# Direct tracking of single proviruses reveals HIV-1/LEDGF complexes excluded from virus-induced membraneless organelles

**Viviana Scoca[1, 2], Marion Louveaux*[3, 4], Renaud Morin*[5], Dmitry Ershov[4, 6], Jean-Yves Tinevez[4], Francesca Di Nunzio[1]**

*[1]Department of Virology, Unit of Molecular Virology and Vaccinology, Pasteur Institute, Paris, France*
*[2] BioSPC Doctoral School, Université de Paris, Paris, France*
*[3] Department of Cell Biology and Infection, Biological Image Analysis, Pasteur Institute, Paris, France*
*[4] Image Analysis Hub/ C2RT, Pasteur Institute, Paris, France*
*[5] Imactiv-3D, 1 Place Pierre Potier, 31106 Toulouse, France*
*[6] Department of Computational Biology, Hub of Bioinformatics and Biostatistics, Pasteur Institute, USR 3756 CNRS, Paris, France.*

*****equal contribution

Running title: HIV-1 MLOs retain integrase and release viral DNA

## Abstract

The ultimate goal of HIV-1 is integration into the host chromatin to optimize the release of high levels of viral progeny and discretely coexist with the host. To uncover the HIV-1 DNA fate in the nuclear landscape we directly tracked the viral DNA (vDNA) and the viral RNA (vRNA) by coupling HIV-1 ANCHOR technology with RNA FISH or MCP-MS2 RNA-tagging bacterial system. Our computational imaging analysis revealed that proviral forms are early located in proximity of the nuclear periphery of mitotic and non-mitotic cells. We also observed that HIV-1 infection prompts clustering formation of the host factor CPSF6 restructuring membraneless organelles enriched in both viral proteins and speckle factors. Interestingly, we observed that integrase proteins are retained in CPSF6 clusters, while the late retrotranscribed DNA was excluded from HIV-induced membranelless organelles (HIV-1 MLOs), indicating that those structures are not proviral sites, but orchestrate viral events prior to the integration step. HIV-1 MLOs are in the vicinity of pre-existing LEDGF clusters. Importantly, we identified actively transcribing proviruses localize, outside HIV-1 MLOs, in LEDGF-abundant regions, known to be active chromatin sites. This study highlights single functional host-proviral complexes in their nuclear landscape, which is markedly restructured by HIV-1 to favor viral replication.

## Introduction

Immediately after fusion at the plasma membrane, HIV-1 cores are released into the cytoplasm and move towards the nucleus, while the viral RNA genome begins the process of reverse transcription into double-stranded DNA (dsDNA) (Campbell and Hope, 2015; Di Nunzio, 2013). Once imported in the nucleus, the interplay between the HIV-1 DNA genome and the host chromatin compartment is crucial for the fate of the virus-host coexistence. HIV-1 can adopt either an episomal or a proviral form (Butler et al., 2001). The choice between these two possibilities and their surrounding chromatin landscape dictates the evolution of HIV infection (Maldarelli, 2016; Sharkey et al., 2011; Sharkey et al., 2000). The mechanisms determining the fate of the proviral DNA in the nucleus are still under investigation (Liu et al., 2020; Olson et al., 2019) also because of the limits imposed by available technologies. Real-time imaging approaches provide new insight into unprecedented spatial information of individual HIV-1 infected cells. Several strategies have enabled the visualization of the viral components during nuclear entry: Integrase (IN)-TC/FlAsH, APOBEC 3F (A3F)-YFP, IN-YFP, Gag-iGFP, Cyclophilin A (CypA)-DsRed/ Capsid (CA), INsfGFP, and CA-GFP (Burdick et al., 2017; Hubner et al., 2009; Lelek et al., 2012) (Francis et al., 2016; Francis and Melikyan, 2018) (Mamede et al., 2017) (Albanese et al., 2008; Francis et al., 2014). Previously, biochemistry studies supported only a limited and secondary role of the viral CA in the early steps of infection (Farnet and Haseltine, 1991). State-of-the-art imaging technologies developed in recent years have shed light on the central role of the viral CA in nuclear import(Blanco-Rodriguez et al., 2020; Burdick et al., 2020) dictated by the CA interplay with host factors, such as CPSF6 (Buffone et al., 2018; Francis et al., 2020; Price et al., 2014) and Nup153 (Di Nunzio et al., 2013; Lelek et al., 2015). Although these techniques improved our knowledge in the HIV field, the labeling of viral proteins only allows the visualization of steps prior to the establishment of proviral or episomal forms. Mechanisms underlying viral post-nuclear entry steps and the role of nuclear structures induced by HIV-1 during viral replication could be revealed only by the direct co-labeling of viral DNA (vDNA) forms and host factors. Thus far, the vDNA can be visualized either by supplying 5-Ethynyl-2´-deoxyuridine (EdU) during infection (Peng et al., 2014), which is limited to fixed non-dividing cells (De Wit et al., 2019) or by DNA-FISH (Marini et al., 2015), which is characterized by harsh conditions of fixed sample preparation that limits its coupling

with other techniques, impeding the study of the interplay between the vDNA and the host factors. Here, we were able to live track episomal and proviral forms, showing their divergent intranuclear behavior. In parallel, we detect the location of functional proviruses in the nuclear space of the main target cells of the virus, such as $CD4^+$T cells and macrophages, by using HIV-1 ANCHOR methodology (Blanco-Rodriguez et al., 2020) to label the vDNA coupled to RNA FISH or to the live bacterial approach MCP-MS2 (Tantale et al., 2016) to identify the vRNA foci. Next, we deepened the study of the nuclear landscape that surrounds proviruses. We observed that viral infection reprogramed the nuclear location of a viral partner, CPSF6. CPSF6 is a paraspeckle factor (Fox et al., 2002; Naganuma and Hirose, 2013) involved in viral nuclear entry though its interaction with the viral CA (Buffone et al., 2018; Francis et al., 2020; Lee et al., 2010; Price et al., 2014). Here we found that CPSF6 relocates in HIV-1 induced membraneless organelles (HIV-1 MLOs), enriched in speckle factors, SC35 (Spector and Lamond, 2011) and SON (Sharma et al., 2010), near LEDGF clusters. Previous studies reported the importance of LEDGF in HIV-1 integration sites distribution (Ciuffi et al., 2005), however a complex between the transcribing proviral DNA and LEDGF has never been visualized. In this study, we determined the nuclear position of several viral forms, including actively transcribing proviruses, with respect to HIV-1- MLOs or the nuclear envelope (NE). In addition, we found that the late retro-transcribed DNA co-exists, but remains separated, from viral INs in the host nucleus. We observed that the viral INs are sequestered in HIV-1 MLOs, suggesting the role of these nuclear structures in the Pre-Integration Complex (PIC) maturation. On the other hand, the late reverse transcripts are fully excluded from HIV-1 MLOs, in fact we found that HIV-1 MLOs are not the sites of viral integration. Finally, we identified functional host-viral complexes, formed by vDNA/vRNA associated to LEDGF, excluded from the inner cavity of HIV-1 MLOs (avg distances from HIV-1 MLOs < 0.6 µm) and located in the vicinity of the NE (avg of distances from NE < 2 µm). In this study we show that HIV-1 DNA localization is key to identifying partners of HIV-1 that aid the virus to hijack cellular mechanisms to persist in the host cells and to highlight the nuclear landscape surrounding single viral genomes. Our results provide new insights into how HIV reprograms and markedly restructures the nuclear environment to orchestrate viral replication steps.

## Results

### Live tracking of HIV-1 DNA forms in the nucleus of target cells.

With the aim of studying HIV-1 DNA nuclear fate in different cell types, we benefited of the HIV-1 ANCHOR DNA labeling system (Fig. 1A) (Blanco-Rodriguez et al., 2020). Once fully retrotranscribed, the HIV-1 genome carrying the bacterial sequence ANCH3 is specifically recognized by OR-GFP protein, which is expressed in the target cells, genetically modified by the lentiviral vector (LV) OR-GFP (Blanco-Rodriguez et al., 2020). The accumulation of OR-GFP, which is a modified version of the bacterial ParB protein (Graham et al., 2014; Sanchez et al., 2015), on the ANCH3 sequence generates bright nuclear signals (Fig.1A). Importantly, OR protein binds exclusively double stranded DNA (Saad et al., 2014). We cloned ANCH3 in the nef gene, thus, HIV-1 ANCHOR system exclusively detects late reverse transcripts, because only late retrotranscripts contain ANCH3 sequence as double stranded DNA (Fig. 1A). Therefore, HIV-1 ANCHOR system allows only the visualization of potential functional vDNA. We optimized and set up our imaging approach in HeLa P4R5 cells (Charneau et al., 1994), a widely used cellular model for HIV-1 infection. Previous studies showed that HeLa cells enable single particle tracking of HIV-1 proteins (Burdick et al., 2020; Francis and Melikyan, 2018). However, since HIV-1 genome dissociates from viral proteins to become a proviral DNA or an extrachromosomal DNA, the track of these viral forms is the only direct measurement to investigate the fate of the viral genome into the host nuclear space. Unlike previous studies, we are now able to directly visualize by live imaging the nuclear vDNA as GFP puncta appearing in the host nucleus once retrotranscribed (Fig. 1A). On the contrary, bright individual spots are rarely visualized in the cytoplasm, where the vDNA is only partially accessible to OR-GFP, as we previously demonstrated (Blanco-Rodriguez et al., 2020). HIV-1 ANCHOR is able to specifically label the vDNA as shown by the correlation between the nuclear GFP puncta and the MOI (multiplicity of infection) used (Fig. S1A). This specificity was corroborated using the reverse transcription inhibitor, Nevirapine (NEV), which completely prevents the detection of vDNA (Fig. 1B, S1B). We then investigated whether HIV-1 ANCHOR allows the detection of different intranuclear vDNAs, such as episomal and integrated forms. To generate episomal forms, either we infected cells with HIV-1 ANCH3 in the presence of the integration inhibitor Raltegravir (RAL), or using an integration-deficient virus, which is mutated at the catalytic site of the

integrase (IN) enzyme (HIV-1 ANCH3 $IN_{D116A}$). We detected intranuclear GFP puncta under both conditions (Fig. 1B), meaning that HIV-1 ANCHOR permits the detection of nuclear episomal vDNA forms. The experiments were validated by qPCR, which confirmed the extrachromosomal forms generated during infection when the integration step is impeded (Fig. 1B). We next assessed whether HIV-1 ANCHOR is sufficiently sensitive to allow the detection of integrated vDNA at the level of a single provirus. We selected single HeLa P4R5 clones carrying the HIV-1 ANCH3 genome. The imaging data correlates with results obtained by ALU PCR (Fig. 1C), demonstrating that this technology is powerful enough to detect a single provirus. Thus, we used this system to live track proviruses and episomal forms to investigate their behaviour in the nucleus (movies S1, S2, and S3). We performed live tracking of vDNAs in cells at six days post infection (p.i.). (movie S1), which contain a negligible amount of episomal forms, likely lost during cell division (Fig. S1C), in the single-cell clone (movie S2) or in cells containing only the episomal vDNA (movie S3). Then, we compared vDNA trajectories in the aforementioned experimental conditions (Fig. 1D). The stable clone and the cells at six days p.i. show vDNA with similar trajectories and diffusion coefficient (D ($nm^2/s$) < 5). Instead, infected cells carrying unintegrated viruses harbored two populations composed of episomal forms with divergent behaviors, one with a low D (D ($nm^2/s$) < 5) and the other with a high D (5 < D ($nm^2/s$) < 20) (Fig. 1D). Our data hint that most, but not all, episomal forms are retained in particular nuclear regions, most likely interacting with host factors or chromatin compartments, consistent with the findings of other studies (Geis and Goff, 2019; Zhu et al., 2018), whereas some episomal forms are free to move in the nuclear space. On the other hand, the integrated forms showed a uniform, low-diffusive behavior.

**Detection of actively transcribing viral DNA at single-cell level in HIV-1 target cells.**

The identification of functional vDNA in the nucleus of infected cells has been one of our goals. For this purpose, we combined HIV-1 ANCHOR technology with RNA FISH (single molecule inexpensive FISH, smiFISH) against HIV-1 polymerase transcripts, to detect vRNA foci of transcription (Tsanov et al., 2016). As expected, in cells treated with RAL, where the integration step is blocked, vRNA signal is highly depleted. In untreated cells it is possible to observe vDNA-vRNA associations, highlighting the proviruses location (Fig. 2A), in fact proviruses are the only transcribing forms. We

strengthen our results on the ability to visualize proviruses by showing the vDNA-vRNA association by live imaging. HIV-1 transcriptional activity was tracked by MCP-GFP system (kind gift from E. Bertrand) to label vRNA (Tantale et al., 2016) coupled with HIV-1 ANCHOR technology (Fig. S2A, B). Next, we compared the intensity of ANCHOR signals of the episomal forms (RAL) with those of the integrated forms (untreated cells) and observed that the signal of unintegrated DNA was two-fold brighter than the proviral DNA spot (Fig. 2A). Interestingly, we also remarked, during live imaging, that the vDNA of the integration-deficient virus (HIV-1 ANCH3 $IN_{D116A}$) aggregated to form “flower-shaped structures” in the nucleus (Fig. S3, movie S4). Aggregation of extrachromosomal forms could prompt a brighter appearance of viral episomal DNA compared to a dimmer signal of the individual proviral DNA (Fig. 2A). Next, we tested the HIV-1 ANCHOR system in the context of the main target cell of HIV-1, primary $CD4^+$ T cells, derived from healthy donors. $CD4^+$ T cells challenged with various MOIs of HIV-1 ANCH3 showed that the number of nuclear spots increased with the increasing of the MOI (Fig. S4), meaning that HIV-1 ANCHOR specifically detects the vDNA in the nucleus of primary lymphocytes. The correlation between different MOIs used and the number of proviruses (r=0.9412), analysed by qPCR, indicates that OR-GFP does not impede the viral integration step (Fig. S4). We tested, as well, the co-detection of viral transcriptional foci and vDNA in T lymphocytes (primary $CD4^+$ T and Jurkat cells) through RNA FISH (Fig. 2B, Fig. S5). We observed that not all vDNAs associated to a transcriptional focus at 48 h p.i., most likely because of the presence of silent episomal forms. On the other hand, the majority of cells analyzed contain vRNA foci colocalized with vDNA (Fig. 2B), meaning that we are able to detect single transcribing proviruses. We also performed similar experiments in non-dividing cells, THP-1, which are macrophage-like cells. Thus, we imaged and analyzed THP-1 cells at 3 days p.i. and we observed that also in these cells we could distinguish individual transcribing proviruses from silent viral forms (Fig. 2C). Of note, the RNA signal considered in our images was exclusively due to the detection of viral transcription foci, since samples treated with NEV did not show any significant intra-nuclear signal (Fig.S6). Taken together, our results suggest that we are able to distinguish episomal forms (mainly silent) from integrated forms that are the only ones able to replicate.

**Nuclear distance of actively transcribing proviruses from the nuclear envelope at single-cell level in dividing and non-dividing cells.**

The location of actively transcribing proviruses in the nuclear space of dividing cells is still a subject of debate (Achuthan et al., 2018; Burdick et al., 2017; Chin et al., 2015; Di Primio et al., 2013; Lelek et al., 2015; Marini et al., 2015), whereas their nuclear location in terminally differentiated cells, as far as we know, has never been studied before. In dividing cells, the 3D analysis of the nuclear location of the vDNA showed the vDNA signals at 48 h p.i. to be more randomly distributed throughout the nucleus compared to the vDNA spots at seven days p.i. (Avg distances from the NE: ~2.5 µm at 48 h p.i. vs ~1.4 µm at 7 days p.i.) (Fig. 3A). The difference in the average distance of the vDNA spots from the NE between the early and late time points p.i. may be due to the co-existence of heterogeneous viral forms at 48 h p.i. (episomal and proviral forms), whereas the vDNA population is more homogeneous at seven days p.i. (Fig. 1D, 3A). In fact, at seven days p.i., almost all remaining vDNAs in the nucleus consisted of integrated forms, as suggested by the imaging of dividing cells following prolonged treatment with RAL (Fig. S1C). Interestingly, we did not detect differences between the distances of the vRNA foci from the NE (~ 1.5 µm) at both time points (48h p.i. and 7d p.i.) and the vDNA spots at seven days p.i. (~ 1.4 µm), indicating that in dividing cells the integration of the proviral DNA occurs early during infection near to the NE (Fig. 3A). We performed similar 3D analysis to calculate the distances of vDNAs and vRNAs from the NE in non-dividing cells at three days p.i.. The lack of mitosis hinders the loss of the episomal forms in non-dividing cells, so we could not discriminate between integrated and unintegrated vDNA forms. However, when we compared vDNAs (episomal and proviral forms) versus vRNA foci (actively transcribing proviral forms) their analysis of distances from the NE revealed a major vicinity of the proviruses to the NE compared to the total vDNA (~2.2 µm for the vDNA vs ~1.8 µm for the vRNA) (Fig.3B). Overall, our data show a similar nuclear spatial distance of actively transcribing proviral forms from the NE at single-cell level in both, dividing and non-dividing cells (distance of vRNA foci from the NE: ~1.7 µm in HeLa cells and ~1.8 µm in THP-1 cells) (Fig.3C).

**The majority of IN proteins are retained in CPSF6 clusters and separate from the late retrotranscribed DNA**

A long-standing question in the early steps of HIV life cycle is whether viral IN proteins remain all associated with late reverse transcribed DNA for long time or only for a short time after the nuclear entrance. This association may not be detectable because of the short-time nature of the event or due to the technical limitations of vDNA and IN visualization. Thus, we exploited our vDNA tracking system to investigate about IN proteins localization in primary $CD4^+$T cells in comparison to terminally differentiated cells, THP-1. We challenged both cell types with HIV-1 ANCH3 and we analysed them at 20h and 30h p.i., respectively. We found that in both cells the vDNA/IN association event was very rare (only one single event in 1 cell for THP-1 and in 3 cells for CD4+T cells) (Fig. 4A). Nuclear IN protein clusters were found in the nucleus, but not at the site of the late reverse transcripts (Fig. 4A). To confirm our data we performed live-imaging, asking whether we could pinpoint the dynamic evolution of the association between the vDNA and the IN in the nucleus of infected cells by coupling the HIV-1 ANCHOR system with the GIR virus (generous gift from Edward Campbell)(Dharan et al., 2016; Hulme et al., 2015). We produced the GIR virus using a viral genome carrying the IN with a mutation in the catalytic site (HIV-1 ANCH3 $IN_{D116A}$), as consequence, the only active IN in these viral particles was the IN-Ruby (Fig. S7). Analysis of the dynamics of the vDNA / IN association shows the two signals to be separated (~ 0.6 to 0.8 µm) within the nucleus for several hours (Fig. S7, movie S5). Our results suggest that, after nuclear entry, the main localization of IN is not at the exact site of the viral DNA forms, even though we can sometimes spot them in proximity. Next, we investigated whether the viral IN was retained in particular nuclear regions by some nuclear host factors. We first found that the viral IN occupies the same positions as the viral CA and its host nuclear partner, CPSF6 (Fig. S8A). We confirmed that HIV-1 reprograms the nuclear localization of CPSF6 from being randomly dispersed in the nucleoplasm to form particular aggregates in both primary lymphocytes and macrophages (Francis et al., 2020) (Rensen et al., BioRxiv doi.org/10.1101/2020.04.12.038067) (Fig. 4B). We also observed that CPSF6-IN-CA clustering is independent of DNA synthesis (Fig. S8B). Then, we quantified the percentage of association per nucleus between the viral IN and CPSF6. We found an average of association of IN to CPSF6 ~ 94.4% in CD4+T cells and ~92.5% in THP-1 cells (Fig. 4B). Taken together our results show that in dividing and in non-dividing cells there is an intra-nuclear separation of the viral DNA and of the IN proteins. Interestingly, the viral IN proteins persist after viral nuclear entry associated to CPSF6

clusters. This observation suggests that the late reverse transcripts are released from the majority of IN proteins which instead are retained in a separated nuclear location occupied by CPSF6. Thus, CPSF6 clusters could serve as sites to release a mature PIC that could contain just few IN proteins necessary for the integration step (Ballandras-Colas et al., 2017; Hare et al., 2009).

**Proviruses are excluded from HIV-1-induced membraneless organelles.**

CPSF6 participates to the viral nuclear translocation and post-nuclear entry steps by interacting with the viral CA (Bejarano et al., 2019; Burdick et al., 2017; Lee et al., 2010; Price et al., 2014) (Francis et al., 2020). Recent studies reported a high level of co-localization between the paraspeckle factor, CPSF6, or the speckle factor and splicing regulator, SC35, (Spector and Lamond, 2011) and the vDNA labeled with EdU (Francis et al., 2020) (Bejarano et al., 2019)(Rensen et al., BioRxiv doi.org/10.1101/2020.04.12.038067). Importantly, it has been shown that EdU labels nuclear clusters composed of multiple viral episomal forms (Francis et al., 2020) (Rensen et al.,BioRxiv doi.org/10.1101/2020.04.12.038067). Therefore, EdU labeling can detect multiple nascent vDNAs but it is likely not sufficiently sensitive to detect an individual viral genome. Contrary to EdU labeling, the HIV-1 ANCHOR system allows the visualization of a single provirus (Fig.1C) and does not interfere with viral transcription (Blanco-Rodriguez et al., 2020) (Rensen et al.,BioRxiv doi.org/10.1101/2020.04.12.038067). Thus, we used HIV-1 ANCHOR technology to visualize the late retrotranscripts in infected macrophage-like cells at 3 days p.i.. We observed that late retrotranscripts do not colocalize with CPSF6 (Fig. 5A). However, upon viral infection, CPSF6 relocates and fully colocalizes with speckle-simile organelles labeled with SC35 (Fig. 5A). SC35 is strongly associated to another speckle factor, SON, independently of viral infection (r pearson ~0.995). More specifically, we found that the nuclear reorganization of CPSF6 is not required for the integration step, as the nuclear organization of CPSF6 was similar in cells treated or not with RAL (Fig. S9). Then, we investigated whether the vDNA colocalizes with SC35. We computed vDNA distances in 3D space from the boundary of SC35 bodies in cells treated with RAL (avg ~ 0.4 µm (Fig. S9)) or in absence of RAL (avg ~ 0.6 µm (Fig. 5A)). Therefore, similarly to CPSF6 also SC35 clusters exclude late reverse transcripts (Fig. 5A). In addition, we found that viral transcriptional foci where also located outside CPSF6/SC35 organelles (Fig. 5B). These results are independent of the presence of

vpx (Fig. S10) that we used to increase viral infectivity in THP-1 (Hrecka et al., 2011; Laguette et al., 2011). Taken together analysis of the nuclear positions of vDNA (episomal and proviral forms) and vRNA foci (proviruses) showed that these viral nuclear forms are excluded from HIV-1 MLOs (Fig. 5C), indicating that those organelles cannot be sites for viral integration or replication.

**Late reverse transcripts DNA and viral RNA foci form complexes with LEDGF, spatially independent of CPSF6 clusters.**

We extended our study on THP-1 cells at 3 days p.i. to investigate whether Lens Epithelium-Derived Growth Factor (LEDGF), which is a partner of the viral IN and leads HIV-1 integration sites selection (Cherepanov et al., 2003; Ciuffi et al., 2005; Emiliani et al., 2005; Ferris et al., 2010; Hare et al., 2009; Lelek et al., 2015; Llano et al., 2006; Shun et al., 2008; Shun et al., 2007; Wang et al., 2014), is associated to the proviruses forming a complex visible by imaging. We observed that CPSF6 proteins cluster in proximity of LEDGF sites (~ 43%), without co-localizing (Fig. 6A). There is a similar number of LEDGF clusters in both infected and uninfected cells (Fig. 6A), while CPSF6 clusters are mainly formed upon infection, suggesting that CSPF6 is prompted by the virus to localize near the pre-existing LEDGF regions (Fig.6A). Of note, the PWWP domain of LEDGF has been shown to interact with H3K36me3 (Pradeepa et al., 2014; Pradeepa et al., 2012), which is a marker of euchromatin. In particular, this post-translational modification is enriched in HIV-1 integration sites (Ciuffi et al., 2005; Lelek et al., 2015; Vansant et al., 2020; Wang et al., 2009). LEDGF also interacts with splicing factors (Singh et al., 2015). Therefore, we analysed the location of proviral foci of transcription in relation to LEDGF and CPSF6. Our previous data shows that CPSF6 co-localizes with SC35, however the distance between the vDNA and SC35 has been estimated to be ~ 0.6 µm (Fig. 5A), therefore, the late retrotranscribed DNA is too distant from those HIV-1 MLOs to be considered part of the same complex. While, the vDNA of the late retrotranscripts is mostly excluded from CPSF6 clusters, we found that ~64% of the vDNA and ~73% of the vRNA foci associate with LEDGF (Fig. 6B). Next we have characterized in more detail the viral genome associations with LEDGF and CPSF6. We calculated that ~ 87.9% of the vDNA/LEDGF complexes and ~80% of vRNA foci/LEDGF complexes are uncoupled to CPSF6 (Figure 6B). We found only few vDNA (~12.1%) and vRNA (~20%) foci associated to LEDGF clusters coupled to CPSF6 (Fig. 6B). Our results suggest that the two cellular factors, CPSF6 and LEDGF,

orchestrate two sequential steps of viral replication, the maturation of the PIC and its integration. Our data indicate that the functional vDNAs are not located inside CPSF6/SC35 clusters, but in external active chromatin domains, enriched in LEDGF, likely to benefit of a favorable environment for viral replication.

## Discussion

Studies of parasites that permanently integrate into the host genome show that they have evolved to target chromatin in such a way as to optimize their coexistence with the host and favor the release of new viral progeny (Bushman, 2003; Craig and Marszalek, 2002). HIV-1 needs to rapidly release high levels of newly generated viruses after infection, intuitively explaining the preference of active genes as integration targets (Ciuffi et al., 2005; Di Nunzio et al., 2013; Lelek et al., 2015; Maldarelli, 2016; Schroder et al., 2002). However, the nuclear location of individual proviruses actively transcribing at the dividing or non-dividing single cell level has never been visualized before. Thus far, the direct detection of HIV-1 genomes after reverse transcription and the distribution of proviral transcriptional foci has been technically challenging. Recent studies have shown the possibility to follow HIV infection in live using surrogate viruses (IN GFP virus (Albanese et al., 2008), a mix of viruses carrying fluorescent and non-fluorescent IN proteins (Hulme et al., 2011), or indirectly follow the viral journey by labeling host viral partners, such as CypA (Francis and Melikyan, 2018) or APOBEC 3F (Burdick et al., 2017)). All these approaches allow to follow viral or host proteins but not viral replication complexes. HIV-1 ANCHOR system allows the live tracking of different nuclear late retrotranscribed DNA forms (Fig.1A), which makes possible to probe their behavior via the measurement of the mean-squared displacement (MSD). We observed a homogenous population formed by proviral DNA with low diffusion coefficient. Episomal forms showed a heterogeneous behavior, characterized by two different populations, implying that these two viral forms can interact with different factors in the nuclear space (Fig. 1D). Moreover, imaging analysis showed that viral episomal forms identified by the accumulation of OR-GFP on ANCH3 sequences gave rise to brighter signals than proviral forms (Fig. 2A). This phenomenon could account for the differences in behavior between integrated and unintegrated forms. Thus, unintegrated forms can be brighter for the following reasons: a) the absence of transcriptional machinery

associated with the vDNA may reduce the steric hindrance that can interfere with the accumulation of OR-GFP on the HIV-1 tag sequence (ANCH3) (Fig. 2A), b) episomal forms can aggregate in a flower shape, generating a brighter signal (Fig. S3, movie S4), c) episomal forms are differently chromatinized than the integrated vDNAs, becoming more accessible to OR-GFP (Geis and Goff, 2019). Next, to track the nuclear location of functional proviruses we co-labeled vDNA and vRNA by combining HIV-1 ANCHOR technology with RNA FISH in the main target cells of HIV-1, $CD4^+$T cells and macrophages (Fig.2A,B,C). We also corroborated results on the visualization of functional proviruses using a live track method, MCP-MS2 system (Tantale et al., 2016), coupled to HIV-1 ANCHOR (Fig.S2). Once the performance of our approach for labeling actively transcriptional proviruses was validated, we investigated the spatial location of proviruses in dividing and non-dividing cells. Three-dimensional analysis of the distribution of vDNAs during early (48 h p.i.) infection in mitotic cells displays that they are more randomly distributed than the proviral DNA from the single clone or from cells at 7days p.i. (Fig. 3A). We also observed that the location of viral transcription sites (vRNA) is highly similar between early and late time points of infection, indicating that proviral DNAs, the only transcribing forms (Fig. 2A), immediately localize to near the NE (~1.5 µm) early after infection (Fig. 3A) in dividing and non-dividing cells (Fig.3B,C). We next investigated the interplay among viral components and their association with host factors during post-nuclear entry steps. We corroborated our published data (Rensen et al.,BioRxiv doi.org/10.1101/2020.04.12.038067) showing that IN proteins strongly co-localize with CPSF6 clusters in macrophage-like cells and in this study we observed a similar association between IN and CPSF6 also in $CD4^+$ T lymphocytes at early time of infection (Fig. 4B). Contrary to CPSF6, the late vDNA products of retrotranscription rarely co-localize with IN in both types of cells (Fig. 4A). These results are in line with the dynamic coexistence of the IN and vDNA, which remain separated but in the proximity for several hours, as it has been shown by real-time imaging in HeLa cells (Fig. S7, movie S5). This result can appear in contrast with recent data from other groups and from us obtained with another DNA labeling technology, EdU (Francis et al., 2020) (Rensen et al., BioRxiv, 2020), showing the vDNA and the IN strongly colocalized with CPSF6. This apparent discrepancy can be explained by the intrinsic features of vDNA labeling tools. EdU system is capable of detecting nascent episomal forms of vDNA abundant in CPSF6 clusters, sites of nuclear reverse transcription (Rensen et al.,BioRxiv

doi.org/10.1101/2020.04.12.038067). Of note, EdU labeling cannot provide direct information on viral replication, as EdU reduces viral transcription (Rensen et al., BioRxiv doi.org/10.1101/2020.04.12.038067). Contrariwise, HIV-1 ANCHOR system allows the detection of late reverse transcripts (Fig.1A) that can generate functional proviral DNA, as we observed in HeLa cells, human lymphocytes, and macrophage-like cells (Fig. 2A,B,C, Fig.S2A,B, Fig. S5). Thus, we combined HIV-1 DNA and HIV-1 RNA labeling to detect the nuclear landscape where functional viral forms are located. Interestingly, we observed that HIV-1 induces the remodeling of speckles characterized by the presence of SC35 and SON, in novel MLOs enriched in viral proteins, such as IN and CA, as well in cellular factors, like the paraspeckle member, CPSF6, (Fig. 4B, 5A,B, S8). Of note, we observed that CPSF6 is fully included inside organelles marked by SC35 during viral infection, otherwise CPSF6 is usually randomly dispersed in the host nucleus of uninfected cells (Fig.4B, Fig.5A). Likely, CPSF6 relocates to speckles prompted by the virus to usurp functions linked to these MLOs. We show that the formation of CPSF6 cluster is independent of viral DNA integration or viral DNA synthesis (Fig. S8B, S9) and that late reverse transcribed products are excluded from them (Fig. 5A). It is possible that once the vRNA is fully retrotranscribed in vDNA (Rensen et al., BioRxiv doi.org/10.1101/2020.04.12.038067), it will quickly leave the HIV-1 MLOs. In fact, vDNAs were located on average at a distance of ~ 600 nm from the boundary of SC35 (Fig. 5A). MCP-labelled vRNA foci shows that actively transcribing proviruses are also located outside HIV-1 MLOs, indicating that those are not sites of viral integration and replication (Fig.5B,C). We also observed that CPSF6 relocates to the vicinity of LEDGF during HIV infection (Fig. 6A). LEDGF is required for the targeting of highly-spliced transcriptional units through its direct interaction with several splicing factors (Singh et al., 2015). LEDGF is also known to determine HIV-1 integration sites distribution in active chromatin sites (Ciuffi et al., 2005; Ferris et al., 2010; Shun et al., 2007), however the visualization of a complex formed by the transcribing provirus and LEDGF remained utopian for a long time. Here, we were able to visualize by fluorescence microscopy a complex formed by the vDNA or vRNA and LEDGF. More than 80% of those complexes are formed by LEDGF uncoupled to CPSF6 (Fig.6B). However, we observed that less than 20% of them can include CPSF6 associated to LEDGF, suggesting that few proviruses need to be near to HIV-1 MLOs without co-localizing, against the majority of them that can replicate farther (Fig. 6B). Whether there is a functional difference between these two

complexes it can be further investigated in future studies. Overall, our study highlights the proviral location in the nuclear landscape remodelled by the viral infection (Fig.7). We found that HIV-1-MLOs retain the viral IN but exclude the late retrotranscribed DNA, indicating a role for HIV-1 MLOs prior to the integration step. In particular, we found foci of viral transcription outside HIV-1 MLOs, located in the nuclear periphery, in complex with LEDGF, indicating that HIV-1 MLOs cannot be sites of viral integration and viral transcription. Furthermore, the visualization of the formation of an actively transcribing provirus in complex of a pre-existing cluster of LEDGF, which is known to lay on active chromatin sites (Ciuffi et al., 2005; Singh et al., 2015) (Fig.7), indicates that LEDGF is not part of the viral complex during nuclear entry (Fig.6A). Taken together our results show that functional proviruses establish near to the NE, excluded from CPSF6 clusters, but locate in a favorable transcriptional environment ~ 600nm from HIV-1 MLOs (Fig.7). Our study supports how new single-cell level approaches are pivotal in the study of functional viral nuclear dynamics, to discriminate the cells susceptible to fuel viremia or that concur to viral persistence.

# Materials and Methods

**Cell lines, primary cells and generation of genetically modified single-cell clones.**

HEK 293T cells (ATCC) are human embryonic kidney cells used to produce lentiviral vectors and HIV-1 viruses. HeLa P4R5 reporter cells are HeLa $CD4^+$ $CCR5^+$ $CXCR4^+$, carrying the LacZ gene under the control of the HIV-1 LTR promoter (Charneau et al., 1994). HeLa MCP-GFP is a HeLa clone stably expressing MCP-GFP bacterial fusion protein (kind gift from E. Bertrand) (Tantale et al., 2016). Jurkat cells (ATCC) are human immortalized T lymphocytes derived from acute T cells leukemia. THP-1 cells (ATCC) are human immortalized monocyte cells derived from acute monocytic leukemia, which, once seeded, were differentiated into macrophage-like cells under Phorbol 12-myristate 13-acetate (PMA) treatment (167nM). HEK293T cells and HeLa cells were cultivated in DMEM medium supplemented with 10% Fetal Bovine Serum (FBS) and 1% penicillin-streptomycin. Jurkat and THP-1 cells were cultivated in RPMI 1640 medium supplemented with 10 % Fetal Bovine Serum (FBS) and 1% penicillin-streptomycin. Primary $CD4^+$T cells were purified from healthy donors' blood obtained

via EFS (Etablissement Français du Sang, Paris), through density gradient centrifugation with Ficoll 400 and isolation of PBMCs buffy coat. Subsequent positive selection was performed with human $CD4^+$ Microbeads (Miltenyi Biotec #130-045-101). The day after $CD4^+$ T cells were activated with T cell Activation/Expansion kit (Miltenyi Biotec #130-091-441) according to protocol, complete RPMI 1640 medium was supplemented with interleuykin-2 (IL-2) (400 IU/mL). HIV-1 ANCH3 single-cell clone was generated starting from HeLa P4R5 cells which were infected with HIV-1 ANCH3 virus at MOI 1. Cells were diluted to 1 cell per well, in 96-well plates. Cell-clone colonies were tested for β-galactosidase expression to check viral expression (kit Roche #11758241001). Positive clones were transduced with lentiviral vector (LV) CMV-OR-GFP for imaging analyses of HIV-1 ANCH3 provirus. All cells were kept in incubator at 37°C and 5% $CO_2$.

**Plasmids and lentiviral vector productions**

HIV-1ΔEnvIN$_{HA}$ΔNef plasmid encodes for the ΔEnvHIV-1 LAI (BRU) viral genome where the Integrase (IN) protein is fused to the hemagglutinin (HA) tag (kindly provided by F. Mammano)(Petit et al., 1999; Petit et al., 2000). HIV-1ΔEnvIN$_{HA}$ΔNef ANCH3 (HIV-1 ANCH3) was obtained through ANCH3 insertion: ANCH3 sequence was cloned by PCR using the template plasmid pANCH3 as we previously described in Blanco et al. (Blanco-Rodriguez et al., 2020). The ANCHOR$^{TM}$ technology and sequences are exclusive property of NeoVirTech (Germier et al., 2017; Mariame et al., 2018). HIV-1ΔEnv IN$_{HA}$(D116A) ΔNef ANCH3 was obtained by insertional mutagenesis using the QuikChange II XL Site-Directed Mutagenesis kit (Agilent #200522), for integrase (IN) mutation at D116. The lentiviral vector plasmids CMV OR-GFP and CMV OR-SANTAKA were obtained by cloning OR-GFP/OR-SANTAKA (plamids from NeoVirTech) in a pTripCMV (ΔU3) plasmid through PCR and cloned using restriction site, AgeI and SgrDI**.** HIV-1ΔEnvIN$_{HA}$ΔNef ANCH3 MS2 plasmid was obtained by inserting the MS2x64 sequence amplified by PCR from pMK123**-**MS2x64 (Tantale et al., 2016) in HIV-1ΔEnvIN$_{HA}$ΔNef ANCH3. Lentiviral vectors and HIV-1 viruses were produced by transient transfection of HEK293T cells through calcium chloride coprecipitation. Co-transfection was performed as following, LV carrying OR-GFP/OR-SANTAKA: 10 µg of OR-GFP/OR-SANTAKA LV, 10 µg of NDK plasmid (gag-pol-tat-rev) and 2.5 µg of pHCMV-VSV-G envelope expression plasmid; LV carrying MCP-GFP (kindly provided by E. Bertand): 10 µg of MCP-GFP, 10 µg of NDK plasmid (gag-

pol-tat-rev) and 2.5 µg of pHCMV-VSV-G envelope expression plasmid; HIV-1ΔEnv pseudotyped with VSV-G: 10 µg HIV-1ΔEnvIN$_{HA}$ΔNef (IN WT or IN(D116A)) plasmid and 2.5 ug of pHCMV-VSV-G plasmid. VSV-G/HIV-1ΔEnv IN$_{HA}$(D116A) ΔNef ANCH3 has been also produced in combination with GIR (Gag-IN-Ruby plasmid, kindly provided by E. Campbell) (Dharan et al., 2016; Hulme et al., 2015). For THP-1 infection, HIV-1 ANCH3 or HIV-1 ANCH3 MS2 viruses and LV OR-GFP or LV MCP-GFP were all produced adding 3 µg of SIV$_{MAC}$ Vpx (Durand et al., 2013) at the moment of HEK293T transfection. After the collection of the supernatant 48h post-transfection, lentiviral particles were concentrated by ultracentrifugation for 1 h at 22000 rpm at 4°C and stored at −80°C. Lentiviral vectors and viruses were tittered through late reverse transcripts (LTR) qPCR in HEK293T cells or HeLa P4R5 cells.

**Quantitative PCR and primers**

Total DNA of infected cells was extracted at 6 hours post infection for Late Reverse Transcripts (LRTs) qPCR and 24 hours post infection for 2LTR-circles and ALU-PCR, through QIAmp DNA micro kit (QIAGEN #56304). Real-Time PCR of LRTs was performed to assess DNA synthesis and used to normalize qPCR data on viral input. The reactions were carried on in 20 µL, in iTaqUniversal SYBR Green Supermix (Bio-Rad #1725124) using primers for U3 sequence: U3 FX: 5'-TTCCGCTGGGGACTTTCCAGGG-3', U3 RX: 5'-AGGCTCAGATCTGGTCTAACC-3'. Real-Time PCR of 2LTR-circles was used to asses nuclear import efficiency. Reactions were performed in 20 µL, in Maxima Probe/ROX qPCR Mastermix (ThermoFisher #K0232) using primers for 2LTR-circle junction: 2LTR FX: 5'-AACTAGGGAACCCACTGCTTAAG-3', 2LTR RX: 5′-TCCACAGATCAAGGATATCTTGTC-3′, 2-LTR probe: 5'-(FAM)-ACACTACTTGAAGCACTCAAG-GCAAGCTTT-(TAMRA)-3'. Real-Time PCR of proviral integrations consisting in a first non-kinetic PCR in 50 µL in Platinum SuperFi DNA Polymerase (ThermoFisher #12351250) and in a second step of qPCR reaction in 20 µL in Maxima Probe/ROX qPCR Mastermix (ThermoFisher #K0232). Primer Alu 166: 5'-TCCCAGCTACTCGGGAGGCTGAGG-3', Alu 2: 5'-GCCTCCCAAAGTGCTGGGATTACAG-3', LambdaU3: 5′-ATGCCACGTAAGCGAAACTTTCCGCTGGGGACTTTCCAGGG-3′ for ALU PCR; Lambda: 5'-ATGCCACGTAAGCGAAACT-3', U5: 5'-CTGACTAAAAGGGTCTGAGG-3', Probe: 5'-(FAM)- TTAAGCCTCAATAAAGCTTGCCTTGAGTGC-(TAMRA) for ALU

546 qPCR(Di Nunzio et al., 2013). In all experiments β-actin detection was used for
547 normalization. β-actin FX: 5'-AACACCCCAGCCATGTACGT-3', β-actin RX: 5-
548 CGGTGAGGATCTTCATGAGGTAGT-3', β-actin probe: (FAM)-
549 CCAGCCAGGTCCAGACGCAGGA-(BHQ1).
550
551 **Epifluorescence and Immunofluorescence**
552 For imaging studies in adherent cell-lines, the cells were plated on coverslips (12mm
553 diameter in 24 well plates or 18 mm diameter in 12 well plate, #1, ThermoFisher) at
554 least 24h before fixation. For HIV-1 ANCH3 imaging, HeLa P4R5 cells were
555 transduced with OR-GFP LV (MOI 0.2) and then infected within the following 3 days
556 with HIV-1 ANCH3 or HIV-1 $IN_{D116A}$ ANCH3 at different MOIs. NEVIRAPINE (NEV) 10
557 µM or RALTEGRAVIR (RAL) 20 µM were used to block respectively DNA synthesis
558 and integration. For HIV-1 MS2 RNA imaging HeLa-MCP cells were infected with HIV
559 ANCH3 MS2 and fixed 24h post infection. For co-visualization of vDNA and vRNA,
560 HeLa MCP-GFP cells were transduced with OR-SANTAKA LV (MOI 1) and 24h later
561 infected with HIV-1 ANCH3 MS2. All THP-1 cells studies were carried on after 48h of
562 differentiation with PMA (167 nM), then, they were transduced with LV OR-GFP Vpx
563 (MOI 5) or LV MCP-GFP Vpx (MOI 20) and 48 h later they were infected with HIV-1
564 ANCH3 Vpx (30 h or 3 days p.i) or HIV-1 ANCH3 MS2 ± Vpx (3 days p.i.). The medium
565 was always supplemented with PMA. For HIV-1 ANCH3 imaging in primary
566 lymphocytes, $CD4^+$T cells, 3 days after activation, were transduced with OR-GFP LV
567 (MOI 0.5) for 72 h and then infected with HIV-1 ANCH3 for 20 h. The day of fixation
568 cells in suspension were seeded on poly-L-lysine (Sigma #P4707) coated coverslips
569 and centrifuged to allow attachment. The day of fixation all cells were washed with
570 PBS and fixed with 4% PFA for 15 minutes. For protein staining, cells were treated
571 with glycine 0.15% for 10 min, permeabilized with Triton X-100 0.5% for 30 min and
572 blocked with 1% bovine serum albumin (BSA) for 30 min. All incubations were carried
573 out at room temperature, in the dark, in humid chamber, 1h with primary antibodies
574 and 45 min with secondary antibodies. Washes between antibody incubations and
575 antibodies dilution were done in 1% BSA. Primary antibodies were diluted as follows:
576 anti-HA 1:500 (Roche #11867423001), anti-HIV-1 p24 (CA) 1:400 (NIH #3537), anti-
577 CPSF6 1:400 (Novus Biologicals #NBP1-85676), anti-SC35 1:200 (Abcam #ab11826),
578 anti-LEDGF 1:200 (BD Bioscience #611715). Secondary antibodies used were the

following: Goat anti-Mouse Alexa-647 1:300 (1:100 for CA) (Invitrogen #A21235); Donkey anti-Rat Alexa-488 1:100 (Invitrogen A21208) for IN-HA) or Goat anti-Rat Alexa-647 1:300 (Invitrogen #A21247) (1:100 for IN-HA); Donkey anti-Rabbit Cy3 1:500 (Jackson Lab #711-165-152). Finally, cells were stained with Hoechst 33342 1:5000 (Invitrogen #H3570) for 5 minutes. Coverslips were mounted on glass slides (Star Frost) with Prolong Diamond Antifade Mountant (Life Technologies #P36970). Confocal microscopy was carried out with a Zeiss LSM700 inverted microscope, with a 63X objective (Plan Apochromat, oil immersion, NA=1.4).

**RNA FISH**

For RNA FISH studies in adherent cell-lines, the cells were seeded on coverslips. HeLa P4R5 cells expressing OR-GFP were infected with HIV-1 ANCH3 (MOI 30) with or without RAL 20 µM and fixed 24h post infection or infected with HIV-1 ANCH3 (MOI 5) for 48h or 7 days. Jurkat T cells expressing OR-GFP were infected with HIV-1 ANCH3 (MOI 5) and fixed 48h post infection. Primary activated $CD4^+$T cells, 3 days after activation, were together transduced with OR-GFP LV (MOI 10) and infected with HIV-1 ANCH3 (MOI 5) for 3 days,.THP-1 cells were differentiated with PMA for 48h and, after 48h transduction with LV OR-GFP Vpx (MOI 5), the cells were infected with HIV-1 ANCH3 Vpx (MOI 20) for 3 days. The day of fixation Jurkat and $CD4^+$T cells were seeded on poly-L-lysine (Sigma #P4707) coated coverslips and centrifuged to allow attachment. The day of fixation all cells were washed with PBS and fixed with 4% PFA for 15 minutes and incubated in 70% ethanol at -20°C at least for one night. Primary smiFISH probes have a targeting sequence against HIV-1 *pol* transcript and a shared readout sequence for secondary probe alignment. Twenty-four smiFISH probes (table I) against HIV pol were designed with Oligostan (Tsanov et al., 2016) and purchased from Integrated DNA Technologies (IDT). Primary probes were pre-hybridized with a secondary FLAP probe conjugated to Cy5 fluorophore through pairing with the readout sequence. Washes and hybridization were performed with Stellaris Buffers (WASH buffer A, WASH buffer B, Hybridization Buffer; LGC Biosearch Technologies), following the manufacturer protocol. Hybridization with the probe was carried out at 37°C in a dark humid chamber for 5 hours. Finally, cells were stained with Hoechst 33342 1:5000 (Invitrogen #H3570) for 5 minutes. Coverslips were mounted on glass slides (Star Frost) with Prolong Diamond Antifade Mountant (Life Technologies

#P36970). Confocal microscopy was carried out with a Zeiss inverted LSM700 microscope, with a 63X objective (Plan Apochromat, oil immersion, NA=1.4).

**Time-lapse microscopy**

For all live imaging studies, the cells were plated on a polymer-coverslip bottom µ-Dish 35 mm (ibidi #81156) or in µ-Dish 35 mm Quad (10,000 cells per chamber, ibidi #80416) and 2D or 3D videos were acquired with an UltraView VOX Spinning Disk Microscope (Perkin-Elmer), based on a CSU-X spinning-disk (Yokogawa), and using a 63X objective (Plan Apochromat, oil immersion, NA=1.4). For integrated provirus imaging, HIV-1 ANCH3 single-cell clone was transduced with OR-GFP LV (MOI 0.2) for 3 days and then plated on the polymer-bottom dish ($10^5$ cells) the day before acquisition. On the other hand, HeLa P4R5 cells were infected with HIV-1 ANCH3 (MOI 10) and five days post infection cells were plated on the polymer-coverslip bottom dish and transduced with LV OR-GFP for 24h. For non-integrated viral DNA, HeLa P4R5 were firstly transduced with OR-GFP LV (MOI 0.2) and the day after cells were plated on polymer-bottom dish and infected with HIV-1 ANCH3 (MOI 10) with RAL 20 µM or with the mutated HIV-1 $IN_{HA}$ (D116A) ANCH3. Cells were imaged in continue for 10 minutes. Experiments of co-live tracking of IN-Ruby (GIR) and vDNA labeled by OR-GFP were performed in HeLa P4R5 cells transduced with LV OR-GFP (MOI 0.2). The following day, cells were infected with HIV-1 $IN_{HA}$(D116A) ANCH3 complemented with the GIR plasmid using an MOI 50. Fluorescence images were taken every 30 seconds for up to 9h post infection. For co-live-imaging of vDNA and vRNA foci, HeLa MCP-GFP cells were transduced with OR-SANTAKA LV (MOI 1) and 24 h later infected with HIV-1 ANCH3 MS2 (MOI 50).

**Imaging Analysis and Statistics**

All data were analyzed in GraphPad Prism 8 (GraphPad Software, La Jolla California USA, www.graphpad.com), computing Student's t test for 2 comparisons, One-way ANOVA for more than two comparisons and Pearson's correlation. P values symbols: ns = non significant, * = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$, **** = $p \leq 0.0001$. All images and videos were analyzed in Fiji (Schindelin et al., 2012) and Icy version 2.0.3.0 (de Chaumont et al., 2012) software.

643 All images and video were processed and analyzed in Fiji (Schindelin et al., 2012) or
644 in Icy software version 2.0.3.0 (de Chaumont et al., 2012). Here below the custom-
645 made protocols:
646 *Particle tracking and MSD measurement*
647 3D videos were analyzed in Fiji (Schindelin et al., 2012) and viral nuclear particles
648 were tracked along time in 3D using TrackMate plugin (Tinevez et al., 2017). The
649 resulting tracks were exported in MATLAB (The MathWorks, Natick, USA) and then
650 used for mean-squared displacement (MSD) analysis with @msdanalyzer(Tarantino
651 et al., 2014). MSD curves were fitted by a straight line from which we derived the value
652 of the diffusion coefficient: D = a / (2*nDim), where « a » is the slope of the MSD curve
653 and « nDim » the dimensionality of the image. MSD curves for which the $R^2$ value was
654 lower than 0.8 were discarded from analysis. Only the initial 25% of the MSD curves
655 were used for the linear fit.
656 *vDNA-vRNA association analysis*
657 For the analysis of RNA FISH images ± RAL in HeLa P4R5 cells, 2D confocal images
658 were analysed with Icy software version 2.0.3.0 (de Chaumont et al., 2012) using a
659 custom made protocol (http://icy.bioimageanalysis.org/protocol/batch-spots-
660 detections-and-statistics and doi: 10.5281/zenodo.3925262, example of HeLa
661 confocal image available at doi: 10.5281/zenodo.3925213). vDNA spots were detected
662 extracting the green channel and applying first a median filter of half size = 1 and then
663 3, the Spot Detector with a scale of 1 (sensitivity to 55) on the nuclear ROI only. Data
664 from the spot detection (e.g. mean fluorescence intensity) were exported for analysis.
665 vRNA mean fluorescence intensity signal was computed from the nuclei ROI. For
666 vDNA - vRNA association studies in Jurkat cells, 3D confocal images were analyzed
667 in Icy version 2.0.3.0 using two successive custom made protocols (first protocol:
668 http://icy.bioimageanalysis.org/protocol/nuclei-segmentation-from-maximum-
669 intensity-projections/ with doi: 10.5281/zenodo.3925181) and second protocol:
670 http://icy.bioimageanalysis.org/protocol/spots-detection-and-colocalisation-analysis-
671 with-soda/ with doi: 10.5281/zenodo.3925191, example of Jurkat confocal image
672 available at doi: 10.5281/zenodo.3925213). Nuclei were segmented on a maximum
673 intensity projection of the Hoechst channel using the Icy block Gaussian filter with a
674 radius of sigma x=5, y=5 and z=5 pixels, the Thresholder block with a manual threshold
675 of 5000 and the Label Extractor block, which uses connected components to separate
676 the nuclei. Cy5 spots were detected with the HK Means block (Dufour et al.,

doi.org/10.1109/ICPR.2008.4761748).The logical operation block was used to keep only the spots detected in the nucleus. These spots were added as input parameter "List detection 1" of the SODA block. EGFP spots were detected using the log3D block to enhance the spots, inverting the log transformed image using the Math Operation Expression block with the following expression: A*B with A scalar=-1 and B the log transformed image, detecting the spots with the Local Extrema block and segmenting the spots with a dilation and a connected components operations. The logical operation block was used to keep only the spots detected in the nucleus. These spots were added as input parameter "List detection 2" of the SODA block. The nucleus ROI was kept as input parameter for "ROI of Analysis (Cell's shape)". The Cy5 and GFP spots ROIs were then imported in the SODA block (Lagache et al., 2018) to analyse colocalization of spots within the nucleus only with a maximum radius of 10, a step of 1 and no fixed search distance.

*vDNA/vRNA nuclear envelope distance analysis*

For the analysis of HeLa P4R5 cells 48 h p.i. and 7 days p.i. and in THP-1 3 days p.i. 3D RNA FISH images, the nuclei were segmented on the Hoechst channel using background subtraction followed by a multilevel Otsu thresholding. The resulting nuclei masks were then directly applied on the vDNA and vRNA channels to select the relevant 3D information inside each labelled region. Inside these regions, vRNA spots were identified using 3D non-local means denoising (Buades et al., "Non-local means denoising." *Image Processing On Line* 1 (2011): 208-212) and automated segmentation, while vDNA spots were extracted with the image analysis pipeline previously described (Komatsu et al., 2018). For each vDNA and vRNA spot in was computed the distance to the nuclear envelope in the 2D z-plane corresponding to the spot centroid location.

*SC35-vDNA distance analysis*

The 3D confocal images were processed for multi-channel image splitting, vDNA and SC35 segmentation. The automated 3D segmentation of cell nuclei included a preliminary non-local means denoising step to cope with the strong signal heterogeneity and the segmentation of SC35 was readjusted to low contrast conditions. The 3D boundary-to-boundary distance (rather than centroid-to-centroid which was done for vDNA/vRNA-NE analysis) between each vDNA spot and its closest SC35 speckle was computed.

## Acknowledgements

We wish to thank Guillermo Blanco-Rodriguez, Stella Frabetti, Philippe Souque and Blandine Monel for experimental help. We thank Nicoletta Casartelli for critical reading of the manuscript. We thank Fabrizio Mammano, Edouard Bertrand, Florian Mueller, and Edward Campbell for sharing reagents. We gratefully acknowledge the UtechS Photonic BioImaging (Imagopole), C2RT, Institut Pasteur, supported by the French National Research Agency (France BioImaging; ANR-10–INSB–04; Investments for the Future). We thank the NIH AIDS Reagents program to support us with precious reagents. This work was funded by the ANRS (Agence Nationale de Recherche sur le SIDA) grant ECTZ88162 with a nominative PhD student fellowship ECTZ88177 for V.S., the Sidaction/FRM grant VIH20170718001, the Pasteur Institute.

**Competing financial interests:** The authors declare no competing financial interests.

## References

Achuthan, V., J.M. Perreira, G.A. Sowd, M. Puray-Chavez, W.M. McDougall, A. Paulucci-Holthauzen, X. Wu, H.J. Fadel, E.M. Poeschla, A.S. Multani, S.H. Hughes, S.G. Sarafianos, A.L. Brass, and A.N. Engelman. 2018. Capsid-CPSF6 Interaction Licenses Nuclear HIV-1 Trafficking to Sites of Viral DNA Integration. *Cell Host Microbe*. 24:392-404 e398.

Albanese, A., D. Arosio, M. Terreni, and A. Cereseto. 2008. HIV-1 pre-integration complexes selectively target decondensed chromatin in the nuclear periphery. *PLoS One*. 3:e2413.

Ballandras-Colas, A., D.P. Maskell, E. Serrao, J. Locke, P. Swuec, S.R. Jonsson, A. Kotecha, N.J. Cook, V.E. Pye, I.A. Taylor, V. Andresdottir, A.N. Engelman, A. Costa, and P. Cherepanov. 2017. A supramolecular assembly mediates lentiviral DNA integration. *Science*. 355:93-95.

Bejarano, D.A., K. Peng, V. Laketa, K. Borner, K.L. Jost, B. Lucic, B. Glass, M. Lusic, B. Muller, and H.G. Krausslich. 2019. HIV-1 nuclear import in macrophages is regulated by CPSF6-capsid interactions at the nuclear pore complex. *Elife*. 8.

Blanco-Rodriguez, G., A. Gazi, B. Monel, S. Frabetti, V. Scoca, F. Mueller, O. Schwartz, J. Krijnse-Locker, P. Charneau, and F. Di Nunzio. 2020. Remodeling of the core leads HIV-1 pre-integration complex in the nucleus of human lymphocytes. *J Virol*.

Buffone, C., A. Martinez-Lopez, T. Fricke, S. Opp, M. Severgnini, I. Cifola, L. Petiti, S. Frabetti, K. Skorupka, K.K. Zadrozny, B.K. Ganser-Pornillos, O. Pornillos, F. Di Nunzio, and F. Diaz-Griffero. 2018. Nup153 Unlocks the Nuclear Pore Complex for HIV-1 Nuclear Translocation in Nondividing Cells. *J Virol*. 92.

749 Burdick, R.C., K.A. Delviks-Frankenberry, J. Chen, S.K. Janaka, J. Sastri, W.S. Hu, and V.K. Pathak. 2017.
750 Dynamics and regulation of nuclear import and nuclear movements of HIV-1 complexes. *PLoS*
751 *Pathog*. 13:e1006570.
752 Burdick, R.C., C. Li, M. Munshi, J.M.O. Rawson, K. Nagashima, W.S. Hu, and V.K. Pathak. 2020. HIV-1
753 uncoats in the nucleus near sites of integration. *Proc Natl Acad Sci U S A*. 117:5486-5493.
754 Bushman, F.D. 2003. Targeting survival: integration site selection by retroviruses and LTR-
755 retrotransposons. *Cell*. 115:135-138.
756 Butler, S.L., M.S. Hansen, and F.D. Bushman. 2001. A quantitative assay for HIV DNA integration in vivo.
757 *Nat Med*. 7:631-634.
758 Campbell, E.M., and T.J. Hope. 2015. HIV-1 capsid: the multifaceted key player in HIV-1 infection. *Nat*
759 *Rev Microbiol*. 13:471-483.
760 Charneau, P., G. Mirambeau, P. Roux, S. Paulous, H. Buc, and F. Clavel. 1994. HIV-1 reverse
761 transcription. A termination step at the center of the genome. *J Mol Biol*. 241:651-662.
762 Cherepanov, P., G. Maertens, P. Proost, B. Devreese, J. Van Beeumen, Y. Engelborghs, E. De Clercq,
763 and Z. Debyser. 2003. HIV-1 integrase forms stable tetramers and associates with LEDGF/p75
764 protein in human cells. *J Biol Chem*. 278:372-381.
765 Chin, C.R., J.M. Perreira, G. Savidis, J.M. Portmann, A.M. Aker, E.M. Feeley, M.C. Smith, and A.L. Brass.
766 2015. Direct Visualization of HIV-1 Replication Intermediates Shows that Capsid and CPSF6
767 Modulate HIV-1 Intra-nuclear Invasion and Integration. *Cell Rep*. 13:1717-1731.
768 Ciuffi, A., M. Llano, E. Poeschla, C. Hoffmann, J. Leipzig, P. Shinn, J.R. Ecker, and F. Bushman. 2005. A
769 role for LEDGF/p75 in targeting HIV DNA integration. *Nat Med*. 11:1287-1289.
770 Craig, E.A., and J. Marszalek. 2002. A specialized mitochondrial molecular chaperone system: a role in
771 formation of Fe/S centers. *Cell Mol Life Sci*. 59:1658-1665.
772 de Chaumont, F., S. Dallongeville, N. Chenouard, N. Herve, S. Pop, T. Provoost, V. Meas-Yedid, P.
773 Pankajakshan, T. Lecomte, Y. Le Montagner, T. Lagache, A. Dufour, and J.C. Olivo-Marin. 2012.
774 Icy: an open bioimage informatics platform for extended reproducible research. *Nat Methods*.
775 9:690-696.
776 De Wit, F., S.R. Pillalamarri, A. Sebastian-Martin, A. Venkatesham, A. Van Aerschot, and Z. Debyser.
777 2019. Design of reverse transcriptase-specific nucleosides to visualize early steps of HIV-1
778 replication by click labeling. *J Biol Chem*. 294:11863-11875.
779 Dharan, A., S. Talley, A. Tripathi, J.I. Mamede, M. Majetschak, T.J. Hope, and E.M. Campbell. 2016.
780 KIF5B and Nup358 Cooperatively Mediate the Nuclear Import of HIV-1 during Infection. *PLoS*
781 *Pathog*. 12:e1005700.
782 Di Nunzio, F. 2013. New insights in the role of nucleoporins: a bridge leading to concerted steps from
783 HIV-1 nuclear entry until integration. *Virus Res*. 178:187-196.
784 Di Nunzio, F., T. Fricke, A. Miccio, J.C. Valle-Casuso, P. Perez, P. Souque, E. Rizzi, M. Severgnini, F.
785 Mavilio, P. Charneau, and F. Diaz-Griffero. 2013. Nup153 and Nup98 bind the HIV-1 core and
786 contribute to the early steps of HIV-1 replication. *Virology*. 440:8-18.
787 Di Primio, C., V. Quercioli, A. Allouch, R. Gijsbers, F. Christ, Z. Debyser, D. Arosio, and A. Cereseto. 2013.
788 Single-cell imaging of HIV-1 provirus (SCIP). *Proc Natl Acad Sci U S A*. 110:5636-5641.
789 Durand, S., X.N. Nguyen, J. Turpin, S. Cordeil, N. Nazaret, S. Croze, R. Mahieux, J. Lachuer, C. Legras-
790 Lachuer, and A. Cimarelli. 2013. Tailored HIV-1 vectors for genetic modification of primary
791 human dendritic cells and monocytes. *J Virol*. 87:234-242.
792 Emiliani, S., A. Mousnier, K. Busschots, M. Maroun, B. Van Maele, D. Tempe, L. Vandekerckhove, F.
793 Moisant, L. Ben-Slama, M. Witvrouw, F. Christ, J.C. Rain, C. Dargemont, Z. Debyser, and R.
794 Benarous. 2005. Integrase mutants defective for interaction with LEDGF/p75 are impaired in
795 chromosome tethering and HIV-1 replication. *J Biol Chem*. 280:25517-25523.
796 Farnet, C.M., and W.A. Haseltine. 1991. Determination of viral proteins present in the human
797 immunodeficiency virus type 1 preintegration complex. *J Virol*. 65:1910-1915.
798 Ferris, A.L., X. Wu, C.M. Hughes, C. Stewart, S.J. Smith, T.A. Milne, G.G. Wang, M.C. Shun, C.D. Allis, A.
799 Engelman, and S.H. Hughes. 2010. Lens epithelium-derived growth factor fusion proteins
800 redirect HIV-1 DNA integration. *Proc Natl Acad Sci U S A*. 107:3135-3140.

Fox, A.H., Y.W. Lam, A.K. Leung, C.E. Lyon, J. Andersen, M. Mann, and A.I. Lamond. 2002. Paraspeckles: a novel nuclear domain. *Curr Biol*. 12:13-25.

Francis, A.C., C. Di Primio, V. Quercioli, P. Valentini, A. Boll, G. Girelli, F. Demichelis, D. Arosio, and A. Cereseto. 2014. Second generation imaging of nuclear/cytoplasmic HIV-1 complexes. *AIDS Res Hum Retroviruses*. 30:717-726.

Francis, A.C., M. Marin, J. Shi, C. Aiken, and G.B. Melikyan. 2016. Time-Resolved Imaging of Single HIV-1 Uncoating In Vitro and in Living Cells. *PLoS Pathog*. 12:e1005709.

Francis, A.C., M. Marin, P.K. Singh, V. Achuthan, M.J. Prellberg, K. Palermino-Rowland, S. Lan, P.R. Tedbury, S.G. Sarafianos, A.N. Engelman, and G.B. Melikyan. 2020. HIV-1 replication complexes accumulate in nuclear speckles and integrate into speckle-associated genomic domains. *Nat Commun*. 11:3505.

Francis, A.C., and G.B. Melikyan. 2018. Single HIV-1 Imaging Reveals Progression of Infection through CA-Dependent Steps of Docking at the Nuclear Pore, Uncoating, and Nuclear Transport. *Cell host & microbe*. 23:536-548 e536.

Geis, F.K., and S.P. Goff. 2019. Unintegrated HIV-1 DNAs are loaded with core and linker histones and transcriptionally silenced. *Proc Natl Acad Sci U S A*. 116:23735-23742.

Germier, T., S. Kocanova, N. Walther, A. Bancaud, H.A. Shaban, H. Sellou, A.Z. Politi, J. Ellenberg, F. Gallardo, and K. Bystricky. 2017. Real-Time Imaging of a Single Gene Reveals Transcription-Initiated Local Confinement. *Biophys J*. 113:1383-1394.

Graham, T.G., X. Wang, D. Song, C.M. Etson, A.M. van Oijen, D.Z. Rudner, and J.J. Loparo. 2014. ParB spreading requires DNA bridging. *Genes Dev*. 28:1228-1238.

Hare, S., F. Di Nunzio, A. Labeja, J. Wang, A. Engelman, and P. Cherepanov. 2009. Structural basis for functional tetramerization of lentiviral integrase. *PLoS Pathog*. 5:e1000515.

Hrecka, K., C. Hao, M. Gierszewska, S.K. Swanson, M. Kesik-Brodacka, S. Srivastava, L. Florens, M.P. Washburn, and J. Skowronski. 2011. Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. *Nature*. 474:658-661.

Hubner, W., G.P. McNerney, P. Chen, B.M. Dale, R.E. Gordon, F.Y. Chuang, X.D. Li, D.M. Asmuth, T. Huser, and B.K. Chen. 2009. Quantitative 3D video microscopy of HIV transfer across T cell virological synapses. *Science*. 323:1743-1747.

Hulme, A.E., Z. Kelley, D. Foley, and T.J. Hope. 2015. Complementary Assays Reveal a Low Level of CA Associated with Viral Complexes in the Nuclei of HIV-1-Infected Cells. *J Virol*. 89:5350-5361.

Hulme, A.E., O. Perez, and T.J. Hope. 2011. Complementary assays reveal a relationship between HIV-1 uncoating and reverse transcription. *Proceedings of the National Academy of Sciences of the United States of America*. 108:9975-9980.

Komatsu, T., C. Quentin-Froignant, I. Carlon-Andres, F. Lagadec, F. Rayne, J. Ragues, R.H. Kehlenbach, W. Zhang, A. Ehrhardt, K. Bystricky, R. Morin, J.M. Lagarde, F. Gallardo, and H. Wodrich. 2018. In Vivo Labelling of Adenovirus DNA Identifies Chromatin Anchoring and Biphasic Genome Replication. *J Virol*. 92.

Lagache, T., A. Grassart, S. Dallongeville, O. Faklaris, N. Sauvonnet, A. Dufour, L. Danglot, and J.C. Olivo-Marin. 2018. Mapping molecular assemblies with fluorescence microscopy and object-based spatial statistics. *Nat Commun*. 9:698.

Laguette, N., B. Sobhian, N. Casartelli, M. Ringeard, C. Chable-Bessia, E. Segeral, A. Yatim, S. Emiliani, O. Schwartz, and M. Benkirane. 2011. SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature*. 474:654-657.

Lee, K., Z. Ambrose, T.D. Martin, I. Oztop, A. Mulky, J.G. Julias, N. Vandegraaff, J.G. Baumann, R. Wang, W. Yuen, T. Takemura, K. Shelton, I. Taniuchi, Y. Li, J. Sodroski, D.R. Littman, J.M. Coffin, S.H. Hughes, D. Unutmaz, A. Engelman, and V.N. KewalRamani. 2010. Flexible use of nuclear import pathways by HIV-1. *Cell Host Microbe*. 7:221-233.

Lelek, M., N. Casartelli, D. Pellin, E. Rizzi, P. Souque, M. Severgnini, C. Di Serio, T. Fricke, F. Diaz-Griffero, C. Zimmer, P. Charneau, and F. Di Nunzio. 2015. Chromatin organization at the nuclear pore favours HIV replication. *Nat Commun*. 6:6483.

852 Lelek, M., F. Di Nunzio, R. Henriques, P. Charneau, N. Arhel, and C. Zimmer. 2012. Superresolution
853 imaging of HIV in infected cells with FlAsH-PALM. *Proc Natl Acad Sci U S A*. 109:8564-8569.
854 Liu, R., F.R. Simonetti, and Y.C. Ho. 2020. The forces driving clonal expansion of the HIV-1 latent
855 reservoir. *Virol J*. 17:4.
856 Llano, M., D.T. Saenz, A. Meehan, P. Wongthida, M. Peretz, W.H. Walker, W. Teo, and E.M. Poeschla.
857 2006. An essential role for LEDGF/p75 in HIV integration. *Science*. 314:461-464.
858 Maldarelli, F. 2016. The role of HIV integration in viral persistence: no more whistling past the proviral
859 graveyard. *J Clin Invest*. 126:438-447.
860 Mamede, J.I., G.C. Cianci, M.R. Anderson, and T.J. Hope. 2017. Early cytoplasmic uncoating is
861 associated with infectivity of HIV-1. *Proc Natl Acad Sci U S A*. 114:E7169-E7178.
862 Mariame, B., S. Kappler-Gratias, M. Kappler, S. Balor, F. Gallardo, and K. Bystricky. 2018. Real-Time
863 Visualization and Quantification of Human Cytomegalovirus Replication in Living Cells Using
864 the ANCHOR DNA Labeling Technology. *J Virol*. 92.
865 Marini, B., A. Kertesz-Farkas, H. Ali, B. Lucic, K. Lisek, L. Manganaro, S. Pongor, R. Luzzati, A. Recchia, F.
866 Mavilio, M. Giacca, and M. Lusic. 2015. Nuclear architecture dictates HIV-1 integration site
867 selection. *Nature*. 521:227-231.
868 Naganuma, T., and T. Hirose. 2013. Paraspeckle formation during the biogenesis of long non-coding
869 RNAs. *RNA Biol*. 10:456-461.
870 Olson, A., B. Basukala, W.W. Wong, and A.J. Henderson. 2019. Targeting HIV-1 proviral transcription.
871 *Curr Opin Virol*. 38:89-96.
872 Peng, K., W. Muranyi, B. Glass, V. Laketa, S.R. Yant, L. Tsai, T. Cihlar, B. Muller, and H.G. Krausslich.
873 2014. Quantitative microscopy of functional HIV post-entry complexes reveals association of
874 replication with the viral capsid. *Elife*. 3:e04114.
875 Petit, C., O. Schwartz, and F. Mammano. 1999. Oligomerization within virions and subcellular
876 localization of human immunodeficiency virus type 1 integrase. *J Virol*. 73:5079-5088.
877 Petit, C., O. Schwartz, and F. Mammano. 2000. The karyophilic properties of human immunodeficiency
878 virus type 1 integrase are not required for nuclear import of proviral DNA. *J Virol*. 74:7119-
879 7126.
880 Pradeepa, M.M., G.R. Grimes, G.C. Taylor, H.G. Sutherland, and W.A. Bickmore. 2014. Psip1/Ledgf p75
881 restrains Hox gene expression by recruiting both trithorax and polycomb group proteins.
882 *Nucleic Acids Res*. 42:9021-9032.
883 Pradeepa, M.M., H.G. Sutherland, J. Ule, G.R. Grimes, and W.A. Bickmore. 2012. Psip1/Ledgf p52 binds
884 methylated histone H3K36 and splicing factors and contributes to the regulation of alternative
885 splicing. *PLoS Genet*. 8:e1002717.
886 Price, A.J., D.A. Jacques, W.A. McEwan, A.J. Fletcher, S. Essig, J.W. Chin, U.D. Halambage, C. Aiken, and
887 L.C. James. 2014. Host cofactors and pharmacologic ligands share an essential interface in HIV-
888 1 capsid that is lost upon disassembly. *PLoS Pathog*. 10:e1004459.
889 Saad, H., F. Gallardo, M. Dalvai, N. Tanguy-le-Gac, D. Lane, and K. Bystricky. 2014. DNA dynamics during
890 early double-strand break processing revealed by non-intrusive imaging of living cells. *PLoS*
891 *Genet*. 10:e1004187.
892 Sanchez, A., D.I. Cattoni, J.C. Walter, J. Rech, A. Parmeggiani, M. Nollmann, and J.Y. Bouet. 2015.
893 Stochastic Self-Assembly of ParB Proteins Builds the Bacterial DNA Segregation Apparatus. *Cell*
894 *Syst*. 1:163-173.
895 Schindelin, J., I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden,
896 S. Saalfeld, B. Schmid, J.Y. Tinevez, D.J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A.
897 Cardona. 2012. Fiji: an open-source platform for biological-image analysis. *Nat Methods*.
898 9:676-682.
899 Schroder, A.R., P. Shinn, H. Chen, C. Berry, J.R. Ecker, and F. Bushman. 2002. HIV-1 integration in the
900 human genome favors active genes and local hotspots. *Cell*. 110:521-529.
901 Sharkey, M., D.Z. Babic, T. Greenough, R. Gulick, D.R. Kuritzkes, and M. Stevenson. 2011. Episomal viral
902 cDNAs identify a reservoir that fuels viral rebound after treatment interruption and that
903 contributes to treatment failure. *PLoS Pathog*. 7:e1001303.

Sharkey, M.E., I. Teo, T. Greenough, N. Sharova, K. Luzuriaga, J.L. Sullivan, R.P. Bucy, L.G. Kostrikis, A. Haase, C. Veryard, R.E. Davaro, S.H. Cheeseman, J.S. Daly, C. Bova, R.T. Ellison, 3rd, B. Mady, K.K. Lai, G. Moyle, M. Nelson, B. Gazzard, S. Shaunak, and M. Stevenson. 2000. Persistence of episomal HIV-1 infection intermediates in patients on highly active anti-retroviral therapy. *Nat Med*. 6:76-81.

Sharma, A., H. Takata, K. Shibahara, A. Bubulya, and P.A. Bubulya. 2010. Son is essential for nuclear speckle organization and cell cycle progression. *Mol Biol Cell*. 21:650-663.

Shun, M.C., Y. Botbol, X. Li, F. Di Nunzio, J.E. Daigle, N. Yan, J. Lieberman, M. Lavigne, and A. Engelman. 2008. Identification and characterization of PWWP domain residues critical for LEDGF/p75 chromatin binding and human immunodeficiency virus type 1 infectivity. *J Virol*. 82:11555-11567.

Shun, M.C., N.K. Raghavendra, N. Vandegraaff, J.E. Daigle, S. Hughes, P. Kellam, P. Cherepanov, and A. Engelman. 2007. LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. *Genes Dev*. 21:1767-1778.

Singh, P.K., M.R. Plumb, A.L. Ferris, J.R. Iben, X. Wu, H.J. Fadel, B.T. Luke, C. Esnault, E.M. Poeschla, S.H. Hughes, M. Kvaratskhelia, and H.L. Levin. 2015. LEDGF/p75 interacts with mRNA splicing factors and targets HIV-1 integration to highly spliced genes. *Genes Dev*. 29:2287-2297.

Spector, D.L., and A.I. Lamond. 2011. Nuclear speckles. *Cold Spring Harb Perspect Biol*. 3.

Tantale, K., F. Mueller, A. Kozulic-Pirher, A. Lesne, J.M. Victor, M.C. Robert, S. Capozi, R. Chouaib, V. Backer, J. Mateos-Langerak, X. Darzacq, C. Zimmer, E. Basyuk, and E. Bertrand. 2016. A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting. *Nat Commun*. 7:12248.

Tarantino, N., J.Y. Tinevez, E.F. Crowell, B. Boisson, R. Henriques, M. Mhlanga, F. Agou, A. Israel, and E. Laplantine. 2014. TNF and IL-1 exhibit distinct ubiquitin requirements for inducing NEMO-IKK supramolecular structures. *J Cell Biol*. 204:231-245.

Tinevez, J.Y., N. Perry, J. Schindelin, G.M. Hoopes, G.D. Reynolds, E. Laplantine, S.Y. Bednarek, S.L. Shorte, and K.W. Eliceiri. 2017. TrackMate: An open and extensible platform for single-particle tracking. *Methods*. 115:80-90.

Tsanov, N., A. Samacoits, R. Chouaib, A.M. Traboulsi, T. Gostan, C. Weber, C. Zimmer, K. Zibara, T. Walter, M. Peter, E. Bertrand, and F. Mueller. 2016. smiFISH and FISH-quant - a flexible single RNA detection approach with super-resolution capability. *Nucleic Acids Res*. 44:e165.

Vansant, G., H.C. Chen, E. Zorita, K. Trejbalova, D. Miklik, G. Filion, and Z. Debyser. 2020. The chromatin landscape at the HIV-1 provirus integration site determines viral expression. *Nucleic Acids Res*.

Wang, G.P., B.L. Levine, G.K. Binder, C.C. Berry, N. Malani, G. McGarrity, P. Tebas, C.H. June, and F.D. Bushman. 2009. Analysis of lentiviral vector integration in HIV+ study subjects receiving autologous infusions of gene modified CD4+ T cells. *Mol Ther*. 17:844-850.

Wang, H., M.C. Shun, X. Li, F. Di Nunzio, S. Hare, P. Cherepanov, and A. Engelman. 2014. Efficient Transduction of LEDGF/p75 Mutant Cells by Gain-of-Function HIV-1 Integrase Mutant Viruses. *Mol Ther Methods Clin Dev*. 1.

Zhu, Y., G.Z. Wang, O. Cingoz, and S.P. Goff. 2018. NP220 mediates silencing of unintegrated retroviral DNA. *Nature*. 564:278-282.

# FIGURE LEGENDS

**Figure 1. Detection and live tracking of nuclear HIV-1 DNA forms. A)** Scheme of vDNA labeled by the HIV-1 ANCHOR system. ANCH3 is the target sequence of the OR protein. ANCH3 has been cloned in the HIV-1 genome in place of Nef gene (HIV-1 ANCH3). The late retrotranscribed DNA of HIV-1 ANCH3 can be visualized in cells stably expressing the OR-GFP protein, thanks to the accumulation of OR protein to the double-stranded ANCH3 sequence. Viral DNA detection occurs mostly in the nucleus, where late reverse transcripts are abundant and are more likely to be exposed to the binding of OR-GFP. Starting from early time points of infection, it is possible to follow HIV-1 ANCH3 in target cells. The labelling results in the appearance of a bright fluorescent spot that can be tracked over time (time post-infection is indicated in the upper-left corner). The time-lapse images are representative of three independent experiments. The cartoon model was created with BioRender.com. **B)** Confocal images of HeLa P4R5 cells expressing OR-GFP infected at MOI 30 with HIV-1 ANCH3, HIV-1 ANCH3 + RAL (20 µM), HIV-1 ANCH3 $IN_{D116A}$ and HIV-1 ANCH3 + NEV (10 µM), 24 h p.i. On the right, histogram plots of 2-LTR circles and ALU-PCR, comparing the different infection conditions, represent the nuclear import and integration rates, respectively. All graphs contain ± SD. **C)** Confocal image of HIV-1 ANCH3 provirus in a single-cell clone. The clone was transduced with OR-GFP LV for the detection of HIV-1 integrated DNA. On the right, histogram plot of ALU-PCR and β-galactosidase expression of the HIV-1 ANCH3 single-cell clone compared to non-infected cells (NI). All graphs contain ± SD. **C)** On the left, Tukey's box plots of the diffusion coefficient (D ($nm^2/s$)) of episomal and integrated HIV-1 DNA forms tracked in 3D time lapses (movies S1, S2, S3). On the right, an example of live track for each experimental group is shown. Low D corresponds to D ($nm^2/s$) < 5 and high D to 5 < D ($nm^2/s$) < 20. The color range indicates the time up to 10 min. Statistics: Student's t test for two comparisons, one-way ANOVA for more than two comparisons. ns=non-significant, *=p ≤ 0.05, **=p ≤ 0.01, ***=p ≤ 0.001, ****=p ≤ 0.0001. Scale bars: 5 µm.

**Figure 2. HIV-1 transcriptional activity in dividing and non-dividing cells. A)** Labeling of vDNA and vRNA in HeLa cells expressing OR-GFP obtained by coupling of HIV-1 ANCHOR (vDNA) with RNA FISH (vRNA). On the left, confocal images of

HIV-1 RNA FISH with or without RAL (20 µM), 24 h p.i. On the right, scatter plot with bars showing the mean fluorescence intensity of the vRNA signal (red) per nucleus and of vDNA spots detection (green) (for vRNA: n = 13 cells (-RAL), 16 cells (+RAL); for vDNA: 95 (-RAL) and 141 (+RAL) spots, n = 13 cells (-RAL), 16 cells (+RAL)). The analysis was performed in 2D. **B)** Labeling of vDNA and vRNA in $CD4^+$T cells expressing OR-GFP. On the left, confocal image of HIV-1 RNA FISH in primary activated $CD4^+$ T cells, 3 days p.i.. On the right, scatter plot with bars showing the percentage of vDNA associated to vRNA signal and vice versa, per nucleus (n= 262 (vDNA spots); 64 (vRNA foci) in infected Jurkat cells, 48 h p.i. (some examples of cells are shown in Fig. S5). The analysis was performed in 3D. **C)** The same techniques have been used to label vDNA and vRNA in macrophage-like cells, THP-1, expressing OR-GFP. On the left, confocal image of infected THP-1, 3 days p.i. On the right, scatter plot with bars showing the percentage of vDNA associated to vRNA signal and vice versa, per nucleus (n=130 (vDNA spots), 99 (vRNA foci). The analysis was performed in 3D. All graphs contain ± SD. Statistics: Student's t test. ns=non-significant, *=p ≤ 0.05, **=p ≤ 0.01, ***=p ≤ 0.001, ****=p ≤ 0.0001. Scale bars: 5 µm.

**Figure 3. Distance of HIV-1 provirus from the nuclear envelope in dividing and non-dividing cells. A)** On the left, confocal images of HIV-1 DNA labeling by HIV-1 ANCHOR system and HIV-1 RNA labeling by RNA FISH in HeLa cells expressing OR-GFP infected with HIV-1 ANCH3 at 48 h p.i. compared to 7 days p.i.. On the right, scatter plot with bars indicating the distance from the nuclear envelope (NE) per detected vDNA spot (green) and vRNA focus (red) (n=vDNA spots: 83 (48 h p.i.) and 16 (7 d p.i.); n=vRNA foci: 52 (48h p.i.) and 18 (7d p.i.)). **B)** On the left, confocal images to detect vDNA and vRNA in THP-1 cells expressing OR-GFP infected with HIV-1 ANCH3, 3 days p.i.. On the right, scatter plot with bars indicating the distance from the nuclear envelope (NE) per detected vDNA spot (green) and vRNA focus (red) (n=139 vDNA spots; n= 126 vRNA foci). **C)** Scatter plot with bars comparing the distances from the nuclear envelope (NE) per detected vRNA focus in HeLa P4R5, 48h p.i. and THP-1, 3 days p.i. (vRNA foci: 52 (HeLa), 126 (THP-1)). All the analyses were performed in 3D. All graphs contain ± SD. Statistics: Student's t test for two comparisons, one-way ANOVA for more than two comparisons. ns=non-significant, *=p ≤ 0.05, **=p ≤ 0.01, ***=p ≤ 0.001, ****=p ≤ 0.0001. Scale bars: 5 µm, inset 1 µm.

**Figure 4. HIV-1 integrase location in dividing and non-dividing cells. A)** Confocal images of primary activated $CD4^+$ T and THP-1 cells expressing OR-GFP infected with HIV-1 ANCH3, 20h and 30h p.i, respectively, compared to uninfected cells. Co-visualization of vDNA (green) and viral IN (magenta). On the bottom scatter plot with bars comparing the percentages of vDNAs associated to IN per nucleus (n=vDNA: 54 (CD4+T), 100 (THP-1)). The analysis was performed in 3D. **B)** Confocal images of primary activated $CD4^+$ T and THP-1 cells infected with HIV-1 ANCH3, 20h and 30h p.i, respectively, compared to uninfected cells. Co-visualization of CPSF6 (red) and viral IN (cyan). The graphs show the percentages of IN-CPSF6 association events and vice versa, per nucleus (n= for CD4+T: 46 (CPSF6), 33 (IN); for THP-1: 206 (CPSF6), 132 (IN)). All graphs contain ± SD. Statistics: Student's t test. ns=non-significant, *=p ≤ 0.05, **=p ≤ 0.01, ***=p ≤ 0.001, ****=p ≤ 0.0001. Scale bars: 5 µm.

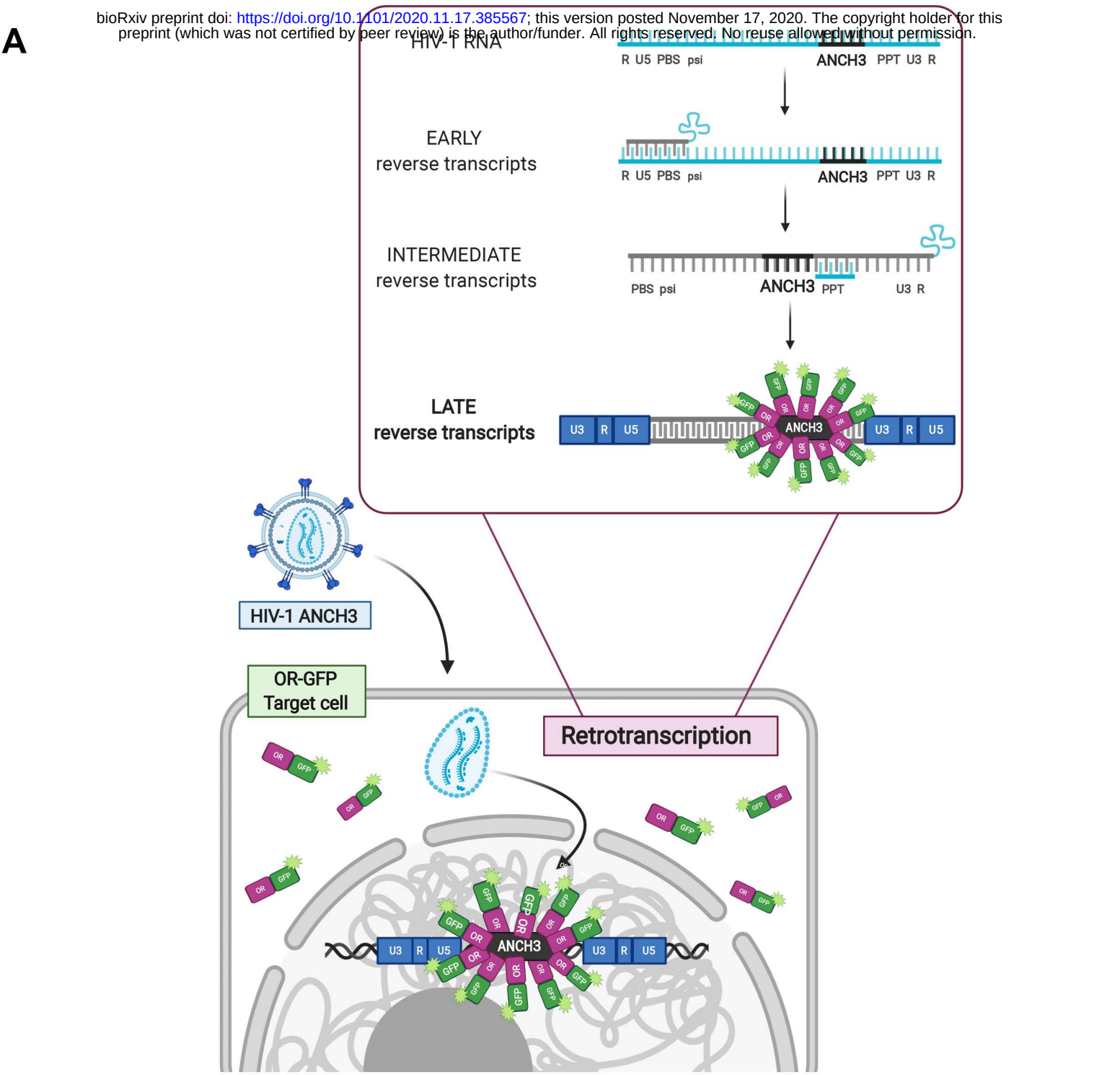
**Figure 5. HIV-1 proviruses are excluded from CPSF6/SC35 MLOs. A)** Confocal images of THP-1 cells expressing OR-GFP infected with HIV-1 ANCH3, 3 days p.i., to co-visualize vDNA (green), CPSF6 (red), SC35 (cyan) or to co-visualize vDNA (green), SON (red), SC35 (cyan), compared to uninfected cells. On the right, the graphs show: the percentages of CPSF6 clusters associated to vDNA or to SC35, per nucleus (n= CPSF6 clusters > 50); the comparison of Pearson's r coefficients of SON-SC35 colocalization between infected and uninfected cells, per nucleus (n = 25 cells (HIV-1 ANCH3), 19 cells (NI)); the average distance of each vDNA from the closest SC35-speckle border (n= vDNAs: 83). **B)** Confocal images of THP-1 cells expressing MCP-GFP infected with HIV-1 ANCH3 MS2, 3 days p.i., to co-visualize vRNA (green), CPSF6 (red), SC35 (cyan), compared to uninfected cells. The graph shows the percentages of CPSF6 clusters associated to vRNA foci or to SC35 (n= CPSF6 clusters: 42). **C)** Scatter plot with bars indicating the percentage of vDNA or vRNA associated to CPSF6, per nucleus (n= 69 (vDNA), 30 (vRNA)). All the analyses were performed in 3D. All graphs contain ± SD. Statistics: Student's t test. ns=non-significant, *=p ≤ 0.05, **=p ≤ 0.01, ***=p ≤ 0.001, ****=p ≤ 0.0001. Scale bars: 5 µm, inset 1 µm.

**Figure 6. HIV-1 functional proviruses are in complex with LEDGF. A)** Confocal images of THP-1 cells expressing OR-GFP infected with HIV-1 ANCH3, 3 days p.i., to co-visualize CPSF6 (red) and LEDGF (white), compared to uninfected cells. On the

left scatter plot of the percentages of CPSF6 found associated to LEDGF clusters per nucleus; on the right, scatter plot of the LEDGF and CPSF6 count per nucleus (n = 16 cells). Scale bar: 5 µm, inset 2 µm. **B)** Confocal images of THP-1 cells expressing OR-GFP infected with HIV-1 ANCH3 and expressing MCP-GFP infected with HIV-1 ANCH3 MS2, to co-visualize vDNA or vRNA (green), CPSF6 (red) and LEDGF (white). The scatter plot shows the percentage of vDNA or vRNA associated to LEDGF (n= 33 (vDNA), 15 (vRNA)). The pie charts show the percentages of the 2 populations of LEDGF clusters associated to vDNA or vRNA: LEDGF alone or LEDGF in proximity of CPSF6. Scale bar: 5 µm, inset 1 µm. All the analyses were performed in 3D. Statistics: Student's t test. ns=non-significant, *=p ≤ 0.05, **=p ≤ 0.01, ***=p ≤ 0.001, ****=p ≤ 0.0001.

**Figure 7. HIV-1 restructures the nuclear landscape for efficient replication.** Once retrotranscribed and imported in the nucleus, HIV-1 genome integrates in open chromatin regions enriched in LEDGF proteins, in less than 2 micrometers form the nuclear envelope. In addition, upon infection, CPSF6 clusters in SC35 nuclear speckles with the viral IN restructuring the nuclear landscape and inducing HIV-1 MLOs formation. Late retrotranscribed DNA and proviruses are excluded from CPSF6 clusters. HIV-1 transcription is favored by the nuclear environment surrounding proviral sites that may profit from the proximity to SC35 and other splicing factors for post-transcriptional processes. Created with BioRender.com.

A



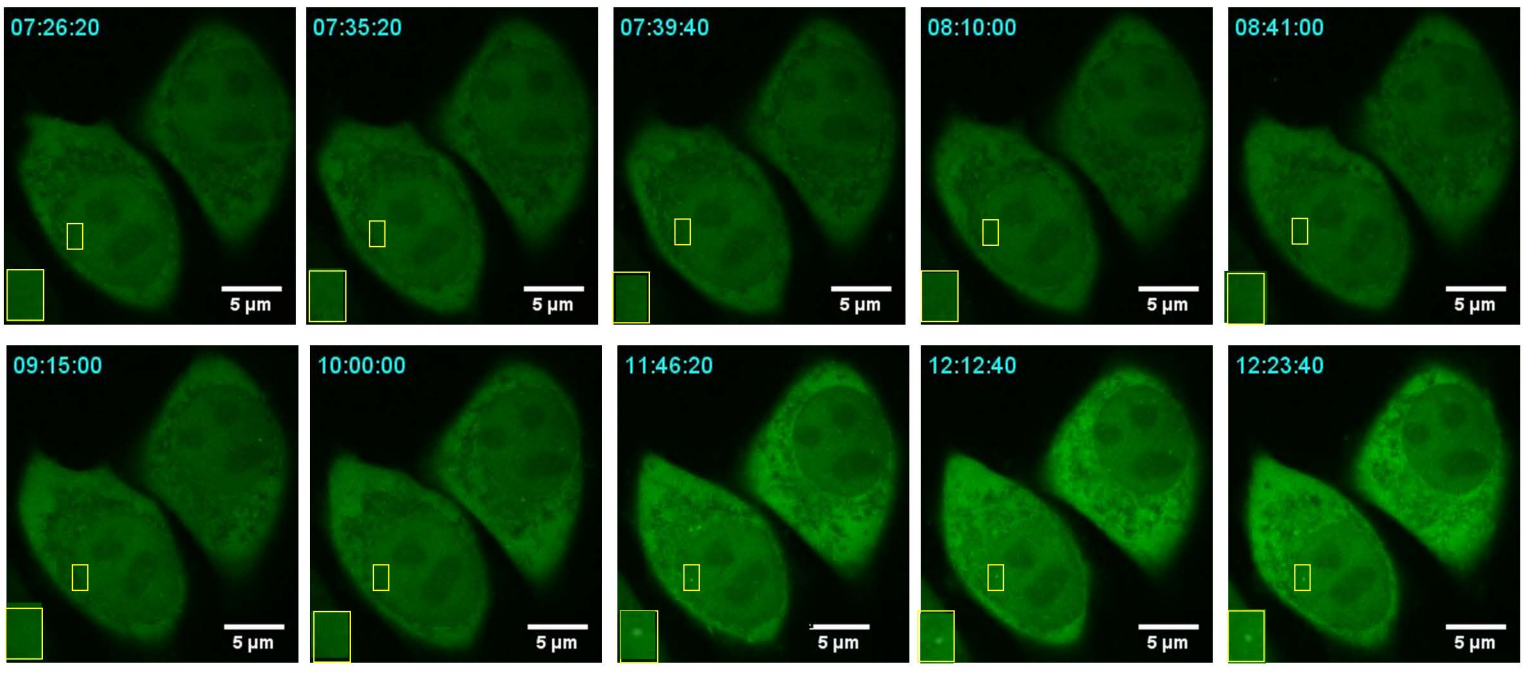




Figure 1

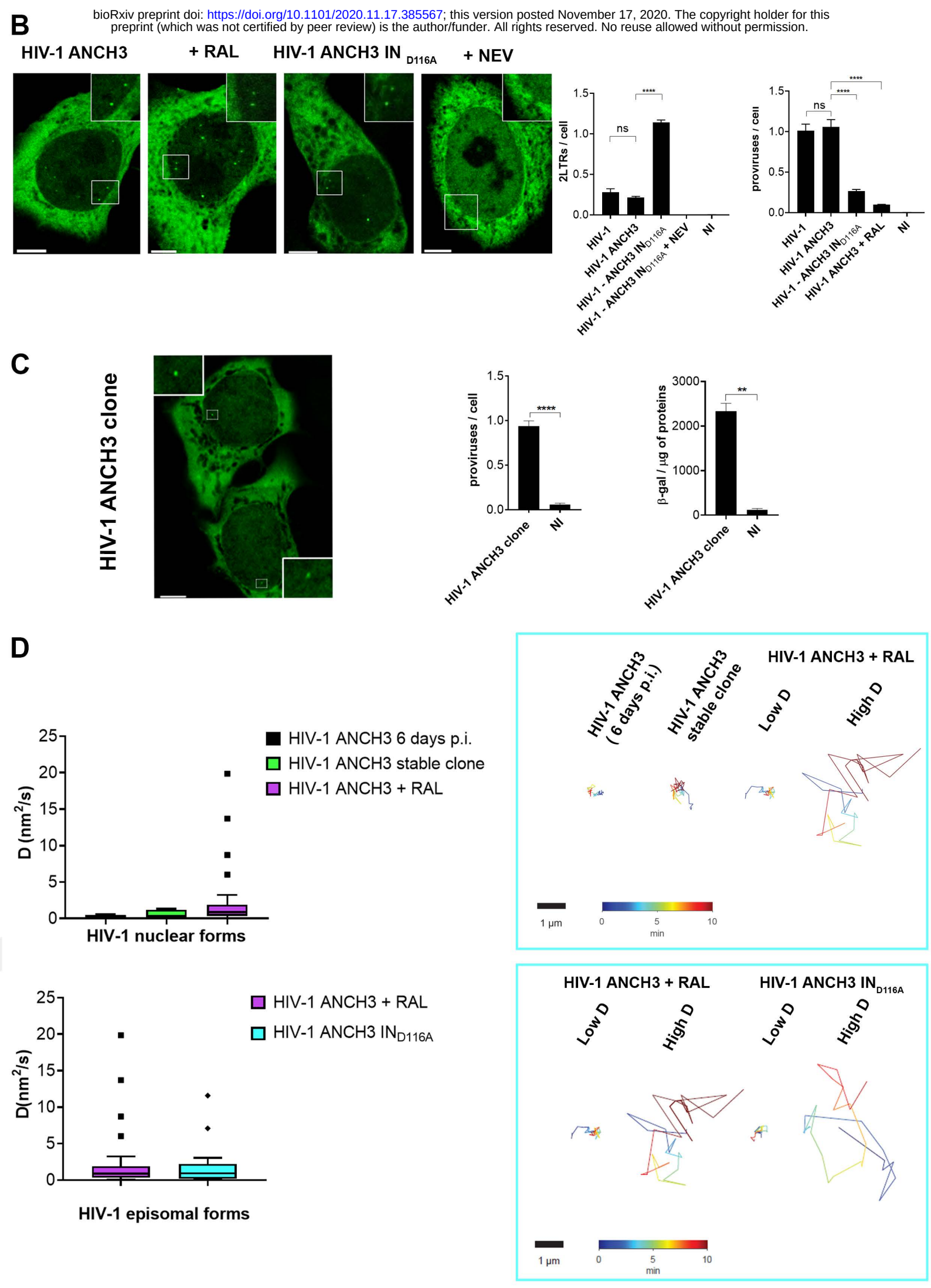



**Figure 1**

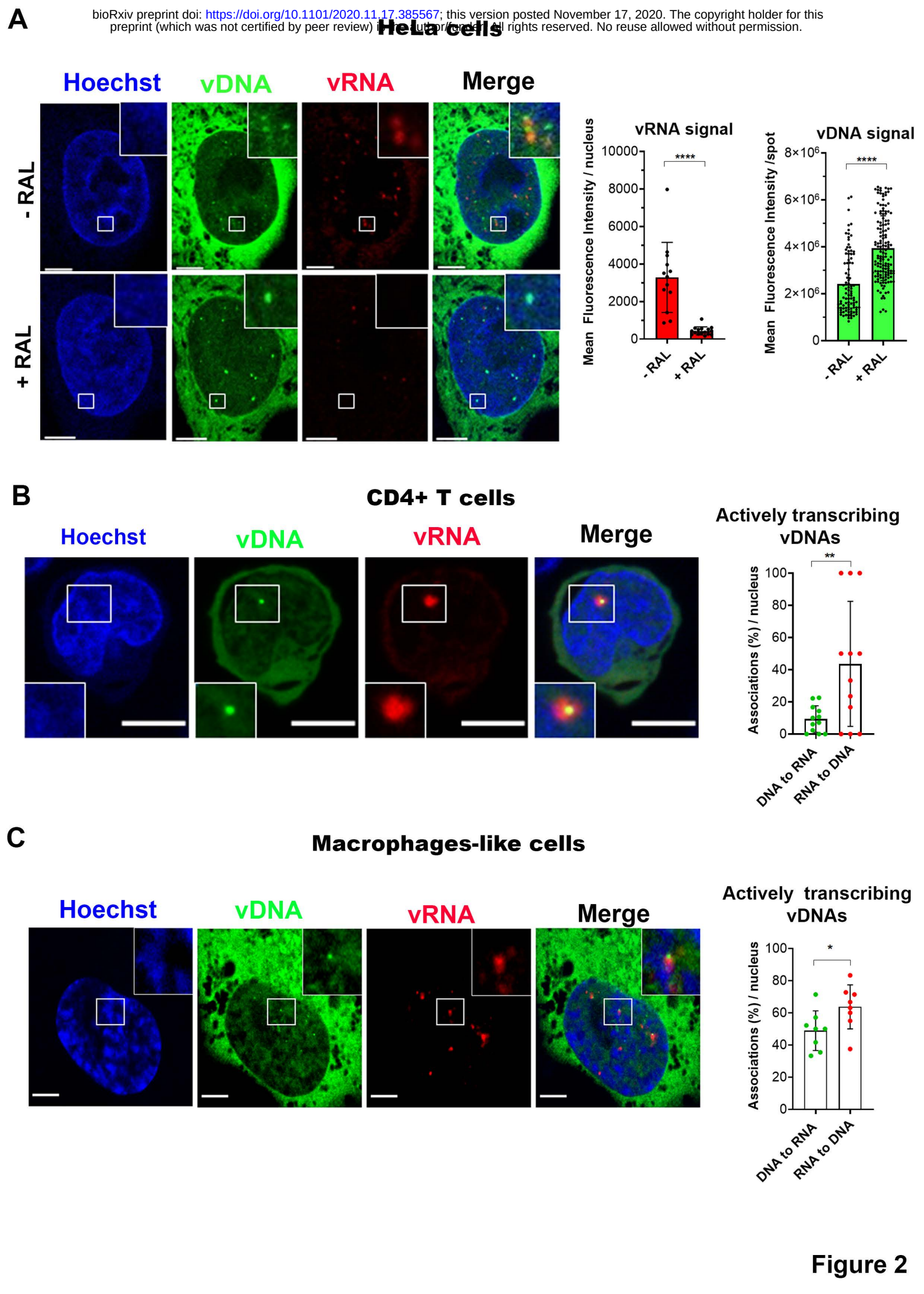
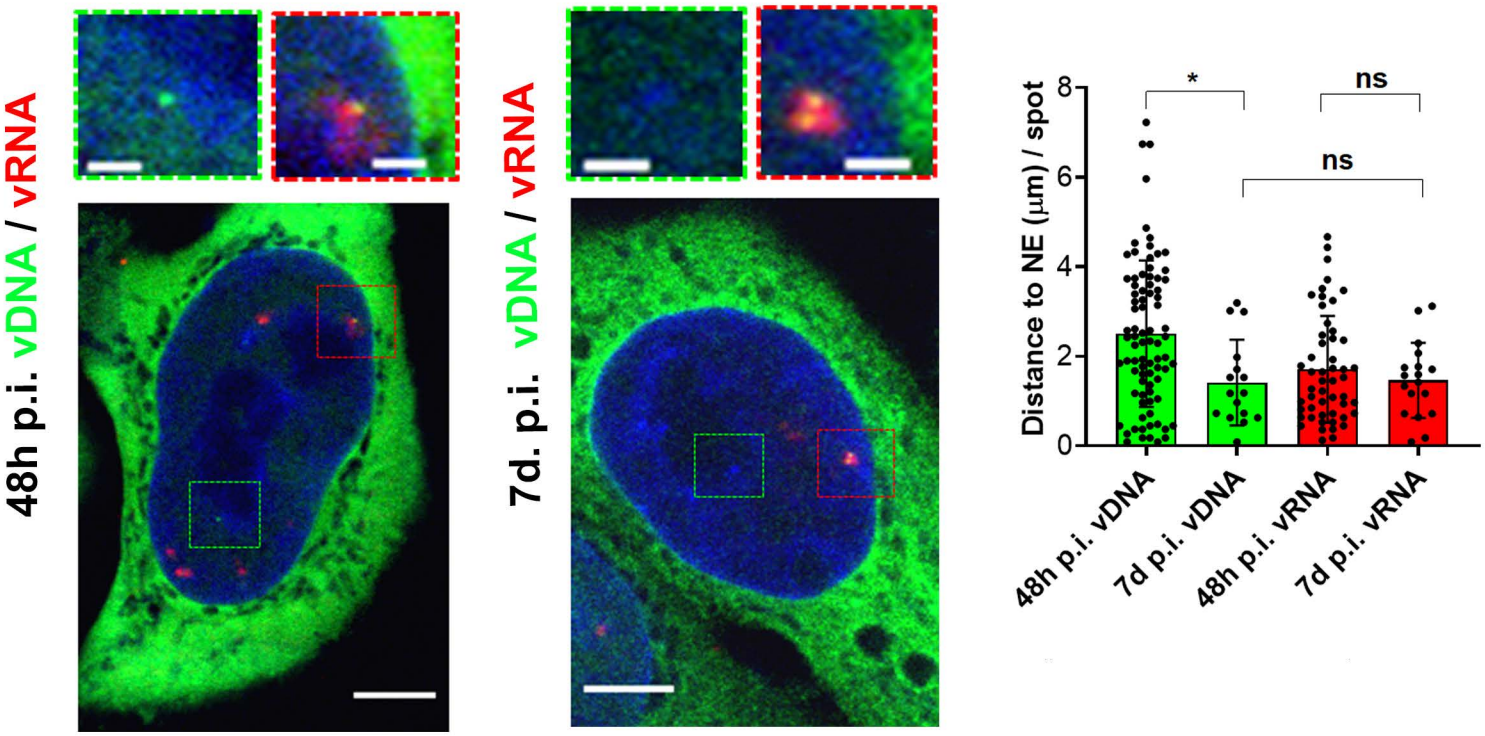



Figure 2

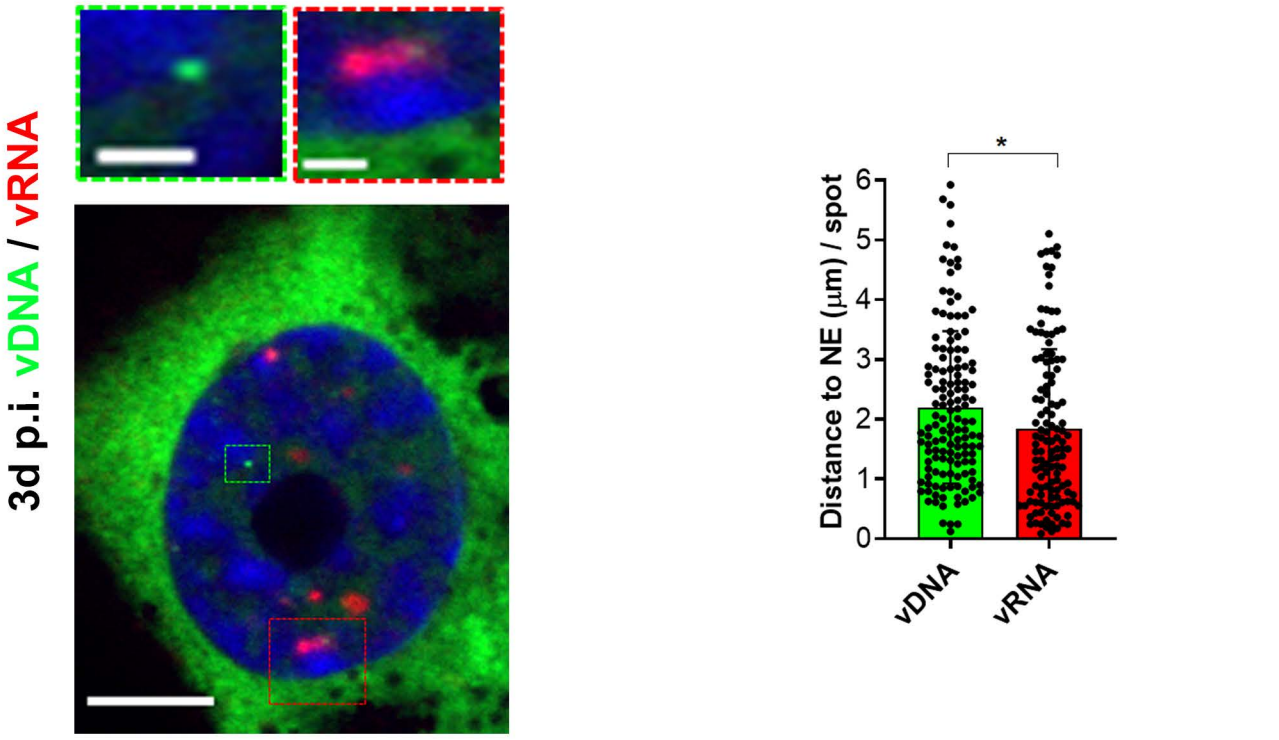
A

## Dividing cells: HeLa




B

## Non-dividing cells: THP-1




C

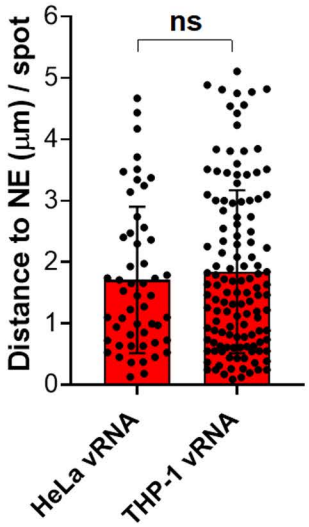
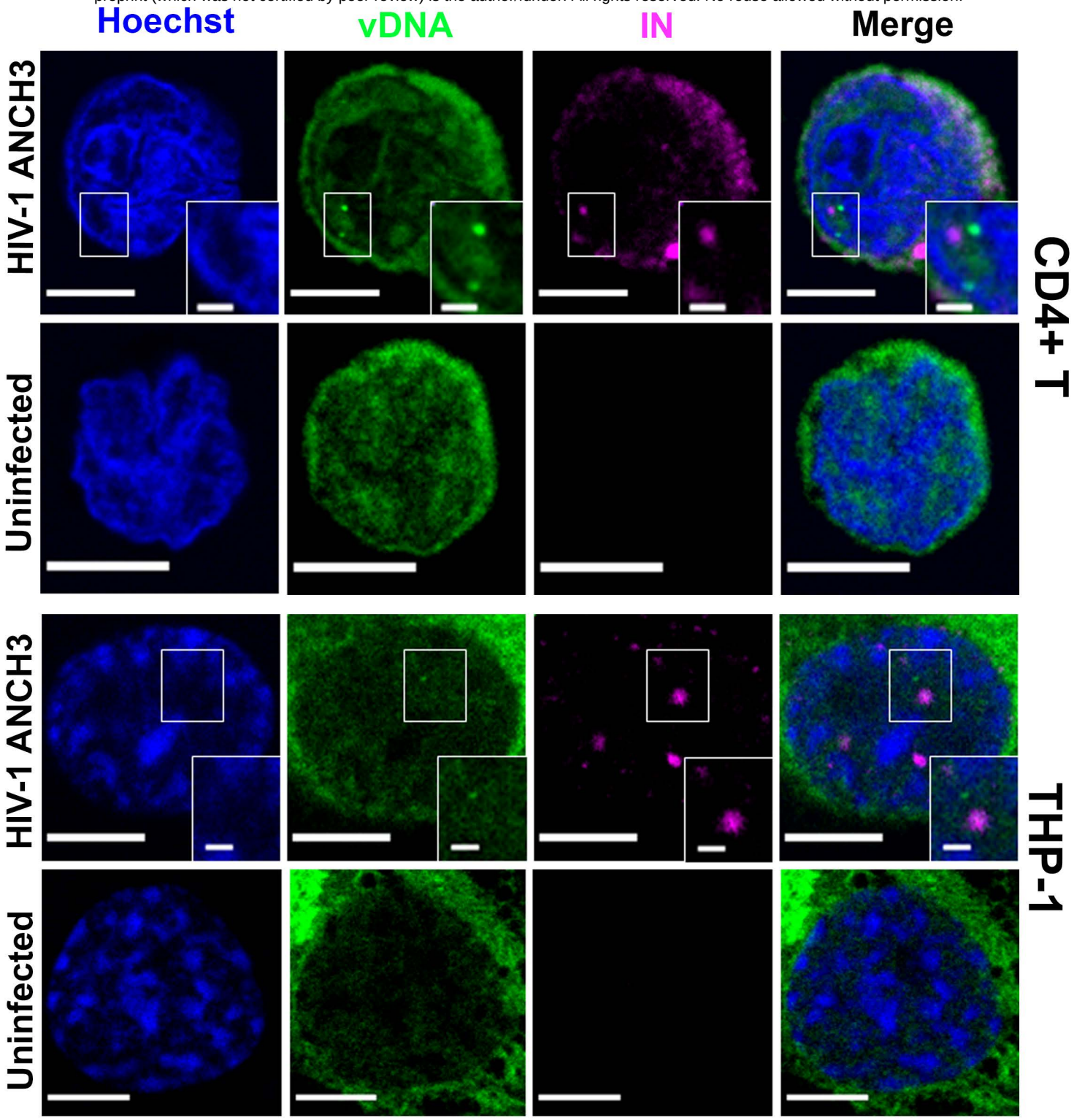



Figure 3

A



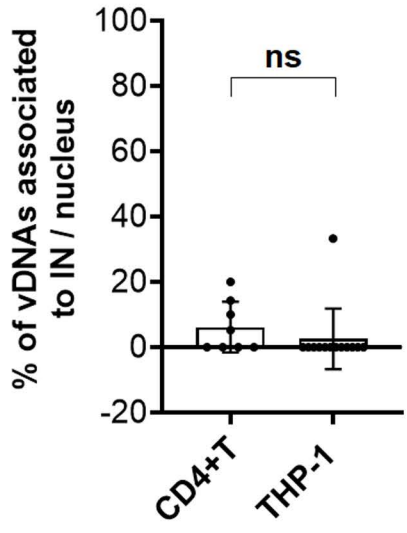
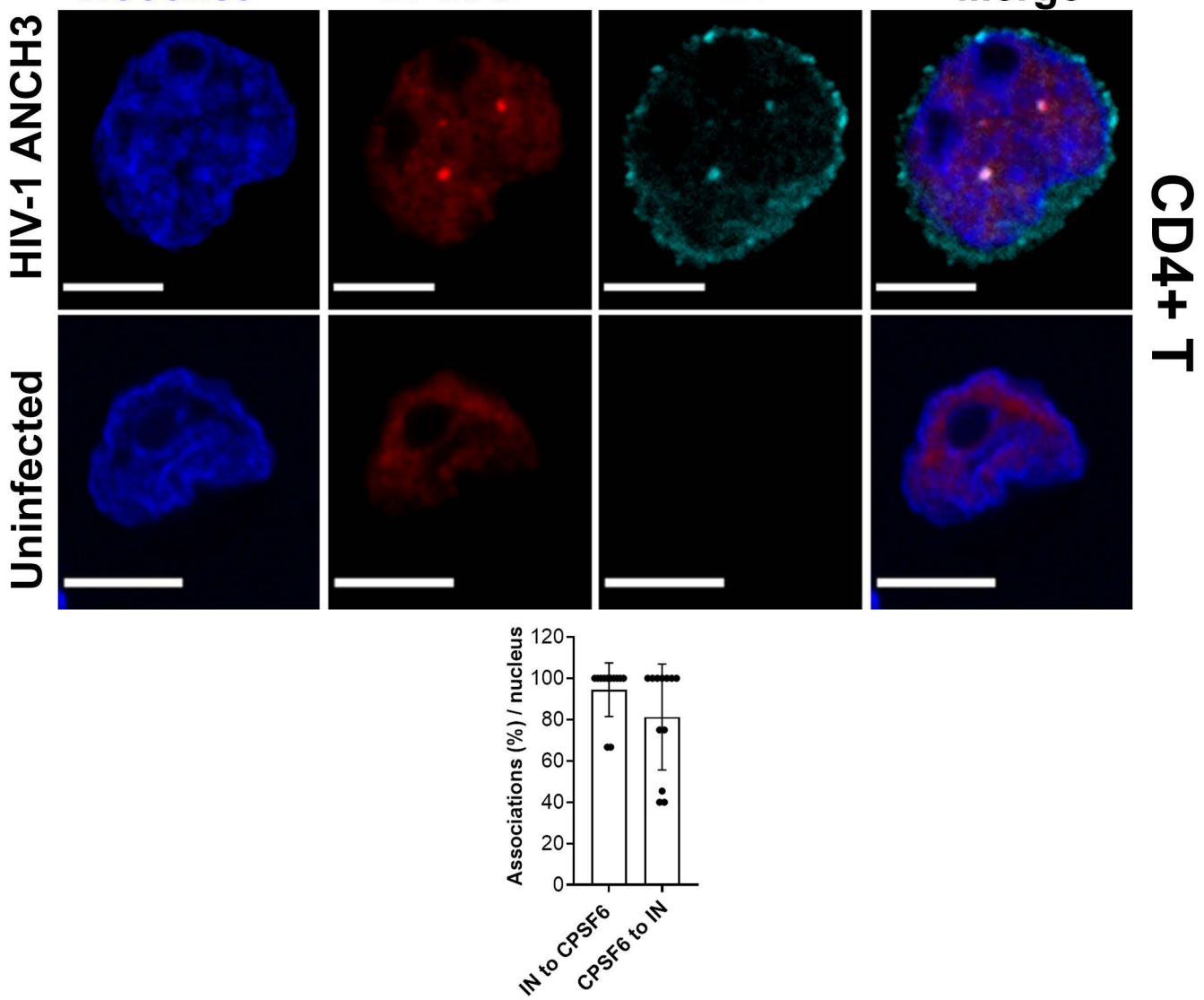




Figure 4

B



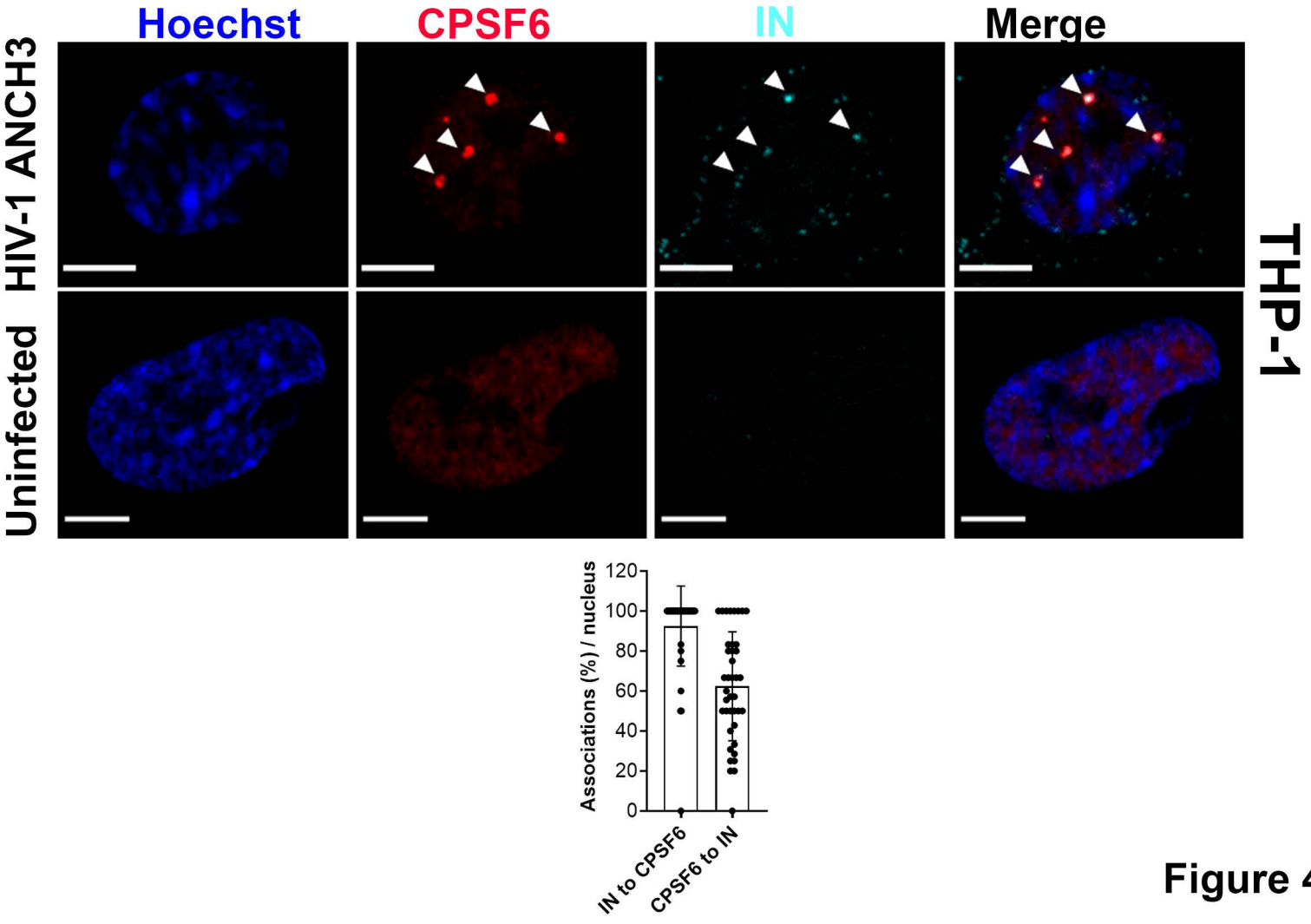




Figure 4

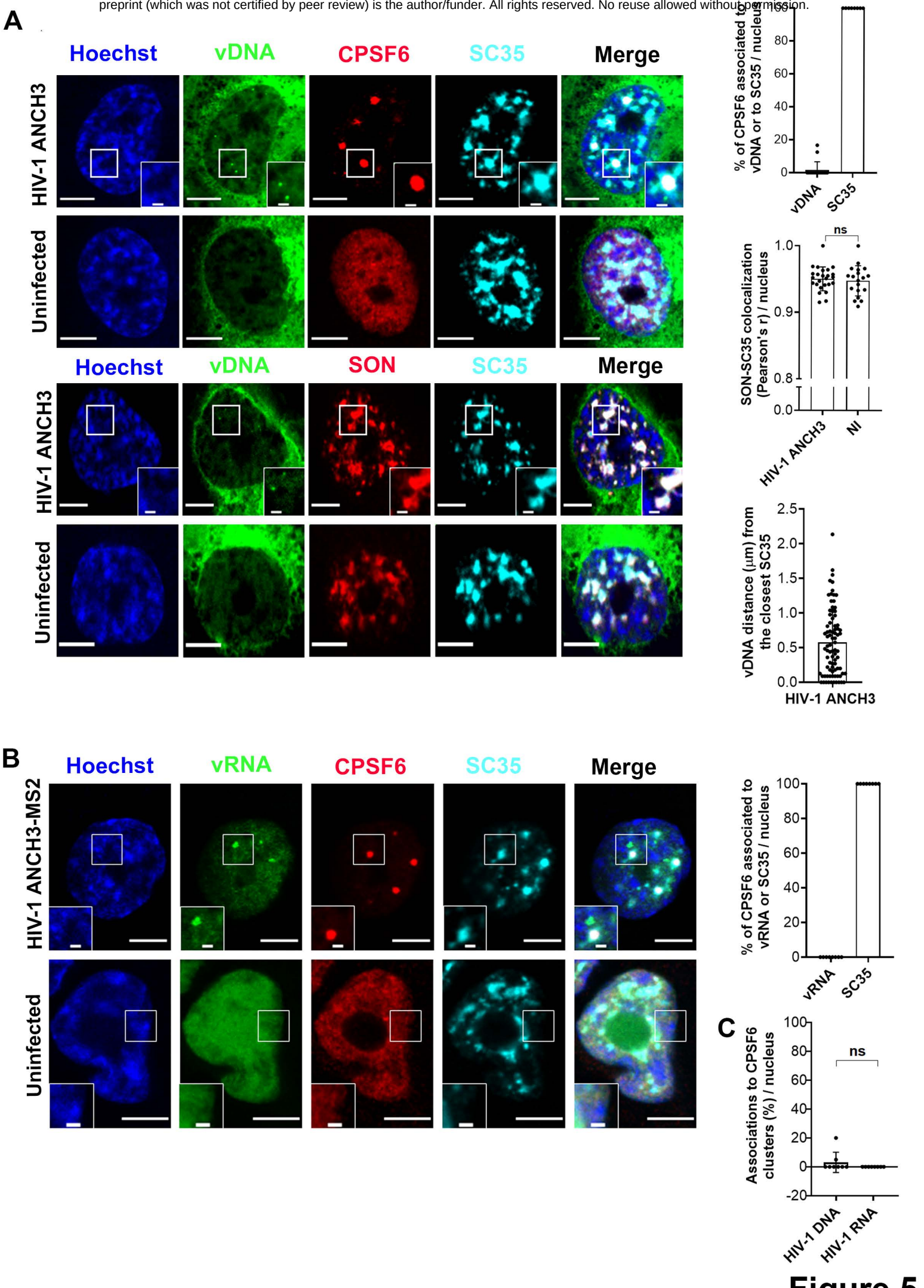
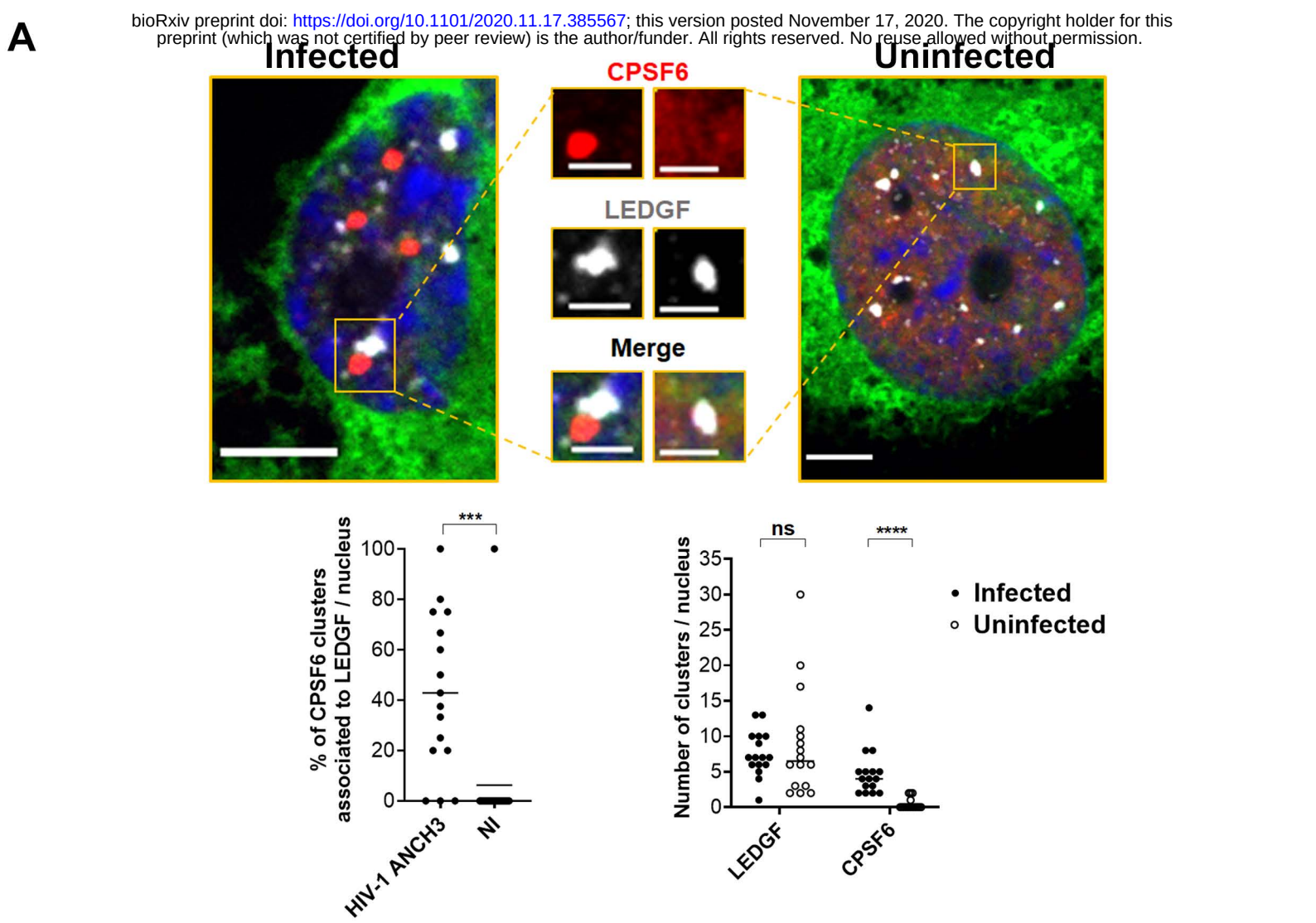



**Figure 5**



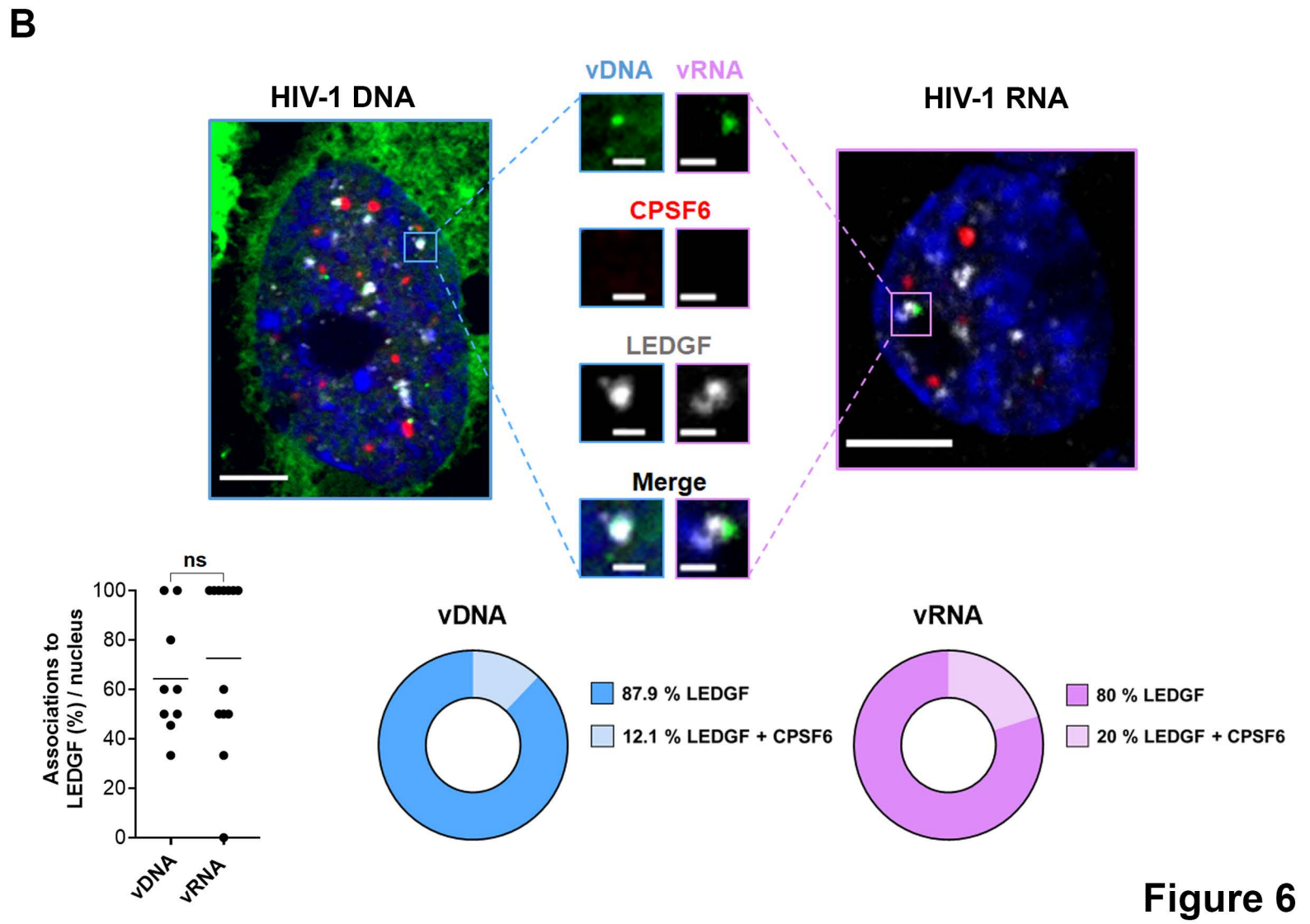




Figure 6

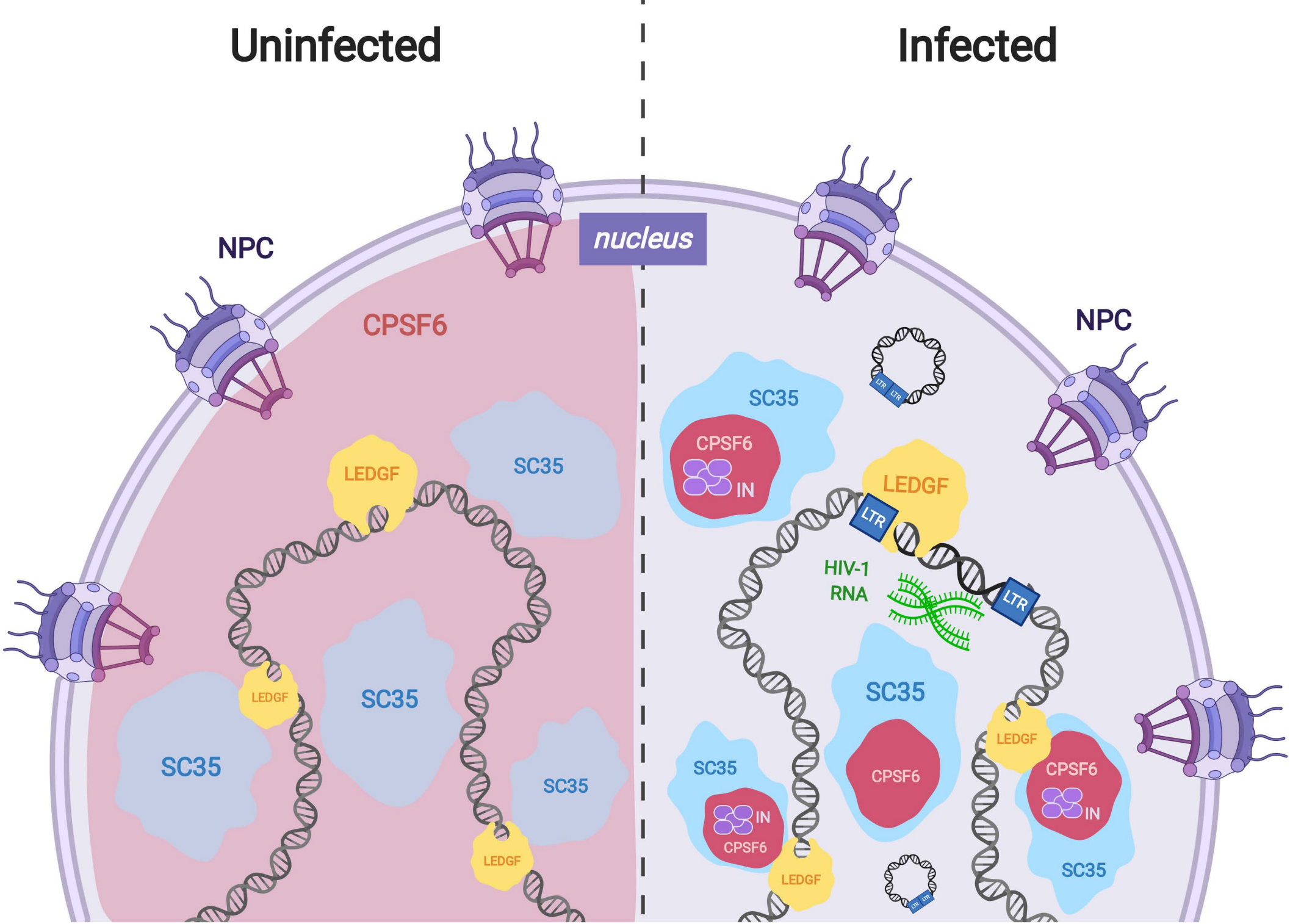

Uninfected
Infected
NPC
nucleus
CPSF6
LEDGF
SC35
LEDGF
SC35
SC35
SC35
LEDGF
NPC
SC35
CPSF6
IN
LTR
LTR
LEDGF
HIV-1
RNA
LTR
SC35
SC35
CPSF6
IN
CPSF6
LEDGF
LEDGF
CPSF6
IN
SC35

Figure 7