In vivo detection of HIV-1 antisense transcripts in untreated and ART-treated individuals

Adam A. Capoferri[1]*†, Rachel Sklutuis[1]†, Toluleke O. Famuyiwa[1], Sachi Pathak[1], Rui Li[2], Jason W. Rausch[1], Brian T. Luke[3], Rebecca Hoh[4], Steven G. Deeks[4], John W. Mellors[5], John M. Coffin[6], Jennifer L. Groebner[1], Fabio Romerio[2], Mary F. Kearney[1]*

[1] HIV Dynamics and Replication Program, National Cancer Institute, Frederick, MD, USA

[2] Department of Molecular and Comparative Pathobiology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[3] Leidos Biomedical Research, Inc, Frederick National Laboratories for Cancer Research, Frederick, MD, USA

[4] Department of Medicine, University of California, San Francisco, CA, USA

[5] Division of Infectious Diseases, Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

[6] Department of Molecular Biology and Microbiology, Tufts University, Boston, MA, USA

† Contributed Equally

* Corresponding Author

Corresponding Author:

Mary F. Kearney, Ph.D.
HIV Replication and Dynamics Program
National Cancer Institute at Frederick
1050 Boyles Street, Building 535, Room 310
Frederick, Maryland 21702-1201
Tel: 301 305 8197
Email: kearneym@nih.gov

Adam A. Capoferri, Ph.D.
HIV Replication and Dynamics Program
National Cancer Institute at Frederick
1050 Boyles Street, Building 535, Room 111
Frederick, Maryland 21702-1201
Tel: 301 846 7365
Email: adam.capoferri@nih.gov

Abstract

Natural antisense transcripts are expressed in eukaryotes, prokaryotes, and viruses and can possess regulatory functions at the transcriptional and/or post-transcriptional levels. *In vitro* studies have shown that HIV-1 antisense transcripts (AST) promote viral latency through epigenetic silencing of the proviral 5′ long terminal repeat (LTR). However, expression of HIV-1 AST *in vivo* have not been convincingly demonstrated. Here, we used single RNA template amplification, detection, and sequencing to demonstrate expression of AST in unstimulated PBMC collected from people with HIV-1 (PWH). We found that AST had high genetic diversity that matched proviruses in cells from blood and lymph nodes. We measured a median of 26 copies of AST per 100 infected cells in PWH on ART and a median of 2 copies per 100 infected cells in PWH not on ART. The expression of HIV-1 AST *in vivo* is consistent with a potential regulatory role in regulation of HIV-1 expression.

Antisense transcripts (AST) are RNA molecules transcribed from the opposite strand of a protein-coding gene that can have protein-coding and/or non-coding activities (1). AST have been identified in eukaryotes, prokaryotes, and viruses and have been shown to possess regulatory functions at both the transcriptional and post-transcriptional levels via multiple mechanisms (2). AST have previously been documented to be encoded by several viruses that infect eukaryotes including members of the *Herpesviridae* (*e.g.* herpes simplex virus-1 and cytomegalovirus) (3, 4) and *Retroviridae* (5-7) families. One of the best characterized AST in a viral system is the *Hbz* gene in Human T-cell Leukemia Virus Type 1 (HTLV-1). The interplay between *Hbz* (antisense) and *Tax* (sense) RNA expression and their protein products is thought to modulate the regulation of cellular pathways that promote survival and proliferation of HTLV-1 infected cells, thereby influencing the progression into adult T cell leukemia/lymphoma or HTLV-1-associated myelopathy/tropical spastic paraparesis (8).

R.H. Miller first provided evidence of an antisense gene (named *asp*) overlapping the *env* gene in the HIV-1 genome (9). Several *in vitro* studies have since demonstrated the presence of a Tat-independent negative sense promoter in the HIV-1 3′ long terminal repeat (LTR) (5, 10, 11), which drives the expression of multiple antisense transcripts (12) with both protein coding (13) and non-coding functions (14). The presence of HIV-1 antisense protein (ASP)-specific antibodies in serum (15) and cytotoxic CD8+ T-lymphocytes (CTLs) in blood samples (16) have been detected from people living with HIV (PWH), thus providing indirect evidence for the expression of ASP *in vivo*. Detection of ASP has been reported in eight chronically infected lymphoid and myeloid cell lines during latent and productive infection (17). HIV-1 AST have been shown to inhibit viral replication and to promote and maintain latency in stably expressing CD4+ T cell lines (18, 19). Although *in vitro* studies suggest that AST function in viral latency, only a few reports have examined its expression *in vivo*. Zapata *et al.* reported low levels of AST in 3 PWH on antiretroviral therapy (ART) (>2 years with undetectable levels of viremia) using real-time PCR in unstimulated CD4+ T cells (19) and Mancarella *et al.* (20) detected AST in CD4+ T cells collected from both untreated and ART-treated PWH after *ex vivo* stimulation with anti-CD3/CD28. However, no studies have provided sequence evidence for expression of AST in unstimulated cells collected from PWH.

Therefore, we set out to determine if we could measure and genetically characterize HIV-1 AST in unmodified peripheral blood mononuclear cells (PBMC) collected from donors on and off ART, and to compare thier expression levels to that of HIV-1 sense transcripts in the same donors. Investigating expression of HIV-1 AST *in vivo* may contribute to our understanding of HIV-1 persistence and reveal new targets for controlling HIV-1 expression without ART.

Results

Participants and samples

To determine the levels of AST during chronic infection, PBMC were collected from 3 PWH on ART (*21*) and 5 PWH who were not on ART (**Table 1**) and who were enrolled either at University of California of San Francisco under the SCOPE trial (clinical trial # NCT00187512) or at University of Pittsburgh under the Optimization of Immunologic and Virologic Assays for HIV Trial (IRB# STUDY20040215). All donors provided written informed consent for the study. For one donor (PID 2669), samples collected at four longitudinal timepoints before and after an ART interruption and re-initiation were available.

Participants were mostly assigned male at birth (n=7/8). Race/ethnicity as reported by donors was Black (n=4/8), White (n=3/8), and Latino (n=1/8). Donors were diagnosed with HIV-1 subtype B with a minimum duration of infection median of 8 years [IQR 5-13 years] prior to sample collection, however this information was unknown for one participant (PID 2669), who was on study more than 5 years. Donors not on ART (untreated) were either ART naïve or not currently on an ART regimen due to a planned or unplanned interruption, and their plasma viremia levels were detectable (median 139,945 range: 128 - 275,685 HIV-1 RNA copies/mL). All but one of the samples collected from PWH on ART had undetectable levels of viremia (<50 HIV-1 RNA copies/mL) with durations of treatment ranging from 2 weeks to 12.8 years. The one sample with detectable viremia on ART (112 HIV-1 RNA copies/mL) was collected from PID 2669 2 weeks after ART re-initiation after an unplanned 4-week treatment interruption. Levels of

plasma viremia were measured as HIV-1 RNA copies/mL by either Abbott® real-time assay or COBAS® HIV-1 test.

HIV-1 AST were detected in ART-treated and untreated PWH, with higher levels in samples collected during ART

Our methods for detecting and measuring levels of HIV-1 AST were modified from Wiegand *et al.* (22), Capoferri *et al.* (23), and Zapata *et al.* (19) and are described in detail in the **Supplementary Materials**. Briefly, to measure levels of HIV-1 AST *in vivo*, we extracted total cell-associated RNA from unmodified donor PBMC with known estimated numbers of HIV-1 infected cells (24), synthesized cDNA using participant-specific exogenous oligo-tagged gene-specific primers targeting AST in the *env* coding region (19), and quantified the number of AST in each sample with participant-specific anti-*env* primers and probes (primer/probe sequences in **Tables S1-S5**) in a digital PCR format. Prior to testing the donor PBMC, we optimized the AST assay on antisense RNA in the ACH-2 cell line and found these cells to express a median of 41 AST per 100 infected ACH-2 cells (**Supplementary Materials**). Extensive controls were performed to ensure complete degradation of HIV-1 DNA (22, 23) and to determine the cut-off for HIV-1 AST (details in **Supplemental Materials**). Equal numbers of no reverse transcriptase wells (negative controls) and experimental wells were included on each PCR plate to ensure no HIV-1 DNA contamination.

AST were detected in all 8 donors (**Fig. 1A & Table S6**). The level of AST in ART-treated samples was a median of 26 copies/100 infected PBMC [IQR 16-47] and the level in untreated samples was a median of ≤ 2 copies/100 infected PBMC [IQR 1-19] (p=0.05, exact p=0.048, Mann-Whitney U-test). Since PID 2669 was overrepresented in the treated group, we also performed the comparison by aggregating the AST in the 2 timepoints prior to the ART interruption, to include only one data point per donor. The re-analysis resulted in a median of 17 AST/100 infected PBMC in the donors on ART vs. ≤ 2 AST/100 infected PBMC in the untreated donors (p=0.21, Mann-Whitney U-test). In PID 3611 (untreated), only 1 AST molecule

was detected in 244 infected cells, which was below the cut-off for the assay (details for cut-off in **Supplementary Materials**), indicating that the AST in this donor sample was ≤ 0.4 copies/100 infected PBMC. In PID 1775 (untreated), detection of 33 copies/100 infected PBMC was an outlier among the other untreated PWH (≤ 0.4-5 copies/100 infected PBMC), as confirmed by the Grubbs' Test for Outliers (G=1.8, α=0.05) (*25*). In PID 2669 (treated), AST were detected in samples collected at all 4 timepoints: 4.3 and 5.5 years after ART initiation and 2 weeks and 4 weeks after ART re-initiation following a brief treatment interruption. Samples after ART re-initiation had higher levels of AST than samples collected prior to ART interruption (49% of 133 infected PBMC vs. 25% of 155 infected PBMC) ($p=3.4x10^{-9}$, binomial test).

HIV-1 AST were detected at modestly higher levels than HIV-1 sense env transcripts in samples on ART

In the samples on ART, we also measured the levels of sense *env* RNA targeting the same sub-genomic region targeted for AST (**Fig. 1B**). Levels of sense *env* RNA (aggregate of unspliced and partially spliced) were quantified similarly as above but as a separate reaction from HIV-1 AST. Levels of AST were modestly higher than sense *env* RNA (median 26 copies/100 infected PBMC [IQR 16-47] vs. median 16 copies/100 infected PBMC [IQR 5-25]) (paired t-test t(5)=2.79, p=0.04).

Long fragments of HIV-1 AST were amplified from single infected cells from donors on ART

Because of the higher levels of AST in the donors on ART, we were able to amplify longer fragments in this subset. We amplified 1.7-kb fragments of AST spanning from the negative sense promoter in the 3′ LTR to *env* (referred to here as "long AST"). Using a modified version of the CARD-SGS assay (*22*, *23*), we used the sequence data to estimate the fraction of infected cells with "long AST" and the levels of AST in single infected cells (**Table S2**; details in **Supplementary Materials**). We found that a median of 4.1% [IQR 1.6 – 5.2%] of infected PBMC had detectable levels of "long AST" with a median of 1.1 copies/cell [IQR 1.0-1.7] at a given point in time, indicating that the frequency of detection of "long AST" was lower than the short AST fragments detected by the digital PCR approach (above). In PID 2669, we

found no significant difference in the fraction of cells with "long AST" or the levels of AST in single infected cells across the four time points (Kruskal-Wallis test, H(3)=3.02, p=0.39). Additionally, we found no significant difference when we aggregated the samples prior to ART interruption and after ART-reinitiation (p=0.83, Mann-Whitney U-test).

Phylogenetic analysis of AST reveals that their expression originates from a diverse population of proviruses

Having successfully amplified the 1.7-kb segments of HIV-1 AST in the donors on ART by RT-PCR, we sequenced the resulting PCR products and performed phylogenetic analyses to compare the genetics of the AST to the proviruses in the same population of infected cells (**Fig. 2**). Standard HIV-1 *env* DNA single-genome sequencing using PBMC from all 3 donors and lymph node mononuclear cells (LNMC) from 2 of the donors was performed previously (*21*) and the data used here as a reference. Neighbor-joining trees were reconstructed using HIV-1 *env* DNA from PBMC and LNMC and ~1.0-kb of the AST in the same genetic region. Trees were rooted on the HIV-1 subtype B consensus sequence. Symbols on the trees show HIV-1 *env* DNA from PBMC (black triangles), HIV-1 *env* DNA from LNMC (blue triangles), and AST (multicolored squares where each color is obtained from a different aliquot of PBMC). AST matching HIV-1 *env* DNA in PBMC or LNMC are indicated with black arrows; AST from infected probable T cell clones are indicated with blue arrows. In PID 1079, a cell expressing high levels of AST is indicated with a red arrow. The genetic diversity of AST was measured by average pairwise distance (APD) with predicted hypermutant sequences removed (*26*).

We detected a high genetic diversity of AST (ranging from 0.7–2.5%) matching a diverse population of proviruses in both PBMC and in LNMC (1.1-2.1%), indicating that a wide variety of proviruses can express AST (**Fig. 2**). In one aliquot of PBMC from PID 1079, we found a rake of 30 identical AST, suggesting that they may have originated from the same infected cell (red arrow, **Fig. 2A**). Further, we found identical AST across multiple aliquots of infected PBMC in all 3 donors, indicating that these AST may have originated from infected T cell clones (blue arrows). In PID 1683, we also found

identical PBMC AST matching proviruses in both PBMC and LNMC (**Fig. 2B**). While we did not directly assess the presence of AST in LNMC due to limited sampling with fine needle aspirates, these data suggest that AST may be expressed in cell clones that are present in tissues as well as in blood. AST were highly diverse in all 4 samples collected from PID 2669: T1 (2.9% APD), T2 (2.4%), T3 (2.6%), and T4 (2.4%). We also identified 11 likely T cell clones that contained some cells with AST (blue arrows). Eight of the 11 clones were found to persist across multiple timepoints, mostly either before or after the ART interruption, and only very rarely persisting both before and after the interruption, suggesting that ART interruption may influence the populations of T cells that express HIV-1 AST (**Fig. 2C**).

AST can be detected in gag, pol, and env coding regions in vivo

In vitro studies have detected AST of varying lengths, including "full-length" (*i.e.*, from the U3/R of the 3′ LTR to *gag*) (*5*, *11*, *12*) (**Fig. S1A, Supplementary Materials**). We asked if AST could be detected, not only in *env*, but in other HIV genomic regions *in vivo*. We measured AST in *gag* (HXB2: 764-2,281) (*27*), *gag*/*pol* (HXB2: 1,849-3,500) (*22*) and *pol*/*vif* (HXB2: 3,996-5,270) in one donor sample (PID 2669 Timepoint #4) (*28*) (**Fig. S3A**). Cell-associated RNA from three aliquots of ~90 infected PBMC was extracted and cDNA targeting the antisense strand of *gag*, *gag/pol*, and *pol/vif* was synthesized using exogenous oligo-tagged gene-specific primers, followed by endpoint PCR amplification and Sanger sequencing (primers in **Table S1,4**). We detected AST in all genomic regions assayed. However, we found lower levels of genetic diversity for AST in *gag*, *gag/pol*, and *pol/vif* regions compared to AST in *env* at the same time points (1.5% vs 3.2%) (**Fig. S3B**). Comparable to levels of "long AST" in *env* (1.7kb described above), levels of AST in *gag*, *gag/pol*, and *pol/vif* were 1.0, 1.1, and 1.0 copies/cell, respectively. Although the sub-genomic regions cannot be genetically linked, these findings suggests that AST may span the entire HIV-1 genome in some infected cells.

Discussion

In this study, we used established quantitative PCR methods (*19*), extensive positive and negative controls, and sequencing to demonstrate the expression of HIV-1 antisense RNA in people living with HIV who are either untreated or are on long or short-term ART. Prior to our study, HIV-1 antisense RNA had been shown to promote and maintain viral latency in stably expressing cell lines by recruiting the enhancer of zeste homolog 2 (EZH2), a core component of the polycomb repressive complex 2 (PRC2), to the HIV-1 5′ LTR (*18, 19*). Recruitment of EZH2 catalyzes trimethylation of lysine 27 on histone H3 (H3K27me3), a suppressive epigenetic mark that promotes nucleosome assembly and suppression of viral transcription. HIV-1 AST are inefficiently polyadenylated and predominately retained in the nucleus to act as a lncRNA (*29*). Although HIV-1 AST have been shown to inhibit viral replication and promote the establishment and maintenance of latency *in vitro* (*19, 30*), studies investigating AST *in vivo* have been limited. Therefore, we sought to determine if AST are expressed *in vivo* in untreated and/or ART-treated PWH, to quantify their levels in bulk and in single infected cells, and to characterize their genetics. To achieve this, we used a digital PCR assay for the detection and quantification of AST in unstimulated PBMC and we used a modified version of CARD-SGS (*22, 23*) to measure the fraction of infected cells with longer fragments of AST (1.7kb from 3′ LTR to *env*), the levels of AST in single infected cells, and the genetic diversity of AST in the unstimulated PBMC populations. Although methods for the CARD-SGS assay have been published previously (*22, 23*), we included detailed methods in the **Supplementary Materials**.

We detected HIV-1 AST in all 8 donors independent of treatment status. However, we found 13-fold higher levels of AST in PBMC samples collected on ART compared to those not on ART, perhaps consistent with AST functioning in the maintenance of viral latency (*19, 30*). An alternative interpretation could be that, in untreated donors, there are higher levels of transcription from the 5′ LTR. RNA polymerase collision (*i.e.*, transcriptional interference, either "sitting duck" or "roadblock") or RNA:RNA hybrid formation between 5′ LTR-driven and 3′ LTR-driven transcription may suppress the expression of AST. Except for one donor on ART (PID 1079) where we identified one cell that may have had about 30 copies of AST, the level of AST in single cells was very low, with a mean of 1.2 copy/cell (range 1-30 copies/cell).

The low levels of HIV-1 AST are consistent with those reported for other eukaryotic AST. In about 25% of protein-coding genes, AST are expressed at approximately 1 or a few copies per cell at any given time (31). In contrast, some AST can be expressed at very high levels, as observed for human *MALAT1* at about 150 TPM (32). This range of antisense lncRNA expression has been well documented across cellular and tissue types in the Functional Annotation of the Mammalian Genome (FANTOM) (33-35), the Genotype-Tissue Expression (GTEx) consortium (32), the Encyclopedia of DNA Elements (ENCODE) project (31), and the Long non-coding RNA Knowledgebase (36).

In the 3 donors on ART, we amplified and sequenced the 1.7-kb fragment of HIV-1 AST overlapping *env* from about 80 infected cells per sample. We found high genetic diversity of the AST including matches to proviruses in peripheral blood and lymph nodes. In some instances, the AST matched both PBMC and LNMC proviral DNA, suggesting that at least some cells in infected CD4+ T cell clones can express HIV-1 AST. In the one donor on ART with samples from multiple timepoints (PID 2669), we identified probable T cell clones with detectable levels of AST that persisted over time, and one that matched proviral DNA found in both PBMC and LNMC. Interestingly, although identical AST were found across the timepoints before the ART interruption and across the timepoints after the ART interruption and re-initiation, only rarely were identical AST detected both before and after the interruption, suggesting that the treatment interruption may have influenced the populations of T cell clones expressing HIV-1 AST. Further supporting the influence of ART interruption on AST expression was the observation that levels of AST were significantly elevated in samples collected within weeks after ART re-initiation relative to samples collected on longer-term ART (4-5 years on ART). It is possible that levels of AST in donors after ART re-initiation may be different than in long-term treated individuals or that relatively short-lived latently infected cells express AST at higher levels, thus, the effect of ART interruption on levels of AST should be further investigated. Together these data show that HIV-1 AST expression, although typically at low levels, can be detected in probable infected T cell clones in both blood and tissues.

Previous *in vitro* studies have reported multiple HIV-1 AST species using northern blot and 5′ Rapid Amplification of cDNA Ends (5′ RACE): Class I (10kb) (12), II (5.5kb) (11), III-iii (3kb) (12), and

IV-ii (2kb) (5) (**Fig. S1A**). Our detection of HIV-1 AST in genes other than in *gag*, *gag/pol*, and *pol/vif* at levels comparable to those we observed in *env*, together with the work by Kobayashi-Ishihara *et al*. (12), suggest that we detected numerous Class I transcripts, which are full-length, expressed at low levels, and not typically polyadenylated. Relatedly, it is important to note that neither our digital PCR nor our single-genome sequencing assays targeting *env* is informative as to which Class the identified AST transcripts belong.

There are some limitations of this study to consider. Due to the fragility and sparsity of HIV-1 AST, we almost certainly underestimate their frequency. We cannot rule out that HIV-1 AST are short-lived and may not always be detectable. Similarly, there is a potential loss of AST from inefficient reverse transcription due to the thermal stability of secondary and tertiary RNA structures (although the high temperature denaturing step leading into cDNA synthesis is designed to help mitigate such effects), or during the sodium acetate/ethanol precipitation of the antisense cDNA. These recovery limitations are supported by the controls included in the **Supplementary Materials** demonstrating recoveries of only ~50% for AST spiked samples. Detecting a single copy of HIV-1 AST naturally presents challenges, particularly in determining if the molecules are due to expression by an HIV-1 or host promoter. Readthrough generation of HIV-1 AST can occur when the provirus integrates opposite to a host gene promoter, allowing transcription into the HIV-1 3′ LTR to produce AST. Previous research identifying the transcription start site of HIV-1 AST *in vitro* used infected cell lines and 5′ RACE (11, 12). We attempted 5′ RACE on our donor samples to determine if the HIV-1 AST originated from an HIV-1 or host promoter at a single RNA molecule level, but without success due to limitations in assay sensitivity. More sensitive 5′ RACE technologies are needed to determine the transcription start site of single RNA molecules *in vivo*, including HIV-1 AST. However, our finding that there are differences in levels of AST across different cells of a T cell clone may favor expression from the HIV 3′ LTR promoter.

This study is the first to show, with sequence confirmation, the expression of HIV-1 AST in PWH. Additional studies are needed to confirm our observation of higher levels of AST in PWH on ART vs. PWH

who are not on ART, and to determine if HIV-1 AST *in vivo* are driven by viral or host promoters. This study, together with those showing the role of HIV-1 AST in inducing and maintaining HIV-1 latency *in vitro* and *ex vivo* (Li *et al.* Romerio, submitted), highlights a previously underexplored potential determinant of HIV-1 persistence both before and during ART, and may lead to new directions for the development of approaches to controlling HIV-1 viremia without ART.

Acknowledgements: We would like to thank the donors for participating in this study. We acknowledge our collaborative interactions with the Behavior of HIV in Viral Environments Center (U54AI170855). We thank Teresa Burdette and Ann Wiegand for administrative support.

Financial Support: This work was supported by intramural NCI funding (ZIA BC 011697) to the HIV Dynamics and Replication Program (MFK) and by the Office of AIDS Research. This work was also supported through an Intramural AIDS Research Fellowship (AAC) from the Office of AIDS Research. Other funders include NCI contract No. 75N91019D00024 to BTL, Leidos Biomedical Research, Inc. subcontract 12XS547 to JWM and 13SX110 to JMC. JMC was a Research Professor of the American Cancer Society and supported in part by Research Grants CA R35 200421 and AI R01 184043. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute, National Institutes of Health, or the Department of Health and Human Services.

Author Contributions: AAC, RS, JWR, JMC, JLG, MFK designed the study. AAC, RS, TOF, SP, JLG performed experiments. AAC, RS, TOF, JWR, BTL, JLG, MFK analyzed results. AAC, RS, JLG performed sequence data analyses. SGD and JWM provided patient samples. AAC, RS, TOF, FR, MFK wrote the manuscript. All authors reviewed the manuscript. AAC and RS contributed equally to this work.

Competing Interest Statement: JWM is a consultant to Gilead Sciences, has received research grants from Gilead Sciences to the University of Pittsburgh, and owns share options in Infectious Disease Connect (co-founder) and Galapagos, NV, unrelated to the current work on HIV. JMC is a member of the Scientific Advisory Board and a Shareholder of ROME Therapeutics, Inc. and Generate Biomedicine, Inc. The remaining authors have no potential conflicts.

Data and Materials Availability: All sequence data are available on GenBank (in process of submission). Sequences from the donors on ART were previously published by McManus and colleagues and can be found in (*21*).

Supplementary Materials

Materials and Methods

Figs. S1-S4

Tables S1 to S6

References: 37-40

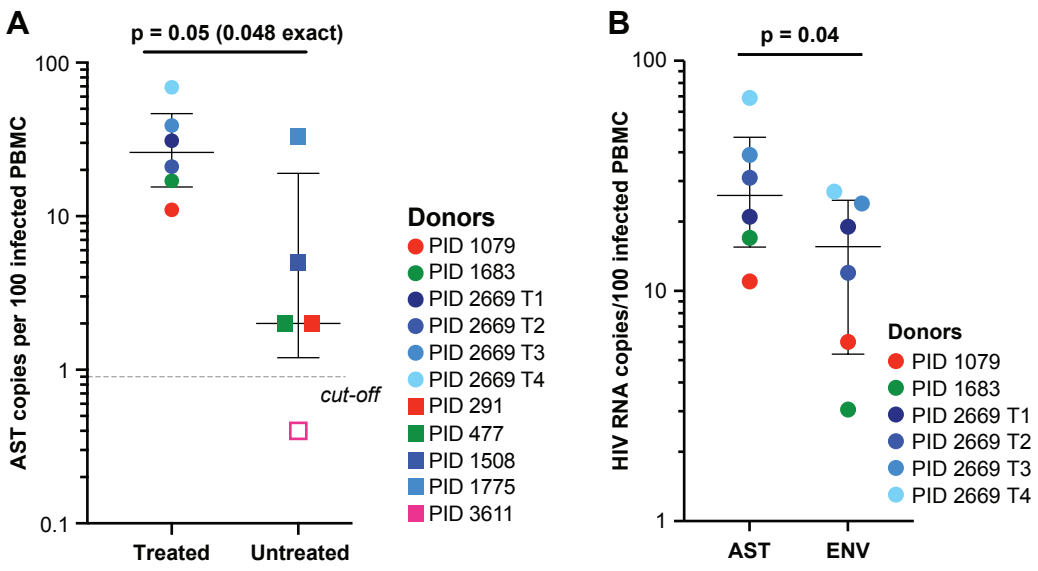


Fig. 1. Levels of HIV-1 AST and sense *env* transcripts in PWH. (**A**) Levels of HIV-1 AST in ART-treated and untreated PWH. PWH on ART (circles) and PWH not on ART (squares), each color represents an individual, and PID 2669 has multiple shades of one color to indicate longitudinal sampling. P value determined with Mann-Whitney U-test. Open shape indicates a sample where AST was detected but was below the assay cut-off. The cut-off was determined as the number of positive RT- wells per number of infected cells assayed across 16 replicates of 100 ACH-2 in a background of 10^5 CEM cells (see **Supplementary Materials**). (**B**) Detection of AST vs. sense *env* in ART-treated PWH. P value determined with paired t-test.

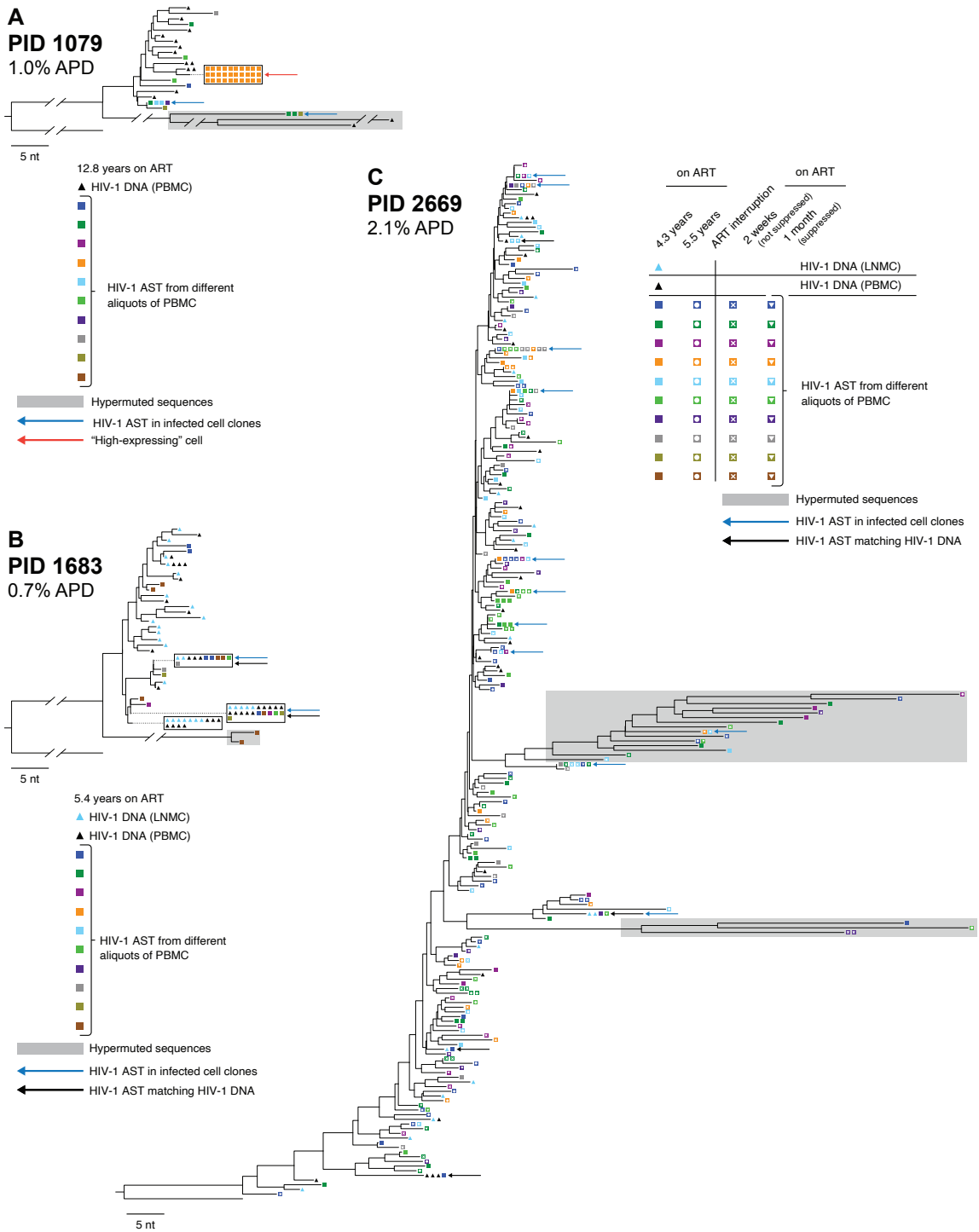


Fig. 2. Distance trees of AST in donors on ART. HIV-1 AST sequenced using the modified CARD-SGS assay (**Supplementary Materials**). Percent average pairwise distance (APD) was calculated without hypermutants. HIV-1 AST sequences were aligned with the HIV-1 DNA *env* sequences (previously reported in McManus *et al*. (*21*)). Black triangles show proviruses from PBMC, light blue triangles show proviruses from lymph node mononuclear cells (LNMC), and squares are the intracellular AST from PBMC

with each color representing a different aliquot of PBMC. Blue arrows indicate identical AST sequences found in >1 aliquot of PBMC, black arrows indicate identical AST sequences that matched either PBMC or LMNC HIV-1 DNA, and the red arrow indicates a "high AST-expressing" cell. Predicted hypermutant sequences are shaded in gray. Trees are rooted on consensus subtype B *env.*

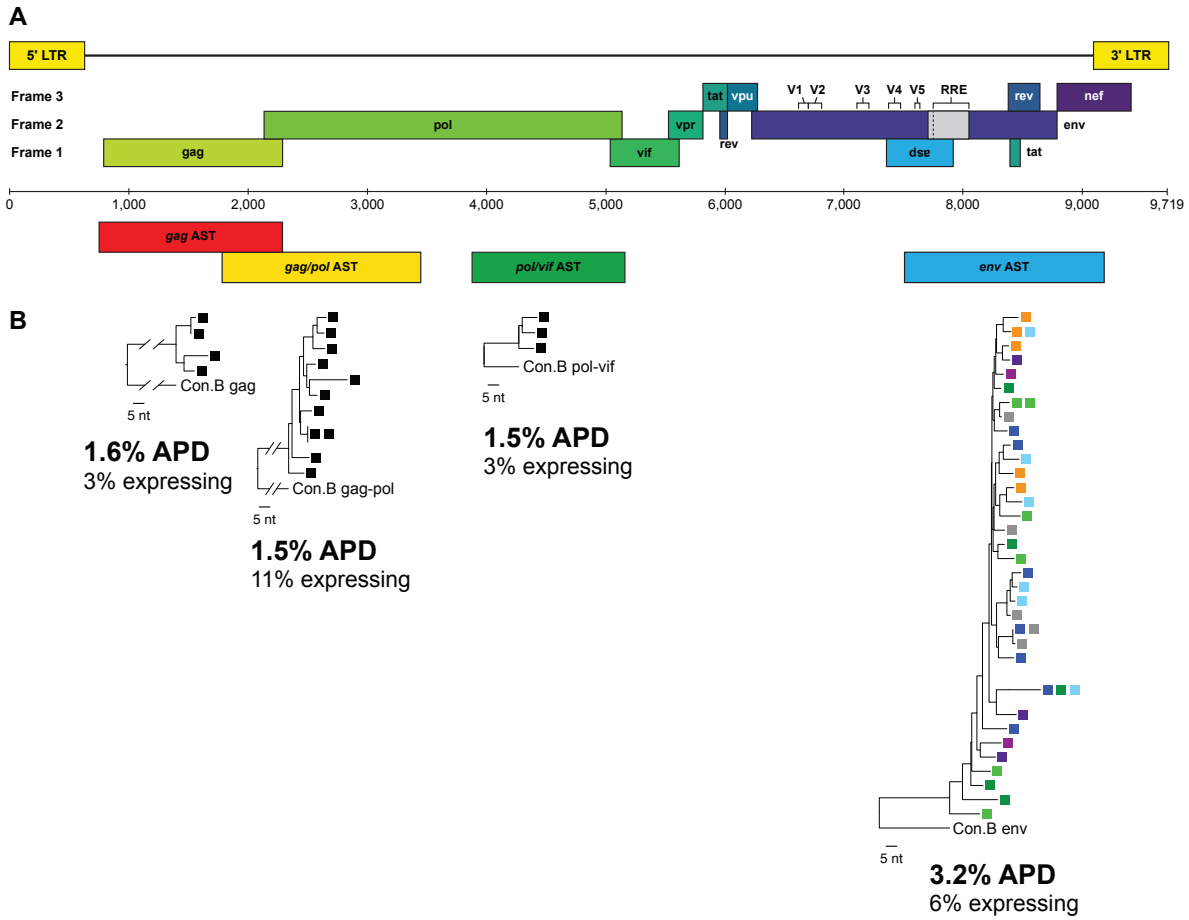


Fig. 3. Detection of antisense transcripts along the proviral genome. (**A**) HIV-1 genome map indicating sub-genomic regions for AST amplification. (**B**) Modified AST single-genome sequencing using exogenous oligo-tagged primers to synthesize cDNA and to PCR amplify antisense transcripts along the genome. RNA from aliquots of 90 infected PBMC were extracted from PID 2669 Timepoint #4 for each sub-genomic region (anti-*gag*, *gag/pol*, *pol/vif*) and denoted by a single colored square. p-distance neighbor-joining phylogenetic reconstruction of positive amplicons were generated and percent APD and fraction of expressing cells are reported. The AST tree for anti-*env* is extracted from **Fig. 2C**, where up to 10 aliquots of PBMC were used. Each color square represents a different aliquot of PBMC from Timepoint #4.

Table 1. Donor demographics.

Group	Participant Identifier (PID)	Cohort Location	Age (years)	Sex[A]	Race/Ethnicity[B]	Minimum duration of infection (years)	HIV plasma RNA copies/mL at sampling[C]	ART regimen at sampling[D]	Duration on ART at sampling
Donors on ART	1079	SCOPE, USCF	64	M	Latino	17.0	<50	FTC/TDF, ETV	12.8 years
	1683	SCOPE, USCF	45	M	White	8.0	<50	FTC/TDF, DRV/r	5.4 years
	2669	SCOPE, USCF	50	M	White	Unknown	<50	ABC/3TC/DTG	4.3 years
			51				<50		5.5 years
			56				112		2 weeks[E]
			56				<50		1 month[E]
Untreated	291	University of Pittsburgh	53	M	Black	9.0	184,243	Not on ART	0[F]
	477	University of Pittsburgh	26	F	Black	5.0	139,845	Not on ART	0[F]
	1508	SCOPE, USCF	53	M	Black	30.0	128	Not on ART	0[F]
	1775	SCOPE, USCF	37	M	White	5.0	1,704	Not on ART	0[F]
	3611	SCOPE, USCF	37	M	Black	4.0	275,685	Not on ART	0[F]

[A] Sex assigned at birth: M (male) and F (female)
[B] Race/Ethnicity reported by donor
[C] Level of plasma viremia was determined by either Abbott® Real-time HIV-1 Assay (SCOPE, UCSF) or COBAS® HIV-1 Test (University of Pittsburgh)
[D] ABC (abacavir), DTG (dolutegravir), DRV (darunavir), ETV (etravirine), FTC (emtricitabine), TDF (tenofovir disoproxil fumarate), /r (ritonavir-boosted), 3TC (lamivudine)
[E] After 5.5 years on ART, the participant had an unexpected ART interruption for approximately 4 weeks. They reinitiated ART with the first timepoint post-ART interruption at 2 weeks with low but detectable HIV-1 plasma viremia. Then 1 month post-ART interruption with plasma viremia suppressed
[F] Donors were not on ART

Table 2. Fraction of infected cells with HIV-1 "long AST" in donors on ART.

Participant Identifier (PID)	Duration on ART at sampling	Estimated number of infected cells assayed	Number of HIV-1 AST sequences obtained	Estimated number of infected cells with HIV-1 AST[A]	Estimate % of infected cells with HIV-1 AST	Average number of HIV-1 AST copies per cell [range][B]
1079	12.8 years	720	43	12	1.7	3.6 [1-30]
1683	5.4 years	1,880	22	20	1.1	1.1 [1-2]
2669	4.3 years	1,440	67	63	4.4	1.1 [1-3]
	5.5 years	1,040	54	52	5.0	1 [1-2]
	2 weeks[#]	2,000	86	74	3.7	1.2 [1-3]
	1 month	720	41	40	5.6	1 [1-2]
	Median	**1,240**	**49**	**46**	**4.1**	**1.1**
	IQR	**720-1,910**	**36-72**	**18-66**	**1.6-5.2**	**1.0-1.7**

"long AST" = 1.7kb from 3′ LTR to *env*

[A] Assumes AST with identical sequences are produced in the same single infected cell

[B] Cells without HIV-1 AST were excluded

[#] Not fully suppressed after 2 weeks on ART following 4-week unplanned ART interruption (ATI). PID 2669 was on ART for 5.5 years prior to an ATI. Samples were collected at 4.3 years and 5.5 years prior to the ATI and at 2 weeks and 1 month after ART re-initiation after the ATI

References

1. O. Khorkova, A. J. Myers, J. Hsiao, C. Wahlestedt, Natural antisense transcripts. *Hum Mol Genet* **23**, R54-63 (2014).
2. R. Li, R. Sklutuis, J. L. Groebner, F. Romerio, HIV-1 Natural Antisense Transcription and Its Role in Viral Persistence. *Viruses* **13**, (2021).
3. K. Kondo, J. Xu, E. S. Mocarski, Human cytomegalovirus latent gene expression in granulocyte-macrophage progenitors in culture and in seropositive individuals. *Proc Natl Acad Sci U S A* **93**, 11137-11142 (1996).
4. J. Zwaagstra, H. Ghiasi, A. B. Nesburn, S. L. Wechsler, In vitro promoter activity associated with the latency-associated transcript gene of herpes simplex virus type 1. *J Gen Virol* **70 (Pt 8)**, 2163-2169 (1989).
5. N. L. Michael, M. T. Vahey, L. d'Arcy, P. K. Ehrenberg, J. D. Mosca, J. Rappaport, R. R. Redfield, Negative-strand RNA transcripts are produced in human immunodeficiency virus type 1-infected cells and patients by a novel promoter downregulated by Tat. *J Virol* **68**, 979-987 (1994).
6. M. H. Rasmussen, B. Ballarín-González, J. Liu, L. B. Lassen, A. Füchtbauer, E. M. Füchtbauer, A. L. Nielsen, F. S. Pedersen, Antisense transcription in gammaretroviruses as a mechanism of insertional activation of host genes. *J Virol* **84**, 3780-3788 (2010).
7. D. Larocca, L. A. Chao, M. H. Seto, T. K. Brunck, Human T-cell leukemia virus minus strand transcription in infected T-cells. *Biochem Biophys Res Commun* **163**, 1006-1013 (1989).
8. C. R. M. Bangham, M. Miura, A. Kulkarni, M. Matsuoka, Regulation of Latency in the Human T Cell Leukemia Virus, HTLV-1. *Annu Rev Virol* **6**, 365-385 (2019).
9. R. H. Miller, Human immunodeficiency virus may encode a novel protein on the genomic DNA plus strand. *Science* **239**, 1420-1422 (1988).
10. K. Bentley, N. Deacon, S. Sonza, S. Zeichner, M. Churchill, Mutational analysis of the HIV-1 LTR as a promoter of negative sense transcription. *Arch Virol* **149**, 2277-2294 (2004).
11. S. Landry, M. Halin, S. Lefort, B. Audet, C. Vaquero, J. M. Mesnard, B. Barbeau, Detection, characterization and regulation of antisense transcripts in HIV-1. *Retrovirology* **4**, 71 (2007).
12. M. Kobayashi-Ishihara, M. Yamagishi, T. Hara, Y. Matsuda, R. Takahashi, A. Miyake, K. Nakano, T. Yamochi, T. Ishida, T. Watanabe, HIV-1-encoded antisense RNA suppresses viral replication for a prolonged period. *Retrovirology* **9**, 38 (2012).
13. Z. Gholizadeh, M. S. Iqbal, R. Li, F. Romerio, The HIV-1 Antisense Gene ASP: The New Kid on the Block. *Vaccines (Basel)* **9**, (2021).
14. G. Ma, J. Yasunaga, M. Matsuoka, Multifaceted functions and roles of HBZ in HTLV-1 pathogenesis. *Retrovirology* **13**, 16 (2016).
15. D. G. Caetano, P. Napoleão-Pêgo, L. M. Villela, F. H. Côrtes, S. W. Cardoso, B. Hoagland, B. Grinsztejn, V. G. Veloso, S. G. De-Simone, M. L. Guimarães, Patterns of Diversity and Humoral Immunogenicity for HIV-1 Antisense Protein (ASP). *Vaccines (Basel)* **12**, (2024).
16. A. Bet, E. A. Maze, A. Bansal, S. Sterrett, A. Gross, S. Graff-Dubois, A. Samri, A. Guihot, C. Katlama, I. Theodorou, J. M. Mesnard, A. Moris, P. A. Goepfert, S. Cardinaud, The HIV-1 antisense protein (ASP) induces CD8 T cell responses during chronic infection. *Retrovirology* **12**, 15 (2015).
17. Y. Affram, J. C. Zapata, Z. Gholizadeh, W. D. Tolbert, W. Zhou, M. D. Iglesias-Ussel, M. Pazgier, K. Ray, O. S. Latinovic, F. Romerio, The HIV-1 Antisense Protein ASP Is a Transmembrane Protein of the Cell Surface and an Integral Protein of the Viral Envelope. *J Virol* **93**, (2019).
18. M. Kobayashi-Ishihara, K. Terahara, J. P. Martinez, M. Yamagishi, R. Iwabuchi, C. Brander, M. Ato, T. Watanabe, A. Meyerhans, Y. Tsunetsugu-Yokota, HIV LTR-Driven Antisense RNA by Itself Has Regulatory Function and May Curtail Virus Reactivation From Latency. *Front Microbiol* **9**, 1066 (2018).
19. J. C. Zapata, F. Campilongo, R. A. Barclay, C. DeMarino, M. D. Iglesias-Ussel, F. Kashanchi, F. Romerio, The Human Immunodeficiency Virus 1 ASP RNA promotes viral latency by recruiting

the Polycomb Repressor Complex 2 and promoting nucleosome assembly. *Virology* **506**, 34-44 (2017).

20. A. Mancarella, F. A. Procopio, T. Achsel, E. D. Crignis, B. T. Foley, G. Corradin, C. Bagni, G. Pantaleo, C. Graziosi, Detection of antisense protein (ASP) RNA transcripts in individuals infected with human immunodeficiency virus type 1 (HIV-1). *Journal of General Virology*, 863–876 (2019).
21. W. R. McManus, M. J. Bale, J. Spindler, A. Wiegand, A. Musick, S. C. Patro, M. D. Sobolewski, V. K. Musick, E. M. Anderson, J. C. Cyktor, E. K. Halvas, W. Shao, D. Wells, X. Wu, B. F. Keele, J. M. Milush, R. Hoh, J. W. Mellors, S. H. Hughes, S. G. Deeks, J. M. Coffin, M. F. Kearney, HIV-1 in lymph nodes is maintained by cellular proliferation during antiretroviral therapy. *J Clin Invest* **129**, 4629-4642 (2019).
22. A. Wiegand, J. Spindler, F. F. Hong, W. Shao, J. C. Cyktor, A. R. Cillo, E. K. Halvas, J. M. Coffin, J. W. Mellors, M. F. Kearney, Single-cell analysis of HIV-1 transcriptional activity reveals expression of proviruses in expanded clones during ART. *Proc Natl Acad Sci U S A* **114**, E3659-e3668 (2017).
23. A. A. Capoferri, A. Wiegand, F. Hong, J. L. Jacobs, J. Spindler, A. Musick, M. J. Bale, W. Shao, M. D. Sobolewski, A. R. Cillo, B. T. Luke, C. M. Fennessey, R. J. Gorelick, R. Hoh, E. K. Halvas, S. G. Deeks, J. M. Coffin, J. W. Mellors, M. F. Kearney, HIV-1 control in vivo is related to the number but not the fraction of infected cells with viral unspliced RNA. *Proc Natl Acad Sci U S A* **121**, e2405210121 (2024).
24. F. Hong, E. Aga, A. R. Cillo, A. L. Yates, G. Besson, E. Fyne, D. L. Koontz, C. Jennings, L. Zheng, J. W. Mellors, Novel Assays for Measurement of Total Cell-Associated HIV-1 DNA and RNA. *J Clin Microbiol* **54**, 902-911 (2016).
25. F. E. Grubbs, Sample Criteria for Testing Outlying Observations. *The Annals of Mathematical Statistics* **21**, 27-58 (1950).
26. P. P. Rose, B. T. Korber, Detecting hypermutations in viral sequences with an emphasis on G --> A hypermutation. *Bioinformatics* **16**, 400-401 (2000).
27. E. Sanders-Buell, M. Salminen, F. McCutchan, Sequencing primers for HIV-1. *Human retroviruses and AIDS*, (1995).
28. P. Swanson, S. G. Devare, J. Hackett, Jr., Molecular characterization of 39 HIV isolates representing group M (subtypes A-G) and group O: sequence analysis of gag p24, pol integrase, and env gp41. *AIDS Res Hum Retroviruses* **19**, 625-629 (2003).
29. G. Ma, J. I. Yasunaga, K. Shimura, K. Takemoto, M. Watanabe, M. Amano, H. Nakata, B. Liu, X. Zuo, M. Matsuoka, Human retroviral antisense mRNAs are retained in the nuclei of infected cells for viral persistence. *Proc Natl Acad Sci U S A* **118**, (2021).
30. S. Saayman, A. Ackley, A. W. Turner, M. Famiglietti, A. Bosque, M. Clemson, V. Planelles, K. V. Morris, An HIV-encoded antisense long noncoding RNA epigenetically regulates viral transcription. *Mol Ther* **22**, 1164-1175 (2014).
31. S. Djebali, C. A. Davis, A. Merkel, A. Dobin, T. Lassmann, A. Mortazavi, A. Tanzer, J. Lagarde, W. Lin, F. Schlesinger, C. Xue, G. K. Marinov, J. Khatun, B. A. Williams, C. Zaleski, J. Rozowsky, M. Röder, F. Kokocinski, R. F. Abdelhamid, T. Alioto, I. Antoshechkin, M. T. Baer, N. S. Bar, P. Batut, K. Bell, I. Bell, S. Chakrabortty, X. Chen, J. Chrast, J. Curado, T. Derrien, J. Drenkow, E. Dumais, J. Dumais, R. Duttagupta, E. Falconnet, M. Fastuca, K. Fejes-Toth, P. Ferreira, S. Foissac, M. J. Fullwood, H. Gao, D. Gonzalez, A. Gordon, H. Gunawardena, C. Howald, S. Jha, R. Johnson, P. Kapranov, B. King, C. Kingswood, O. J. Luo, E. Park, K. Persaud, J. B. Preall, P. Ribeca, B. Risk, D. Robyr, M. Sammeth, L. Schaffer, L. H. See, A. Shahab, J. Skancke, A. M. Suzuki, H. Takahashi, H. Tilgner, D. Trout, N. Walters, H. Wang, J. Wrobel, Y. Yu, X. Ruan, Y. Hayashizaki, J. Harrow, M. Gerstein, T. Hubbard, A. Reymond, S. E. Antonarakis, G. Hannon, M. C. Giddings, Y. Ruan, B. Wold, P. Carninci, R. Guigó, T. R. Gingeras, Landscape of transcription in human cells. *Nature* **489**, 101-108 (2012).

32. A. Battle, C. D. Brown, B. E. Engelhardt, S. B. Montgomery, Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213 (2017).

33. A. R. R. Forrest, H. Kawaji, M. Rehli, J. Kenneth Baillie, M. J. L. de Hoon, V. Haberle, T. Lassmann, I. V. Kulakovskiy, M. Lizio, M. Itoh, R. Andersson, C. J. Mungall, T. F. Meehan, S. Schmeier, N. Bertin, M. Jørgensen, E. Dimont, E. Arner, C. Schmidl, U. Schaefer, Y. A. Medvedeva, C. Plessy, M. Vitezic, J. Severin, C. A. Semple, Y. Ishizu, R. S. Young, M. Francescatto, I. Alam, D. Albanese, G. M. Altschuler, T. Arakawa, J. A. C. Archer, P. Arner, M. Babina, S. Rennie, P. J. Balwierz, A. G. Beckhouse, S. Pradhan-Bhatt, J. A. Blake, A. Blumenthal, B. Bodega, A. Bonetti, J. Briggs, F. Brombacher, A. Maxwell Burroughs, A. Califano, C. V. Cannistraci, D. Carbajo, Y. Chen, M. Chierici, Y. Ciani, H. C. Clevers, E. Dalla, C. A. Davis, M. Detmar, A. D. Diehl, T. Dohi, F. Drabløs, A. S. B. Edge, M. Edinger, K. Ekwall, M. Endoh, H. Enomoto, M. Fagiolini, L. Fairbairn, H. Fang, M. C. Farach-Carson, G. J. Faulkner, A. V. Favorov, M. E. Fisher, M. C. Frith, R. Fujita, S. Fukuda, C. Furlanello, M. Furuno, J.-i. Furusawa, T. B. Geijtenbeek, A. P. Gibson, T. Gingeras, D. Goldowitz, J. Gough, S. Guhl, R. Guler, S. Gustincich, T. J. Ha, M. Hamaguchi, M. Hara, M. Harbers, J. Harshbarger, A. Hasegawa, Y. Hasegawa, T. Hashimoto, M. Herlyn, K. J. Hitchens, S. J. Ho Sui, O. M. Hofmann, I. Hoof, F. Hori, L. Huminiecki, K. Iida, T. Ikawa, B. R. Jankovic, H. Jia, A. Joshi, G. Jurman, B. Kaczkowski, C. Kai, K. Kaida, A. Kaiho, K. Kajiyama, M. Kanamori-Katayama, A. S. Kasianov, T. Kasukawa, S. Katayama, S. Kato, S. Kawaguchi, H. Kawamoto, Y. I. Kawamura, T. Kawashima, J. S. Kempfle, T. J. Kenna, J. Kere, L. M. Khachigian, T. Kitamura, S. Peter Klinken, A. J. Knox, M. Kojima, S. Kojima, N. Kondo, H. Koseki, S. Koyasu, S. Krampitz, A. Kubosaki, A. T. Kwon, J. F. J. Laros, W. Lee, A. Lennartsson, K. Li, B. Lilje, L. Lipovich, A. Mackay-sim, R.-i. Manabe, J. C. Mar, B. Marchand, A. Mathelier, N. Mejhert, A. Meynert, Y. Mizuno, D. A. de Lima Morais, H. Morikawa, M. Morimoto, K. Moro, E. Motakis, H. Motohashi, C. L. Mummery, M. Murata, S. Nagao-Sato, Y. Nakachi, F. Nakahara, T. Nakamura, Y. Nakamura, K. Nakazato, E. van Nimwegen, N. Ninomiya, H. Nishiyori, S. Noma, T. Nozaki, S. Ogishima, N. Ohkura, H. Ohmiya, H. Ohno, M. Ohshima, M. Okada-Hatakeyama, Y. Okazaki, V. Orlando, D. A. Ovchinnikov, A. Pain, R. Passier, M. Patrikakis, H. Persson, S. Piazza, J. G. D. Prendergast, O. J. L. Rackham, J. A. Ramilowski, M. Rashid, T. Ravasi, P. Rizzu, M. Roncador, S. Roy, M. B. Rye, E. Saijyo, A. Sajantila, A. Saka, S. Sakaguchi, M. Sakai, H. Sato, H. Satoh, S. Savvi, A. Saxena, C. Schneider, E. A. Schultes, G. G. Schulze-Tanzil, A. Schwegmann, T. Sengstag, G. Sheng, H. Shimoji, Y. Shimoni, J. W. Shin, C. Simon, D. Sugiyama, T. Sugiyama, M. Suzuki, N. Suzuki, R. K. Swoboda, P. A. C. ’t Hoen, M. Tagami, N. Takahashi, J. Takai, H. Tanaka, H. Tatsukawa, Z. Tatum, M. Thompson, H. Toyoda, T. Toyoda, E. Valen, M. van de Wetering, L. M. van den Berg, R. Verardo, D. Vijayan, I. E. Vorontsov, W. W. Wasserman, S. Watanabe, C. A. Wells, L. N. Winteringham, E. Wolvetang, E. J. Wood, Y. Yamaguchi, M. Yamamoto, M. Yoneda, Y. Yonekura, S. Yoshida, S. E. Zabierowski, P. G. Zhang, X. Zhao, S. Zucchelli, K. M. Summers, H. Suzuki, C. O. Daub, J. Kawai, P. Heutink, W. Hide, T. C. Freeman, B. Lenhard, V. B. Bajic, M. S. Taylor, V. J. Makeev, A. Sandelin, D. A. Hume, P. Carninci, Y. Hayashizaki, F. C. The, R. P. the, Clst, A promoter-level mammalian expression atlas. *Nature* **507**, 462-470 (2014).

34. J. A. Ramilowski, C. W. Yip, S. Agrawal, J. C. Chang, Y. Ciani, I. V. Kulakovskiy, M. Mendez, J. L. C. Ooi, J. F. Ouyang, N. Parkinson, A. Petri, L. Roos, J. Severin, K. Yasuzawa, I. Abugessaisa, A. Akalin, I. V. Antonov, E. Arner, A. Bonetti, H. Bono, B. Borsari, F. Brombacher, C. J. Cameron, C. V. Cannistraci, R. Cardenas, M. Cardon, H. Chang, J. Dostie, L. Ducoli, A. Favorov, A. Fort, D. Garrido, N. Gil, J. Gimenez, R. Guler, L. Handoko, J. Harshbarger, A. Hasegawa, Y. Hasegawa, K. Hashimoto, N. Hayatsu, P. Heutink, T. Hirose, E. L. Imada, M. Itoh, B. Kaczkowski, A. Kanhere, E. Kawabata, H. Kawaji, T. Kawashima, S. T. Kelly, M. Kojima, N. Kondo, H. Koseki, T. Kouno, A. Kratz, M. Kurowska-Stolarska, A. T. J. Kwon, J. Leek, A. Lennartsson, M. Lizio, F. López-Redondo, J. Luginbühl, S. Maeda, V. J. Makeev, L. Marchionni, Y. A. Medvedeva, A. Minoda, F. Müller, M. Muñoz-Aguirre, M.

Murata, H. Nishiyori, K. R. Nitta, S. Noguchi, Y. Noro, R. Nurtdinov, Y. Okazaki, V. Orlando, D. Paquette, C. J. C. Parr, O. J. L. Rackham, P. Rizzu, D. F. Sánchez Martinez, A. Sandelin, P. Sanjana, C. A. M. Semple, Y. Shibayama, D. M. Sivaraman, T. Suzuki, S. C. Szumowski, M. Tagami, M. S. Taylor, C. Terao, M. Thodberg, S. Thongjuea, V. Tripathi, I. Ulitsky, R. Verardo, I. E. Vorontsov, C. Yamamoto, R. S. Young, J. K. Baillie, A. R. R. Forrest, R. Guigó, M. M. Hoffman, C. C. Hon, T. Kasukawa, S. Kauppinen, J. Kere, B. Lenhard, C. Schneider, H. Suzuki, K. Yagi, M. J. L. de Hoon, J. W. Shin, P. Carninci, Functional annotation of human long noncoding RNAs via molecular phenotyping. *Genome Res* **30**, 1060-1072 (2020).

35. C.-C. Hon, J. A. Ramilowski, J. Harshbarger, N. Bertin, O. J. L. Rackham, J. Gough, E. Denisenko, S. Schmeier, T. M. Poulsen, J. Severin, M. Lizio, H. Kawaji, T. Kasukawa, M. Itoh, A. M. Burroughs, S. Noma, S. Djebali, T. Alam, Y. A. Medvedeva, A. C. Testa, L. Lipovich, C.-W. Yip, I. Abugessaisa, M. Mendez, A. Hasegawa, D. Tang, T. Lassmann, P. Heutink, M. Babina, C. A. Wells, S. Kojima, Y. Nakamura, H. Suzuki, C. O. Daub, M. J. L. de Hoon, E. Arner, Y. Hayashizaki, P. Carninci, A. R. R. Forrest, An atlas of human long non-coding RNAs with accurate 5′ ends. *Nature* **543**, 199-204 (2017).

36. F. Seifuddin, K. Singh, A. Suresh, J. T. Judy, Y.-C. Chen, V. Chaitankar, I. Tunc, X. Ruan, P. Li, Y. Chen, H. Cao, R. S. Lee, F. S. Goes, P. P. Zandi, M. S. Jafri, M. Pirooznia, lncRNAKB, a knowledgebase of tissue-specific functional annotation and trait association of long noncoding RNA. *Scientific Data* **7**, 326 (2020).

37. L. Josefsson, S. von Stockenstrom, N. R. Faria, E. Sinclair, P. Bacchetti, M. Killian, L. Epling, A. Tan, T. Ho, P. Lemey, W. Shao, P. W. Hunt, M. Somsouk, W. Wylie, D. C. Douek, L. Loeb, J. Custer, R. Hoh, L. Poole, S. G. Deeks, F. Hecht, S. Palmer, The HIV-1 reservoir in eight patients on long-term suppressive antiretroviral therapy is stable with few genetic changes over time. *Proc Natl Acad Sci U S A* **110**, E4987-4996 (2013).

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

39. K. M. Bruner, Z. Wang, F. R. Simonetti, A. M. Bender, K. J. Kwon, S. Sengupta, E. J. Fray, S. A. Beg, A. A. R. Antar, K. M. Jenike, L. N. Bertagnolli, A. A. Capoferri, J. T. Kufera, A. Timmons, C. Nobles, J. Gregg, N. Wada, Y. C. Ho, H. Zhang, J. B. Margolick, J. N. Blankson, S. G. Deeks, F. D. Bushman, J. D. Siliciano, G. M. Laird, R. F. Siliciano, A quantitative approach for measuring the reservoir of latent HIV-1 proviruses. *Nature* **566**, 120-125 (2019).

40. L. M. Mansky, H. M. Temin, Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J Virol* **69**, 5087-5094 (1995).

Supplementary Materials for

In vivo detection of HIV antisense transcripts in untreated and ART-treated individuals

Adam A. Capoferri[1]*†, Rachel Sklutuis[1]†, Toluleke O. Famuyiwa[1], Sachi Pathak[1], Rui Li[2], Jason W. Rausch[1], Brian T. Luke[3], Rebecca Hoh[4], Steven G. Deeks[4], John W. Mellors[5], John M. Coffin[6], Jennifer L. Groebner[1], Fabio Romerio[2], Mary F. Kearney[1*]

Corresponding author: kearneym@mail.nih.gov and adam.capoferri@nih.gov

The PDF file includes:

Materials and Methods
Figs. S1 to S4
Tables S1 to S6
References 37-40

Materials and Methods

Participant cohorts, sample collection, and study approval

PBMC were collected from PWH on ART who were enrolled at the University of California of San Francisco in the SCOPE trial (clinical trial # NCT00187512). PBMC were also collected from PWH who were not currently on ART and were enrolled at the University of Pittsburgh in the Optimization of Immunologic and Virologic Assays for HIV Trial (IRB# STUDY20040215). The studies were approved by the University of California San Francisco Institutional Review Board and the University of Pittsburgh Institutional Review Board. All donors provided written informed consent for HIV-1 quantification and sequencing. PBMC were separated using Ficoll, resuspended in FBS with 10% DMSO, and stored in liquid nitrogen (LN_2) until testing.

Overview of HIV-1 antisense transcript (AST) quantification and sequencing approaches

HIV-1 AST quantification and sequencing assays were performed by modifying our cell-associated RNA and DNA single-genome sequencing assay (CARD-SGS) (*22*, *23*) (detailed methods below). Briefly, total nucleic acid was extracted from serially diluted aliquots of PBMC collected from PWH, the DNA was digested, and AST cDNA synthesized using a gene-specific primer in the opposite orientation of the *env* coding region (**Fig. S1B**). The cDNA from each aliquot was spread across 96-well PCR plates for amplification (~50bp) and probe detection to determine the dilution that yielded <30% positive PCR products, indicating that cDNA was primarily amplified from single molecules. By determining the endpoint dilution, we could quantify the number of AST in each aliquot and determine the fraction of the infected cells with AST by using the number of HIV-1 DNA molecules in replicate aliquots for the denominator. HIV-1 DNA was measured with the integrase cell-associated DNA (iCAD) assay (*24*). The number of HIV-1 DNA molecules is a surrogate for the number of infected cells since previous studies showed that most infected cells carry only a single provirus both before and during ART (*37*). For a subset of the samples, a 1.7-kb fragment of AST was also amplified at an endpoint and Sanger sequenced to assess the genetic diversity of the transcripts.

Overview of the development and optimization of AST quantification and sequencing

During the development and optimization of the HIV-1 AST quantification and sequencing assays, we controlled for endogenous priming during cDNA synthesis and contamination of HIV-1 DNA. We also determined the sensitivity and background of the AST assays. An overview of these developmental steps is provided here, but detailed protocols for each set of experiments are provided below.

1. To eliminate endogenous self-priming during AST cDNA synthesis (*19*, *38*), we used a 5′ end exogenous oligo-tagged gene specific primer (**Table S1**). The tag generates cDNA carrying the

exogenous tag sequence to function as the target for the forward primer during first-round PCR amplification; thus, allowing for specific amplification of cDNA molecules containing the exogenous oligo-tag sequence, rather than endogenous self-primed templates.

2. To ensure complete DNA digestion, we used various amounts of the ACH-2 infected cell line (~1 provirus/cell), spiked into $1x10^5$ or $1x10^6$ uninfected CEM cells (**Fig. S2**). Nucleic acid was extracted, DNase treated, and "cDNA synthesized" without reverse transcriptase. Nested PCR was performed to determine the maximum number of cells that can be assayed to ensure that HIV DNA digestion was complete. We found complete digestion of viral DNA at 100 infected cells per aliquot, in agreement with prior findings with the CARD-SGS assay where <300 infected cells per aliquot was the upper limit (*22*).
3. To optimize the AST assay, we used *in vitro* transcribed AST generated with a pMiniT-AST vector containing a T7 promoter (**Fig. S3**). The concentration of AST was measured by spectrophotometry and 10^4 copies were spiked into nucleic acid extracted from $1x10^5$ uninfected CEM cells. The number of positives detected in the assay was compared to the known number of transcripts. The optimized AST detection assay had a sensitivity of ~50% (average of 5,589 copies detected), which could be a consequence of including the exogenous oligo-tag on the cDNA primer reducing cDNA synthesis efficiency or loss of RNA template from the purification steps of the assay.
4. After optimizing the assay using *in vitro* transcribed AST, we tested the assay on 100 ACH-2 cells spiked into 10^5 uninfected CEM cells (**Fig. S4**). Using the AST assay (43bp amplicon) at endpoint with probe detection, we found a median of 41 AST molecules per 100 infected ACH-2 cells [IQR 25-69] (n=16 replicates). We detected 13 AST /100 infected ACH-2 cells when amplifying and sequencing a 1.7-kb fragment of anti-*env*. The less frequent detection of the 1.7-kb amplicon is not unexpected compared to the 43bp amplicon.
5. To determine the cut-off for the digital (probe detection at an endpoint) AST assay, we performed replicates of "no reverse-transcriptase (RT)" controls on 100 ACH-2 cells spiked into 10^5 uninfected CEM cells. We found 4 potential false positive signals in 650 infected cells assayed when RT was not included in the cDNA reaction, making our assay cut-off 0.6 AST/100 infected cells. We also tested $5x10^5$ uninfected CEM cells spread across 182 PCR wells and detected 1 false positive, making the false-positive assay background <0.0002/100 cells assayed. To overcome the background, we limited each digital assay to <100 infected cells and included equal numbers of "no RT" controls on each PCR plate.

Cell lines, maintenance, and storage

Cell lines used for assay development were uninfected CEM/C1 cells (American Type Culture Collection, #CRL-2265) and HIV-1 infected ACH-2 cells (HIV Reagent Program, #ARP-349). ACH-2 cells are A3.01 cells infected with a single integrated provirus (HIV-1_{LAI}) with a defect in TAR (C37T). Complete Medium was prepared as RPMI 1640 supplemented with 1% Penicillin-Streptomycin-Glutamine (Gibco, #10378016) and 10% heat-inactivated fetal bovine serum (HI-FBS). Cell Freeze Medium was prepared as RPMI 1640 supplemented with 10% HI-FBS (CEM/C1 cells) or 10-40% HI-FBS (ACH-2 cells), and 10% DMSO.

Cells were maintained in Complete Medium (as described above) and seeded in 2x vented T-25 flasks (1:3 and 2:3 from vial contents) with a total of 6mL in each flask. Cells were incubated overnight at 37°C, 5% CO_2 with flasks upright. The following day, the medium was aspirated without disturbing the cells and were replenished with warmed Complete Medium. The cells were subcultured and transferred to a 15mL conical tube for centrifugation at 150x*g* for 10 minutes, then media was aspirated. The cells were resuspended in 1mL Complete Medium and counted by hemocytometer at 1:10 and 1:100 dilutions. Viable cells were seeded at 2-4x10^5 in new T-25 flasks with 6mL of warmed Complete Medium. If needed, cells were expanded in T-75 flasks (seeding at 4-6x10^5) with medium volume between 10-20mL. The cells were split every 2-3 days as needed.

The cells were prepared in Cell Freeze Medium at a minimum of 1x10^6 cells/mL by gentle pipetting of cells and aliquoted into labeled cryotubes. They were placed in a freezing container with 100% isopropanol in outer container to freeze cells at -80°C overnight. The following day, the cells were transferred to LN_2 storage.

Sample Preparation

To thaw viably frozen cells, RPMI 1640 was warmed to 37°C and added dropwise to thaw PBMC for nucleic acid extraction. Each viably frozen cell vial was warmed for ~2 minutes at 37°C prior to RPMI being added dropwise to the vial. Thawed donor PBMC were aliquoted and centrifuged at 500×*g* for 5 minutes. Supernatant was removed and the pelleted cells were used immediately for downstream assays or were frozen on dry ice and stored back in LN_2 until ready for use.

Generation of control RNA

RNA controls were generated to determine the efficiency of the HIV-1 AST cDNA synthesis and PCR amplification. AST (*anti-env*) were amplified from a pUC57-AST plasmid obtained from Fabio Romerio (Department of Molecular and Comparative Pathobiology, The Johns Hopkins University School of Medicine, Baltimore, MD) using M13 Fwd and M13 Rev primers (**Table S2**). Using Platinum II Taq (ThermoFisher, #14966025), 50µL PCR reactions were carried out with 50ng of plasmid DNA and 10µL

of 5X Platinum II PCR Buffer, 1µL of 10mM dNTPs, 1µL of each 10µM primer, 0.4µL of Platinum II Taq polymerase, and molecular-grade water to bring to 50µL. PCR cycling was performed as follows: 94°C for 2 minutes and 45 cycles of 94°C for 15 seconds, 60°C for 15 seconds, and 68°C for 45 seconds, followed by a final extension at 68°C for 1 minute. The PCR product was then purified using a QIAquick PCR Purification Kit (QIAGEN, #28104). The PCR product was cloned using NEB® PCR Cloning Kit (New England Biolabs, #E1202S) where 80ng of the purified PCR product was ligated into a linearized pMiniT 2.0 vector and transformed into the NEB® stable competent 10-beta *E. coli* following manufacturer's instructions (New England Biolabs, #C3019H). Transformed colonies were picked, grown overnight according to manufacturer's instructions, and miniprepped using the QuickLyse Miniprep Kit (QIAGEN, #27406). Plasmids were screened for the insert by selective PCR screening using Cloning Analysis Fwd and Rev primers (**Table S2**) with PCR cycling performed as follows: 94°C for 2 minutes and 45 cycles of 94°C for 15 seconds, 60°C for 15 seconds, and 68°C for 45 seconds, followed by a final extension at 68°C for 1 minute. The plasmid was then Sanger sequenced to confirm orientation and sequence (**Table S2**).

To generate control transcripts, 1µg of the AST clone (mentioned above) was digested by restriction enzyme PmeI (New England Biolabs, #R0560S) for AST (*anti-env*) or ZraI (New England Biolabs, #R0659S) for sense *env* transcripts following manufacturer's instructions. Digested DNA was then purified using the QIAquick PCR Purification Kit (QIAGEN, #28104) and concentrated using ethanol precipitation (described below). RNA was transcribed using HiScribe T7 Quick High Yield RNA Synthesis Kit (New England Biolabs, #E2050S) for AST or the HiScribe SP6 Quick High Yield RNA Synthesis Kit (New England Biolabs, #E2070S) for sense *env* transcripts. Transcribed RNA was purified using a Monarch RNA Cleanup Kit (New England Biolabs, #T2030) and quantified by Nanodrop spectrophotometer. The RNA was diluted to $1x10^6$ copies/mL with 5mM Tris-HCl pH 8.0 and stored at -80°C for temporary storage and LN_2 for longer storage to help avoid degradation of RNA.

To assess the completeness of digestion of plasmid DNA, approximately 10^4 copies of the transcripts were spiked into extracted nucleic acid from 10^5 CEM cells and DNase I treated (described below). The control transcripts were used for cDNA synthesis using random hexamer primers (ThermoFisher, #SO142) in the following reaction: 2.5µL 10mM dNTPs and 2.5µL 2µM random hexamer primer were added to each 20µL RNA sample. The RNA was denatured at 65°C for 10 minutes, then immediately cooled at -20°C for 1 minute. Next, 25µL of SuperScript III Reverse Transcriptase (Invitrogen, #18080-044) master mix was added to the denatured RNA. Master mix was made by combining 10µL 5X first strand buffer, 1µL 0.1M DTT, 12µL RNase-free water, 1µL 40U/µL RNaseOUT recombinant ribonuclease inhibitor (Invitrogen, #10777-019), and 1µL 200U/µL SuperScript III Reverse Transcriptase. cDNA was synthesized at 25°C for 5 minutes, 55°C for 1 hour, 70°C for 15 minutes, then cooled to 4°C. To remove hybridized RNA, 1µL RNase H (5 units) (New England Biolabs, #M0297S) was added and

incubated at 37°C for 20 minutes followed by an enzyme heat-inactivation at 75°C for 10 minutes, and the cDNA was cooled to 4°C and used immediately or stored at -80°C.

DNA and RNA Extraction

Total nucleic acid was extracted from endpoint diluted aliquots of HIV-infected cells with AST RNA, determined by performing serial dilutions of PBMC until 96-well PCR plates yielded <30% AST positives. Nucleic acid was extracted by adding 100µL 3M guanidine HCl solution (18.75mL 8M guanidinium HCl, 2.5mL 1M Tris-HCl pH 8.0, 0.5mL 100mM $CaCl_2$) and 5µL 20mg/mL Proteinase K to a cell pellet, vortexed, and incubated at 42°C heat block for 1 hour. Addition of 400µL 6M guanidine isothiocyanate solution (25mL 6M guanidinium isothiocyanate, 14mL 1M Tris-HCl pH 8.0, 53µL EDTA pH 8.0) and 6µL 20mg/mL glycogen to lysed cell pellet was vortexed, and incubated at 42°C heat block for 10 minutes. Five hundred microliters of 100% isopropanol was added with mixing, followed by centrifugation at 21,000x*g* for 10 minutes at room temperature to precipitate nucleic acids. Supernatant was removed without disturbing the pellet and washed with 750µL 70% ethanol. Precipitated nucleic acid was air-dried. For DNA, the pellet was resuspended in 100µL 5mM Tris-HCl pH 8.0 and ready for use. For RNA, the DNA was digested by either (i) DNase I (Roche, #04716728001) or (ii) ezDNase (ThermoFisher, #11766051):

(i) DNase I: The air-dried extracted nucleic acid was resuspended in 38µL DNase buffer and 2µL of 10units/µL DNase I and incubated for 20 minutes in a 37°C water bath. After incubation, 200µL of 6M GuSCN (Sigma, #50983) was added and mixed well, followed by the addition of 250µL of 100% isopropanol. Sample was vortexed for 10 seconds, then centrifuged at 21,000×*g* for 10 minutes to pellet nucleic acid. Supernatant was removed and pellet was washed with 1mL 70% ethanol, followed by additional centrifugation at 21,000×*g* for 10 minutes. Finally, the DNA-digested RNA pellet was air-dried, resuspended in 20µL of 5 mM Tris-HCl pH 8.0, and used immediately for cDNA synthesis. Alternatively, RNA could be temporarily stored in 70% ethanol at -80°C.

(ii) ezDNase: The air-dried extracted nucleic acid was resuspended in 32µL 5mM Tris-HCl pH 8.0. Sample was split in equal volumes to control for successful DNA digestion. To ensure successful DNA digestion, manufacturer's instructions were followed, with volumes of reagents doubled and maximum incubations used. Briefly, 2µL 10X ezDNase Buffer and 2µL ezDNase were added to each sample and incubated at 37°C for 5 minutes. To inactivate the ezDNase, the mixture was incubated at 55°C for 5 minutes in the presence of 2µL 0.1M DTT. Incubations were performed in a thermocycler. Samples were immediately cooled at -20°C and brought up to room temperature before directly moving into cDNA synthesis.

cDNA Synthesis

The cDNA was synthesized by adding 2.5µL 10mM dNTPs and 2.5µL 2µM gene-specific tagged cDNA primer (AST cDNA, **Table S1**) to each sample. The RNA was denatured at 85°C for 10 minutes, then immediately cooled at -20°C for 1 minute. Next, 25µL SuperScript III Reverse Transcriptase (Invitrogen, #18080-044) master mix was added to the denatured RNA. Master mix was 10µL 5X first strand buffer, 0.5µL 0.1M DTT, 13.5µL RNase-free water, 0.5µL 40 U/µL RNaseOUT recombinant ribonuclease inhibitor (Invitrogen, #10777-019), and 1µL 200 U/µL SuperScript III Reverse Transcriptase. cDNA was synthesized at 45°C for 1 hour, then cooled to 4°C. To remove hybridized RNA, 1µL RNase H (5 units) (New England Biolabs, #M0297S) was added and incubated at 37°C for 20 minutes followed by an enzyme heat-inactivation at 75°C for 10 minutes. The cDNA was cooled to 4°C and was transferred to a low bind Eppendorf tube and 0.1 vol 3M Sodium Acetate pH 5.5 (Invitrogen, #AM9740) and 1µL 20mg/mL glycogen (Roche, #1090139300) were added. After mixing, 3 volumes 95% ethanol was added, and the entire mixture was vortexed for 10 seconds and incubated overnight at -20°C. The next day, the precipitated cDNA was centrifuged at 21,000x*g* for 20 minutes at room temperature, and supernatant was removed. The pellet was washed with 600µL 70% ethanol, pulse vortexed, and centrifuged for 15 minutes at 21,000x*g* at room temperature. Supernatant was removed and cDNA pellet was air dried and then resuspended in 20µL or 50µL (for digital PCR quantitation) in 5mM Tris-HCl pH 8.0. The cDNA was then used immediately or stored at -80°C.

Quantification of HIV-1 infected cells

The number of HIV infected cells was estimated from the levels of HIV DNA in PBMC. For donors on ART, HIV-1 DNA levels were previously measured using the integrase cell-associated single-copy DNA (iCAD) assay (primers in **Table S3**) (*21*, *24*). We also measured HIV-1 DNA levels using a modified primer probe set in the Rev Response Element (RRE) adapted from the Intact Proviral DNA Assay (*39*) (**Table S4**). Digital PCR was performed using Lightcycler 480 Probes Master hot-start reaction mix (Roche, #4707494001) in 20µL volumes per reaction. A 2mL master mix was prepared by combining 1mL 2X Lightcycler 480 Probes Master hot-start reaction mix, 12µL 100µM RRE Fwd primer, 12µL 100µM RRE Rev, 2µL 100µM RRE probe, 774µL RNase-free water, and 200µL diluted DNA (DNA was diluted up to 200µL with 5mM Tris-HCl pH 8.0). Digital PCR was performed on the Roche LightCycler480 at a denaturation temperature of 95ºC for 10 minutes and 55 cycles (95ºC for 15 seconds and 60ºC for 1 minute). Equal numbers of "no RT" controls and experimental wells were included on each plate to ensure complete HIV-1 DNA digestion. Two negative (no template, 5mM Tris-HCl pH 8.0) and two positive (100 ACH-2 in a background $1x10^5$ CEM cells of DNA) controls were also included on each PCR plate. The number of

positive wells was used to calculate the number of infected cells per million PBMC in each of the donor samples.

Design of gene-specific primers and probes

Due to intra- and inter-host HIV-1 sequence variation, we extracted HIV-1 DNA from PBMC (as described above) and performed PCR at endpoint targeting *env* (primers in **Table S1**) and performed Sanger sequencing (primers in **Table S4**). The designed primers for were used accordingly (**Table S5**, donor-specific mutations are shown in bold red).

AST quantification by digital PCR

AST digital PCR was performed using Lightcycler 480 Probes Master hot-start reaction mix (Roche, #4707494001) in 20µL volumes per reaction. A master mix was prepared by combining 2X Lightcycler 480 Probes Master hot-start reaction mix, 100µM Fwd primer, 100µM Rev primer, 100µM AST probe, RNase-free water, and cDNA synthesized as described above. Digital PCR was performed on the Roche LightCycler480 at a denaturation temperature of 95ºC for 10 minutes and 55 cycles (95ºC for 15 seconds and 60ºC for 1 minute). Equal numbers of "no RT" controls and experimental wells were included on each plate to ensure complete HIV-1 DNA digestion. Two negative (no template, 5mM Tris-HCl pH 8.0) and two positive (100 ACH-2 in a background $1x10^5$ CEM cells of DNA) controls were also included on each PCR plate.

AST quantification and sequencing

AST quantification and sequencing was modified from the original CARD-SGS assay protocol as previously described (*22, 23*) and was used to sequence and quantify HIV-1 DNA and sense and antisense HIV RNA in *gag, pol,* and *env*. Briefly, cDNA or DNA was diluted to a near endpoint in a volume of 200µL and added to 800µL Platinum II Taq (ThermoFisher Scientific, #14966025) master mix containing 200µL 5X Platinum II PCR Buffer, 20µL 10mM dNTPs, 8µL primers (4µL 50µM Forward and 4µL 50µM Reverse primers, for respective region, **Table S1**), 564µL molecular-grade water, and 8µL Platinum II Taq polymerase. The total volume of cDNA and master mix was spread across a 96-well plate (10µL per well). PCR cycling was performed with respect to sub-genomic region (**Table S1**). Following the first round of PCR, each well was diluted with 50µL 5mM Tris-HCl pH 8.0. For nested PCR, 2µL diluted PCR 1 product was transferred from each well to a 96-well plate containing 8µL Platinum II Taq master mix (described above). Nested PCR was performed with respect to sub-genomic region with respective thermocycling conditions (**Table S1**). Nested PCR product was diluted with 50µL 5mM Tris-HCl pH 8.0. Positive wells

were identified by GelRed detection (Biotium, #41003) using UV imaging. Positive PCR reactions were sent for Sanger sequencing (**Table S4**).

The levels of HIV expression in single cells were determined by the number of identical sequences in each aliquot. Because the reverse transcription step is known to introduce errors at a rate of about 10^{-4} per sequenced nucleotide of viral cDNA sequences (*40*), RT-PCR variants that differed by a single nucleotide ('fuzz') from a group of 7 or more identical sequences within the same aliquot were, conservatively, considered to belong to the same rake of identical sequences. The R script for "defuzzing" is available at https://github.com/michaelbale/RStuff/blob/master/fPECS_fncs.R.

Sequence Analysis

Sequences were aligned with MAFFT v7.450 (FFT-NS-1 200PAM/k=2 algorithm) in Geneious Prime ® 2020.2.4. Minor adjustments were performed manually. P-distance neighbor joining trees were reconstructed using MEGA X and rooted on the consensus B HIV sequence. Population genetic diversity was calculated as average pairwise distance (APD) using MEGA 11. APOBEC3-F/G mediated hypermutation was predicted using the Los Alamos National Laboratory HIV Sequence database tool, Hypermut (https://www.hiv.lanl.gov/content/sequence/HYPERMUT/hypermut.html).

Data availability:

All sequence data are available on GenBank (in process of submission). Sequences from the donors on ART were previously published by McManus and colleagues and can be found in (*21*).

Statistical analyses

Mann-Whitney U-test, Kruskal-Wallis test, Spearman's correlation, Grubbs' Outlier test, and Shapiro-Wilk test for normality for informed Paired T-test were performed using Prism GraphPad 9.2. Spearman's correlation was selected due to the random variation in biological systems.

To test if there was a difference in the levels of AST pre- and post-ART interruption we used a binomial test. From the two sample donations pre-interruption, a total of 155 infected PBMC were assayed and 39 infected PBMC were found to contain AST (25%). Post-interruption, 133 infected PBMC were assayed and 65 infected PBMC contained AST (49%). A Binomial distribution was used to show that 65 out of 133 is significantly larger than 39 out of 155.

Statistical tests applied are indicated in the text and/or figures/table legends. Additional statistical analyses were performed using R (version 4.3.1).

Fig. S1

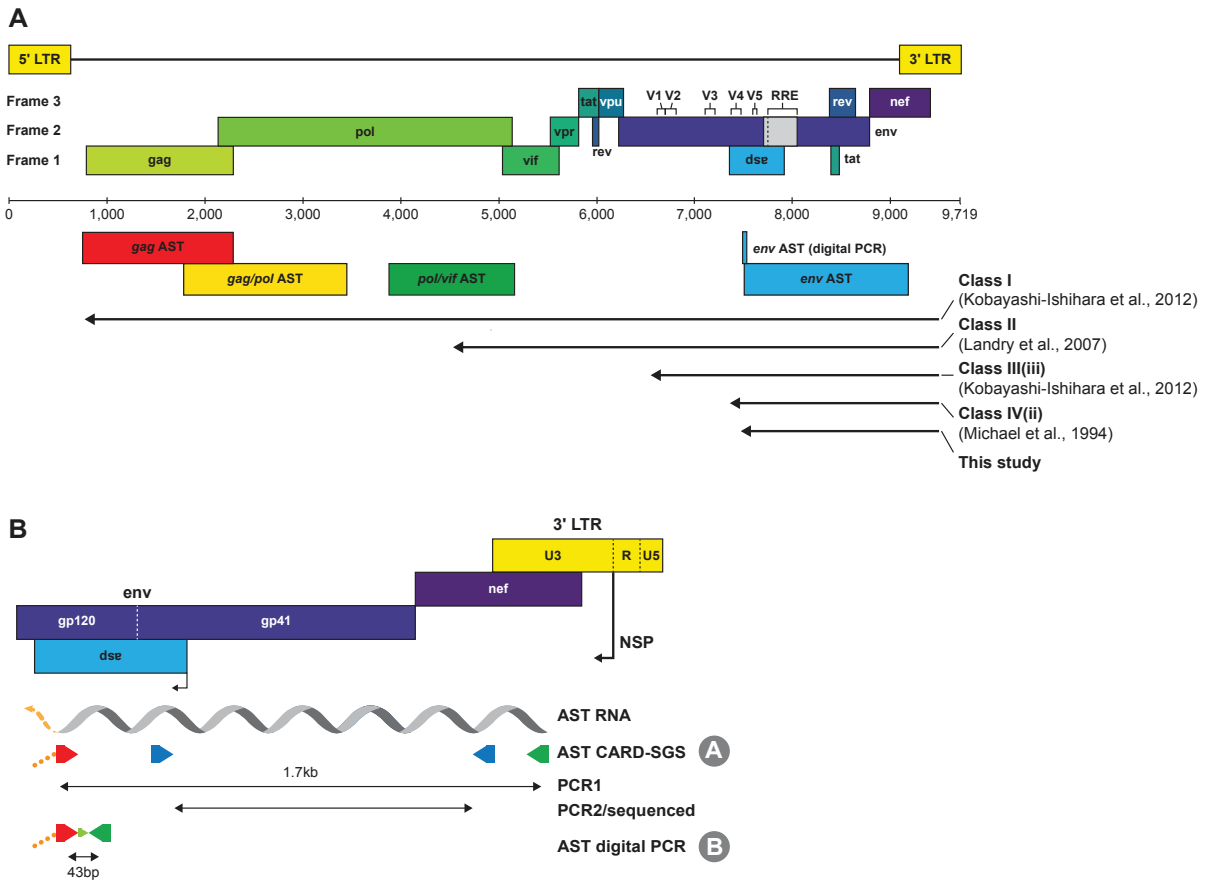


Fig. S1. Assays for measuring levels and sequencing AST *in vivo*. (**A**) Measuring AST along the HIV-1 genome. The antisense regions for single-genome sequencing of *gag, gag/pol*, *pol/vif*, and *env* are shown with the anti-*env* location for the digital PCR assay indicated as well. The genome map is relative to HXB2. Different length Classes of AST are defined for Class I (*12*), Class II (*11*), Class III(iii) (*12*), Class IV (*5*), and in our study. (**B**) A detailed view of the locations for the AST CARD-SGS assay (protocol A) and the digital PCR assay (protocol B). The negative sense promoter (nsp) for AST is shown at the 3′ LTR U3/R junction. The orange segment of the AST RNA indicates the potential for the transcript to be longer such as Class I & II. The tag for the primers is bold dashed orange that stems from the cDNA synthesis primer (red). Protocol A for AST CARD-SGS includes PCR1 with the forward primer to the tag to amplify a ~1.7-kb fragment from gp120 *env* to *nef*/U3, and a nested PCR before sequencing. Protocol B for AST digital PCR, uses the forward primer binding to the tag, and amplifies 43bp. The AST digital PCR assay amplification region overlaps the AST CARD-SGS assay region in *env*.

Fig. S2

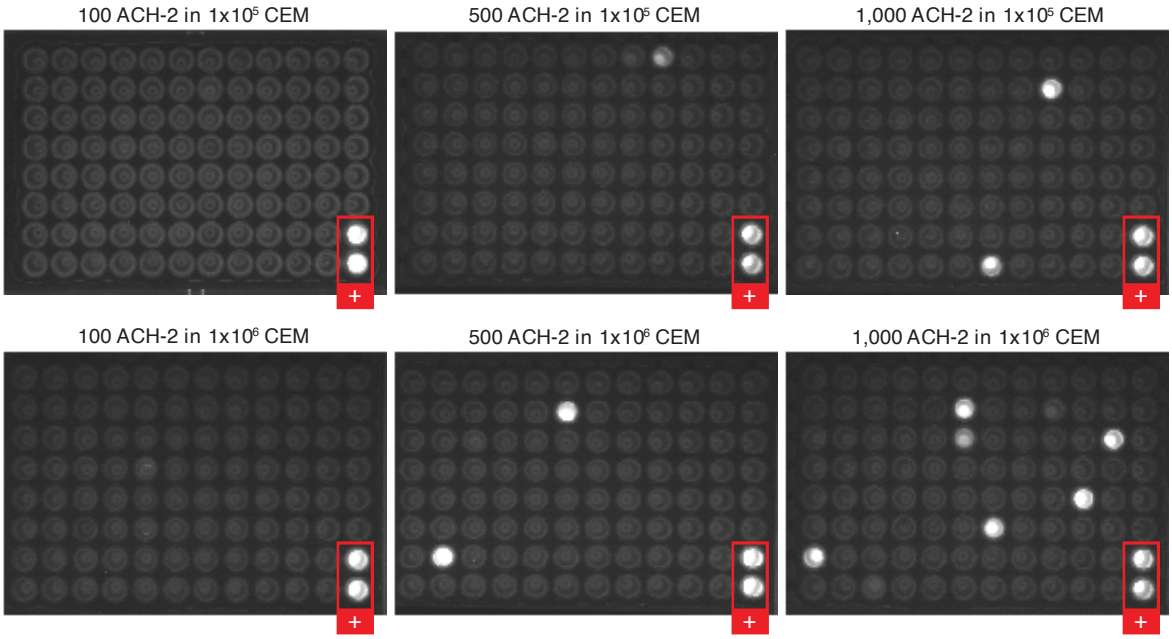
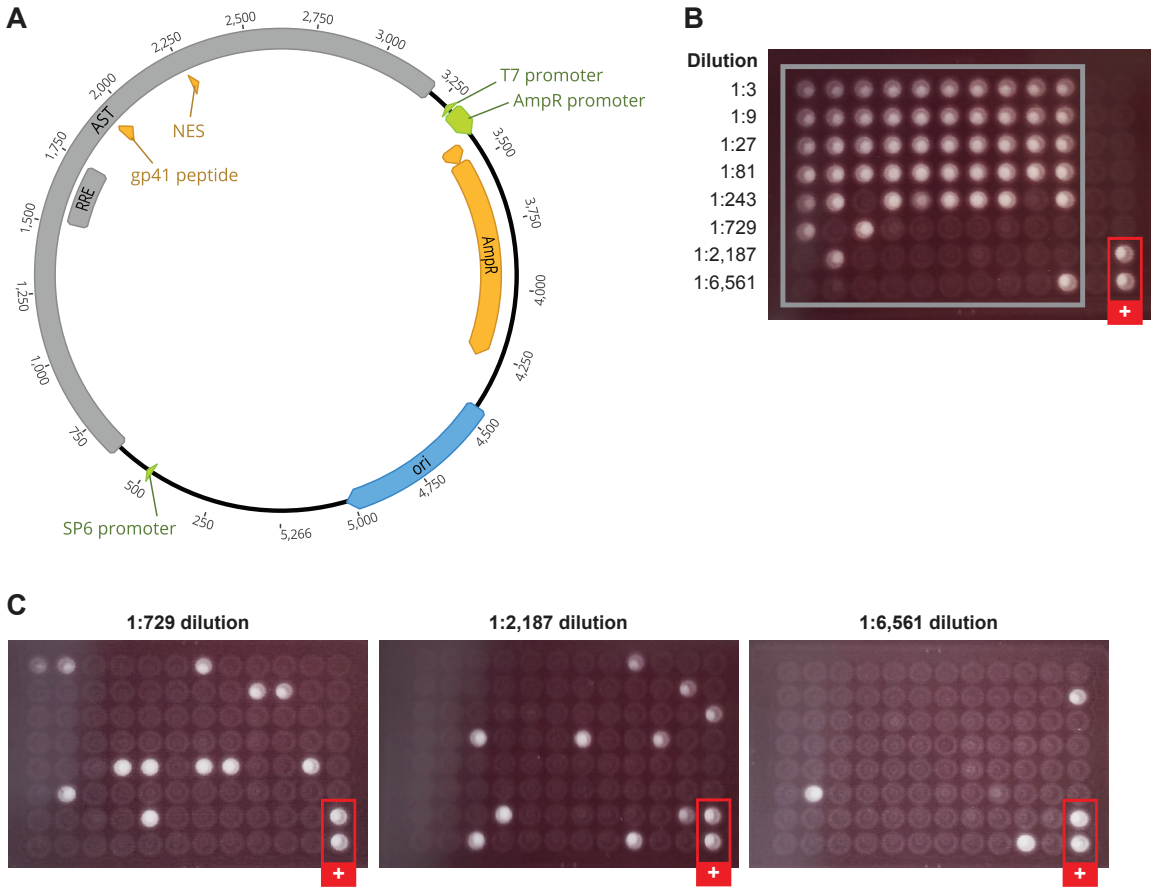


Fig. S2. Development of the Antisense Transcripts Cell-Associated RNA and DNA Single Genome Sequencing (AST CARD-SGS) assay – control for DNA digestion. cDNA reactions were performed with the exogenous oligo-tagged RT primer but excluding the reverse transcriptase enzyme, SuperScript III. The reaction was then spread across a 96-well plate and nested PCR was completed. Each plate included positive control wells (red box) that contained ACH-2 DNA.

Fig. S3



D

Dilution	Positive wells in 50uL	Copies in 150uL	Total copies	Average
1:729	12	36	2,916	5,589
1:2,187	10	3	7,290	
1:6,561	3	9	6,561	

Fig. S3. Determining the sensitivity of the AST CARD-SGS. (**A**) Plasmid used for *in vitro* expression of AST (pMiniT Expression Vector). (**B**) PCR replicates of serially diluted *in vitro* transcribed, then reverse transcribed AST for quantification and endpoint determination (10^4 copies of AST RNA, determined by spectrophotography, were spiked into nucleic acid extracted from 10^5 uninfected CEM cells (to mimic the 1/1,000 infected cells:uninfected cells in PWH). 2µL of each dilution was added to each PCR reaction well. (**C**) Detection plates of 50µL of select dilutions determined at endpoint. Detection plates were performed with 150µL as well (not shown, but results provided in **Fig. S3D**). (**D**) Average total number of detected AST copies based on the number of positives in the detection plates. Of 10^4 AST RNA copies, 5,589 were recovered, showing that the sensitivity of our methods for AST detection is about 50%.

Fig. S4

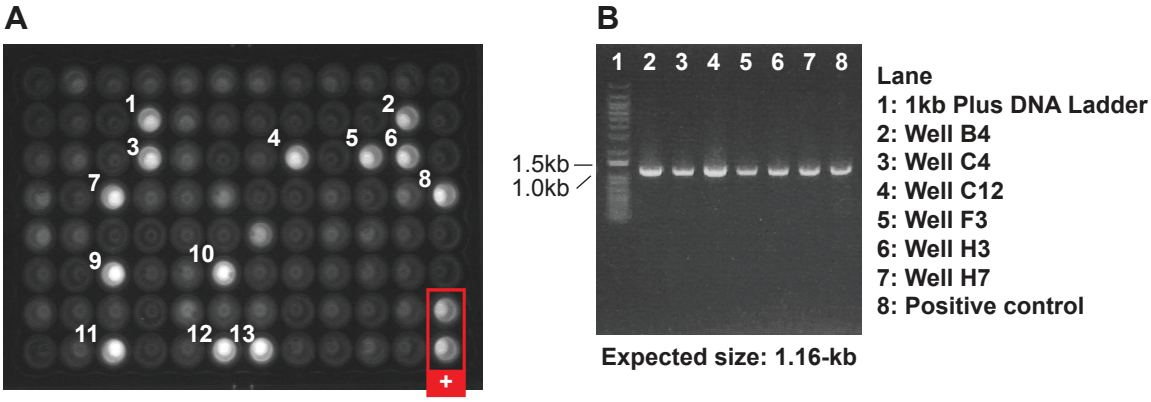


Fig. S4. Optimized AST CARD-SGS Assay. (**A**) The *env* AST CARD-SGS assay was applied to an aliquot of 100 ACH-2 cells in a background of $1x10^5$ CEM cells. (**B**) Select positive amplicons from the detection plate in **Fig. S4A** were run on 0.8% agarose gel with the expected amplicon size of ~1.16-kb.

Table S1. Primers for AST cDNA synthesis and PCR amplification.

Primer Name	Primer Sequence (5′→3′)	Primer Use
Tag Fwd[A]	**CTGATCTAGAGGTACCGGATCC**	Tag acting as Fwd primer
AST cDNA	**CTGATCTAGAGGTACCGGATCC**AACATGTGGCAGGAAGTAGG	*env* AST cDNA primer
Env cDNA	**CTGATCTAGAGGTACCGGATCC**ACATCTAATTTGTCCACTGA	*env* cDNA primer
1849Tag+	**CTGATCTAGAGGTACCGGATCC**GATGACAGCATGTCAGGGAG	*gag/pol* AST cDNA primer
3996Tag+	**CTGATCTAGAGGTACCGGATCC**CATCTAGCTTTGCAGGATTCG	*pol/vif* AST cDNA primer
G00-Tag+	**CTGATCTAGAGGTACCGGATCC**GACTAGCGGAGGCTAGAAG	*gag* AST cDNA primer
AST-OF	**CTGATCTAGAGGTACCGGATCC**	*env* AST PCR1[B]
AST-OR	TGGTACTAGCTTGTAGCACCATCC	
AST-IF	AGCAGAACAATTTGCTGAGGGC	*env* AST PCR 2[C]
AST-IR	GTCATTGGTCTTAAAGGTACCTGAGG	
1849+	GGATCCGATGACAGCATGTCAGGGAG	*gag/pol* AST PCR1[D]
3500-	CTATTAAGTCTTTTGATGGGTCATAA	
1870+	GAGTTTTGGCTGAAGCAATGAG	*gag/pol* AST PCR2[E]
3410-	CTGTTAGTGGTATTACTTCTGTTAGTGCTT	
5270-	CTGACCCAAATGCCAGTCTC	*pol/vif* AST PCR1[F]
4133+	GGAAAAGGTCTATCTGGCATG	*pol/vif* AST PCR2[G]
5248-	TCTCCTGTATGCAGACCCCA	
G01	AGGGGTCGTTGCCAAAGA	*gag* AST PCR1[H]
G10	CAGTATTAAGCGGGGGAGAATT	*gag* AST PCR2[I]
G15	CTTTGCCACAATTGAAACACTT	
envB5out	TAGAGCCCTGGAAGCATCCAGGAAGT	*env* PCR1[J]
9538r	AGAGAGACCCAGTACAGGCAAAA	
envB5in	TTAGGCATCTCCTATGGCAGGAAGAAG	*env* PCR2[K]
9418r	CAAGCTCGATGTCAGCAGTTCT	

[A] Exogenous oligo-tag in black bold
[B] *env* AST PCR1: 94°C for 2 min, 45 cycles of 94°C for 15 sec, 60°C for 15 sec, and 68°C for 30 sec
[C] *env* AST PCR2: 94°C for 2 min, 40 cycles of 94°C for 15 sec, 60°C for 15 sec, and 68°C for 20 sec
[D] *gag*/*pol* AST PCR1: 94°C for 2 min, 45 cycles of 94°C for 15 sec, 60°C for 15 sec, and 68°C for 30 sec
[E] *gag*/*pol* AST PCR2: 94°C for 2 min, 40 cycles of 94°C for 15 sec, 60°C for 15 sec, and 68°C for 30 sec
[F] *pol*/*vif* AST PCR1: 94°C for 2 min, 45 cycles of 94°C for 15 sec, 60°C for 15 sec, and 68°C for 20 sec
[G] *pol*/*vif* AST PCR2: 94°C for 2 min, 40 cycles of 94°C for 15 sec, 60°C for 15 sec, and 68°C for 20 sec
[H] *gag* AST PCR1: 94°C for 2 min, 45 cycles of 94°C for 15 sec, 60°C for 15 sec, and 68°C for 25 sec
[I] *gag* AST PCR2: 94°C for 2 min, 40 cycles of 94°C for 15 sec, 60°C for 15 sec, and 68°C for 20 sec
[J] *env* PCR1: 94°C for 2 min, 45 cycles of 94°C for 15 sec, 60°C for 30 sec, and 68°C for 3 min
[K] *env* PCR2: 94°C for 2 min, 40 cycles of 94°C for 15 sec, 60°C for 30 sec, and 68°C for 20 sec

Table S2. Primers used for AST controls.

Primer Name	Primer Sequence (5′→3′)
AST-NEF2R	GTCATTGGTCTTAAAGGTACCTGAGG
AST-1586R	AGCAGAACAATTTGCTGAGGGC
AST-SEQ2R	GGTGAATATCCCTGCCTAACTCTAT
AST-SEQ2F	GGTTTAACATAACAAATTGGCTGTGGTATATAA
AST-SEQ3F	ATGGGTGGCAAGTGGTCAAA
Cloning Analysis Fwd	ACCTGCCAACCAAAGCGAGAAC
Cloning Analysis Rev	TCAGGGTTATTGTCTCATGAGCG
Rev16	ATGGGAGGGGCATACATTGCT
Rev17	CCTGGAGCTGTTTAATGCCCCAGAC
M13 Fwd	GTAAAACGACGGCCAGT
M13 Rev	CAGGAAACAGCTATGAC

Table S3. Digital PCR Primers and Probes.

Primer Name	Primer/Probe Sequence (5′→3′)	Primer Use
AST-Tag	**ATCTAGAGGTACCGGATCCAAC**	*env* AST
AST Rev-qPCR	TGATGAACATCTAATTTGTCCACTGA	
AST Probe	/56-FAM/AGCAATGT/ZEN/ATGCCCCTCCCA/3IaBkFQ/	
AST-Tag	**ATCTAGAGGTACCGGATCCACA**	*env*
env Fwd-qPCR	ACAAATTATAAACATGTGGCAGAAAGTAGG	
env Probe	/56-FAM/TGGGAGGGG/ZEN/CATACATTGCT/3IaBkFQ/	
LTR U5 Fwd	CTTAAGCCTCAATAAAGCTTGCC	LTR U5
LTR U5 Rev	GGATCTCTAGTTACCAGAGTC	
LTR U5 Probe	/5′HEX/AGTAGTGTG/ZEN/TGCCCGTCTG/3IaBkFQ/	
RRE Fwd	GCAGAGAGAAAAAAGAGC	RRE
RRE Rev	GCCTGTACCGTCAGC	
RRE Probe	/5′HEX/TTCCTTGGG/ZEN/TTCTTGGGAGCAG/3IaBkFQ/	
iCAD Fwd	TTTGGAAAGGACCAGCAAA	IN DNA Assay (iCAD)
iCAD Rev	CCTGCCATCTGTTTTCCA	
iCAD Probe	/56-FAM/AAAGGTGAA/ZEN/GGGGCAGTAGTAATACA/3IaBkFQ/	

Exogenous oligo-tag sequence in bold

Table S4. Sequencing Primers.

Primer Name	Primer Sequence (5′→3′)	Primer Use
AST-Seq2 Rev	GGTGAATATCCCTGCCTAACTCTAT	*env* AST
AST-Seq2 Fwd	GGTTTAACATAACAAATTGGCTGTGGTATATAA	
AST-Seq3 Fwd	ATGGGTGGCAAGTGGTCAAA	
AST-IF	AGCAGAACAATTTGCTGAGGGC	
AST-IR	GTCATTGGTCTTAAAGGTACCTGAGG	
2030+	TGTTGGAAATGTGGAAAGGAAGGAC	*gag/pol* AST
2600+	ATGGCCCAAAAGTTAAACAATGGC	
2610-	TTCTTCTGTCAATGGCCATTGTTTAAC	
3330-	TTGCCCAATTCAATTTTCCCACTAA	
Poli9D	AAAATTAGCAGGAMGATGGCCAG	*pol/vif* AST
Poli10B	TATTCATAGATTCYACTACTCCTTG	
4133+	GGAAAAGGTCTATCTGGCATG	
5248-	TCTCCTGTATGCAGACCCCA	
G30	CAGTAGCAACCCTCTATTGTGT	*gag* AST
G25	ATTGCTTCAGCCAAAACTCTTGC	
G10	CAGTATTAAGCGGGGGAGAATT	
G15	CTTTGCCACAATTGAAACACTT	
For16	TTTAATTGTGGAGGAGAATTTTTCTA	*env*
For17	AGCAGCAGGAAGCAGTATGGGCGC	
For18	CATATCAAATTGGCTGTGGTATAT	
Rev15	CTGCCATTTAACAGCAGTTGAGTTGA	
Rev16	ATGGGAGGGGCATACATTGCT	
envB5in	TTAGGCATCTCCTATGGCAGGAAGAAG	
envB5out	TAGAGCCCTGGAAGCATCCAGGAAGT	
9418r	CAAGCTCGATGTCAGCAGTTCT	

Table S5. Donor specific primers.

Primer Name	Primer Sequence (5′→3′)
2669_ASTI-F	AGCAGAACAATTTGCTGAG**R**GC
2669_ASTI-R	GTCATTGGTCTTAAAGGTACCTG**R**GG
2669 AST-R-qPCR	TGATGAACATCTAATT**A**GTCC**TT**TGA
2669_*env*-Tagged	**CTGATCTAGAGGTACCGGATCC**TGATGAACATCTAATT**A**GTCC**TT**TGA
1079_ASTR-qPCR	TGATGAACATCTAATTT**S**TCC**M**CTGA
1079_*env* -Tagged	**CTGATCTAGAGGTACCGGATCC**ACATCTAATTT**S**TCC**M**CTGA
1683_AST-Tagged	**CTGATCTAGAGGTACCGGATCC**AACATGTGGCA**AG**AA**A**TAGG
1683_AST-OR	TGGTACTAGCTTGTAGCACCA**C**CC
1683_AST-IF	AGCAGAACAATTTGCTGAG**A**GC
1683_AST-R-qPCR	TGATGAACATCTAAT**CA**GTCC**G**CTGA
1683_*env* -Tagged	**CTGATCTAGAGGTACCGGATCC**ACATCTAAT**CA**GTCC**G**CTGA
1683_*env* F-qPCR	ACAAATTATAAACATGTGGCA**AG**AA**A**TAGG
291_AST-R	TGATGAACA**RY**TAATTT**K**TCC**TY**TGA
1508_AST-Tagged	**CTGATCTAGAGGTACCGGATCC**AA**T**ATGTGGCAG**GG**AGTAGG
1508_AST-R	TGATGAACA**K**CTAATTTGTCC**KYK**GA
1508_AST-P	/56-FAM/AGC**C**ATGTA/ZEN/TGCCCCTCCCA/3IaBkFQ/
3611_AST-Tagged	**CTGATCTAGAGGTACCGGATCC**AACATGTGGCA**AG**AAGTAGG
3611_AST-R	TGATG**C**ACA**G**CTAATT**Y**GTCC**T**CTGA
3611_AST-P	/56-FAM/A**G**GCAATGT/ZEN/ATGCCCCTCCCA/3IaBkFQ/
1226_AST-Tagged	**CTGATCTAGAGGTACCGGATCC**AACATGTGGCA**AC**AAGTAGG
1226_AST-R	TGATGAACA**G**CTAAT**C**TGTCC**T**C**G**GA
1775_AST-Tagged	**CTGATCTAGAGGTACCGGATCC**AACATGTGGCA**R**AAAGTAGG
1775_AST-R	TGATGAACAT**T**TAATTTGTCC**T**CTGA
477_AST-R	TGATGAACA**Y**CTAATTAGTCCACTGA

Exogenous oligo-tag sequence in black bold
Bold red nucleotide reflects donor specific base
Degenerate bases: R (A/G), S (S/G), M (A/C), Y (C/T), and K (G/T) following IUPAC nucleotide code

Table S6. Levels of HIV-1 AST in ART-treated and untreated donors.

Participant Identifier (PID)	Duration on ART at sampling	Estimated number of infected PBMC assayed[B]	Number of HIV-1 AST molecules detected	Number of HIV-1 AST molecules per 100 infected PBMC
1079	12.8 years	80	9	11
1683	5.4 years	188	32	17
2669	4.3 years	90	19	21
	5.5 years	65	20	31
	2 weeks[A]	88	34	39
	1 month[A]	45	31	69
	Median	**84**	**26**	**26**
	IQR	**60-115**	**17-33**	**16-47**
291	0	104	2	2
477	0	83	2	2
1508	0	22	1	5
1775	0	40	13	33
3611	0	244	1	≤0.4[C]
	Median	**83**	**2**	**2**
	IQR	**31-174**	**1-8**	**1-19**

[A] After 5.5 years on ART, the participant had an unexpected ART interruption for approximately 4 weeks They reinitiated ART with the first timepoint post-ART interruption at 2 weeks with low but detectable HIV-1 plasma viremia. Then 1 month post-ART interruption with plasma viremia suppressed

[B] Estimated the number of HIV DNA levels using the integrase cell-associated single-copy DNA (iCAD) assay (*24*)

[C] Detectable but below the cut-off