: <http://dx.plos.org/10.1371/journal.ppat.1007802> PMID: 31116788