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

2 Jeffrey Kuniholm¹, Elise Armstrong², Brandy Bernabe^{3,5}, Carolyn Coote², Anna Berenson⁴,

3 Samantha D. Drinan⁴, Alex Olson², Xianbao He², Nina Lin², Juan I. Fuxman Bass⁴, Andrew J.

4 Henderson^{1,2,3*}

5 ¹Boston University School of Medicine, Department of Microbiology, Boston, MA, USA; ²Boston
6 University School of Medicine, Department of Medicine, Section of Infectious Diseases; ³Boston
7 University School of Medicine Graduate Medical Sciences, ⁴Boston University, Department of
8 Biology

9 ⁵Current Address: Emory University, Laney Graduate School, Atlanta, GA

10

11 **Short Title: Defective HIV-1 proviral transcription**

12 ***Corresponding Author**

13 hender@bu.edu (AJH)

14

15 **Keywords**

16 HIV transcription

17 HIV latency

18 Defective Proviruses

19

ABSTRACT

HIV- 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 digital drop 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 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

45 paper, we demonstrate that acute infection of macrophages and resting T cells is biased towards
46 generating defective viruses which are expressed by DNA regulatory elements in the HIV
47 genome. These studies describe an alternative mechanism for chronic expression of HIV
48 genomes.

49

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 patients treated with antiretroviral drugs that have undetectable HIV-1 expression (3–10). People living with HIV-1 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 HIV patients 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, many that lack 5' long terminal repeats and despite these deletions still express HIV-1 RNAs and proteins. Importantly, we detect these mRNAs in a subset of people living with HIV 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), Figure 1A, *Supplemental Figure 1A*). CD4+ T cells were either unstimulated or activated with anti-CD3/anti-CD28 beads prior to infection with HIV-1_{VSvgNL4-3}, whereas monocyte derived macrophages (MDMs) were infected with R5-tropic HIV-1_{NL4-3 BaL}. Infected cells were cultured for at least 48-hours to allow the completion of reverse transcription and integration (*Supplemental Figure 1B*). 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 (Figure 1A,B) 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'-deletions. 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 (*Supplemental Figure 2*). 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.

110

Figure 1. Acute HIV-1 Infection Generates Defective Provirus 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_{NL43-BaL} for 48 hours prior to DNA isolation. Resting and anti-CD3/CD-28 bead activated CD4+ T cells were infected with HIV-1_{NL43-VSVg} 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 is patient matched.

111

112 Proviruses lacking 5'LTRs are transcribed. Several studies demonstrated that defective HIV
113 proviruses are transcribed and translated (19,20). To determine if the provirus populations
114 detected in HIV-1 infected CD4+ T cells and MDMs were transcriptionally competent, we adapted
115 a reverse transcription ddPCR assay (RT-ddPCR) originally described by Yukl et al (21) by
116 multiplexing probes for simultaneous detection of transcripts spanning the R-U5-Gag junction
117 ("LTR" probe), distal transcripts ("Nef" probe), and sequences that included both upstream and
118 distal sequences (5' UTR/Nef double positives, *Supplemental Figure 3*) (22,23). Multiplex RT-
119 ddPCR performed using infected CD4+ T cell cDNA detected transcripts which included 5'
120 untranslated region (UTR) sequence only and double-positive 5' UTR-Nef transcripts. We also
121 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 (Figure 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 (*Supplemental Figure 3*). 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'UTRs 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. (Figure 2B, 2C). We also observed approximately 25% of RNAs in infected MDMs lacked 5' UTRs (Figure 2D, E, F). 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 (*See Supplemental Figure 4*). 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.

138

139 Chronically infected individuals on ART demonstrate a gradual decrease in intact HIV proviruses

140 concurrent with selection and expansion of defective proviruses due to sustained immunological

141 pressure and clearance (15,17–19,24). Notably, defective HIV proviruses are observed to decay

142 at a slower rate than intact provirus clones, presumably due to their relatively low HIV-1

143 expression levels (13,23–26). We posit that HIV-1 mRNA lacking 5' UTRs should be detected in

144 people living with HIV (PLWH) on ART. Using DNA isolated from PBMCs of PLWH on ART

145 (*Supplemental Table 2*), we first confirmed the presence of HIV-1 proviruses in chronically

146 infected individuals by IPDA with most proviruses detected in these individuals being defective

147 (Figure 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 LTR 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' UTR start sites (Figure 3 B, C). Env sequences lacking 5'UTRs were detected with signal above uninfected PBMC background controls in 10/11 RNA samples from HIV-positive PBMC donors (Figure 3B, *Supplemental Table 2*). 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.

Figure 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 5' LTR and Env transcripts 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 of single positive transcripts per μg of RNA. Dotted line represents the mean assay background signal calculated from five HIV-negative PBMC samples. Each point represents an independent patient sample. A total of 11 patient samples were assessed.

Defective proviruses 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).

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* (*Supplemental Table #1*) 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 43 products were sequenced to map their 5' ends. MDM and CD4+ T cell sequences demonstrated that 46.5% of the transcripts were splice products which initiated at the transcriptional start site within the 5' LTR sequence (Figure 4A, B). The remaining 53.4% of sequences initiated transcription within the *Env* gene, across a span of approximately 200 nucleotides located between +7760 to +8178 bp downstream of the canonical 5' LTR transcriptional start site. 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 TATA box motifs (7483-7488 bp and 7662-7669 bp) within the *Env* sequence, adjacent to the start sites identified by 5' RACE (Figure 4B). 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_{NL4-3} genome into 17 overlapping fragments (*Supplemental Table 3*). 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

190 histidine and were blue when provided X-Gal. The eY1H screen detected 60 transcription factor-
 191 DNA interactions, representing 43 unique transcription factors which bound to six regions within
 192 the HIV-1 genome (*Supplemental Table 4*; Figure 4C, (designated 2, 10, 11, 14, 15 and 16)).
 193 Transcription factor binding hot spots were clustered within and near HIV *Env* gene, whereas,
 194 transcription factor binding was not detected in *Pol* (Regions 3-9). Identified transcription factors
 195 included homeobox, basic helix-loop-helix, and zinc finger families (*Supplemental Table 4*). The
 196 distribution of transcription factor binding hotspots near the HIV *Env* gene correlated with the
 197 transcriptional start site suggested by the 5' RACE results. To investigate if these intragenic
 198 sequences that bound host transcription factors function as cis-regulatory elements, reporter
 199 vectors were generated with HIV-1 fragments cloned upstream of the luciferase reporter gene.
 200 Reporter plasmids were transiently transfected in HEK293T cells, and firefly luciferase was
 201 assessed 24 hours post-transfection (Figure 4D). Most of the intragenic sequences, including
 202 Regions 11 and 16, which bound several transcription factors in the eY1H assays, modestly
 203 influence transcription of the reporter gene. However, regions 14 and 15, which spanned
 204 sequences 7001 to 8521 bp in *Env* and bound several transcription factors in the eY1H assays,
 205 exhibited 5-10 fold higher luciferase signal compared to pcDNA3 control. These sequences also
 206 spanned sites that included putative promoter elements and the RACE transcriptional start sites.

Figure 4. Identification of intragenic transcriptional elements. (A) Schematic of HIV-1 genome with putative promoters indicated as dotted red lines as predicated 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 NL43 sequence (N=43 sequences). (Middle) 5' RACE sequences which lacked 5' UTR aligned to NL43 sequence (N=23). (Bottom) 5' RACE sequences which contained 5' UTR aligned to NL43 sequence (N=20). (C) 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 Supplemental Table 3). 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 specify HIV-1 genomic sequence. (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 only. HIV-LTR-Luciferase was used as a positive control. These data include three independent transfection experiments 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 $p < 0.001$.

207

208 In addition, we generated a viral expression construct, pNL4-3Δ5' in which we deleted

209 approximately 6,000 bp of the 5' end of the provirus, including the 5' LTR, *Gag* and *Pol* genes,

210 and transfected this construct into HEK293T cells (Figure 5A). Transcription of *Env* containing

211 sequences was detected by RT-ddPCR (Figure 5B). Furthermore, 35 kd and 19 kd proteins were

212 detected by immunoblotting when whole cell lysates from pNL4-3Δ5' transfected cells were

213 probed with human HIV anti-serum suggesting that these RNAs are translated (Figure 5C). It

214 should be noted that immunoblots required longer exposure times to visualize the proteins

215 expressed from pNL4-3Δ5' pNL4-3Δ5' probably reflecting that these intragenic elements are

216 much weaker than the HIV-1 5'-LTR which increased luciferase expression on the order of ~200-

217 fold compared to pcDNA control in transfected HEK293T cells (Figure 4E). 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.

Figure 5. Truncated HIV-1 proviral construct expresses RNA and protein. (A) Schematic of HIV-1_{NL4-3} in pNL43 expression plasmid and location of StuI-restriction digest site used to generate pNL43Δ5' (bottom). (B) RT-ddPCR data from HEK-293T cells transfected with either full length pNL43 (black) or pNL43Δ5' (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 pNL43Δ5'. Data include 4 independent pNL43 transfections and three pNL43 Δ5' transfections. (C) Western blot images using cell lysate of HEK293T cells transfected with either full length pNL43 plasmid or pNL43Δ5'. 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.

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 (17,18,38), 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 people living with HIV (PLWH) on ART and long-term controllers which demonstrated a relatively rapid decay of intact HIV-1 provirus compared to defective proviruses (23–25,39). These dynamics result in the progressive accumulation of pauciclonal 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 (13–16,23,24,39). In this study, we provide evidence that a subset of these defective proviruses are expressed 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 (40–42). 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 large 5' deletions. Our results suggest that upon acute infection, even in the absence of immune selection, a foundation of cells harboring defective viruses is rapidly established (43).

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 (17,38). 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 (44–48). 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 (27,31,32,49,50). 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(50). 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 (51–53). 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 (19,20,38). 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 (54–56). 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 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 (11,12,57). 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.

Acknowledgements

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. We thank the Prov/Bos CFAR Basic Science and Biobehavioral Cores (P30 AI042853) and Dr. Nelson Lau of the Genome Science Institute for assistance in digital drop 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.

METHODS

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 (41). 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 20ng/mL M-CSF (BioLegend, Cat: 574802) at 37°C and 5% CO₂. 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_{NL4-3} and HIV-1_{NL4-3 BaL} (AIDS Reagents Program) were generated by transfecting HEK293T cells using PEI transfection reagent (Sigma, Cat: 408727). Pseudotyped HIV-1_{VSVg-NL4-3}, HEK293T cells were co-transfected with an expression vector containing VSVg and HIV-1_{NL4-3}. HIV-1_{NL4-3} virus stocks titers were estimated using CEM-GFP cells obtained from the AIDS Reagent Program (58). HIV-1_{NL4-3 BaL} virus stocks were titrated on primary MDMs and IU/mL were estimated from IPDA results. All infections were performed using MOI < 1 (HIV-1_{VSVg-NL4-3} MOI = 0.1, HIV-1_{NL4-3} MOI = 0.02, HIV-1_{NL4-3 BaL} MOI = 0.01). DNase/MgCl₂⁺ 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 (59). MDMs were infected by adding HIV-1_{NL4-3BaL} directly to the cells at 37°C for 4 hr, washed twice with PBS and maintained in fresh media. 1 μM 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 Qiashreder 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 (See *Supplementary Table 1*). 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. Provirus 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 H₂O 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 μ l of 5x SuperScript III buffer, 7 μ l of 50 mM MgCl₂, 3.5 μ l of random hexamers (50 ng/ μ l), 1.35 μ l of oligo dT15 (500ug/mL), 3.5 μ l of 10 mM deoxynucleoside triphosphates (dNTPs), 1.75 μ l of RNaseOUT (40 U/ μ l; Invitrogen), 2 μ l of SuperScript III RT (200 U/ μ l; Invitrogen), and nuclease-free dH₂O to bring the final reaction volume to 70 μ 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. *Supplemental Table 1* 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.

ART-Patient Samples

PBMC from patients enrolled in a study assessing inflammation and immune dysfunction in people living with HIV and aging were used (Boston University 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. See *Supplemental Table #2* for details. PBMC were isolated using Ficoll-Hypaque density gradient separation and cryopreserved at -150°C until use.

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 µg), 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 µL) 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/µl), 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 µL of Tricine-EDTA Buffer before RACE-PCR.

For 5' RACE PCR a nef specific reverse primer was designed (*See Supplemental Table 1*). The RACE reaction mixture consisted of 5' RACE PCR sample: PCR-Grade H₂O (15.5 µL), 2X SeqAmp Buffer (25 µL), SeqAmp DNA Polymerase (1 µL), 5' RACE-Ready cDNA (2.5 µL), 10X Universal Primer Mix (5 µL), 5' gene-specific Primer (10 µM). 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 µL), Gel-purified RACE product (7 µL), and In-Fusion HD Master Mix (2 µL). 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 (*Supplemental Table 3*). “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).

Transfection with HIV-1 Vectors

HIV proviral fragments, designated regions 1-17 (*Supplemental Table 3*), 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-Δ5’ construct was derived from pNL4-3 by enzymatic digestion with *StuI* and blunt end ligation deleting 6,825 bp of the 5’ end of the proviral genome.

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 hours 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 μl 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- Δ 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 H₂O. 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- β -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 $p < 0.001$.

REFERENCES

1. Crooks AM, Bateson R, Cope AB, Dahl NP, Griggs MK, Kuruc JD, et al. 2015. Precise quantitation of the latent HIV-1 reservoir: implications for eradication strategies. *J Infect Dis* 212:1361–5.
2. Siliciano JD, Kajdas J, Finzi D, Quinn TC, Chadwick K, Margolick JB, et al. 2003. Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nat Med* 9:727–8.
3. Araínga M, Edagwa B, Mosley RL, Poluektova LY, Gorantla S, Gendelman HE. 2017. A mature macrophage is a principal HIV-1 cellular reservoir in humanized mice after treatment with long acting antiretroviral therapy. *Retrovirology* 14:17.
4. Burdo TH, Weiffenbach A, Woods SP, Letendre S, Ellis RJ, Williams KC. 2013. Elevated sCD163 in plasma but not cerebrospinal fluid is a marker of neurocognitive impairment in HIV infection. *AIDS* 27:1387–95.
5. Rappaport J, Volsky DJ. 2015. Role of the Macrophage in HIV-Associated Neurocognitive Disorders and Other Comorbidities in Patients on Effective Antiretroviral Treatment. *J Neurovirol* 21:235.
6. Hendricks CM, Cordeiro T, Gomes AP, Stevenson M. 2021, The Interplay of HIV-1 and Macrophages in Viral Persistence. *Front Microbiol* 12:646447.
7. Porcheray F, Samah B, Léone C, Dereuddre-Bosquet N, Gras G. 2006. 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*. 349:112–20.
8. Deeks SG, Lewin SR, Havlir D V. 2013.The End of AIDS: HIV Infection as a Chronic Disease. *Lancet* 382:1525.
9. Campbell JH, Hearps AC, Martin GE, Williams KC, Crowe SM. 2014. The importance of monocytes and macrophages in HIV pathogenesis, treatment, and cure. *AIDS* 28:2175.
10. Olson A, Coote C, Snyder-Cappione JE, Lin N, Sagar M. 2021. HIV-1 Transcription but Not Intact Provirus Levels are Associated With Systemic Inflammation. *J Infect Dis* 223:1934–42.
11. Akiyama H, Miller CM, Ettinger CR, Belkina AC, Snyder-Cappione JE, Gummuluru S. 2018. HIV-1 intron-containing RNA expression induces innate immune activation and T cell dysfunction. *Nat Commun* 9:1–12.
12. McCauley SM, Kim K, Nowosielska A, Dauphin A, Yurkovetskiy L, Diehl WE, et al. 2018. Intron-containing RNA from the HIV-1 provirus activates type I interferon and inflammatory cytokines. *Nat Commun*. 9:1–10.
13. Pinzone MR, VanBelzen DJ, Weissman S, Bertuccio MP, Cannon L, Venanzi-Rullo E, et al. 2019. Longitudinal HIV sequencing reveals reservoir expression leading to decay which is obscured by clonal expansion. *Nat Commun* 10:728.
14. Liu R, Simonetti FR, Ho Y-C. 2020. The forces driving clonal expansion of the HIV-1 latent reservoir. *Virol J* 17:1–13.
15. Anderson EM, Simonetti FR, Gorelick RJ, Hill S, Gouzoulis MA, Bell J, et al. 2020.

509 Dynamic Shifts in the HIV Proviral Landscape During Long Term Combination
510 Antiretroviral Therapy: Implications for Persistence and Control of HIV Infections. *Viruses*
511 12:136.

512 16. Bruner KM, Murray AJ, Pollack RA, Soliman MG, Laskey SB, Capoferri AA, et al. 2016.
513 Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nat Med*. 22:1043–
514 9.

515 17. Bruner KM, Wang Z, Simonetti FR, Bender AM, Kwon KJ, Sengupta S, et al. 2019. A
516 novel quantitative approach for measuring the reservoir of latent HIV-1 proviruses. *Nature*
517 566:120.

518 18. Pollack RA, Jones RB, Perteu M, Bruner KM, Martin AR, Thomas AS, et al. 2017.
519 Defective HIV-1 Proviruses Are Expressed and Can Be Recognized by Cytotoxic T
520 Lymphocytes, which Shape the Proviral Landscape. *Cell Host Microbe*. 21:494-506

521 19. Imamichi H, Smith M, Adelsberger JW, Izumi T, Scrimieri F, Sherman BT, et al. 2020.
522 Defective HIV-1 proviruses produce viral proteins. *Proc Natl Acad Sci U S A* 117:3704.

523 20. Yukl SA, Kaiser P, Kim P, Telwatte S, Joshi SK, Vu M, et al. 2018. HIV latency in isolated
524 patient CD4+ T cells may be due to blocks in HIV transcriptional elongation, completion,
525 and splicing. *Sci Transl Med* 10:9927.

526 21. Levy CN, Hughes SM, Roychoudhury P, Reeves DB, Amstutz C, Zhu H, et al. 2021. A
527 highly multiplexed droplet digital PCR assay to measure the intact HIV-1 proviral
528 reservoir. *Cell Reports Med*. 2:100243.

529 22. Liu R, Catalano AA, Ho Y-C. 2021. Measuring the size and decay dynamics of the HIV-1
530 latent reservoir. *Cell Reports Med* 2:100249.

531 23. Antar AAR, Jenike KM, Jang S, Rigau DN, Reeves DB, Hoh R, et al. 2020. Longitudinal
532 study reveals HIV-1–infected CD4+ T cell dynamics during long-term antiretroviral
533 therapy. *J Clin Invest* 130:3543.

534 24. Peluso MJ, Bacchetti P, Ritter KD, Beg S, Lai J, Martin JN, et al. 2020. Differential decay
535 of intact and defective proviral DNA in HIV-1–infected individuals on suppressive
536 antiretroviral therapy. *JCI Insight*. 5:e132997.

537 25. Falcinelli SD, Kilpatrick KW, Read J, Murtagh R, Allard B, Ghofrani S, et al. 2021.
538 Longitudinal Dynamics of Intact HIV Proviral DNA and Outgrowth Virus Frequencies in a
539 Cohort of Individuals Receiving Antiretroviral Therapy. *J Infect Dis* 224:92–100.

540 26. Cochrane AW, Jones KS, Beidas S, Dillon PJ, Skalka AM, Rosen CA. 1991. Identification
541 and characterization of intragenic sequences which repress human immunodeficiency
542 virus structural gene expression. *J Virol*. 65: 5305-13.

543 27. Colin J, Libri D, Porrua O. 2011. Cryptic Transcription and Early Termination in the
544 Control of Gene Expression. *Genet Res Int*. 2011:653494.

545 28. Goffin V, Demonté D, Vanhulle C, de Walque S, de Launoit Y, Burny A, et al. 2005.
546 Transcription factor binding sites in the pol gene intragenic regulatory region of HIV-1 are
547 important for virus infectivity. *Nucleic Acids Res* 33:4285–310.

548 29. Lint C Van, Ghysdael J, Paras P, Jr, Burny A, Verdin E. 1994. A transcriptional regulatory
549 element is associated with a nuclease-hypersensitive site in the pol gene of human
550 immunodeficiency virus type 1. *J Virol* 68:2632.

- 551 30. Verdin E. 1991. DNase I-hypersensitive sites are associated with both long terminal
552 repeats and with the intragenic enhancer of integrated human immunodeficiency virus
553 type 1. *J Virol.* 65: 6790-6799.
- 554 31. Verdin E, Becker N, Bex F, Droogmans L, Burny A. 1990. Identification and
555 characterization of an enhancer in the coding region of the genome of human
556 immunodeficiency virus type 1. *Proc Natl Acad Sci U S A.* 87:4874-8.
- 557 32. Knudsen S. 1999. Promoter2.0: for the recognition of PolII promoter sequences.
558 *Bioinformatics* 15:356–61.
- 559 33. Sloutskin A, Danino YM, Orenstein Y, Zehavi Y, Doniger T, Shamir R, et al. 2015.
560 ElemeNT: a computational tool for detecting core promoter elements. *Transcription* 6:41.
- 561 34. Fuxman Bass JI, Sahni N, Shrestha S, Garcia-Gonzalez A, Mori A, Bhat N, et al. 2015.
562 Human Gene-Centered Transcription Factor Networks for Enhancers and Disease
563 Variants. *Cell* 161:661–73.
- 564 35. Pedro KD, Agosto LM, Sewell JA, Eberenz KA, He X, Bass JIF, et al. 2021. A functional
565 screen identifies transcriptional networks that regulate HIV-1 and HIV-2. *Proc Natl Acad*
566 *Sci USA* 118:e2012835118.
- 567 36. Santoso CS, Li Z, Lal S, Yuan S, Gan KA, Agosto LM, et al. 2020. Comprehensive
568 mapping of the human cytokine gene regulatory network. *Nucleic Acids Res* 48:12055–
569 73.
- 570 37. Imamichi H, Dewar RL, Adelsberger JW, Rehm CA, O'Doherty U, Paxinos EE, et al.
571 2016. Defective HIV-1 proviruses produce novel protein-coding RNA species in HIV-
572 infected patients on combination antiretroviral therapy. *Proc Natl Acad Sci U S A*
573 113:8783.
- 574 38. Jiang C, Lian X, Gao C, Sun X, Einkauf KB, Chevalier JM, et al. 2020. A unique viral
575 reservoir landscape in HIV-1 elite controllers. *Nature* 585:261.
- 576 39. Wang Z, Simonetti FR, Siliciano RF, Laird GM. 2018. Measuring replication competent
577 HIV-1: advances and challenges in defining the latent reservoir. *Retrovirology* 15:21.
- 578 40. Pinzone MR, O'Doherty U. 2018. Measuring integrated HIV DNA ex vivo and in vitro
579 provides insights about how reservoirs are formed and maintained. *Retrovirology* 15:22.
- 580 41. Agosto LM, Yu JJ, Dai J, Kaletsky R, Monie D, O'Doherty U. 2007. HIV-1 integrates into
581 resting CD4+ T cells even at low inoculums as demonstrated with an improved assay for
582 HIV-1 integration. *Virology* 368:60.
- 583 42. Agosto LM, Herring MB, Mothes W, Henderson AJ. 2018. HIV-1-Infected CD4+ T Cells
584 Facilitate Latent Infection of Resting CD4+ T Cells through Cell-Cell Contact. *Cell Rep.*
585 24:2088–100.
- 586 43. Abrahams MR, Joseph SB, Garrett N, Tyers L, Moeser M, Archin N, et al. 2019. The
587 replication-competent HIV-1 latent reservoir is primarily established near the time of
588 therapy initiation. *Sci Transl Med.* 11:eaaw5589.
- 589 44. Arrigo S, Yun M, Beemon K. 1987. cis-acting regulatory elements within gag genes of
590 avian retroviruses. *Mol Cell Biol.* 7:388-97.
- 591 45. Campbell M, Eng C, Luciw PA. 1996. The simian foamy virus type 1 transcriptional

transactivator (Tas) binds and activates an enhancer element in the gag gene. *J Virol.* 70: 6847-55.

46. Löchelt M, Flügel RM, Aboud M. 1994. The human foamy virus internal promoter directs the expression of the functional Bel 1 transactivator and Bet protein early after infection. *J Virol.* 68:638-45.

47. Lochelt M, Muranyi W, Flugel RM. 1993. Human foamy virus genome possesses an internal, Bel-1-dependent and functional promoter. *Proc Natl Acad Sci U S A.* 90:7317-21

48. Reuss FU, Coffin JM. 1998. Mouse Mammary Tumor Virus Superantigen Expression in B Cells Is Regulated by a Central Enhancer within the pol Gene. *J Virol.* 74:8183-7.

49. Graf M, Bojak A, Deml L, Bieler K, Wolf H, Wagner R. 2000. Concerted Action of Multiple cis-Acting Sequences Is Required for Rev Dependence of Late Human Immunodeficiency Virus Type 1 Gene Expression. *J Virol.* 74:10822-6.

50. Kint S, Trypsteen W, De Spiegelaere W, Malatinkova E, Kinloch-de Loes S, De Meyer T, et al. 2020. Underestimated effect of intragenic HIV-1 DNA methylation on viral transcription in infected individuals. *Clin Epigenetics* 12:1–11.

51. Deeks SG, Tracy R, Douek DC. 2013. Systemic Effects of Inflammation on Health during Chronic HIV Infection. *Immunity* 39:633.

52. Višković K, Židovec Lepej S, Gorenec A, Grgić I, Lukas D, Zekan Š, et al. 2018. Cardiovascular markers of inflammation and serum lipid levels in HIV-infected patients with undetectable viremia. *Sci Reports* 8:1–8.

53. Lopez Angel CJ, Pham EA, Du H, Vallania F, Fram BJ, Perez K, et al. 2021. 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* 118:e2022928118.

54. Champiat S, Raposo RAS, Maness NJ, Lehman JL, Purtell SE, Hasenkrug AM, et al. 2012. Influence of HAART on Alternative Reading Frame Immune Responses over the Course of HIV-1 Infection. *PLoS One.* 7(6).

55. Bansal A, Carlson J, Yan J, Akinsiku OT, Schaefer M, Sabbaj S, et al. 2010. CD8 T cell response and evolutionary pressure to HIV-1 cryptic epitopes derived from antisense transcription. *J Exp Med.* 207:51–9.

56. Cardinaud S, Moris A, Février M, Rohrlisch PS, Weiss L, Langlade-Demoyen P, et al. 2004. Identification of Cryptic MHC I-restricted Epitopes Encoded by HIV-1 Alternative Reading Frames. *J Exp Med.* 199:1053–63.

57. Akiyama H, Jalloh S, Park S, Lei M, Mostoslavsky G, Gummuluru S. 2020. Expression of HIV-1 intron-containing RNA in microglia induces inflammatory responses. *J. Virol.* 95:e01386.

58. Gervaix A, West D, Leoni LM, Richman DD, Wong-Staal F, Corbeil J. 1997. A new reporter cell line to monitor HIV infection and drug susceptibility in vitro. *Proc Natl Acad Sci U S A.* 94:4653.

59. Gagne M, Michaels D, Schiralli Lester GM, Gummuluru S, Wong WW, Henderson AJ. 2019. Strength of T cell signaling regulates HIV-1 replication and establishment of latency. *PLOS Pathog* 15:e1007802.

634

635

636

637

638

639

Supplemental Materials

640

Supplemental Figure 1. IPDA Assay 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 1×10^6 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.

Supplemental Figure 2. 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-1_{NL4-3}. For CD4+ T cells, EFV was added at a concentration of 10 μ M 30 minutes prior to spinoculation with HIV-1_{NL4-3}. 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 patient matched. Data are from 3 separate donors. (B) IPDA data for EFV pretreated MDMs infected with HIV-1_{NL4-3-BaL}. For MDMs, EFV was added at a concentration of 10 μ M 24-hours prior to addition of HIV-1_{NL4-3-BaL} to the cell cultures. Data are from 3 separate donors.

Supplemental Figure 3. 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.

Supplemental Figure 4. 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_{NL4-3} BaL infected MDMs or HIV-1_{NL4-3} 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 is patient matched between unstimulated and activated cells.

Supplemental Table 1. Primer and Probe Sequences.

IPDA Primers*			
Assay	Primer Name	Function	Sequence (5' to 3')
Psi	Psi-Forward	Forward	CAGGACTCGGCTTGCTGAAG
	Psi-Reverse	Reverse	GCACCCATCTCTCTCTTCTAGC
Env	Env-Forward	Forward	AGTGGTGCAGAGAGAAAAAGAGC
	Env-Reverse	Reverse	GTCTGGCCTGTACCGTCAGC
RPP30-1	RPP30-1-F	Forward	GATTGGACCTGCGAGC
	RPP30-1-R	Reverse	GCGGCTGTCTCCACAAG
RPP30-2	RPP30-2-F	Forward	GACACAATGTTTGGTACATGGTTAA
	RPP30-2-R	Reverse	CTTTGCTTTGTATGTTGGCAGAAA

IPDA Probes*		
Probe Name	Fluorophore	Sequence (5' to 3')
Psi	FAM	TTTTGGCGTACTCACCAGT
Env	VIC	CCTTGGGTTCTTGGGA
Env Hypermutation	None	CCTTAGGTTCTTAGGAGC
RPP30-1	HEX	CTGACCTGAAGGCTCT
RPP30-2	FAM	CCATCTCACCAATCATTCTCCTTCCTTC

RT-ddPCR Primers**			
Assay	Primer Name	Function	Sequence
Long LTR	Kumar F	Forward	GCCTCAATAAAGCTTGCCTTGA
	Kumar R	Reverse	GGGCGCCACTGCTAGAGA
Nef	F8883-03	Forward	GGTGGGAGCAGTATCTCGAGA
	R9040-10	Reverse	TGTAAGTCATTGGTCTTAAAGGTACCTGAGG

RT-ddPCR Probes**			
Assay	Primer Name	Fluorophore	Sequence
Long LTR	Kumar P	FAM	CCAGAGTCACACAACAGACGGGCACA
Nef	P8967-50	FAM	CCAGGCACAAGCAGCATT

*IPDA primer and probe designs first described by Bruner et al.(18)

**RT-ddPCR primer and probe designs first described by Yukl et al.(21)

5' RACE - Gene Specific Nef Primer	
Primer Name	Sequence (5' to 3')
RACE-Nef-R	GATGGGGTGGGAGCAGTATCTCGAGACC

710 **Supplemental Table 2. Patient Details - People Living with HIV on Antiretroviral Therapy**

Subject	Age	Sex	Race	Ethnicity	CD4 at enrollment	Nadir CD4	Years on ART
A014	66	Male	White	Not Hispanic or Latino	1523	508	12.4
A033	60	Female	Black / African American	Not Hispanic or Latino	394	394	6.4
A042	68	Female	Black / African American	Not Hispanic or Latino	494	619	13.2
A044	54	Female	White	Not Hispanic or Latino	932	192	7.2
A058	64	Male	Black / African American	Not Hispanic or Latino	567	567	6.7
A061	59	Female	Black / African American	Not Hispanic or Latino	625	543	12.5
W102	75	Male	Black / African American	Not Hispanic or Latino		182	5.1
W124	67	Male	White	Not Hispanic or Latino	810	284	4.4
W143	33	Male	Other	Hispanic/Latino	471	270	3.2
X101	55	Female	Black / African American	Not Hispanic or Latino	414	259	2.7
X104	51	Male	Black / African American	Not Hispanic or Latino	1055	607	3.3

711

712

713

Supplemental Table 3. HIV-1 Tile Sequences

NL43 Region Number	HXB2 Sequence Range
NL43-1	501-1527
NL43-2	1031-2022
NL43-3	1501-2537
NL43-4	2003-3022
NL43-5	2512-3549
NL43-6	3001-4024
NL43-7	3528-4522
NL43-8	4001-5033
NL43-9	4501-5531
NL43-10	5007-6026
NL43-11	5506-6539
NL43-12	6001-7027
NL43-13	6515-7525
NL43-14	7001-8027
NL43-15	7501-8521
NL43-16	8003-9016
NL43-17	8501-9709

717 **Supplemental Table #4.** Human Transcription Factor Binding Events Identified by Yeast-1-
718 Hybrid Assay with HIV-1 Tile Sequences

Transcription Factor	Family	HIV Tile Sequence
Iroquois homeobox protein 5	HD	HIV1-2, HIV1-11, HIV1-14, HIV1-15
Nuclear factor 1 A-type	MH1	HIV1-2
Iroquois homeobox protein 6	HD	HIV1-2, HIV1-11, HIV1-15
Homeobox protein Hox-C8	HD	HIV1-2
SRY-Box Transcription Factor 14	HMG	HIV1-2, HIV1-16
Nuclear factor 1 C-type	MH1	HIV1-2
PLAG1 Like Zinc Finger 1	ZF-C2H2	HIV1-2
Homeobox protein Hox-A6	HD-HOX	HIV1-2, HIV1-16
Transcription factor AP-2 beta	AP-2	HIV1-10, HIV1-14, HIV1-15
Grainyhead Like Transcription Factor 2	CP2	HIV1-10, HIV1-14, HIV1-16
Forkhead box protein H1	WH - FH	HIV1-11
Twist Family BHLH Transcription Factor 1	bHLH	HIV1-11
Class B basic helix-loop-helix protein 8	bHLH	HIV1-11
Runt-related transcription factor 1	RUNT	HIV1-11
Runt-related transcription factor 3	RUNT	HIV1-11
Zinc finger protein basonuclin-1	ZF - C2H2	HIV1-11
Zinc Finger Protein 710	ZF - C2H2	HIV1-11, HIV1-16
Musculin	bHLH	HIV1-11
Transcription factor 21	bHLH	HIV1-11
Transcription factor EC	bHLH	HIV1-11
MYC Associated Factor X	bHLH	HIV1-11
Atonal BHLH Transcription Factor 1	bHLH	HIV1-11, HIV1-14
T-Box Transcription Factor 22	T-BOX	HIV1-11
POU Class 4 Homeobox 3	HD - POU	HIV1-11
Myogenic Factor 6	bHLH	HIV1-11

Transcription Factor	Family	HIV Tile Sequence
Zinc Finger And BTB Domain Containing 10	ZF - C2H2	HIV1-11, HIV1-14
Glial Cells Missing Transcription Factor 1	GCM	HIV1-11
ETS2 Repressor Factor	WH - ETS	HIV1-14
CAMP Responsive Element Binding Protein 5	bZIP	HIV1-14
Zic Family Member 3	ZF - C2H2	HIV1-14
Zic Family Member 1	ZF - C2H2	HIV1-14
E74 Like ETS Transcription Factor 2	WH - ETS	HIV1-14, HIV1-16
Transcription Factor AP-2 Alpha	AP-2	HIV1-14
T Cell Leukemia Homeobox 3	HD	HIV1-14
Haematopoietically-expressed homeobox protein	HD	HIV1-14, HIV1-16
Early B-cell factor 3	IPT/TIG	HIV1-15, HIV1-16
Forkhead Box J2	WH - FH	HIV1-16
Krüppel Like Factor 4	ZF - C2H2	HIV1-16
SIX Homeobox 6	HD - SIX	HIV1-16
Spi-1 Proto-Oncogene	WH - ETS	HIV1-16
REL Proto-Oncogene, NF-KB Subunit	IPT/TIG, p53	HIV1-16
Early B-cell factor 1	IPT/TIG	HIV1-16
Nuclear Receptor Subfamily 2 Group F Member 2	ZF - NHR	HIV1-16

719

720

721

722

723

724

725

726