Epitranscriptomic Addition of m^5C to HIV-1 Transcripts Regulates Viral Gene Expression

David G. Courtney,[1] Kevin Tsai,[1] Hal P. Bogerd,[1] Edward M. Kennedy,[1,5] Brittany A. Law,[2] Ann Emery,[3] Ronald Swanstrom,[3,4] Christopher L. Holley,[2] and Bryan R. Cullen[1,6,*]

[1]Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, NC 27710, USA
[2]Department of Medicine, Duke University Medical Center, Durham, NC 27710, USA
[3]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, USA
[4]Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, USA
[5]Present address: Oncorus, 50 Hampshire Street, Suite 401, Cambridge, MA 02139
[6]Lead Contact
*Correspondence: bryan.cullen@duke.edu
https://doi.org/10.1016/j.chom.2019.07.005

SUMMARY

How the covalent modification of mRNA ribonucleotides, termed epitranscriptomic modifications, alters mRNA function remains unclear. One issue has been the difficulty of quantifying these modifications. Using purified HIV-1 genomic RNA, we show that this RNA bears more epitranscriptomic modifications than the average cellular mRNA, with 5-methylcytosine (m^5C) and 2′O-methyl modifications being particularly prevalent. The methyltransferase NSUN2 serves as the primary writer for m^5C on HIV-1 RNAs. NSUN2 inactivation inhibits not only m^5C addition to HIV-1 transcripts but also viral replication. This inhibition results from reduced HIV-1 protein, but not mRNA, expression, which in turn correlates with reduced ribosome binding to viral mRNAs. In addition, loss of m^5C dysregulates the alternative splicing of viral RNAs. These data identify m^5C as a post-transcriptional regulator of both splicing and function of HIV-1 mRNA, thereby affecting directly viral gene expression.

INTRODUCTION

RNAs are subject to a range of covalent modifications at the single nucleotide level and over 100 distinct RNA modifications have been described (Gilbert et al., 2016; Li and Mason, 2014; Roundtree et al., 2017). While primarily found on non-coding RNAs (ncRNAs), especially tRNAs, several of these "epitranscriptomic" modifications are also found on eukaryotic mRNAs, and there has been considerable recent interest in defining their function.

The proteins involved in epitranscriptomic regulation of mRNAs can be divided into "writers," which add the modification; "readers," which detect the modification and execute its phenotypic effect(s); and "erasers," which remove the modification. The addition of a methyl group to the *N*6 position of adenosine (m^6A) is the most common mRNA modification in eukaryotes (Meyer and Jaffrey, 2014; Roundtree et al., 2017), and m^6A has therefore attracted considerable attention. The m^6A modification is added co-transcriptionally by a complex consisting of the methyltransferase METTL3 and two co-factors, METTL14 and WTAP (Meyer and Jaffrey, 2014). Once added, m^6A can be read by nuclear YTHDC1 (Hsu et al., 2017; Xu et al., 2014), which may regulate alternative mRNA splicing (Xiao et al., 2016). After nuclear export, m^6A sites are bound by three cytoplasmic reader proteins, YTHDF1, YTHDF2, and YTHDF3, which are thought to regulate mRNA stability and translation (Shi et al., 2017; Wang et al., 2014, 2015).

Less is known about the function of other epitranscriptomic mRNA modifications, of which one of the most common is the addition of a methyl group to the C5 position of cytosine (m^5C) (Li et al., 2017; Squires et al., 2012). The primary writer for m^5C on mRNAs has been proposed to be NSUN2, although human cells express seven other cytosine methyltransferases of which one, DNMT2, may also act on mRNA (Gilbert et al., 2016; Khoddami and Cairns, 2013; Squires et al., 2012). The function of m^5C remains largely undefined, although it has been proposed that m^5C residues, which are often located close to translation initiation codons (Yang et al., 2017), can promote mRNA translation and enhance nuclear RNA export (Li et al., 2017; Xing et al., 2015; Yang et al., 2017). Here, we initially focus on the quantification of the epitranscriptomic modifications found on a single, highly purified RNA species synthesized in human cells, the genomic RNA (gRNA) of HIV-1. Using ultra high-performance liquid chromatography linked to tandem mass spectrometry (UPLC-MS/MS), we demonstrate that HIV-1 gRNAs are extensively epitranscriptomically modified, with an m^5C level ~14× higher than seen on cellular mRNAs. We identify NSUN2 as the primary m^5C writer on HIV-1 gRNA and show that loss of NSUN2 reduces HIV-1 protein expression yet does not affect viral RNA levels, consistent with a role for m^5C in promoting mRNA translation (Li et al., 2017; Xing et al., 2015). Finally, we report that loss of NSUN2 perturbs the alternative splicing of HIV-1 transcripts. Together, these data reveal that HIV-1 has evolved to acquire a high level of epitranscriptomic m^5C modifications and that these, in turn, promote viral gene expression by regulating RNA splicing and promoting the translation of viral mRNAs.

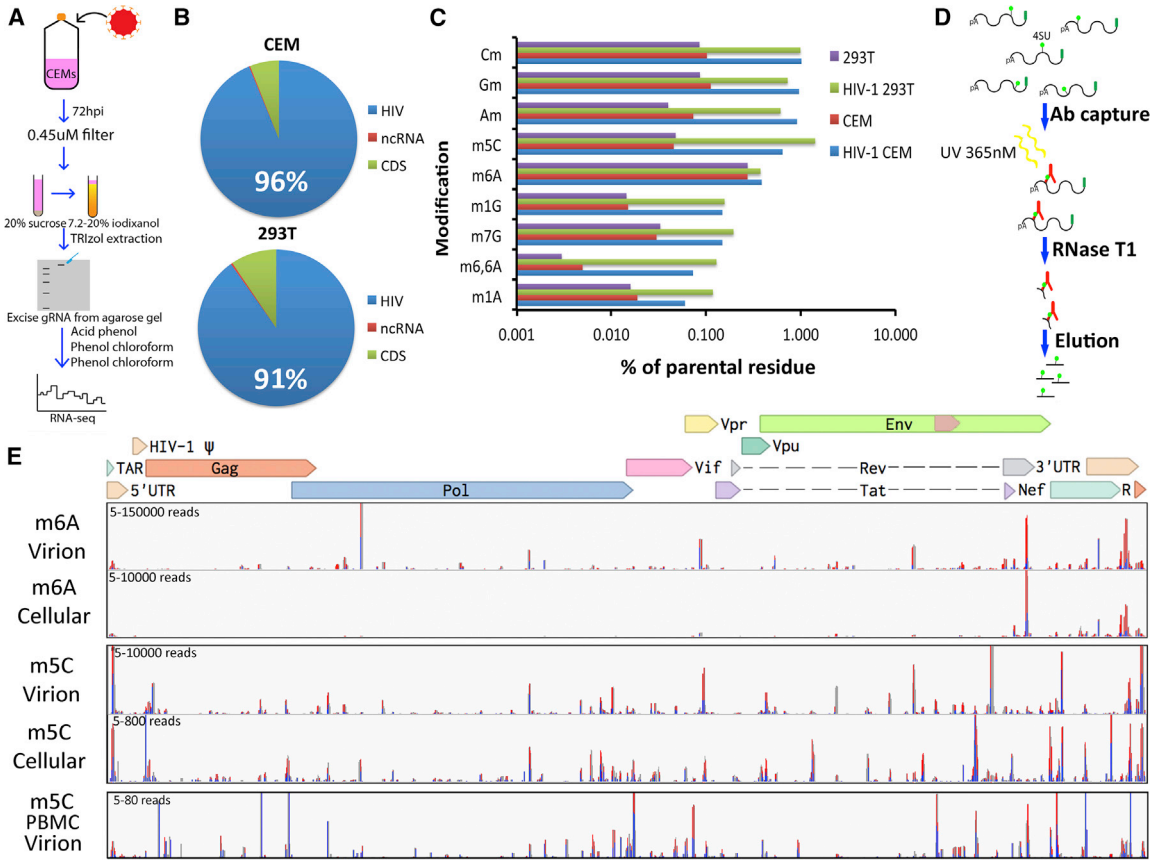

Figure 1. Quantification and Mapping of RNA Modifications Present on HIV-1 Transcripts

(A) Schematic of the purification of the gRNA present in HIV-1 virions released from infected CEM T cells.

(B) Purified HIV-1 gRNA derived from infected CEM T cells or transfected 293T cells was analyzed by RNA-seq to determine the level of contaminating human RNA. CDS; human mRNAs. n = 1.

(C) UPLC-MS/MS was used to determine the percentage of each RNA modification listed, normalized to the level of the parental nucleoside, in the two HIV-1 gRNA samples and in CEM or 293T poly(A)+ RNA. n = 1.

(D) Schematic of the antibody capture technique used to map m^6A or m^5C modifications on HIV-1 RNA.

(E) The top two images show PA-m^6A-seq mapping, while the lower three images show PA-m^5C-seq mapping, of HIV-1 RNA derived from either virions or cell lysates of the matched virus producer CEM cells or purified virions from infected primary CD4+ T cells.

RESULTS

HIV-1 gRNAs Bear Extensive Epitranscriptomic Modifications

While several techniques can be used to map epitranscriptomic modifications on RNAs (Helm and Motorin, 2017; Kennedy et al., 2016; Khoddami et al., 2015; Li et al., 2017), the resultant data do not provide information about the level of modification at the mapped sites, which could range anywhere from 100% to <10%. To quantify the epitranscriptomic modifications present on a single human RNA species, we focused on the RNA genome of HIV-1, which is packaged into virions released by infected cells and is therefore easier to purify to homogeneity than cellular mRNAs. We harvested supernatant media from HIV-1-infected cultures of the human CD4+ T cell line CEM and isolated virions away from cellular debris and exosomes by pelleting through a sucrose cushion and then banding on an iodixanol gradient (Eckwahl et al., 2016) (Figure 1A). While this resulted in a pure HIV-1 virion preparation, virions are known to package several cellular RNAs, including tRNAs (Eckwahl et al., 2016), thus mandating an additional RNA purification step. This was achieved by denaturation of the RNA sample prior to agarose gel electrophoresis followed by excision and recovery of the ~9 kb HIV-1 gRNA (Figure 1A). The resultant gRNA preparation, with a yield of ~30 ng, was then analyzed for purity by RNA sequencing (RNA-seq). This showed that the CEM-derived RNA sample consisted of ~96.0% HIV-1 RNA (Figure 1B), derived from the entire HIV-1 genome, ~3.8% human mRNAs and ~0.2% human ncRNA, of which <0.01% was of tRNA origin. Because human tRNAs and rRNAs are highly modified, the almost complete elimination of ncRNAs was critical. In contrast, human mRNAs bear few epitranscriptomic modifications (Li et al., 2016) and, at this low level, would not affect the overall pattern of modified nucleosides detected on the HIV-1 gRNA. To assess whether the level of epitranscriptomic modification of HIV-1 gRNAs varied across different producer cell lines, we also purified HIV-1 virions released from the human embryonic kidney cell line 293T after transfection with an expression plasmid, pNL4-3, that encodes a wild-type

Table 1. Relative Level of the Indicated RNA Modifications Detected on Purified HIV-1 Virion RNA and Total poly(A)+ Cellular RNA samples

Residue	Absolute Number per CEM HIV-1 gRNA	% Parent Nucleotide in CEM HIV-1 gRNA	% Parent Nucleotide CEM Poly(A)+	% Parent Nucleotide in 293T HIV-1 gRNA	% Parent Nucleotide 293T Poly(A)+
Am	30.1	0.915	0.073	0.613	0.040
Gm	21.3	0.961	0.112	0.727	0.086
Cm	16.6	1.020	0.102	0.993	0.085
m^6A	12.7	0.387	0.274	0.376	0.275
m^5C	10.5	0.645	0.046	1.422	0.048
m^1G	3.3	0.149	0.015	0.157	0.014
m^7G	3.3	0.149	0.030	0.195	0.033
$m^{6,6}A$	3.1	0.073	0.005	0.129	0.003
m^1A	2.4	0.060	0.019	0.118	0.016

(WT) NL4-3 provirus. Virion-derived gRNA was purified as described above and the resultant gRNA preparation consisted of ~90.5% HIV-1 gRNA, ~9.2% human mRNAs, ~0.3% ncRNA, and <0.01% tRNA.

To quantify the level of modified RNA nucleosides, we digested the RNA sample to ribonucleosides and then performed UPLC-MS/MS (Basanta-Sanchez et al., 2016). Because identification requires the use of appropriate molecular standards, some modified nucleosides, such as the recently identified N4-acetylcytidine modification (Arango et al., 2018) could not be identified in our analysis. Moreover, due to the inefficient ionization of uridine compared to other nucleobases, the UPLC-MS/MS method does not effectively measure modified U-residues, such as 2′O-methyluridine (Um), at the low level of viral RNA input (~30 ng) used here.

Figure 1C and Table 1 present the level of a range of modified nucleosides quantified in the purified HIV-1 gRNA derived from virions released by infected CEM T cells or transfected 293T cells. The level of modified nucleosides detected in purified poly(A)+ RNA isolated from uninfected CEM or 293T cells is also presented, and these are similar to the levels previously reported by others (Li et al., 2016). Of note, the level of epitranscriptomic RNA modifications on HIV-1 gRNA isolated from virions produced in infected CEM T cells or transfected 293T cells was similar, and in both cases, several modifications were detected at much higher levels on HIV-1 gRNA than on CEM or 293T poly(A)+ RNA. These include 2′O-methyladenosine Am (~13× higher), 2′O-methylguanosine (Gm) (~9× higher), 2′O-methylcytosine (Cm) (~10× higher), and m^5C (~14× higher), all four of which represent ~1% of the cognate nucleotide on HIV-1 gRNAs. While other more minor, modified nucleosides were also detected at higher levels on the HIV-1 gRNA than on poly(A)+ RNA, they are predicted to only contribute 2 to 3 residues per gRNA (Table 1) and are of unclear functional significance. Inosine, which is derived by editing of adenosine and commonly found on tRNAs and mRNAs, was not detected on HIV-1 gRNAs. This was expected, as inosine is read as guanosine by reverse transcriptase (Doria et al., 2009) so that adenosines present on the HIV-1 genome that are edited to inosine would, in time, become fixed as guanosines. While we, and others, have reported the mapping of m^6A residues on HIV-1 transcripts and proposed that these enhance HIV-1 replication (Kennedy et al., 2016; Lichinchi et al., 2016), the level of m^6A on the HIV-1 gRNA, at ~0.38% of the parental nucleoside, is only a modest ~1.4× higher than seen in total poly(A)+ RNA. Nevertheless, due to the A-rich character of HIV-1 virion RNA, this translates into an average of ~13 m^6A residues per HIV-1 genome (Table 1). In contrast, because the HIV-1 genome includes relatively few C residues, the higher relative level of m^5C on the HIV-1 gRNA translates into a similar number of predicted m^5C residues, i.e., ~11 m^5C residues in CEM-derived virions. Importantly, m^6A, m^5C, and Nm residues are not predicted to inhibit the reverse transcription step in the HIV-1 life cycle.

Mapping of Epitranscriptomic Modifications on the HIV-1 RNA Genome

To map m^5C residues on HIV-1 gRNAs, we used a modification of the previously described photo-crosslinking-assisted m^6A sequencing (PA-m^6A-seq) procedure dubbed PA-m^5C-seq (Chen et al., 2015) (Figures 1D and S2). CEM T cells or primary $CD4^+$ T cells were infected with HIV-1 and then pulsed with 4-thiouridine (4SU), a highly photoactivatable uridine analog. Total cellular RNA was then isolated from CEMs and subjected to one round of poly(A)+ RNA isolation. In parallel, released virions from CEM or primary T cells were recovered and total virion RNA isolated. 10 μg of poly(A)+ cellular RNA or virion RNA was then processed using the PA-m^6A-seq technique, as previously described (Kennedy et al., 2016) or substituting an m^5C-specific antibody for the m^6A-specific antibody.

Because the PA-m^5C-seq technique has not been previously described, we wished to confirm that it detects known m^5C sites. As shown in Figures S1A and S1B, known m^5C sites on $tRNA^{Ser}$ and 5S rRNA (Hussain et al., 2013a) were accurately detected using PA-m^5C-seq (these ncRNAs remained present in this cellular RNA preparation because one round of poly(A)+ RNA isolation does not remove all ncRNA). When mapped over the entire CEM transcriptome, our data confirm the previously observed concentration of m^5C around translation start codons (Yang et al., 2017) (Figures S1D and S1E), with a slight concentration around the translation termination codon (Figure S1F). Overall, m^5C residues were evenly distributed between the 5′ UTR, coding sequences (CDSs), and 3′UTR (Figure S1G), even though 5′ UTRs are generally shorter than CDSs or 3′ UTRs. Analysis of all m^5C sites on cellular RNAs identified an m^5C consensus sequence (Figure S1C) that was, however, only found at ~20% of mapped m^5C residues and was not present at m^5C

sites mapped on HIV-1 transcripts. As noted below, there are several known cytosine methyltransferases and this may represent a consensus site for one of these. Analysis of the prevalence of m^5C residues revealed that most cellular mRNAs bear few or no m^5C residues, though 5 were detected on the EPPK1 mRNA and 4 on three other mRNAs (Figures S1H and S1I). While ~11 m^5C residues were detected on HIV-1 gRNAs (Table 1), we note that the average cellular mRNA is ~2.2 kb in length, while the HIV-1 gRNA is ~9.2kb. Overall, our data mapped 1641 m^5C residues on 1489 cellular mRNAs, which represents an average of only ~1.1 m^5C residues per m^5C-containing mRNA (Figure S1J).

Analysis of PA-m^5C-seq data generated using intracellular and virion gRNA samples revealed ~19 distinct m^5C peaks from infected CEMs that, while observable on both virion and cellular viral RNAs, nevertheless varied in intensity (Figures 1E and S2A). Of the 19 peaks observed in multiple CEM PA-m5C-seq replicates, 18 were also identified on HIV-1 gRNA produced from primary CD4+ T cells (Figures 1E and S2B). m^5C peaks were observed across the entire HIV-1 genome but were more prevalent toward the 3′ end. The observation of ~19 major m^5C peaks on HIV-1 gRNA, while UPLC-MS/MS detected ~11 m^5C residues in total, implies that many of the m^5C peaks visualized in Figure 1E do not result from the complete methylation of mapped residues.

We also performed PA-m^6A-seq analysis in parallel using virion RNA and poly(A)+ cellular RNA (Figure 1E, top). As expected, this analysis revealed no overlap between the antibody binding sites detected using the m^5C-specific versus m^6A-specific antibody. The PA-m^6A-seq data generated using poly(A)+ RNA recovered from HIV-infected T cells confirmed the previously reported concentration of m^6A toward the 3′ end of the HIV-1 genome and identified the same four major m^6A clusters located in the *env:rev* overlap, in *nef*, in the LTR "U3" region and in TAR (Figure 1E, second row) (Kennedy et al., 2016). In contrast, while analysis of virion RNA fully confirmed the location of the same four m^6A clusters detected in intracellular HIV-1 RNA, these data also identified several m^6A clusters in other regions of the HIV-1 genome, including in *pol*, *vpr*, and *env* (Figure 1E, top row). This result can only be partially explained by the fact that virion gRNA is unspliced, while intracellular HIV-1 RNA exists as an almost equal mixture of unspliced, singly spliced and multiply spliced RNAs, almost all of which retain the 3′ UTR region. We note that HIV-1 gRNAs, even though they are capped and polyadenylated, are not thought to be translated in producer cells but rather packaged directly into virions (Kharytonchyk et al., 2016). This raises the interesting possibility that translation may facilitate the loss of m^6A modifications located in open reading frames; for example, due to the ribosomal removal of bound m^6A readers, followed by m^6A removal by eraser proteins. Conversely, it is also possible that m^6A may facilitate gRNA packaging into virions.

Identification of NSUN2 as the Primary HIV-1 m^5C Methyltransferase

Given the large number of mapped m^5C sites on the HIV-1 genome, the lack of a consensus sequence that can identify which "C" residue within any given PA-m^5C-seq read is actually modified, and the location of many reads within functionally important regions of the HIV-1 genome, including CDSs, the interrogation of how m^5C addition affects HIV-1 gene expression by mutagenesis of the HIV-1 genome is not feasible. We therefore next sought to identify the "writer" protein that adds m^5C to viral mRNAs.

Addition of m^5C residues to RNAs is mediated by the seven members of the NSUN family of methyltransferases, NSUN1 through NSUN7, as well as by the DNA methyltransferase homolog DNMT2 (Bohnsack et al., 2019; Reid et al., 1999). Cytosine methylation by NSUN proteins uses two conserved cysteine residues, one of which forms a transient covalent bond to the pyrimidine base, while the second conserved cysteine (Figure 2B) is essential for release (King and Redman, 2002). Therefore, while mutagenesis of the first cysteine to alanine (C321A in NSUN2) results in an inactive mutant, mutagenesis of the second cysteine to alanine (C271A in NSUN2) results in a mutant that spontaneously crosslinks to target cytosines (Hussain et al., 2013b). We reasoned that expression of each NSUN protein, with this second cysteine mutated (Figure 2B), should result in the crosslinking of only the relevant NSUN protein, that mediates m^5C addition to HIV-1 transcripts, to viral gRNA which should, in turn, result in the virion incorporation of that NSUN protein (Figure 2A). We therefore generated cysteine to alanine mutations in the context of FLAG-tagged cDNA expression vectors encoding NSUN1, NSUN2, NSUN4, and NSUN5 (NSUN3, 6, and 7 were not expressed at detectable levels) and transfected these mutant NSUN expression plasmids into 293T cells along with a plasmid expressing the NL4-3 HIV-1 provirus. In parallel, we co-transfected a plasmid expressing FLAG-tagged APOBEC3G (A3G), which is incorporated into HIV-1 virions, as a positive control (Harris et al., 2003). As shown in Figure 2C, the four NSUN mutants, as well as A3G, were readily detected in transfected cells, while analysis of the released HIV-1 virions only detected A3G and NSUN2. This result was not unexpected, as NSUN2 has been reported to add m^5C to several cellular mRNAs. While the N2-C271A mutant is efficiently packaged into HIV-1 virions, WT NSUN2 is not detected (Figure 2D). The subcellular location of NSUN2, as determined by immunofluorescence (IF), is predominantly nuclear, thus suggesting that addition of m^5C to mRNAs, like the addition of m^6A, may be co-transcriptional (Ke et al., 2017).

We next used CRISPR-Cas to generate two NSUN2 knockout (KO) cell lines, N2.4 and N2.16, in the human 293T cell line (Sanjana et al., 2014). Loss of NSUN2 expression, which was confirmed both by DNA sequencing (Figure S3A) and by western blot (Figure S3B), reduced the level of m^5C in total poly(A)+ RNA (Figure S3C). Loss of NSUN2 expression had little effect on the growth of 293T cells (Figure 2F) yet significantly inhibited HIV-1 gene expression and replication (Figures 2G and 2H). In Figure 2G, WT or NSUN2 KO 293T cells were transfected with a CD4 expression plasmid and either a control plasmid or a plasmid expressing WT NSUN2. The cells were then infected with a replication-competent, NL4-3-based indicator virus (NL-NLuc), in which the dispensable *nef* gene had been partially replaced with the Nano luciferase (*NLuc*) indicator gene, and NLuc expression quantified at 24, 48, and 72 hpi. Loss of NSUN2 expression significantly reduced NLuc expression and HIV-1 spread, while complementation with an NSUN2 expression plasmid rescued this inhibition. Similar data are reported

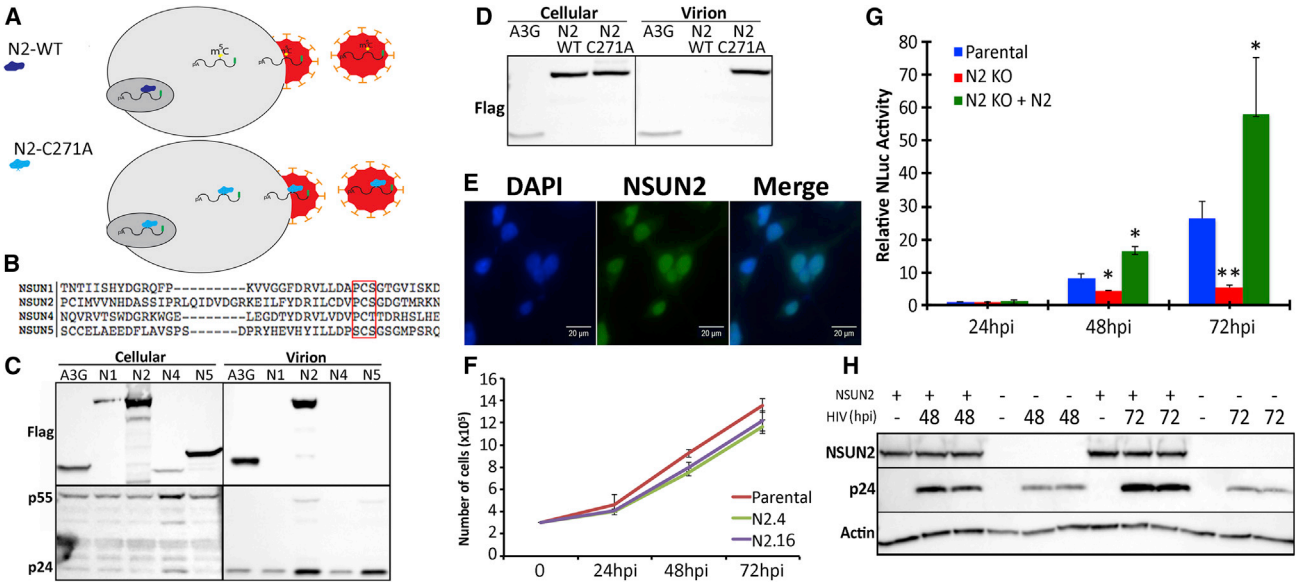

Figure 2. Identification of NSUN2 as the Primary m^{5}C Writer of HIV-1 RNAs

(A) Schematic showing the potential packaging of a covalently bound, mutant NSUN protein into HIV-1 virions along with the HIV-1 gRNA, while WT NSUN protein remains in the infected cell.

(B) Sequence alignment of a segment of the NSUN protein family, with the conserved motif involved in the detachment of covalently bound NSUN proteins from target RNAs boxed in red. A cysteine to alanine mutation in this motif results in NSUN proteins becoming covalently bound to target cytosines.

(C) At left, a western blot showing expression of FLAG-tagged A3G and FLAG-tagged mutants of NSUN1/NOP2 (N1), NSUN2 (N2), NSUN4 (N4), and NSUN5 (N5) in 293T cells co-transfected with pNL4-3. At right is a western blot showing the presence of A3G and N2-C271A in HIV-1 virions. n = 3.

(D) Western blots analyzing the intracellular expression of FLAG-tagged A3G, WT NSUN2 (N2) or N2-C271A in 293T cells co-transfected with HIV-1 (left) or in released virions (right). n = 3.

(E) IF localization of endogenous NSUN2 in 293T cells. Cells were stained with an NSUN2-specific antiserum, shown in green, as well as with DAPI, which stains DNA blue. n = 3.

(F) Growth curves for the two NSUN2 KO, 293T-derived clones, N2.4 and N2.16, showing comparable growth rates to WT 293T cells. n = 3 with SD indicated.

(G) WT 293T cells, NSUN2 KO cells, or NSUN2 KO cells transfected with a WT NSUN2 expression vector, were co-transfected with pCD4 and then infected with NL-NLuc. Induced NLuc levels were determined at 24, 48, and 72 hpi. Expression of virally encoded NLuc was inhibited by loss of NSUN2 but rescued by the NSUN2 expression plasmid. n = 3, *$p < 0.05$, **$p < 0.01$.

(H) Western blot from an experiment similar to (G) but using WT NL4-3. This experiment measured HIV-1 p24 Gag expression at 48 and 72 hpi in WT 293T cells and in the NSUN2 KO cell lines N2.4 and N2.16. Again, HIV-1 protein expression was reduced in the NSUN2 KO cells. n = 3.

in Figure 2H, which looks at WT or NSUN2 KO 293T cells transfected with a CD4 expression plasmid and then infected with WT NL4-3. Again, loss of NSUN2 reduced HIV-1 gene expression, as measured by analysis of HIV-1 Gag expression.

Spontaneous Crosslinking of NSUN2 C271A to Viral RNA

If NSUN2 is responsible for the addition of all m^{5}C modifications to HIV-1 RNAs, then a loss of NSUN2 expression should result in the loss of m^{5}C from the HIV-1 genome. Conversely, and as previously proposed (Hussain et al., 2013b), if the N2-C271A mutant spontaneously crosslinks to C residues that are normally converted to m^{5}C, then cross-linking and immunoprecipitation followed by sequencing (CLIP-seq) analysis using N2-C271A should identify the same residues identified by PA-m^{5}C-seq. As shown in Figure 3A, CLIP-seq using HIV-infected, CD4-expressing WT 293T cells overexpressing WT NSUN2 (upper band) or the N2-C271A mutant (second band) failed to detect a significant level of spontaneous crosslinking of WT NSUN2 to HIV-1 RNA but did detect site-specific crosslinking of the N2-C271A mutant to viral RNA. Importantly, PA-m^{5}C-seq performed in parallel using HIV-infected WT 293T cells (Figure 3A, band 3, and Figure S2C) shows that the m^{5}C sites detected by CLIP-seq using N2-C271A or using PA-m^{5}C-seq are indeed congruent (compare Figure 3A band 2 with band 3). Moreover, PA-m^{5}C-seq performed using HIV-infected NSUN2 KO cells (Figure 3, band 4) showed that almost all the m^{5}C sites detected using PA-m^{5}C-seq are lost in these cells, although, remarkably, three prominent m^{5}C sites are retained. These three m^{5}C residues are also notable in that they were not detected by N2-C271A CLIP-seq (compare Figure 3A row 2 with row 4). Therefore, we conclude that while NSUN2 accounts for the majority of the m^{5}C residues that are added to the HIV-1 RNA genome, these three m^{5}C residues are added by a different enzyme(s) that remains unidentified.

If multiple m^{5}C residues are indeed present on HIV-1 gRNAs, and given that crosslinking of N2-C271A to m^{5}C residues appears to be efficient, then N2-C271A expression and virion incorporation should inhibit HIV-1 replication early in the viral life cycle. Moreover, the existence of the NSUN2 KO 293T cell line raised the possibility of more fully analyzing how NSUN2 affects HIV-1 gene expression. We therefore transfected NSUN2 KO cells with a control vector, or vectors expressing WT NSUN2, the N2-C271A mutant or a second mutant, N2-C321A, in which the cysteine that becomes covalently bound to the target cytosine was mutated to alanine, which results in a recessive

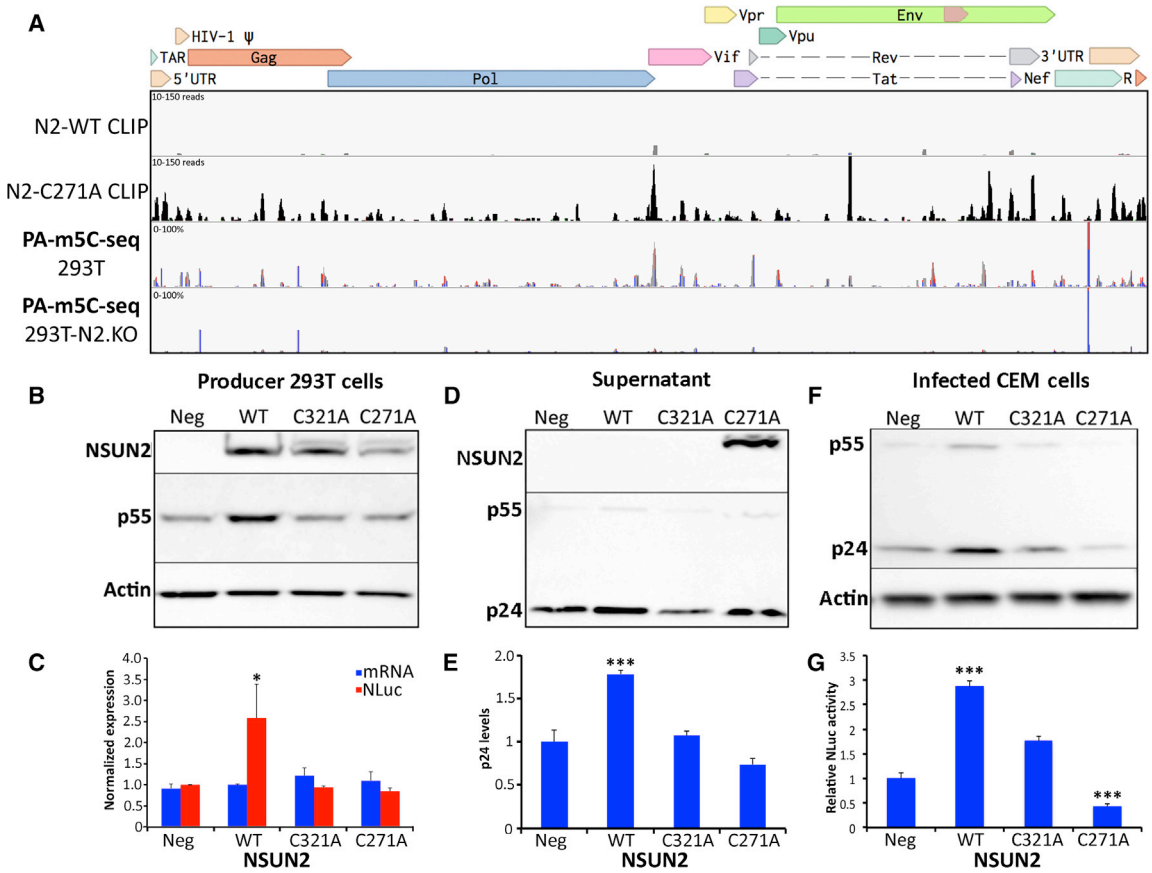

Figure 3. NSUN2-Specific m^{5}C Sites on HIV-1 Transcripts Upregulate HIV-1 Gene Expression

(A) The upper two images show mapping of NSUN2 binding sites on intracellular HIV-1 RNAs in infected 293T cells using CLIP-seq. FLAG-tagged WT NSUN2 was used as a control, while the covalently bound, FLAG-tagged N2-C271A mutant was used to identify NSUN2-binding sites. Lower two images map m^{5}C sites present on intracellular HIV-1 RNAs produced in HIV-infected WT 293T cells, or in the N2.4 KO cells, using PA-m^{5}C-seq.

(B) Western blot of NSUN2 or HIV-1 p55 Gag expression in N2.4 cells transfected with an empty vector, pN2 (WT NSUN2), pN2-C321A, or pN2-C271A, as well as with pNL-NLuc. Actin was used as a loading control. n = 3.

(C) In blue, qRT-PCR was used to quantify the level of HIV-1 RNA in the samples analyzed in (B). The probe used was specific for the LTR U3 region and values were normalized to cellular GAPDH mRNA. In red, NLuc activity from the cells shown in (B). n = 3 with SD indicated; *$p < 0.05$.

(D) Western blot of virions isolated from the supernatant media from the HIV-1 producer cells analyzed in (B), demonstrating packaging of N2-C271A. n = 3.

(E) ELISA was used to measure p24 levels in the supernatant of the HIV-1 producer cells analyzed in (B). n = 3 with SD indicated; ***$p < 0.001$.

(F) Western blot of intracellular p55 and p24 Gag expression levels in CEM cells infected with the NL-NLuc virus generated from the 293T cell cultures analyzed in (B)–(E). Virus was diluted to equal levels, as measured by p24 ELISA in (E), prior to infection. n = 3.

(G) Intracellular NLuc activity was determined using extracts from the same NL-NLuc-infected CEM cell cultures analyzed in (F). n = 3, ***$p < 0.001$.

negative phenotype. As shown in Figure 3B, transfection of NSUN2 KO cells with a plasmid encoding the replication-competent NL-NLuc indicator virus gave rise to a readily detectable level of Gag expression that was increased by co-expression of WT NSUN2, but not of either NSUN2 mutant. This effect is quantified in Figure 3C, which shows that expression of the virally encoded *NLuc* indicator gene was also significantly ($p < 0.05$) enhanced upon expression of WT NSUN2 in *trans*, while expression of either NSUN2 mutant had no effect. Importantly, analysis of HIV-1 RNA expression by qRT-PCR, using primers specific for the HIV-1 U3 region, failed to detect any difference in the level of viral RNA expressed in the presence or absence of NSUN2, even though the level of virally encoded protein expression, as measured using NLuc (Figure 3C) or HIV-1 Gag (Figure 3B), was increased in the presence of WT NSUN2. The increase in HIV-1 Gag expression seen in the presence of WT NSUN2 was quantified by the use of an ELISA to measure the level of secreted p24 Gag protein in the supernatant media harvested from the transfected cells analyzed in Figures 3B and 3C. These data (Figure 3E) revealed a significant increase in the level of p24 Gag release in the presence of WT, but not mutant, NSUN2 ($p < 0.001$). As expected, the N2-C271A mutant, but not WT NSUN2 or the N2-C321A mutant, was packaged into HIV-1 virions (Figure 3D).

Next, we asked if the presence of m^{5}C on HIV-1 gRNAs, or the crosslinking of N2-C271A to that RNA, would affect the ability of HIV-1 to infect the CD4+ T cell line CEM. For this purpose, the supernatant media harvested from the cultures analyzed in Figure 3E were normalized to identical levels of p24 Gag and then used to infect the CD4+ T cell line CEM. Infected cells were then analyzed by western blot for the *de novo* production of HIV-1 Gag (Figure 3F) and for NLuc expression (Figure 3G).

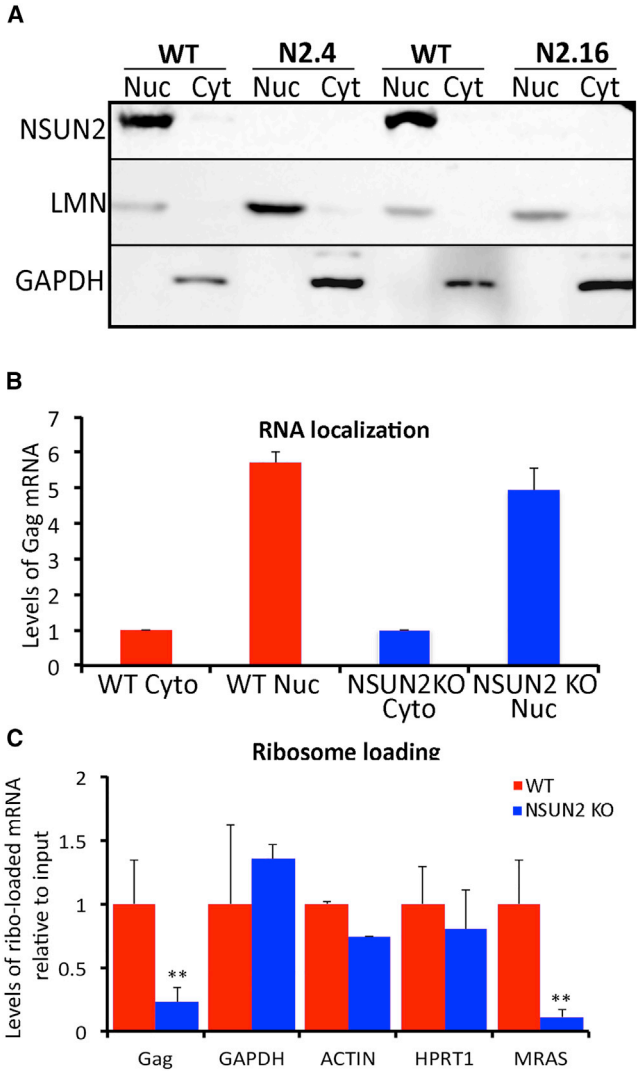

Figure 4. Loss of NSUN2 Affects the Translation of HIV-1 RNAs

(A) Western blot of WT or NSUN2 KO 293T cells separated into nuclear (Nuc) and cytoplasmic (Cyt) fractions 72 h after transfection with pNL4-3. Blots were probed for NSUN2, Lamin A/C and GAPDH. n = 2.

(B) qRT-PCR was used to measure the level of HIV-1 gag mRNA present in the cytoplasmic and nuclear fractions isolated from the pNL4-3 transfected WT or NSUN2 KO 293T cells analyzed in (A). GAPDH mRNA was measured in parallel and used to normalize the HIV-1 RNA data. The relative gag mRNA level in the cytoplasm was set at 1. n = 4 with SD indicated.

(C) WT or NSUN2 KO cells were transfected with pNL4-3 and used to analyze both the total and ribosome associated levels of cellular GAPDH, actin, HPRT1, and MRAS mRNA, as well as HIV-1 gag RNA, at 72 h after infection, with the average level detected in WT cells set at 1. n = 3 with SD indicated; $^{**}p < 0.01$.

Surprisingly, the HIV-1 virus produced in the presence of WT NSUN2 gave rise to significantly more expression of both HIV-1 Gag and NLuc than did virus produced in the absence of NSUN2 or in the presence of the recessive mutant N2-C321A (Figures 3F and 3G). In contrast, and as predicted, expression of N2-C271A in *trans* inhibited infection of CEM T cells, as measured either by expression of HIV-1 Gag protein or of the virally encoded *NLuc* indicator gene (Figures 3F and 3G), and this effect was significant ($p < 0.001$). While the presence of crosslinked N2-C271A on the HIV-1 gRNA would be predicted to inhibit HIV-1 reverse transcription, it is unclear why higher levels of m^5C should exert a significant positive effect in this assay. Possibilities include less efficient reverse transcription, perhaps due to reduced methylation of the tRNA primer, or defects in some aspect of the viral assembly when m^5C is absent from the viral gRNA.

Data presented in Figure 3C demonstrate that the loss of NSUN2 expression results in reduced expression of HIV-1 Gag even though expression of gag mRNA is unaffected. We considered two possible explanations for this effect. One possibility is that unspliced gag mRNAs are not effectively exported from the nucleus to the cytoplasm in the absence of m^5C residues, and it has indeed been proposed that m^5C residues promote mRNA export (Yang et al., 2017). We therefore transfected the WT or NSUN2 KO 293T cells with pNL4-3 and separated the cells into nuclear and cytoplasmic fractions at 72 h after transfection (Malim et al., 1989). We then performed western blot analyses of these fractions for NSUN2, which should be nuclear in WT cells, for lamin A/C, which should be nuclear in all cells, and for GAPDH, which should be cytoplasmic. As shown in Figure 4A, this is indeed the result obtained, thus confirming the integrity of the subcellular fractionation procedure. The level of gag mRNA in these fractions was then determined by qRT-PCR and normalized to GAPDH mRNA quantified in parallel. This analysis demonstrated that the level of gag mRNA in both the nuclear and cytoplasmic fractions was unchanged in the presence and absence of NSUN2 (Figure 4B).

If cytoplasmic viral mRNA levels are constant in the absence of NSUN2, yet less viral protein is produced, then this implies reduced translation. This in turn should be reflected in less efficient loading of ribosomes onto viral mRNAs. To determine whether this is the case, we used a published procedure (Subtelny et al., 2014) to measure the ribosome association of mRNAs in the presence or absence of NSUN2. WT or NSUN2 KO cells transfected with pNL4-3 were treated with cycloheximide to block translation elongation. The cells were then lysed, polysomes sheared by passage through a syringe needle and the ribosomes, together with any bound RNA fragments, then separated from free mRNAs by pelleting through a sucrose cushion. The relative level of ribosome-associated HIV-1 gag mRNA, as determined by qRT-PCR, decreased significantly ($p < 0.01$) in the absence of NSUN2 and, hence, of most m^5C modifications. As controls, we also measured the ribosome loading of cellular GAPDH, actin, and HPRT mRNAs, which lack m^5C sites and therefore should be unaffected by loss of NSUN2 expression, and of cellular MRAS mRNA, which contains 4 m^5C sites and therefore might be expected to also undergo reduced translation in the absence of NSUN2 (Figure 1I). Indeed, and as predicted, loss of NSUN2 had no effect on ribosome loading onto the three mRNAs that lack m^5C sites, while a significant reduction in ribosome loading onto the normally highly m^5C modified MRAS mRNA was detected (Figure 4C). These results confirm that the reduced ribosome association observed with HIV-1 Gag mRNA in the absence of NSUN2 is specific to m^5C modified mRNAs and not due to a global reduction in ribosome loading onto all mRNAs.

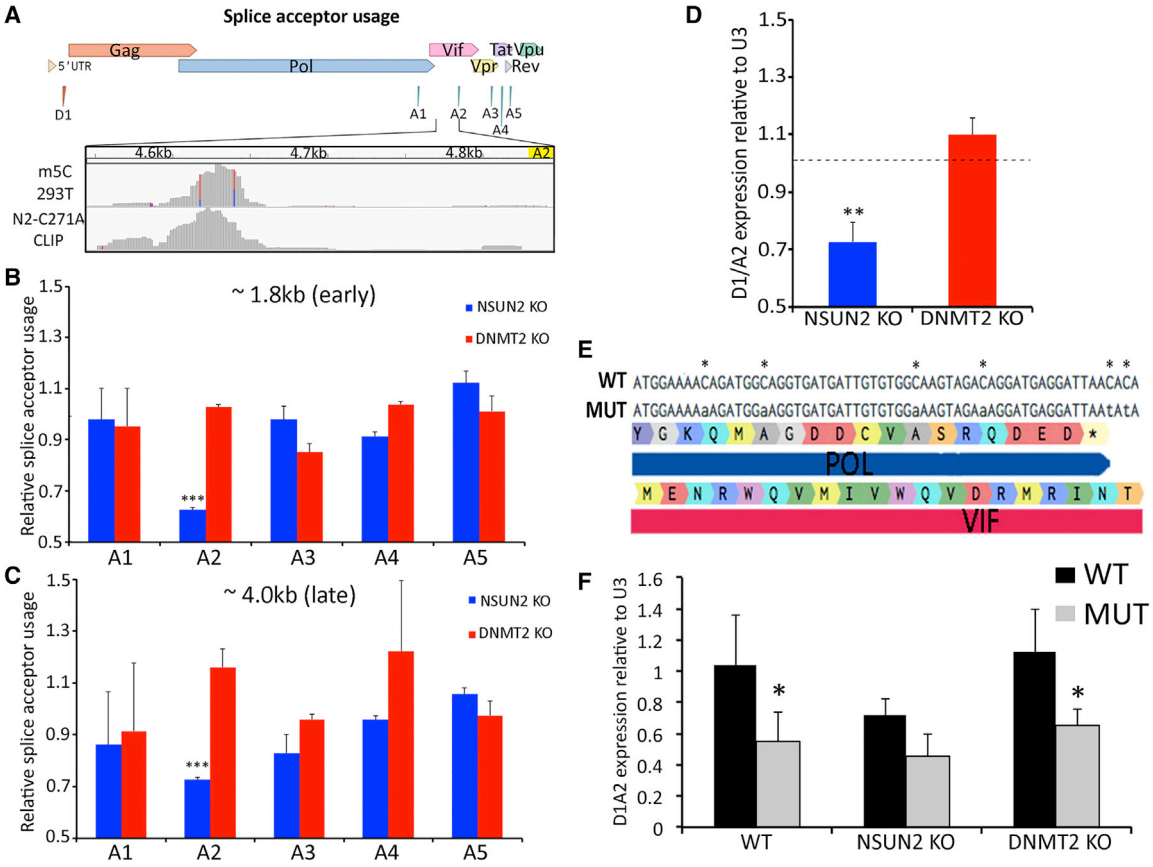

Figure 5. NSUN2 Loss Affects Alternative Splicing of HIV-1 RNAs

(A) Schematic showing the location of the HIV-1 major 5′ splice site D1 and the five 3′ splice sites A1 through A5. PA-m^{5}C-seq and N2-C271A-binding sites located adjacent to the A2 splice site are also shown.

(B) Assessment of 3′ splice site usage from the D1 5′ splice site in completely spliced, ~1.8 kb HIV-1 RNAs harvested from NSUN2 or DNMT2 293T KO cell lines, normalized to WT 293T cells, which was set at 1.0. n = 3, ***$p < 0.001$.

(C) Assessment of 3′ splice site usage from the D1 5′ splice site in incompletely spliced, ~4.0 kb HIV-1 RNAs harvested from NSUN2 or DNMT2 KO 293T cell lines, normalized to WT 293T cells, which was set at 1.0. n = 3, ***$p < 0.001$.

(D) qRT-PCR analysis showing usage of the D1/A2 splice junction relative to total viral RNA levels, measured with a U3-specific probe, using RNA isolated from transfected NSUN2 or DNMT2 293T KO cell lines. Data were normalized to WT 293T cells, which was set at 1.0 (dotted line). n = 3 with SD indicated, **$p < 0.01$.

(E) Nucleotide and protein sequences underlying the m^{5}C peak shown in (D), with mutated C residues indicated.

(F) qRT-PCR quantification of D1/A2 splice site usage in WT, NSUN2, or DNMT2 KO 293T cells transfected with WT pNL4-3 or with the HΔm^{5}C1 proviral mutant. n = 3 with SD indicated, *$p < 0.05$. For the experiments shown in this figure, both the N2.4 and N2.16 NSUN2 KO cell lines and the D2.1 and D2.2 DNMT2 KO cell lines were used.

Loss of NSUN2 Affects HIV-1 Alternative Splicing

HIV-1 RNAs are extensively alternatively spliced, and we therefore wondered if m^{5}C modifications affect this process. We used a previously described method that relies on Primer-ID-tagged deep sequencing to precisely quantify the level of splicing from the major HIV-1 5′ splice site, designated D1 in Figure 5A, to all five potential 3′ splice sites located near the center of the HIV-1 RNA genome, designated A1 to A5 (Emery et al., 2017). In these experiments, we separately quantified the level of splicing from D1 to A1 through A5 in viral mRNAs belonging to the early, ~1.8 kb class of HIV-1 transcripts (Figure 5B) or late, ~4 kb class of transcripts (Figure 5C). We analyzed HIV-1 RNA splicing in WT 293T cells, in the NSUN2 KO cell lines N2.4 and N2.16, and in a third 293T cell line in which the tRNA-specific cytosine methyltransferase DNMT2 had been knocked out by gene editing, as an additional control (Figure S3D) (Bohnsack et al., 2019; Khoddami and Cairns, 2013). All data presented in Figures 5B and 5C are normalized to WT 293T cells with the raw data given in Table S1. As may be observed, we detected a significant ($p < 0.01$) reduction in the utilization of the A2 3′ splice site in the absence of NSUN2, while utilization of the other four 3′ splice sites was not significantly affected (although a trend toward higher utilization of A5 was noted). In contrast, loss of the DNMT2 methyltransferase had no significant effect (Figures 5B and 5C).

To confirm this result, we used qRT-PCR to quantify both the total level of HIV-1 RNA and of viral transcripts bearing the A1 to D2 splice junction, using a junction-specific probe, in WT, NSUN2 KO, or DNMT2 KO 293T cells transfected with pNL4-3. These data (Figure 5D), which were again normalized to WT 293T cells, again revealed a significant ($p < 0.01$) decrease in the utilization of the D1/A2 splice junction in the NSUN2 KO, but not DNMT2 KO, cell lines.

We noted an m^5C site, detected by both PA-m^5C-seq and N2-C271A CLIP-seq, located ~150 nt 5′ to the A2 splice site (Figure 5A), that might be functionally relevant. This site lies in the overlap region between the carboxy terminus of Pol and the amino terminus of Vif and silent mutagenesis was not possible (Figure 5E). However, as neither Pol nor Vif affects gene expression from a transfected HIV-1 provirus, it was nevertheless feasible to examine whether this m^5C modification(s) affected splicing in *cis*. We mutated all six "C" residues under the observed m^5C peak, to generate the HΔm^5C1 mutant (Figure 5E), and then transfected WT or NSUN2 KO cells with WT or mutant pNL4-3. At 72 h post-transfection, total RNA was harvested and utilization of the D1/A2 splice junction quantified by qRT-PCR. If the A2-proximal m^5C site shown in Figure 5A indeed regulates utilization of this splice site, then the HΔm^5C1 mutant should phenocopy the loss of NSUN2 expression and a significant ($p < 0.05$) reduction in D1/A2 splice site usage was indeed observed in both WT and DNMT2 KO cells transfected with the HΔm^5C1 provirus (Figure 5F). In contrast, if the A2-proximal m^5C site fully accounts for the observed effect on splicing to A2, then loss of NSUN2 expression should have no additional effect on the utilization of the D1/A2 splice junction by the HΔm^5C1 mutant. In fact, we observed a modest reduction in utilization of the D1/A2 splice junction when the HΔm^5C1 mutant was transfected into NSUN2 KO cells, when compared to WT cells, though this effect was no longer statistically significant ($p = 0.12$). Overall, these data therefore confirm that the m^5C site indicated in Figure 5A is indeed promoting utilization of the D1/A2 splice junction.

DISCUSSION

Viruses, particularly positive-sense RNA viruses, offer several potential advantages as a model system to study the phenotypic effects and mechanisms of action of epitranscriptomic mRNA modifications. Firstly, these viruses package their RNA genome, which is also a functional mRNA, into virion particles that are released from infected cells. As shown here, this facilitates the purification of that mRNA and, hence, the quantitative analysis of the epitranscriptomic modifications present on that RNA. In contrast, the purification of any single cellular mRNA to homogeneity is difficult if not impossible. Secondly, viruses such as HIV-1 are constantly under intense selective pressure so that, as noted by Coffin, "the influence of small selective forces in populations (of HIV-1) is so strong that no mutation can be assumed to be truly neutral" (Coffin, 1995). Therefore, if a given epitranscriptomic modification confers a selective advantage, such as enhanced mRNA translation, viral mRNAs should have evolved to acquire a level of that epitranscriptomic modification that maximizes this advantage. As a result, one would predict that epitranscriptomic modifications that facilitate some aspect of RNA function would accrue to a higher level on viral mRNAs than on the average cellular mRNA, which is not subject to the same selective pressures. Finally, because viruses should seek to maximize the advantage conferred by a given epitranscriptomic modification, the phenotypic effect of that modification should be close to maximal, thus facilitating the mechanistic dissection of that phenotype.

In this manuscript, we have sought to leverage these predictions as a means to begin to identify the epitranscriptomic modifications that regulate HIV-1 replication. Our first goal was to isolate highly purified HIV-1 gRNA from virions and to then precisely define the level of the epitranscriptomic modifications present on this RNA. These data, as predicted above, revealed that several epitranscriptomic modifications are present at a substantially greater level on HIV-1 gRNAs than seen in total cellular poly(A)+ RNA (Figure 1C; Table 1). Strikingly, this result was unchanged when we analyzed viral gRNA isolated from virions released either by infected CEM T cells or transfected 293T cells, which argues that this high level of epitranscriptomic modification is intrinsic to the viral RNA analyzed and only minimally affected by differences in the viral producer cells used. The biggest differences were seen with 2′O-Me modified bases, which were present at ~10× higher levels in the HIV-1 gRNA, and the m^5C modification, which was ~14× more prevalent. Clearly, these large differences are unlikely to occur by chance and the 2′O-Me modification has indeed been recently reported to facilitate HIV-1 replication by inhibiting innate antiviral immune responses (Ringeard et al., 2019). Here, we have focused on m^5C, another highly prevalent modification detected on HIV-1 gRNAs (Figure 1C).

While the UPLC-MS/MS data (Figure 1C; Table 1) indicate that HIV-1 gRNAs contain ~11 m^5C moieties, the mapping of m^5C residues in CEMs, performed using the PA-m^5C-seq technique in duplicate (Figures 1E and S2A) identified >19 m^5C editing sites. In addition, 18 of these m^5C modified sites were also mapped in infected primary CD4+ T cells (Figure S2B). This not only indicates that many m^5C sites on the HIV-1 gRNA are only partially modified but also implies that it would be difficult to use mutagenesis to define their functional significance, given that many of these m^5C modifications are in open reading frames or in other areas of the HIV-1 genome with known functions. Therefore, we sought to define the factor that adds or "writes" these m^5C residues as an alternative means to discover their functional significance.

Previous work has shown that mutagenesis of a key cysteine residue in members of the NSUN family of cytosine methyltransferases results in the crosslinking of the protein to target cytosines (Hussain et al., 2013b), and we were able to show that when this mutation (C271A) was introduced, then NSUN2 was not only packaged into HIV-1 virions (Figures 2C and 2D) but also crosslinked to specific sites on the HIV-1 RNA genome that coincide with m^5C sites mapped by PA-m^5C-seq (Figure 3A). As expected, if NSUN2 is indeed responsible for m^5C addition to HIV-1 transcripts, KO of the *NSUN2* gene using CRISPR-Cas resulted in the loss of the vast majority of m^5C residues detected on the HIV-1 gRNA, though interestingly, three m^5C sites were retained and therefore must be added by a distinct, currently unknown, methyltransferase (Figure 3A).

Analysis of HIV-1 replication and gene expression revealed that loss of NSUN2 function inhibited HIV-1 Gag protein expression without reducing gag RNA expression (Figure 3C). Because m^5C has been suggested to enhance nuclear mRNA export (Yang et al., 2017) and to enhance mRNA translation (Li et al., 2017; Xing et al., 2015), we analyzed the subcellular distribution of the HIV-1 gag RNA (Figure 4B), and the level of ribosome association of that RNA (Figure 4C), in the presence and absence of NSUN2. These data indicated that while the HIV-1 gag mRNA is efficiently exported from the nucleus regardless of NSUN2 expression, and hence of m^5C addition to that RNA, the exported

gag mRNA bound ribosomes less effectively in the absence of NSUN2. Moreover, our data also show that ribosome recruitment to the cellular mRNA encoding MRAS, which is highly m^5C modified, is also inhibited when NSUN2 is lost, while the mRNAs encoding three cellular housekeeping proteins, GAPDH, HPRT, and actin, that lack m^5C sites are unaffected. Therefore, these data clearly demonstrate a role for m^5C in promoting the translation of not only viral but also some cellular mRNAs.

Because some epitranscriptomic modifications affect RNA splicing (Xiao et al., 2016), and HIV-1 RNAs are alternatively spliced (Emery et al., 2017), we also examined whether loss of NSUN2 function would affect this process. Surprisingly, RNA-seq and qRT-PCR data revealed a specific reduction in the utilization of the A2 splice acceptor when NSUN2 expression was blocked (Figures 4E–4G). Because our mapping data had identified an m^5C site located immediately 5′ to the A2 splice site (Figure 5A), we hypothesized that loss of this m^5C site might also affect the utilization of the D1/A2 splice junction. In fact, mutagenesis of this m^5C site again resulted in a reduction in the utilization of the D1/A2 splice junction, even in the presence of NSUN2 (Figure 5F). Therefore, these data reveal that addition of m^5C to HIV-1 RNAs can also affect the alternative splicing of HIV-1 transcripts.

STAR★METHODS

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

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - HIV-1 Infections
  - HIV-1 Virion Purification
  - RNA Modification Identification by UPLC-MS/MS
  - Purification of CD4+ T Cells from PBMCs
  - Construction of Expression Plasmids
  - Virion Packaging of Mutant NSUN Proteins
  - Western Blot Analyses
  - Immunofluorescence
  - Generation of Gene Knock Out 293T Cells by Genome Editing
  - HIV-1 Infection of NSUN2 Knockout 293T Cells
  - NSUN2-CLIP-seq
  - NSUN2 Mutants and HIV-1 Protein Expression
  - Quantification of HIV-1 RNA Levels Using qRT-PCR
  - Nuclear and Cytoplasmic Fractionation
  - Quantification of Ribosome Loading on to HIV-1 Transcripts
  - Characterization of HIV-1 RNA Splicing
  - Bioinformatic Analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.chom.2019.07.005.

ACKNOWLEDGMENTS

This research was funded in part by NIH grants R01-DA046111 to B.R.C. and U54-GM103297 to B.R.C. and R.S. This research received infrastructure support from the Duke University Center for AIDS Research (CFAR) (P30-AI064518), the UNC CFAR (P30-AI50410), and the UNC Lineberger Comprehensive Cancer Center (P30-CA16068). D.G.C. was funded by Marie Skłodowska-Curie Global Fellowship MSCA-IF-GF:747810.

AUTHOR CONTRIBUTIONS

Conceptualization, D.G.C., E.M.K., and B.R.C.; Methodology, D.G.C., H.P.B, K.T., E.M.K., B.A.L., A.E., and B.R.C.; Formal Analysis, D.G.C., K.T., E.M.K., and A.E.; Investigation, D.G.C., H.P.B., K.T., E.M.K., B.A.L., and A.E.; Writing – Original Draft, D.G.C., and B.R.C.; Writing – Review & Editing, D.G.C., K.T., H.P.B., E.M.K., R.S., C.L.H., and B.R.C.; Supervision, R.S., C.L.H, and B.R.C.; Funding Acquisition, D.G.C., R.S., and B.R.C.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 26, 2019
Revised: May 30, 2019
Accepted: July 16, 2019
Published: August 14, 2019

REFERENCES

Arango, D., Sturgill, D., Alhusaini, N., Dillman, A.A., Sweet, T.J., Hanson, G., Hosogane, M., Sinclair, W.R., Nanan, K.K., Mandler, M.D., et al. (2018). Acetylation of cytidine in mRNA promotes translation efficiency. Cell *175*, 1872–1886.e24.

Basanta-Sanchez, M., Temple, S., Ansari, S.A., D'Amico, A., and Agris, P.F. (2016). Attomole quantification and global profile of RNA modifications: epitranscriptome of human neural stem cells. Nucleic Acids Res. *44*, e26.

Bieniasz, P.D., Fridell, R.A., Aramori, I., Ferguson, S.S., Caron, M.G., and Cullen, B.R. (1997). HIV-1-induced cell fusion is mediated by multiple regions within both the viral envelope and the CCR-5 co-receptor. EMBO J *16*, 2599–2609.

Bohnsack, K.E., Hobartner, C., and Bohnsack, M.T. (2019). Eukaryotic 5-methylcytosine (m(5)C) RNA methyltransferases: mechanisms, cellular functions, and links to disease. Genes (Basel) *10*, E102.

Chen, K., Lu, Z., Wang, X., Fu, Y., Luo, G.Z., Liu, N., Han, D., Dominissini, D., Dai, Q., Pan, T., et al. (2015). High-resolution N(6) -methyladenosine (m(6) A) map using photo-crosslinking-assisted m(6) A sequencing. Angew. Chem. Int. Ed. Engl. *54*, 1587–1590.

Coffin, J.M. (1995). HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. Science *267*, 483–489.

Courtney, D.G., Kennedy, E.M., Dumm, R.E., Bogerd, H.P., Tsai, K., Heaton, N.S., and Cullen, B.R. (2017). Epitranscriptomic enhancement of influenza A virus gene expression and replication. Cell Host Microbe *22*, 377–386.e5.

Doehle, B.P., Schäfer, A., and Cullen, B.R. (2005). Human APOBEC3B is a potent inhibitor of HIV-1 infectivity and is resistant to HIV-1 Vif. Virology *339*, 281–288.

Dominissini, D., Moshitch-Moshkovitz, S., Salmon-Divon, M., Amariglio, N., and Rechavi, G. (2013). Transcriptome-wide mapping of N(6)-methyladenosine by m(6)A-seq based on immunocapturing and massively parallel sequencing. Nat. Protoc *8*, 176–189.

Dominissini, D., Nachtergaele, S., Moshitch-Moshkovitz, S., Peer, E., Kol, N., Ben-Haim, M.S., Dai, Q., Di Segni, A., Salmon-Divon, M., Clark, W.C., et al. (2016). The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. Nature *530*, 441–446.

Doria, M., Neri, F., Gallo, A., Farace, M.G., and Michienzi, A. (2009). Editing of HIV-1 RNA by the double-stranded RNA deaminase ADAR1 stimulates viral infection. Nucleic Acids Res. *37*, 5848–5858.

Eckwahl, M.J., Arnion, H., Kharytonchyk, S., Zang, T., Bieniasz, P.D., Telesnitsky, A., and Wolin, S.L. (2016). Analysis of the human immunodeficiency virus-1 RNA packageome. RNA *22*, 1228–1238.

Emery, A., Zhou, S., Pollom, E., and Swanstrom, R. (2017). Characterizing HIV-1 splicing by using next-generation sequencing. J. Virol. *91*, e02515–e02516.

Gilbert, W.V., Bell, T.A., and Schaening, C. (2016). Messenger RNA modifications: form, distribution, and function. Science *352*, 1408–1412.

Harris, R.S., Bishop, K.N., Sheehy, A.M., Craig, H.M., Petersen-Mahrt, S.K., Watt, I.N., Neuberger, M.S., and Malim, M.H. (2003). DNA deamination mediates innate immunity to retroviral infection. Cell *113*, 803–809.

Helm, M., and Motorin, Y. (2017). Detecting RNA modifications in the epitranscriptome: predict and validate. Nat. Rev. Genet. *18*, 275–291.

Hsu, P.J., Zhu, Y., Ma, H., Guo, Y., Shi, X., Liu, Y., Qi, M., Lu, Z., Shi, H., Wang, J., et al. (2017). Ythdc2 is an N(6)-methyladenosine binding protein that regulates mammalian spermatogenesis. Cell Res. *27*, 1115–1127.

Hussain, S., Aleksic, J., Blanco, S., Dietmann, S., and Frye, M. (2013a). Characterizing 5-methylcytosine in the mammalian epitranscriptome. Genome Biol. *14*, 215.

Hussain, S., Sajini, A.A., Blanco, S., Dietmann, S., Lombard, P., Sugimoto, Y., Paramor, M., Gleeson, J.G., Odom, D.T., Ule, J., et al. (2013b). NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. Cell Rep *4*, 255–261.

Ke, S., Pandya-Jones, A., Saito, Y., Fak, J.J., Vågbø, C.B., Geula, S., Hanna, J.H., Black, D.L., Darnell, J.E., Jr., and Darnell, R.B. (2017). m6A mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover. Genes Dev. *31*, 990–1006.

Kennedy, E.M., Bogerd, H.P., Kornepati, A.V., Kang, D., Ghoshal, D., Marshall, J.B., Poling, B.C., Tsai, K., Gokhale, N.S., Horner, S.M., et al. (2016). Posttranscriptional m(6)A editing of HIV-1 mRNAs enhances viral gene expression. Cell Host Microbe *19*, 675–685.

Kharytonchyk, S., Monti, S., Smaldino, P.J., Van, V., Bolden, N.C., Brown, J.D., Russo, E., Swanson, C., Shuey, A., Telesnitsky, A., et al. (2016). Transcriptional start site heterogeneity modulates the structure and function of the HIV-1 genome. Proc. Natl. Acad. Sci. USA *113*, 13378–13383.

Khoddami, V., and Cairns, B.R. (2013). Identification of direct targets and modified bases of RNA cytosine methyltransferases. Nat. Biotechnol. *31*, 458–464.

Khoddami, V., Yerra, A., and Cairns, B.R. (2015). Experimental approaches for target profiling of RNA cytosine methyltransferases. Methods Enzymol. *560*, 273–296.

King, M.Y., and Redman, K.L. (2002). RNA methyltransferases utilize two cysteine residues in the formation of 5-methylcytosine. Biochemistry *41*, 11218–11225.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

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

Li, Q., Li, X., Tang, H., Jiang, B., Dou, Y., Gorospe, M., and Wang, W. (2017). NSUN2-mediated m5C methylation and METTL3/METTL14-Mediated m6A methylation cooperatively enhance p21 translation. J. Cell. Biochem. *118*, 2587–2598.

Li, S., and Mason, C.E. (2014). The pivotal regulatory landscape of RNA modifications. Annu. Rev. Genomics Hum. Genet. *15*, 127–150.

Li, X., Xiong, X., and Yi, C. (2016). Epitranscriptome sequencing technologies: decoding RNA modifications. Nat. Methods *14*, 23–31.

Lichinchi, G., Gao, S., Saletore, Y., Gonzalez, G.M., Bansal, V., Wang, Y., Mason, C.E., and Rana, T.M. (2016). Dynamics of the human and viral m6A RNA methylomes during HIV-1 infection of T cells. Nat. Microbiol. *1*, 16011.

Malim, M.H., Hauber, J., Le, S.Y., Maizel, J.V., and Cullen, B.R. (1989). The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. Nature *338*, 254–257.

Mefferd, A.L., Bogerd, H.P., Irwan, I.D., and Cullen, B.R. (2018). Insights into the mechanisms underlying the inactivation of HIV-1 proviruses by CRISPR/Cas. Virology *520*, 116–126.

Meyer, K.D., and Jaffrey, S.R. (2014). The dynamic epitranscriptome: N6-methyladenosine and gene expression control. Nat. Rev. Mol. Cell Biol. *15*, 313–326.

Morgenstern, J.P., and Land, H. (1990). Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. Nucleic Acids Res. *18*, 3587–3596.

Olarerin-George, A.O., and Jaffrey, S.R. (2017). MetaPlotR: a Perl/R pipeline for plotting metagenes of nucleotide modifications and other transcriptomic sites. Bioinformatics *33*, 1563–1564.

Qin, Y., Yao, J., Wu, D.C., Nottingham, R.M., Mohr, S., Hunicke-Smith, S., and Lambowitz, A.M. (2016). High-throughput sequencing of human plasma RNA by using thermostable group II intron reverse transcriptases. RNA *22*, 111–128.

Reid, R., Greene, P.J., and Santi, D.V. (1999). Exposition of a family of RNA m(5)C methyltransferases from searching genomic and proteomic sequences. Nucleic Acids Res. *27*, 3138–3145.

Ringeard, M., Marchand, V., Decroly, E., Motorin, Y., and Bennasser, Y. (2019). FTSJ3 is an RNA 2′-O-methyltransferase recruited by HIV to avoid innate immune sensing. Nature *565*, 500–504.

Roundtree, I.A., Evans, M.E., Pan, T., and He, C. (2017). Dynamic RNA modifications in gene expression regulation. Cell *169*, 1187–1200.

Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods *11*, 783–784.

Shi, H., Wang, X., Lu, Z., Zhao, B.S., Ma, H., Hsu, P.J., Liu, C., and He, C. (2017). YTHDF3 facilitates translation and decay of N(6)-Methyladenosine-modified RNA. Cell Res. *27*, 315–328.

Squires, J.E., Patel, H.R., Nousch, M., Sibbritt, T., Humphreys, D.T., Parker, B.J., Suter, C.M., and Preiss, T. (2012). Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. Nucleic Acids Res. *40*, 5023–5033.

Subtelny, A.O., Eichhorn, S.W., Chen, G.R., Sive, H., and Bartel, D.P. (2014). Poly(A)-tail profiling reveals an embryonic switch in translational control. Nature *508*, 66–71.

Wang, X., Lu, Z., Gomez, A., Hon, G.C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G., et al. (2014). N6-methyladenosine-dependent regulation of messenger RNA stability. Nature *505*, 117–120.

Wang, X., Zhao, B.S., Roundtree, I.A., Lu, Z., Han, D., Ma, H., Weng, X., Chen, K., Shi, H., and He, C. (2015). N(6)-methyladenosine modulates messenger RNA translation efficiency. Cell *161*, 1388–1399.

Whisnant, A.W., Bogerd, H.P., Flores, O., Ho, P., Powers, J.G., Sharova, N., Stevenson, M., Chen, C.H., and Cullen, B.R. (2013). In-depth analysis of the interaction of HIV-1 with cellular microRNA biogenesis and effector mechanisms. MBio *4*, e000193.

Xiao, W., Adhikari, S., Dahal, U., Chen, Y.S., Hao, Y.J., Sun, B.F., Sun, H.Y., Li, A., Ping, X.L., Lai, W.Y., et al. (2016). Nuclear m(6)A reader YTHDC1 regulates mRNA splicing. Mol. Cell *61*, 507–519.

Xing, J., Yi, J., Cai, X., Tang, H., Liu, Z., Zhang, X., Martindale, J.L., Yang, X., Jiang, B., Gorospe, M., et al. (2015). NSun2 promotes cell growth via elevating cyclin-dependent kinase 1 translation. Mol. Cell. Biol. *35*, 4043–4052.

Xu, C., Wang, X., Liu, K., Roundtree, I.A., Tempel, W., Li, Y., Lu, Z., He, C., and Min, J. (2014). Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. Nat. Chem. Biol. *10*, 927–929.

Yang, X., Yang, Y., Sun, B.F., Chen, Y.S., Xu, J.W., Lai, W.Y., Li, A., Wang, X., Bhattarai, D.P., Xiao, W., et al. (2017). 5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an m5C reader. Cell Res. *27*, 606–625.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-FLAG	Sigma	Cat#F1804
Anti-m6A	Synaptic Systems	Cat#202111
Anti-m^5C	Diagenode	Cat#C15200081
Anti-NSUN2	Proteintech	Cat#20854-1-AP
Anti-Actin	Proteintech	Cat#60008-1-Ig
Anti-Gag	NIH AIDS Reagent	Cat#6458
Anti-Mouse HRP (1:10000)	Sigma	Cat#A9044
Anti-Rabbit HRP (1:10000)	Sigma	Cat#A6154
Alexa Fluor 488 Anti-mouse	Thermo	Cat#A11029
Alexa Fluor 488 Anti-rabbit	Thermo	Cat#A11034
Bacterial and Virus Strains		
HIV-NL4 3	NIH AIDS Reagent	Cat#114
HIV-NL4 3 Nluc	Mefferd et al., 2018	N/A
DH5α	NEB	Cat#C2988J
Chemicals, Peptides and Recombinant Proteins		
PEI	Polysciences	Cat#23966-2
Puromycin	Gemini Bio-Products	Cat#400-128P
4-thiouridine (4SU)	Carbosynth	Cat#NT06186
16% Paraformaldehyde	Thermo	Cat#28908
Laemmli Buffer	Cold Spring Harbour	http://cshprotocols.cshlp.org/
WesternBright ECL kit	Advansta	Cat#K-12045-D50
Lipofectamine RNAiMAX	Thermo	Cat#13778150
Optiprep	Sigma	Cat#D1556
Nuclease P1	Sigma	Cat#N8630
Vectashield mounting media	Vector Labs	Cat#H-1200
Critical Commercial Assays		
TruSeq Small RNA Sample Preparation Kit	Illumina	Cat#15016911
SuperScript III	Invitrogen	Cat#18080-044
GoTaq green PCR master mix	Promega	Cat#M7123
Power Sybr Green PCR Mastermix	Applied Biosystems	Cat#4367659
PrimeTime® Gene Expression Master Mix	IDT	Cat#1055770
Pierce™ Streptavidin Magnetic Beads	Thermo	Cat#88816
siNSUN2	Origene	Cat#SR310319
Nano luciferase assay kit	Promega	Cat#N1110
Luciferase assay system	Promega	Cat#E1500
Ribo-Zero Gold rRNA Removal Kit	Illumina	Cat#MRZG126
SMARTer® Stranded Total RNA-Seq Kit v2 - Pico Input Mammalian	Takara Bio	Cat#634411
HIV-1 p24 Antigen capture ELISA	ABL	Cat#5421
Deposited Data		
PA-m^6A-seq	This study	GEO#GSE130972
PA-m^5C-seq	This study	GEO#GSE130972
NSUN2-CLIP	This study	GEO#GSE130972

(*Continued on next page*)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Models: Cell Lines		
HEK293T	ATCC	Cat#CRL-11268
CEM	ATCC	Cat#CCL-119
HeLa	ATCC	Cat#CCL-2
HEK293T-N2.4	This Study	N/A
HEK293T-N2.16	This Study	N/A
HEK293T-D2.1	This Study	N/A
HEK293T-D2.2	This Study	N/A
HEK293T-CD4	This Study	N/A
Oligonucleotides		
See Table S2	This study	N/A
Recombinant DNA		
lentiCRISPR v2	Sanjana et al., 2014	Addgene; 52961
ΔCR8.74	unpublished	Addgene; 22036
pMD2.G	unpublished	Addgene; 12259
pcDNA3 NSUN1.CA-Flag	This Study	N/A
pcDNA3 NSUN2.C271A-Flag	This Study	N/A
pcDNA3 NSUN4.CA-Flag	This Study	N/A
pcDNA3 NSUN5.CA-Flag	This Study	N/A
pcDNA3 APOBEC3G-Flag	Doehle et al., 2005	N/A
pcDNA3 NSUN2-Flag	This Study	N/A
pcDNA3 NSUN2.C321A-Flag	This Study	N/A
pBabe-CD4	This Study	N/A
pBS-CMV-gagpol	unpublished	Addgene; 35614
Software and Algorithms		
Bowtie	Langmead et al., 2009	http://bowtie-bio.sourceforge.net/index.shtml
Fastx Toolkit	Cold Spring Harbor	http://hannonlab.cshl.edu/fastx_toolkit/index.html
Samtools	Li et al., 2009	http://samtools.sourceforge.net
IGV	Broad Institute	http://software.broadinstitute.org/software/igv/
Other		
Tris-Glycine SDS PAGE gel	Invitrogen	Cat#WT4202BOX
10% (wt/vol) polyacrylamide gels	Biorad	Cat#3450052

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Bryan R. Cullen (bryan.cullen@duke.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cells lines used in this study include HEK293T (referred to as 293T) cells, a kidney epithelial cell line of human female origin and CEM cells, a peripheral blood T lymphoblast cell line of human female origin. All cell lines were purchased from the American Type Culture Collection. All cells were cultured at 37°C with 5% CO_2. 293T cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and Antibiotic-Antimycotic. CEM cells were cultured in Roswell Park Memorial Institute (RPMI) medium supplemented with 10% FBS and Antibiotic-Antimycotic. Primary CD4+ T cells of female origin were cultured in RPMI medium supplemented with 10% FBS, Antibiotic-Antimycotic and supplemented with IL2 72 h before infection.

METHOD DETAILS

HIV-1 Infections

In most HIV-1 infection experiments the WT HIV-1 NL4-3 isolate was used, except where Nano luciferase (NLuc) activity was measured, which required the use of the replication competent NL-NLuc indicator virus, in which the viral *nef* gene has been

substituted with the *NLuc* indicator gene (Mefferd et al., 2018). A plasmid expressing the NL4-3 (pNL4-3) or NL-NLuc (pNL-NLuc) provirus was transfected into 293T using PEI. After 24 h, the spent media were replaced with fresh media. At 72 h post-transfection, the supernatant media were filtered through a 0.45μm filter before being overlaid onto target cells.

HIV-1 Virion Purification

HIV-1 virions were purified by a two-step method, as previously described (Eckwahl et al., 2016). Briefly, supernatant media were harvested from HIV-infected CEM T-cells or pNL4-3 transfected 293T cells at 72 h post-infection or transfection, passed through 0.45 μm filter and the virions collected by pelleting through a 20% sucrose cushion. The virion pellet was then resuspended and layered onto a 7.2% to 20% iodixanol (OptiPrep, Axis-Shield) gradient prior to ultracentrifugation, which separates virions from cellular debris and exosomes. The virion band was then harvested and total RNA extracted using TRIzol. The isolated RNA was heat denatured in an RNA loading buffer containing urea, and then run on a preparative TBE-agarose gel. A single RNA band of ∼9kb in size, corresponding to the size of HIV-1 gRNA, was visualized and excised and RNA isolated using acid phenol followed by phenol-chloroform extraction. The bulk of the purified HIV-1 RNA was then used for UPLC-MS/MS analysis of RNA modifications while a small aliquot was retained for RNA-seq analysis to determine the purity of the HIV-1 RNA. The gRNA from HIV-1 infected CEM was processed using the SMARTer® Stranded Total RNA-Seq Kit v2 following the manufacturer's protocol. The gRNA from pNL4-3 transfected 293T cells was processed by TGIRT-seq, using a previously described protocol (Qin et al., 2016). Human poly(A)+ RNA samples were isolated from CEM or 293T cells using TRIzol, before undergoing two rounds of purification using oligo-dT cellulose beads followed by ribodepletion using a RiboZero Gold kit (Illumina).

RNA Modification Identification by UPLC-MS/MS

Nucleosides were generated from purified HIV-1 RNA by nuclease P1 digestion (Sigma, 2U) in buffer containing 25mM NaCl and 2.5mM $ZnCl_2$ for 2h at 37°C, followed by incubation with Antarctic Phosphatase (NEB, 5U) for an additional 2h at 37°C (Dominissini et al., 2016). Nucleosides were separated and quantified using UPLC-MS/MS as previously described (Basanta-Sanchez et al., 2016), except acetic acid was used in place of formic acid.

Purification of CD4+ T Cells from PBMCs

Peripheral blood mononuclear cells (PBMCs) were isolated from total blood by density gradient centrifugation (lymphocyte separation medium; Cellgro number 25-072-CV). CD4+ T cells were then isolated using the Dynabead CD4 positive isolation kit (Invitrogen; 1131D) following the manufacturer's instructions. Cells were activated by incubation in phytohemagglutinin (PHA) and mouse monoclonal antibodies specific for human CD28 and CD49d (BD Biosciences number 347690) for three days (Whisnant et al., 2013).

PA-m^{5}C-seq and PA-m^{6}A-seq

For PA-m^{5}C-seq, a protocol closely similar to that previously published (Courtney et al., 2017) was used. 293T cells transfected with a CD4 expressing plasmid, CEM T-cells or primary CD4+ T cells were infected with a stock of NL4-3 generated by transfection of 293T cells with the pNL4-3 vector. At 48 hours post-infection (hpi), the cells were pulsed with 100 mM 4 thiouridine (4SU) in fresh media for a further 24 h. At the end of the 4SU pulse, total cellular RNA was extracted using TRIzol, while HIV-1 genomic RNA (gRNA) was collected from virions that were isolated from the supernatant media by ultracentrifugation through a 20% sucrose cushion. Total cellular RNA was poly(A)+ purified using oligo-dT cellulose beads and 10 μg of poly(A)+ RNA or virion gRNA was then used following the previously reported PA-m^{6}A-seq protocol (Chen et al., 2015) using either an m^{6}A-specific or m^{5}C-specific polyclonal antibody.

Construction of Expression Plasmids

All NSUN expression plasmids were made using the pcDNA3 vector. NSUN1 (NM_006170), NSUN2 (NM_017755), NSUN4 (NM_199044) and NSUN5 (NM_148956) were cloned out of an in-house HeLa cDNA library and the sequence verified by Sanger DNA sequencing. Each NSUN construct was designed with a FLAG epitope tag at the N-terminus of the protein. A single missense mutation of Cysteine to Alanine in the conserved PCS region (Figure 2B) was introduced into each NSUN expression vector to generate NSUN mutants unable to release from target cytosine residues on RNAs, as previously reported (Hussain et al., 2013b). The NSUN2-based form of this mutant is named N2-C271A. A second NSUN2 mutant, generated by introduction of the C321A mutation (N2-C321A), is predicted to lack the ability to covalently bind target RNAs and thus would be unable to methylate target cytosines. The previously described CD4 expression vector pCD4 was used to render 293T cells permissive for HIV-1 infection (Bieniasz et al., 1997). A polyclonal 293T cell derivative stably expressing CD4 was generated by transduction of 293T cells using the MLV-based retroviral vector pBabe (Morgenstern and Land, 1990) modified to encode human CD4, followed by selection for puromycin resistance.

A mutant HIV-1 NL4-3 clone, pHIV-1Δm^{5}C, was generated in pNL4-3 that removes all C residues found in a peak bound by the m^{5}C antibody and by NSUN2-C271A that is located 5' of the A2 splice acceptor site, as shown in Figure 5A, to remove m^{5}C modifications from this region (Figure 5E). U residues were substituted in place of C residues.

Virion Packaging of Mutant NSUN Proteins

Plasmids expressing FLAG tagged, mutant forms of NSUN1, NSUN2, NSUN4 and NSUN5, or FLAG-tagged A3G, were co-transfected at equal concentrations with the HIV-1 expression plasmid pNL4-3 into 293T cells using polyethyleneimine (PEI). At 72 h

post-transfection, the supernatant media were passed through a 0.45μm filter, and virions harvested by ultracentrifugation through a 20% sucrose cushion. Protein from isolated virions was harvested in Laemmli buffer before being analyzed by Western blot. Protein from producer cells was also harvested in Laemmli buffer to demonstrate the expression of transfected constructs.

Western Blot Analyses

Protein samples were extracted using Laemmli buffer, sonicated and denatured at 95°C for 15 min. Proteins were separated on Tris-Glycine-SDS polyacrylamide gels. After electrophoresis, proteins were transferred to a nitrocellulose membrane, and then blocked in 5% milk in PBS + 0.1% Tween. Membranes were incubated in primary and secondary antibodies diluted in 5% milk in PBS + 0.1% Tween for 1 h each and then washed in PBS + 0.1% Tween. Each antibody was used at a 1:5000 dilution. Western blot signals were visualized by chemiluminescence.

Immunofluorescence

Cells were seeded on coverslips in a 12 well plate. After 48 h, the media were removed, cells were washed with PBS and then fixed in 4% paraformaldehyde in PBS for 15 min. Cell membranes were then permeabilized by incubation in 0.1% Triton-X in PBS for 15 mins. The cells were then incubated at 4°C overnight with a rabbit polyclonal anti-NSUN2 antibody at a 1:200 dilution in 7.5% bovine serum albumin in PBS. Next day, cells were washed with PBS before being incubated for 1 h at room temperature with anti-rabbit alexa fluor 488 in 7.5% bovine serum albumin in PBS. Cells were then washed with PBS before being placed on slides with Vectashield mounting media, containing the blue fluorescent DNA stain 4′,6-diamidino-2-phenylindole (DAPI), and visualized using an Axio Imager upright immunofluorescence microscope.

Generation of Gene Knock Out 293T Cells by Genome Editing

The clonal NSUN2 KO 293T cell lines N2.4 and N2.16 (Figure S3B) and DNMT2 KO 293T cell lines D2.1 and D2.2 (Figure S3D) were generated by genome editing using CRISPR/Cas9, as previously described (Sanjana et al., 2014). Briefly, an sgRNA guide sequence specific for the human *NSUN2* or *DNMT2* gene was cloned into LentiCRISPRv2 (a gift from Dr. Feng Zhang) and lentiviral vector produced by transfection into 293T cells, along with the packaging plasmids ΔCR8.74 and pMD2.G. Naive 293T cells were transduced with this lentiviral vector and selected for puromycin resistance. Resistant cells were then single cell cloned by limiting dilution, expanded and assessed for knockout of NSUN2 expression by Western blot. NSUN2 negative cells were then verified by DNA sequencing of the relevant region of the *NSUN2* gene. The predicted reduction in m5C modification of cellular mRNAs was confirmed by UPLC-MS/MS (Figure 3C).

HIV-1 Infection of NSUN2 Knockout 293T Cells

The pCD4 expression plasmid was transfected into WT 293T cells, the NSUN2 knockout line N2.4, or co-transfected into N2.4 together with the NSUN2 expression plasmid pcDNA3-N2. After 48 h, the cells were infected with the NL-NLuc indicator virus. Cells were harvested at 24, 48 and 72 hpi and washed 3 times in PBS before being lysed in Passive Lysis Buffer (PLB; Promega). The NLuc activity from these cell lysates was then quantified using the Nano Luciferase Assay Kit (Promega). For Western blots, CD4 was transfected into 293T cells or N2.4 and cells infected with NL4-3. At 48 and 72 hpi, proteins were harvested from infected cells and visualized by Western blot.

NSUN2-CLIP-seq

293T cells were transfected using PEI with pCD4 and plasmids expressing FLAG-tagged WT NSUN2 or the N2-C271A mutant. Cells were infected with NL4-3 48 h after transfection and harvested for NSUN2-CLIP-seq at 48 hpi. The NSUN2-CLIP-seq experiments were performed using a modified version of a previously described method termed miCLIP (Hussain et al., 2013b). The original protocol was modified to replace the RNase I digestion with an RNase T1 digestion to generate smaller RNA fragments to better ascertain the NSUN2 RNA footprint. After elution of RNA fragments from NSUN2 by proteinase K digestion, Illumina small RNA-seq libraries were generated, following the manufacturer's instructions.

NSUN2 Mutants and HIV-1 Protein Expression

The N2.4 NSUN2 knockout cell line was transfected with pcDNA3, pcDNA3-N2, pcDNA3-N2.C321A or pcDNA-N2.C271A, along with pNL-NLuc, using PEI. Media were changed 24 h later. At 72 h post-transfection, cells were harvested for Western blot, lysed in PLB for a NLuc assay, or RNA was extracted by TRIzol with HIV-1 RNA levels assessed by qRT-PCR with a probe set targeting U3. In addition, media from these cells was filtered through a 0.45μm filter and a small volume was concentrated by ultracentrifugation, as described above, with p24 levels from the rest measured by ELISA following the manufacturer's instructions. After normalization by ELISA, fresh CEM cells were infected with equal p24 levels of NL-NLuc and incubated for a further 72 h. Infected CEM cells were then harvested in either PLB for NLuc assay or in Laemmli buffer for Western blot.

Quantification of HIV-1 RNA Levels Using qRT-PCR

Relative HIV-1 RNA expression levels were determined by qRT-PCR. The level of GAPDH mRNA was used to normalize all qRT-PCR experiments. All primer sequences are listed in the Resource Table. For multicycle infections, RNA was collected at 0, 48 and 72 hpi.

RNA was extracted using the TRIzol method. cDNA was generated using the Ambion cDNA synthesis kit with random primers, following the manufacturer's protocol. All qRT-PCR experiments were performed using PrimeTime® Gene Expression Master Mix following the manufacturer's instructions. All qRT-PCR data were quantified using the ΔΔCT method.

Nuclear and Cytoplasmic Fractionation

Nuclear and cytoplasmic fractionation (Malim et al., 1989) was used to determine the level of HIV-1 RNA in the nucleus and cytoplasm of transfected cells. Briefly, WT or NSUN2 KO 293T cells were transfected with pNL4-3 and incubated for 72 h. Cells were then washed with PBS, scraped off the plate and pelleted at 4°C. The cell pellet was resuspended in NP40 lysis buffer (10mM Tris-HCl pH 7.4, 10mM NaCl, 3mM $MgCl_2$ and 0.5% NP-40), briefly vortexed and then incubated on ice for 5 minutes. Lysates were spun in a microfuge at full speed for 15 seconds to pellet the nuclei. The supernatant was removed and cytoplasmic RNA was extracted using TRIzol LS. The nuclear pellet was washed twice with NP40 lysis buffer and pelleted as before. RNA was then extracted from nuclei using TRIzol. RNA samples underwent reverse transcription followed by qPCR to quantify HIV-1 and GAPDH mRNA levels.

Quantification of Ribosome Loading on to HIV-1 Transcripts

Ribosome-loaded RNA was isolated essentially as previously described (Subtelny et al., 2014). Briefly, WT and NSUN2 KO 293T cells were transfected with pNL4-3 and incubated for 72 h. Cells were then treated with 100μg/ml of cycloheximide for 10 minutes at 37°C, washed in PBS and lysed in ribosome lysis buffer (10mM Tris-HCl pH 7.4, 5mM $MgCl_2$, 100mM KCl, 1% Triton X, Protease inhibitor, 2mM DTT, 100μg/ml cycloheximide and RNase inhibitor). One tenth of the lysate was then used for RNA isolation for the "input" sample. The rest of the lysates were sheared by passage through a 26-gauge needle before clarification by centrifugation at 1300g for 10 minutes. The supernatants were brought up to 10ml in lysis buffer and overlaid on a 30% sucrose cushion, also made up in lysis buffer, and centrifuged at 164,000g for 2 hr at 4°C. RNAs were then extracted from the ribosomal pellet using TRIzol. RNA isolated from the input and ribosome loaded fractions then underwent reverse transcription followed by qPCR to quantify HIV-1, GAPDH, Actin, HPRT1 and MRAS mRNA levels.

Characterization of HIV-1 RNA Splicing

HIV-1 RNA splicing analysis was performed as previously described (Emery et al., 2017). Each sample was numbered and the researcher blinded prior to processing and analysis of these samples. Briefly, a method of Primer-ID tagged deep sequencing was used to quantify the usage of each of the 4 splice donor and 10 splice acceptor sites in HIV NL4-3. Total cellular RNA was isolated from WT, NSUN2 KO or DNMT2 KO 293T cells transfected with a WT NL4-3 proviral expression plasmid and two different cDNA reactions performed. Two distinct RT primers were used, one for capturing incompletely spliced transcripts (~4kb primer) or one for completely spiced transcripts (~1.8kb primer). Each of the cDNA primers includes a common sequence at its 5′ end that serves as the sequence of the reverse primer in PCR. The forward primer is common for both size classes and is used in the PCR amplification incompletely and completely spliced transcripts. After deep sequencing of these cDNA libraries an in-house script reads the forward sequence to identify the donor and acceptor sites in the spliced RNA template before doing the same for the reverse read and combines the information to give a pattern of splice junctions. These patterns are sorted according to Primer ID with each individually identified read counting as only a single observation. Usage of each donor and acceptor site can then be quantified in relation to the total number of unique reads obtained.

An additional method used to determine the usage of the D1/A2 splice junction was based on qRT-PCR analysis. A primer probe set was designed to target the D1/A2 region of HIV-1 (sequences listed in the Resource Table) and results were normalized to two additional HIV-1 specific probe sets specific for the Gag or LTR U3 region, as described above. This protocol was used for both WT NL4-3 and the pHIV-1Δm^5C mutant provirus

Bioinformatic Analysis

Read alignments were performed using Bowtie (Langmead et al., 2009). Reads were first aligned to the human genome build hg19 allowing up to 1 mismatch, then unaligned reads were aligned to the HIV NL4-3 transcriptome, again allowing up to 1 mismatch. Characteristic T>C mutations, resulting from 4SU incorporation and crosslinking, were present among both human and viral aligned reads. All data was processed using in-house Perl scripts and Samtools (Li et al., 2009), and visualized with IGV, as previously described (Courtney et al., 2017). The CEM poly(A)+ PA-m^5C-seq dataset was further assessed for human mRNA alignments using the protocol previously published for PA-m^6A-seq data analysis (Dominissini et al., 2013). This pipeline was used to determine m^5C peaks in the human transcriptome and generate an m^5C motif. Metagene analysis of m^5C site average relative location within the body of cellular mRNAs were done using a previously published script package (Olarerin-George and Jaffrey, 2017). RNA-seq data anaylsis was performed by first aligning reads to the HIV-1 genome, then the human genome build hg19, followed by the ensembl ncRNA database and the tRNAscan-SE database. The raw sequencing data obtained by RNA-seq have been submitted to the NCBI expression omnibus and are available under GenBank accession number (GEO: GSE130972).

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical details of experiments can be found in the figure legends. All error bars displayed indicate standard deviation. Significance was calculated by T-test, comparing test samples to the control, with the significance of a result being determined as $p<0.05$ = *, $p<0.01$ = ** and $p<0.001$ = ***. Blinding of samples was only performed for the characterization of HIV-1 splicing.

DATA AND SOFTWARE AVAILABILITY

All deep sequencing data have been submitted to NCBI GEO. The accession number for the deep sequencing data reported in this paper is GSE130972.