Decoding the biogenesis of HIV-induced CPSF6 puncta and their fusion with the nuclear speckle.

Chiara Tomasini[1*], Celine Cuche[1*], Selen Ay[1], Maxence Collard[1], Bin Cui[2], Mohammad Rashid[2], Shaoni Bhattacharjee[1], Julian Buchrieser[3], Charlotte Luchsinger[2], Cinzia Bertelli[2], Vladimir N. Uversky[4], Felipe Diaz-Griffero[2#], Francesca Di Nunzio[1#]

[1] *Institut Pasteur, Advanced Molecular Virology Unit, Department of Virology, Université Paris Cité, 75015 Paris, France*

[2] *Albert Einstein College of Medicine, Department of Immunology and Microbiology, New York 10461, USA*

[3] *Institut Pasteur,* Virus and Immunity Unit*, Department of Virology, Université Paris Cité, 75015 Paris, France.*

[4] *Department of Molecular Medicine and USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA*

 # Address correspondence to Francesca Di Nunzio, dinunzio@pasteur.fr, or Felipe Diaz-Griffero, felipe.diaz-griffero@einsteinmed.edu

** These authors have equally contributed*

Highlights

- The formation of HIV-induced CPSF6 puncta is critical for restoring HIV-1 nuclear reverse transcription (RT).
- CPSF6 protein lacking the FG peptide cannot bind to the viral core, thereby failing to form HIV-induced CPSF6 puncta.
- The FG peptide, rather than low-complexity regions (LCRs) or the mixed charge domains (MCDs) of the CPSF6 protein, drives the formation of HIV-induced CPSF6 puncta.
- HIV-induced CPSF6 puncta form individually and later fuse with nuclear speckles (NS) via the intrinsically disordered region (IDR) of SRRM2.

Summary

Viruses rely on host cellular machinery for replication. After entering the nucleus, the HIV genome accumulates in nuclear niches where it undergoes reverse transcription and integrates into neighboring chromatin, promoting high transcription rates and new virus progeny. Despite anti-retroviral treatment, viral genomes can persist in these nuclear niches and reactivate if treatment is interrupted, likely contributing to the formation of viral

reservoirs. The post-nuclear entry dynamics of HIV remain unclear, and understanding these steps is critical for revealing how viral reservoirs are established.

In this study, we elucidate the formation of HIV-induced CPSF6 puncta and the domains of CPSF6 essential for this process. We also explore the roles of nuclear speckle scaffold factors, SON and SRRM2, in the biogenesis of these puncta. Through genetic manipulation and depletion experiments, we demonstrate the key role of the intrinsically disordered region of SRRM2 in enlarging nuclear speckles in the presence of the HIV capsid.

We identify the FG domain of CPSF6 as essential for both puncta formation and binding to the viral core, which serves as the scaffold for CPSF6 puncta. While the low-complexity regions (LCRs) modulate CPSF6 binding to the viral capsid, they do not contribute to puncta formation, nor do the disordered mixed charge domains (MCDs) of CPSF6. These results demonstrate how HIV evolved to hijack host nuclear factors, enabling its persistence in the host.

Of note, this study provides new insights into the underlying interactions between host factors and viral components, advancing our understanding of HIV nuclear dynamics and offering potential therapeutic targets for preventing viral persistence.

Introduction

Since the discovery of HIV (Barre-Sinoussi et al., 1983), the initial stages of the viral life cycle have been understood to primarily occur within the host cytoplasm. Only the pre-integration complex, carrying the fully reverse-transcribed viral DNA, was believed to enter the nucleus for integration into the host chromatin (Suzuki and Craigie, 2007). Recent studies highlighted that the viral genome is transported in the nucleus via a shuttle that shields it from the hostile cellular environment (Rasaiyaah et al., 2013). This shuttle is constituted by the viral capsid (Blanco-Rodriguez and Di Nunzio, 2021; Blanco-Rodriguez et al., 2020; Chen et al., 2016; Selyutina et al., 2020; Yamashita and Emerman, 2004; Zila et al., 2021), which comprises 250 hexamers and 12 pentamers (Pornillos et al., 2009). Within the structure of the capsid, hydrophobic pockets exist between hexamers, which serve as targets for various nucleoporins, especially the ones carrying FG repeats, facilitating the translocation of the capsid through the nuclear pore complex (NPC) (Buffone et al., 2018; Di Nunzio, 2013; Di Nunzio et al., 2012; Lelek et al., 2015; Matreyek et al., 2013; Price et al., 2014). Recent studies suggest that HIV uses multiple FG regions of several nucleoporins to translocate through the NPC, acting as a chaperone by itself (Dickson et al., 2024; Fu et al., 2024). Using a reductionist system of Nup98 condensates, it has been demonstrated that FG-mediated phase partitioning identifies specific sites on the capsid that allow it to interact autonomously with these phases. These findings complement the evolving understanding of the early stages of HIV infection, which has been revisited in recent years. They unveil that crucial stages of early viral infection occur within the host nucleus (Burdick et al., 2020; Dharan et al., 2020; Francis et al., 2020; Scoca et al., 2023; Selyutina et al., 2020). Notably, it has been shown that the pre-integration complex forms within the host nucleus (Muller et al., 2021; Scoca et al., 2023), and incoming viral RNA genomes accumulate in nuclear niches containing , such as like cleavage and polyadenylation specificity factor subunit 6 (CPSF6), RNA-binding protein SON and Splicing component, 35 kDa (SC35, also known as serine/arginine-rich splicing factor 2 (SRSF2) (Francis et al., 2020; Rensen et al., 2021; Scoca et al., 2023). However, the mechanistic requirements that govern post-nuclear entry phases, which are crucial for a successful viral infection and the establishment of viral reservoirs, remain enigmatic.

Specifically, the mechanism behinds the formation of nuclear niches containing viral components and nuclear speckle (NS) factors, such as SC35—traditionally used as a marker for nuclear speckles remains unclear. However, recent studies have revealed that SON and serine/arginine repetitive matrix protein 2 (SRRM2) are essential for NS biogenesis (Fu and Maniatis, 1990). Particularly, intrinsically disordered regions play a critical role in the NS formation (Ilik et al., 2020). These membraneless organelles (MLOs) fulfil various cellular

functions besides splicing. Recent evidence indicates a direct role of NSs in cellular transcription regulation, as their spatial proximity correlates with gene expression amplification, as demonstrated by live-cell imaging of heat-shock responsive genes (Chen et al., 2018; Zhang et al., 2021). HIV, being a virus capable of generating new particles through splicing and integrating into active host genes, finds NSs highly conducive for viral replication. HIV particles, along with their RNA genome, accumulate within nuclear MLOs enriched in NS factors (Rensen et al., 2021; Scoca et al., 2023). Notably, CPSF6, a paraspeckle factor first identified as a viral partner by KewalRamani's laboratory (Lee et al., 2010), has been clearly detected in HIV-induced CPSF6 puncta (Francis et al., 2020; Lee et al., 2010; Luchsinger et al., 2023; Rensen et al., 2021; Scoca et al., 2023). These puncta serve as hubs for nuclear reverse transcription (RT) and the formation of pre-integration complexes, which generate active proviruses detected outside but in close proximity to NSs (Li et al., 2021; Scoca et al., 2023).

In our study, we aim to elucidate how these HIV-induced CPSF6 form and identify the NS factors involved in their formation. CPSF6, along with NS factors, contains intrinsically disordered regions that can guide HIV-1 to the correct nuclear location for successful infection or allow the virus to remain sequestered during drug treatment, forming reservoirs. Importantly, viral reservoirs are the major bottleneck for curing the infection.

In this study, we investigate which disordered domain of CPSF6 is responsible for tracking the viral core and generating HIV-induced CPSF6 puncta in the host nucleus. Simultaneously, we elucidate the key component of NSs that, through its intrinsically disordered regions (IDRs), enable fusion with HIV-induced CPSF6 puncta, likely stabilizing them. Notably, we observed that the virus rebounds when anti-reverse transcription drugs are removed, but only if nuclear niches containing HIV, NS factors, and CPSF6 are present. If these niches are pharmacologically dismantled, viral rebound does not occur.

Overall, studying the biogenesis of HIV-induced nuclear niches is crucial for understanding how the virus navigates and persists in the nucleus of infected cells, and for designing new antiretroviral strategies.

Results

Critical Role of HIV-Induced CPSF6 Puncta in Restoring Nuclear RT After Anti-RT Therapy Discontinuation.

Upon nuclear entry, HIV enhances the formation of CPSF6 clusters, where RT ends. The treatment with the reversible RT inhibitor nevirapine (NEV) can trap the viral RNA genome in these nuclear niches (Rensen et al., 2021; Scoca et al., 2023). Once NEV is removed, the trapped vRNA can resume RT entirely within the nucleus, a process we term nuclear RT . Here we demonstrate that this phenomenon is dependent on the presence of CPSF6 puncta, as their disruption by high doses of PF74 (25 µM) significantly impairs nuclear RT, shown by the absence of luciferase expression, similar to full NEV treatment (Figure 1A-B).

CPSF6 FG Domain is Required for HIV-Induced Puncta Formation.

Formation of CPSF6 puncta upon HIV-1 infection hinges into two key events: the entry of the HIV-1 core into the nucleus and the binding of CPSF6 to the HIV-1 core (Blanco-Rodriguez and Di Nunzio, 2021; Blanco-Rodriguez et al., 2020; Buffone et al., 2018; Zila et al., 2021). To determine the contribution of CPSF6's disordered domains for the formation of CPSF6 puncta upon HIV-1 infection, we correlated the binding of CPSF6 to the HIV-1 core with the formation of CPSF6 puncta. To this end, we first generated CPSF6 knockout (KO) THP-1 cells (Figure 2A-B) to eliminate the interference from the endogenous protein, which could affect the interpretation of results regarding the role of the analyzed CPSF6 domains. CPSF6 depletion in THP-1 cells was performed by CRISPR Cas9 technology. To completely eliminate the expression of the CPSF6 gene we selected single clones by limiting dilution. We identified a clone that was completely KO for CPSF6, confirmed through western blot and immunofluorescence (Fig. 2A-B) and we infected this clone and the control clone with HIV. CPSF6 puncta were detected only in the control-infected cells and not in the KO clone (Fig. 2B; Suppl. Fig.3A). The viral integrase (IN) was observed within CPSF6 puncta, consistent with previous studies (Francis et al., 2020; Rensen et al., 2021; Scoca et al., 2023), but absent in CPSF6 KO cells where viral IN was predominantly observed in the cytoplasm (Fig. 2B; Suppl. Fig.3A). Thus, we used KO cells for CPSF6 to assess the role of selected CPSF6 domains in HIV-induced condensates. We designed various CPSF6 deletion mutants (Figure 2C) to specifically assess the significance of the main disordered regions of CPSF6 protein (Suppl. Figure 2A) such as, the FG motif, the low complexity regions (LCRs), and the mixed charge domain (MCD), in the ability of CPSF6 to bind to the core and facilitate the formation of CPSF6 puncta. We investigated the role of the FG peptide by generating a mutant that exclusively lacks the FG peptide (ΔFG). Previous *in vitro* studies have shown that the FG peptide binds to the hydrophobic pocket formed between capsid hexamers (Buffone et al., 2018; Price et al., 2014). Here, we want to investigate the role of FG peptide in the context of the protein.

To further explore this, we developed an alternative plasmid by expanding the FG peptide deletion to include surrounding prion-like LCRs (ΔFG ΔLCR). These regions, outside the CPSF6 context, have been identified as crucial for facilitating strong CPSF6 binding to capsid lattices (Wei et al., 2022). In our study, we aim to evaluate their role within a more physiological setting. Additionally, we assessed a CPSF6 variant that carries the 15-mer FG peptide flanked by non-LCR sequences, such as those derived from Beta-adducin (ADD2), kindly provided by Mamuka Kvaratskhelia (ΔLCR+ADD2). These protein segments are known for their high flexibility, akin to the LCR of CPSF6. Furthermore, to elucidate the contribution of the LCRs of CPSF6 in the formation of CPSF6 puncta, we generated a mutant lacking both LCRs (ΔLCR) (Figure 2C). Analysis of the MCD contribution to both the ability of CPSF6 to bind to the core and formation of CPSF6 puncta was achieved by deleting the MCD and adding 3 nuclear localization signals (3xNLS ΔMCD) since the deletion of the MCD results in a protein that localizes mainly into the cytoplasm (Figure 2C; Suppl. Figure 2B).

To correlate the ability of CPSF6 to bind to the HIV-1 core with formation of CPSF6 puncta, we expressed wild type and mutant CPSF6 constructs in THP-1 cells knockout for CPSF6. Subsequently, we infected these cells with HIV-1 and analyzed the presence or absence of CPSF6 clusters at 24 hours post-infection. Importantly, for the imaging experiment we expressed CPSF6 WT and mutants without tags to avoid the formation of aggregates that could interfere with our conclusions. Our data show that HIV-induced CPSF6 clusters can form extremely rarely with the deletion mutant CPSF6 ADD2ΔLCR and with the mutant lacking the FG or both the FG peptide and the LCRs (Fig. 2D-F; Suppl. Fig.3B). However, when we analyzed the role of the MCD domain in CPSF6 puncta formation, which was indicated to be important for condensing CPSF6 in NS (Greig et al., 2020), comparing the number of CPSF6 WT puncta induced by HIV infection with CPSF6 mutants revealed that the MCD domain does not play a critical role in HIV-induced CPSF6 puncta formation (Fig. 2D-F). In addition, we observed that the majority of analysed CPSF6 3xNLSΔMCD puncta contain vRNA inside, similar to CPSF6 WT puncta (Figure 2G), thus corroborating the lack of a role for this intrinsically disordered domain in HIV-induced CPSF6 puncta. Since the NLS domain from SV40, which replaces the MCD, is highly basic and could potentially induce condensates, we fused CPSF6 with a non-basic NLS (PY-NLS) or removed the NLS entirely. Even though these two proteins do not efficiently enter the nucleus, the few that do manage to reach the nucleus can host viral particles, as evidenced by the presence of IN. Many viruses are typically blocked in the cytoplasm due to the presence of these mutants that are mainly cytoplasmic. However, because we used a high viral dose, the blockage in the cytoplasm was not complete. As a result, the viruses that successfully entered the nucleus induced the formation of clusters associated with CPSF6-deleted mutants, indicating that the

MCD is not critical for the formation of HIV-induced CPSF6 puncta (Fig.2H). Similar to the MCD, when we compared CPSF6 truncated for the LCRs with CPSF6 WT, we observed that the LCRs do not contribute to CPSF6 puncta formation. Therefore, the FG peptide alone, without the LCRs, is the only CPSF6 domain required for their formation (Fig. 2D-F; Suppl. Fig.3B).

Next, we tested the ability of the different CPSF6 deletion mutants for their ability to bind the viral core using a previously described capsid binding assay (Selyutina et al., 2018). Wild type and mutant CPSF6 proteins were expressed in human 293T cells at similar levels (INPUT) (Figure 3A). Extracts containing wild type and mutant CPSF6 proteins were incubated with stabilized HIV-1 capsid tubes for 1 h at 25˚ C in the presence of 10 µM of PF74, which is a small molecule that competes with CPSF6 for binding to the hydrophobic pocket formed between hexamers that constitute the viral core (Buffone et al., 2018; Price et al., 2014). HIV-1 capsid stabilized tubes were washed, and the bound proteins were eluted using Laemli buffer (BOUND). For every construct, the percentage of bound protein relative to input in the presence or absence of PF74 is shown (Figure 3B). Our results revealed that the absence of the FG peptide (ΔFG) entirely abolished CPSF6's ability to bind to the viral core. In agreement, simultaneous deletion of the FG motif and LCRs (ΔFG ΔLCR) resulted in a construct unable to bind to the viral core. Similar outcomes were observed when the LCRs were replaced with sequences derived from ADD2, even if the FG was present.

LCR-FG is notably more disordered than ADD2-FG, containing a high proportion of prolines (48 out of 98 residues), which makes it mostly non-foldable (Figure 3C-L). Since proline is a structure-disrupting residue, LRC-FG is not expected to adopt any secondary structure. In contrast, ADD2-FG contains fewer prolines (15 out of 98 residues) but has many charged residues. It is predicted to form two short α helices and a ß strand, arranged as: α helix - FG – ß strand - α helix. ADD2-FG may form a flexible collapsed state, as its oppositely charged residues are evenly distributed, potentially allowing polyelectrostatic compaction. This suggests that FG within ADD2-FG may be less accessible for the interaction with the viral core's hydrophobic pocket due to its involvement in this collapsed conformational ensemble (Figure 3C-G, suppl. Fig.4). This aligns with the inability of CPSF6 carrying ADD2 in place of the LCRs to induce CPSF6 puncta (Figure 2F). On the other hand, the deletion of only the two LCRs, while keeping the FG peptide intact, resulted in unexpected findings. The ΔLCR mutant exhibited a stronger binding affinity for the viral core when compared to the wild-type protein (Figure 3B). These results suggest that the LCRs surrounding the FG motif are modulating the affinity of CPSF6 to the viral core, which might be important for function. By contrast, deletion of the MCD (ΔMCD) but retention of other regions, such as the FG peptide and the LCRs, demonstrated a binding affinity to the viral core similar to that of the wild-type

protein. These results suggest that the MCD domain is not involved in the binding of CPSF6 to the viral core, which is not surprising since the CPSF6 (1-358), which does not have an MCD, binds to the viral core (Lee et al., 2010).

Thus, the viral capsid, through the FG peptide of CPSF6, constitutes the scaffold of HIV-induced CPSF6 puncta.

In summary, our results suggest that the FG peptide is the main determinant involved in the binding of CPSF6 to the viral capsid. Interestingly, our work implies that the LCRs may be modulating the affinity of the FG motif for the viral core (Figure 3B). Recognition motifs that mediate protein-protein interactions, such as the FG motif of CPSF6, are usually embedded within longer intrinsically disordered regions that can modulate affinity of the interaction (Karlsson et al., 2022). Taken together, our data show that the FG peptide coordinates both the binding to the viral core and the induction of CPSF6 puncta. This coordination suggests that the FG peptide plays a critical dual role in recognizing the viral capsid and facilitating the cellular clustering of CPSF6, which may be part of the cellular response to viral entry.

HIV-induced CPSF6 mutants puncta and Nuclear Speckles.

The NS factor SC35, commonly used as a marker of NS, has been detected in HIV-induced CPSF6 puncta (Figure 4A)**.** In this study, we investigated whether HIV-induced CPSF6 mutant puncta are associated with SC35. We infected cells expressing CPSF6 wild-type (WT), CPSF6 3xNLS ∆MCD, CPSF6 ∆LCR, CPSF6 ∆MCD, and CPSF6 PY NLS ∆MCD, and examined whether the nuclear puncta formed by the various CPSF6 proteins associate with SC35. Imaging analysis revealed no significant difference in the association of SC35 with CPSF6 WT or the CPSF6 mutants (Figure 4B), confirming that the disordered domains, LCRs, and MCDs are dispensable for the formation of HIV-induced CPSF6 puncta that localize in the canonical nuclear niche marked by SC35.

Biogenesis of HIV-Induced CPSF6 Puncta carrying Nuclear Speckle factors.

Nuclear speckle factors, particularly those involved in their biogenesis, such as SON and SRRM2 (Ilik et al., 2020), have been identified as constituents of HIV-induced CPSF6 puncta . However, the specific role of NSs in these puncta remains unclear. Most of the existing results have been obtained through immunofluorescence experiments conducted several days post-infection. In this study, we conducted a time course experiment to investigate the biogenesis of HIV-induced CPSF6 puncta, which contain NS factors. Our goal was to capture the fusion event between the HIV-induced CPSF6 puncta and NS, providing insights into the dynamics of how HIV manipulates host nuclear structures during infection. We hypothesized

that NS factors might either be recruited during the initial formation of HIV-induced CPSF6 puncta, shortly after the virus is released from the nuclear basket of the NPC, or later via the fusion between NSs and HIV-induced CPSF6 puncta. To investigate this, we performed live imaging to track CPSF6-mNeonGreen and NSs in cells expressing endogenous SRRM2 fused with a Halo tag, using CRISPR Paint (courtesy of Roy Parker) (Lester et al., 2021).We fixed and labeled samples at different time points post-infection, ranging from 6 h.p.i. to 30 h.p.i. (Figure 5A, Suppl. Fig.1G). At 6 h.p.i., 27% of HIV-induced CPSF6 puncta were still individual, compared to only 9% at 30 h.p.i. (Figure 5B). Concurrently, 61% of HIV-induced CPSF6 puncta were fused with NS at 6 h.p.i., rising to 75% at 30 h.p.i. (Figure 5B). This indicates a progressive increase in the number of HIV-induced CPSF6 puncta fusing with NS over time (Figure 5C). Overall, we detected individual CPSF6 clusters (green) and NSs (red) that quickly fused, confirming that this fusion occurs within the two independent puncta, CPSF6 and SRRM2, rather than during the formation of HIV-induced CPSF6 puncta (Figure 5A-C, movies 1A, 1B).

Taken together, these results suggest that HIV-induced CPSF6 puncta first form independently of NS and later fuse with NS, causing an enlargement of NS as part of the hijacking process by HIV-1.

Role of SON and SRRM2 in the Fusion and Stabilization of HIV-Induced CPSF6 Puncta within Nuclear Speckles.

Macrophage-like cells, THP-1, were depleted for SON or SRRM2 using AUM*silence*™ ASO technology (Gao et al., 2024; Marasca et al., 2022; Mazzeo et al., 2024; Zhang et al.). The level of depletion of SON and SRRM2 was evaluated by immunofluorescence using antibodies against SON and SRRM2 (Figure 6A). Both depleted cells were analyzed also for the presence of NS, labelled by SC35 (Suppl. Fig.1A-E). However, recent findings suggest that the primary target of the SC35 mAb is SRRM2 . To confirm this, cells depleted of SRRM2 were labelled with antibodies against both SRRM2 and SC35 (Suppl. Figure 5). We have observed that the reduction of SRRM2 resulted in a slight decrease in the mean intensity of SC35, whereas the depletion of SON (Suppl. Fig.1F) did not have the same effect.

Subsequently, we infected THP-1 depleted cells for SRRM2 and SON or control cells with HIV-1 and we fixed them at 48 h post-infection for immunofluorescence. We calculated the percentage of CPSF6 clusters in HIV infected THP-1 control cells (approximately 78%) and in HIV infected THP-1 cells depleted for SRRM2 (about 43%) and SON (around 66%). The results indicate that the partial depletion of SRRM2 affects the formation of HIV-induced puncta, while the depletion of SON slightly reduces their establishment (Figure 6B).

The IDR of SRRM2 is a crucial element for the fusion of HIV-Induced CPSF6 puncta to the nuclear speckles.

CPSF6 contains several disordered regions (Di Nunzio et al., 2023) as well as NS factors(Ilik et al., 2020). Previous studies have established the role of IDR of SRRM2 in the biogenesis of NS (Ilik et al., 2020).

Here, we investigate the role of IDRs within NS factors, with a specific emphasis on SRRM2, in the fusion and stabilization of HIV-induced CPSF6 clusters. To address this inquiry, we utilized previously published HEK293 cell lines generated using the CRISPaint system (Lester et al., 2021), comprising two distinct lines: HaloTag SRRM2, SRRM2 full-length (FL) (1-2748 aa) fused with halo tag(insertion at amino acid [aa] 2,708), and a cell line lacking the C-terminal IDR of SRRM2, known as ∆IDR HaloTag SRRM2(1-429 aa, with halo insertion at aa 430) .

As anticipated, the truncated form of SRRM2 displayed a more diffuse distribution within the nucleus, without recruitment to NS (suppl. Fig.1A-E), and consequently lacked nuclear puncta (Figure 6C). On the other hand, the number of SON puncta was highly similar between the two cell lines (Figure 6D).

Subsequently, we quantified the formation of CPSF6 puncta in both cell lines infected with HIV-1. No significant difference in HIV-induced CPSF6 puncta formation was observed between HEK293 and HEK293 carrying the SRRM2 halo tag (~27%). However, a substantial reduction in HIV-induced CPSF6 puncta was evident in the cell line carrying the SRRM2 form that lacks the IDR (~11%) (Figure 6E; Suppl. Fig.1). Collectively, these results underscore the pivotal role of the SRRM2 IDR in the stabilization of HIV-induced CPSF6 puncta through their fusion with NSs (Figure 5A-C).

Discussion

HIV-1 capsid has transformed long-standing assumptions in the field. Previously considered an undruggable viral target due to the belief that it disassembled shortly after HIV entered target cells, whereas recent findings have revealed its essential role in nuclear import (Ay and Di Nunzio, 2023; Blanco-Rodriguez and Di Nunzio, 2021; Blanco-Rodriguez et al., 2020; Chen et al., 2016; Taylor and Fassati, 2024; Yamashita and Emerman, 2004; Zila et al., 2021). Notably, the first-in-class antiretroviral capsid inhibitor, Lenacapavir, has shown remarkable results in patients, demonstrating that the capsid, contrary to previous belief, can indeed be a viable therapeutic target. While Lenacapavir improves patients' quality of life with only two injections per year (Link et al., 2020; Segal-Maurer et al., 2022), no current antiretroviral drugs provide a cure. This may be due to our incomplete understanding of certain aspects of HIV biology. In this study, we shed light on the post-nuclear entry steps, a critical phase for the establishment of viral reservoirs, which represent the main barrier to a cure. Recent findings have shown that not only the HIV viral capsid translocates through the NPC, but that the viral nuclear entry also enhances the formation of CPSF6 puncta. Additionally, it has been revealed that RT is completed within the nucleus (Burdick et al., 2020; Dharan et al., 2020; Rensen et al., 2021; Scoca et al., 2023; Selyutina et al., 2020). Furthermore, incoming viral RNA has been observed to be sequestered in nuclear niches in cells treated with the reversible reverse transcriptase inhibitor, NEV. When macrophage-like cells are infected in the presence of NEV, the incoming viral RNA is held within the nucleus (Rensen et al., 2021; Scoca et al., 2023). This scenario is comparable to what is observed in patients undergoing antiretroviral therapy. Interestingly, we found that if CPSF6 puncta are pharmacologically dismantled, the nuclear reverse transcription cannot be restored, indicating that HIV-induced CPSF6 puncta play a crucial role in the viral life cycle, particularly for their potential role in forming viral reservoirs. Understanding the biogenesis of these puncta could be a significant step towards deepening our knowledge of HIV biology and providing additional tools to combat this pandemic virus.

Here we identify the disordered FG peptide essential for the binding with the viral capsid as the inducer of HIV-induced CPSF6 puncta. Notably, CPSF6 protein lacking the FG peptide is incapable of forming nuclear puncta. Moreover, we discovered that the two major intrinsically disordered regions of CPSF6, the MCDs and the LCRs, are dispensable for the formation of viral nuclear puncta. The MCDs of CPSF6 have been shown to provide cohesion for NS condensation (Greig et al., 2020). In the context of HIV infection, the viral capsid induces the formation of CPSF6 puncta that are depleted of MCDs carrying the viral RNA genome. These CPSF6 puncta are highly similar to those formed by wild-type CPSF6, suggesting that MCDs do not play a significant role in this process. Additionally, the LCRs of CPSF6 also do not

appear to influence puncta formation, as their depletion does not reduce the number of CPSF6 puncta. This indicates that neither the MCDs nor the LCRs are involved in CPSF6 puncta formation during HIV infection. Surprisingly, when we assessed the ability of CPSF6 domains to bind to the viral capsid, we observed that the deletion of LCRs increases CPSF6's ability to bind to the viral capsid. We hypothesize that a change in charges may alter the binding mechanism of CPSF6 when LCRs are absent. In scenarios where the FG motif is depleted, we observed a dramatic inhibition of HIV-induced CPSF6 puncta formation and a lack of binding to the viral core *in vitro*. On the flip side, the linkage between CPSF6 entities is facilitated by the FG peptides' interaction with certain hydrophobic CA pockets along adjoining hexamers (Wei et al., 2022). Therefore, it's conceivable that FG peptides, not involved in the capsid's binding, could coalesce similarly to FG-Nups. These undergo phase separation, forming condensates with nuclear pore complex-like permeability barrier features (Hülsmann et al., 2012).

We have also identified HIV-induced CPSF6 puncta formation independent of NS at early stages post-infection, with a progressive increase in CPSF6 puncta co-localizing with NS over time. These results were obtained by focusing our studies on two scaffold proteins involved in NS biogenesis: SON and SRRM2. Over the past approximately 0.6–1.2 billion years of metazoan evolution, these two factors have undergone significant lengthening, unlike many other proteins involved in splicing. This extension primarily occurred within their IDRs, which are commonly associated with liquid-liquid phase separation (LLPS) and the formation of biomolecular condensates (Rai et al., 2018). Co-depleting SRRM2 with SON, or depleting SON in a cell line where it is deleted the intrinsically disordered C-terminus of SRRM2, abolished the formation of NSs (Ilik et al., 2020). However, the depletion of only one factor does not abolish NSs. Consistent with this finding, we observed that the depletion of SRRM2 does not affect the presence of SON nuclear clusters. Similar results were obtained with cells genetically modified to express the truncated form of SRRM2 lacking the IDRs. When the ΔIDR SRRM2-halo tag was detected by the halo ligand, there was no recruitment of the truncated SRRM2 form in NSs. However, if SC35 is used as a target for antibodies, it can still be detected, albeit with a much lower intensity signal than in cells expressing the full-length SRRM2. Additionally, we observed a significant reduction in the detection of HIV-induced CPSF6 puncta. The few CPSF6 clusters detected in these cells colocalized with the weak SC35 signal. These results suggest that the IDR of SRRM2 plays an important role in HIV-induced CPSF6 puncta stabilization, but rare clusters can appear, likely induced by redundant NS factors. This indicates that further investigation is needed to better understand the hijacking of NS by HIV.

Taken together, our results reveal the intricate interplay between individual CPSF6 domains and the viral capsid in dictating the nuclear fate of the virus. Concurrently, the IDR domain of SRRM2 contributes to the enlargement of particular nuclear speckles with the fusion of HIV-induced CPSF6 puncta. Lastly, these HIV-induced CPSF6 puncta necessitate of FG peptide to engage the viral core.

Overall, this study could provide insights into the understanding of viral invasion and persistence within the host.

Materials and methods

Cell lines

THP-1 cells are immortalized monocytic cells, which, once seeded, differentiate into macrophage-like cells under phorbol 12-myristate 13-acetate (PMA) treatment (160 nM). THP-1 cells were also engineered knocking out CPSF6. These cell populations These cells were cultivated in RPMI 1640 medium supplemented with 10% fetal bovine serum (FBS) and 1% penicillin–streptomycin solution (100 U/ml). HEK293T cells are human embryonic kidney cells used to produce LVs. For figure 2, we used two engineered HEK293 strains, Halo tagged SRRM2 HEK293 cells and Halo tagged SRRM2 ∆IDR HEK293 cells, kind gifts from Roy Parker's lab (Lester et al., 2021), characterized by the Halo tagged SRRM2 protein. In Halo tagged SRRM2 ∆IDR HEK293 cells Parker's lab deleted also the SRRM2 sequence encoding for aa 430-2748. All the HEK293 strains were cultivated in Dulbecco's modified Eagle medium supplemented with 10% FBS and 1% penicillin–streptomycin (100 U/ml).

Bacteria strains

All E. coli bacteria strains were grown in Luria-Bertani (LB) medium at 37°C. DHα competent cells and Stellar Competent Cells were used for molecular cloning, while E.coli One-Shot BL21star (DE3) cells were exploited for protein production.

Plasmids

To express the WT CPSF6/WT CPSF6-mNeonGreen and the mutant CPSF6/mutant CPSF6-mNeonGreen clones, the correspondent coding sequences were engineered in pSICO plasmids. The two original plasmids used were pSICO CPSF6-mNeonGreen and pLPCX

CPSF6 ADD2, gift from Mamuka's lab. HIV-1ΔEnvINHAΔNef plasmid encodes the ΔEnvHIV-1 LAI (BRU) viral genome where the IN protein is fused to the HA tag while pNL4.3 ∆env ∆Nef IRES GFP plasmid encodes the ΔEnvHIV-1 NL4.3 viral genome and contains also a GFP sequence headed by an IRES. The pNL4.3 ∆env ∆NefLuc has the Luciferase cDNA as reporter gene.

CRISPR-Cas9 knockout in THP-1 cells

To target CPSF6, three different crRNAs were used simultaneously (specific sequence: 5'-TCGGGCAAATGGCCAGTCAAAGG-3', 5'-AGGACGGGGCCGTTTTCCAGGGG-3', and 5'-CATGTAATCTCGGTCTTCTGGGG-3', all ordered from Integrated DNA Technologies, IDT). Pre-designed unspecific crRNA was used as control (IDT). crRNA and tracrRNA were resuspended in IDT Duplex Buffer according to the manufacturer's instructions. On the day of the nucleofection, duplexes were formed by mixing equimolar concentration of crRNA and tracrRNA, followed by 5-min annealing at 95°C. RNA duplexes were then mixed (1:2) with TrueCut™ Cas9 Protein v2 for 10 min at RT to generate ribonucleoprotein (RNP) complexes. 2×10^5 THP1 cells were resuspended in P3 Primary Cell Nucleofector™ Solution, mixed with RNP and Alt-R® Cas9 Electroporation Enhancer (90 pmol, IDT), and nucleofected in a 4D-Nucleofector™ System using the P3 Primary Cell 4D-Nucleofector™X Kit S (program FI-110). After nucleofection, cells were seeded in complete RPMI medium with 20% FBS. Three days after nucleofection, cells were plated for clonal selection.

Clonal selection of KO cell lines

Seventy-two hours post nucleofection, cells were diluted in RPMI medium containing 20% FBS and plated in five 96-well plates at 1 and 5 cells/well condition. After one-month, selected microcolonies (50-100) are placed into 24- wells plate. Once that wells were near confluence, cells were transferred into the well of a 6- well plate. After growing for another one-month, cells were proceeded for western blot.

Western blot

Proteins were extracted on ice from THP-1 cells using RIPA buffer (20 mM HEPES, pH 7.6, 150 mM NaCl, 1% sodium deoxycholate, 1% Nonidet P-40, 0.1% SDS, 2mM EDTA, complete protease inhibitor), and protein concentration was quantified using a detergent-compatible (DC) protein assay (Pierce™ BCA Protein Assay Kit) with bovine serum albumin

(BSA) as a standard. 90 µg of total protein lysate were loaded onto SDS-PAGE 4- 12% Bis-Tris gels (Invitrogen); an Ab rabbit anti-CPSF6 (1:500) and an anti-rabbit HRP-conjugated (1:5000) were used for the detection of CPSF6, whereas the normalization was done by an Ab anti- actin HRP-conjugated (1:3000). Visualization was carried out using an ECL solution .

AUMsilence™ ASO

AUMsilenceTM 352 ASOs were synthesized by AUM BioTech, LLC (Philadelphia, USA). THP-1 negative control, THP-1 KD SRRM2 and THP-1 KD SON cells were differentiated with PMA (160 nM) for 48 h then incubated 72 h with 10 µM of a AUMsilenceTM ASOs complementary to the mRNA of SRRM2 and SON, respectively (scramble control AUMscrambleTM). All cells were kept in incubator at 37°C and 5% CO2. These cells were cultivated in RPMI 1640 medium supplemented with 10% fetal bovine serum (FBS) and 1% penicillin–streptomycin solution (100 U/ml).

Cloning

pSICO CPSF6-mNeonGreen was used to generate deletion mutants by exceeding different regions from the original sequence according to Table2. The mutants were produced with and without the mNeonGreen tag, except for pSICO CPSF6-ΔMCD and pSICO CPSF6-ΔMCD PY NLS. All primers are specified in Table1.

For pSICO CPSF6-mNeonGreen ΔLCR ADD2, CPSF6 ΔLCR ADD2 sequence was amplified by PCR from pLPCX CPSF6 ADD2 and specific primers were designed to add also BamHI restriction sites at the extremities. Phusion Flash High-Fidelity PCR Master Mix was used and the reaction was performed according to the manufacture datasheet (100 ng template DNA and 56° annealing temperature). PCR products were treated with DpnI for 1h at 37°C and then digested with BamHI restriction enzyme at 37°C for 1h. 2 ul of the backbone pSICO CPSF6-mNeonGreen were digested with BamHI at 37°C for 3h and further treated with CIP. After gel extraction of the DNA and purification, insert and backbone were ligated with T4 Ligase for 2 h at 22°C. 5 ul of the product were used to transform 50 ul of DH5α bacteria (30 min at 4°C, 45 sec at 42°C, 2 min at 4°C, incubation in SOC medium for 1h at 37°C and plating on LB agar dishes).

All the other mutants were obtained using In-Fusion®Snap Assembly protocol, which allowed to amplify the original plasmid deleting the small region of interest. The different primers used are reported in Table1. As reported in the In-Fusion®Snap Assembly protocol (https://www.takarabio.com/documents/User%20Manual/In/InFusion%20Snap%20Assembly%20User%20Manual.pdf), 5 ng of the original plasmid were amplified with PrimeSTARMax DNA polymerase in 35 PCR cycles (10 sec at 98°C, 15 sec at 55°C, 5sec/kb at 72°C). The PCR products were then digested for 1h at 37°C using DpnI to get read of the original

plasmid. After plasmid cleaning up, it was circularized through a ligation of 15 min at 50°C. 2.5 ul of products were used in Stellar™ Competent Cells' transformation, following the same procedure used for DH5α transformation, previously explained.

Lentiviral vectors and viral productions

LVs and HIV-1 viruses were produced by transient transfection of HEK293T cells through calcium chloride co-precipitation. Co-transfection was performed as follows: for LVs, 10 µg of transfer vector, 10 µg of packaging plasmid (gag-pol-tat-rev), and 2.5 µg of pHCMV-VSV-G envelope plasmid; for VSV- HIV-1ΔEnvIN$_{HA}$ΔNef-VPX viruses and VSV-pNL4.3 ∆env ∆Nef IRES GFP-VPX, 10 µg HIV-1ΔEnvIN$_{HA}$ΔNef plasmid or pNL4.3 ∆env ∆Nef IRES GFP, 2.5 µg of pHCMV-VSV-G plasmid and 3 ug of SIVMAC Vpx (Durand et al., 2013). After the collection of the supernatant 48 h post-transfection, lentiviral particles were concentrated by ultracentrifugation for 1 h at 22000 rpm at 4°C and stored at −80°C. LVs and viruses were tittered by qPCR in HEK293T cells 3 days post-transduction.

Cell transduction and infection

THP-1 ctrl CRISPR clone 2 cells and THP-1 (duplex1-2-3 CRISPR) KO clone 4 cells were differentiated with PMA (160 nM) for 72 h then transduced with different mutants of CPSF6 for 72h (MOI=1) and then infected for 30h with HIV-1ΔEnvIN$_{HA}$ΔNef Vpx (MOI=10) in presence of Nevirapine (10uM). The medium was always supplemented with PMA (160 nM) .

For Halo tagged SRRM2 HEK 293 cells and Halo tagged SRRM2 ∆IDR HEK 293 cells, 2×10^5 cells were seeded on coverslips coated with polylysine in complete growth medium (DMEM, GlutaMAX^{TM}-I, 10% FBS, and 1% P/S) and incubated at 37°C (5% CO_2) for 24 h. Cells were then infected with the HIV-1ΔEnvIN$_{HA}$ΔNef (LAI) Bru (MOI 10) in complete growth medium supplemented with Nevirapine (10µM) for 24h.

THP-1 control (scramble), THP-1 KD SRRM2 Cells and THP-1 KD SON cells were seeded on coverslips and differentiated with PMA (160 nM) for 72 h. Then incubated 48 h with 10 µM of a FANA ASOs (scramble control FANA (SCR-FANA), SRRM*2*-FANA and *SON*-FANA). ASOs used in this study were designed and synthesized by AUM LifeTech (Philadelphia, PA, USA). Next, cells were infected with the pNL4.3 ∆env ∆Nef IRES GFP-VPX (MOI 25) in complete growth medium and incubated at 37°C in 5% CO_2 for 4 days.

In the four timepoints experiment, THP-1 cells were differentiated with PMA (160 nM) for 72 h. Cells were then infected with HIV-1ΔEnvIN$_{HA}$ΔNef Vpx (MOI=10) in presence of Nevirapine (10uM), in complete growth medium supplemented with PMA (160 nM), and incubated at 37°C in 5% CO_2 for 6 h, 9 h, 12 h, 30 h post infection.

Immunofluorescense Microscopy

Immunostaining

On the day of fixation, the cells were washed with PBS and fixed with 4% PFA for 15 min. Cells were treated with glycine 0.15% for 10 min, permeabilized with 0.5% Triton X-100 for 30 min, and blocked with 1% BSA for 30 min. All antibody incubations were carried out at room temperature in a dark humid chamber, for 1 h with primary antibodies and for 45 min with secondary antibodies. Washes between antibody incubations and antibody dilutions were done in 1% BSA.

Primary antibodies were diluted as follows: anti-HA 1:500, anti-CPSF6 1:400, anti-SC35 1:200, anti-SON 1:200, anti-SRRM2 1:200, Secondary antibodies used were goat anti-rabbit Alexa-488 1:300 and donkey anti-Rabbit Cy3 1:1000 for CPSF6, SON and SRRM2, goat anti-rat Alexa-647 1:100 for IN-HA, goat anti-mouse Alexa-647 1:300 or goat anti-mouse Alexa-488 1:300 for SC35, donkey anti-mouse Alexa-647 1:300 for SRRM2.

Finally, cells were stained with Hoechst 33342 1:10000 for 5 min. Coverslips were mounted on glass slides (Star Frost) with Prolong Diamond Antifade Mountant.

HaloTag Labelling

To detect the HaloTag in Halo tagged SRRM2 HEK 293 cells and Halo tagged SRRM2 ΔIDR HEK 293 cells, we used HaloTag®TMR Ligand following the Technical Manual available at www.promega.com/protocols. Specifically, cells were incubated in DMEM supplemented with HaloTag®TMR Ligand (5 µM) for 15 min at 37°C. The ligand-containing medium was then removed and replaced with an equal volume of 1X PBS, repeating the step twice and ending with warm complete medium. Cells were incubated in complete culture medium for 30 min at 37°C. The medium was then removed and replaced with an equal volume of fresh warm culture medium.

Immuno-RNA FISH

On the day of fixation, fixed cells were incubated in Permeabilization/Blocking buffer (1% BSA, 0.3% Triton X-100, 2mM Vanadyl Ribonucleoside complexes (VRCs) in RNase-free PBS) for 1 h before the antibodies' incubations. Antibodies were diluted in Permeabilization/Blocking buffer. After the primary and secondary antibody staining (respectively of 1h and 45 min) with the respective washes, cells were fixed for a second time in PFA 4% (in RNase-free PBS) for 10 min at RT with subsequent washes with RNase-free PBS. In the meantime, 40 pmol of primary smiFISH probes (Tsanov et al., 2016) (24

smiFISH probes designed against HIV-1 pol sequence (Rensen et al., 2021) were hybridized with 50 pmol of secondary FLAP probe conjugated to a Cy5 fluorophore (Cy5/AATGCATGTCGACGAGGTCCGAGTGTAA/Cy5Sp/) in 1X NEBuffer 3 (diluted in RNase-free H_2O) using a thermocycler. The program setting follows: 3 min at 85°C, 3 min at 65°C and 5 min at 25°C. FISH-probe solution was then diluted 1:50 in Hybridization buffer (90% Stellaris® RNA-FISH Hybridization Buffer, 10% Deionized Formamide). After the samples were washed in Wash A buffer (20% Stellaris® RNA-FISH Wash Buffer A, 10% Deionized Formamide, in RNase-free H_2O) at RT for 5 min, they were placed on parafilm, covered with 50 ul of FISH-probe solution in Hybridization buffer and incubated overnight at 37°C. The next day, cells were washed with Wash A buffer in the dark at 37°C for 30 min. Afterwards, the samples were incubated for 10 min in Hoechst 333342 diluted 1:10000 in RNase-free H_2O and then washed with Wash B buffer for 5 min at RT in the dark. Finally, the cells were washed in RNase-free H_2O before the coverslips were mounted on microscopy slides using ProLongTM Diamond Antifade mounting medium. The mounting medium was cured overnight at RT under the chemical hood and away from light.

Images acquisition

Images were acquired using a Zeiss LSM700 confocal inverted microscope, with a 63× objective (Plan Apochromat, oil immersion, NA = 1.4), using diode lasers at 405 nm, 488 nm, 555 nm and 639 nm for the respective fluorophores. A pixel size of 0.07 µm was used.

Live imaging on HIV-1 infected Halo tagged SRRM2 HEK 293 cells

Halo tagged SRRM2 HEK293 cells were seeded and transduced with CPSF6-mNeonGreen lentiviral vector (MOI 0.5) for 24 h. 0.3 x 10^6 Cells were then transferred on poly-L-Lysin coated Ibidi-dishes and infected with HIV-1ΔEnvIN$_{HA}$ΔNef (MOI = 40) in presence of Nevirapine (10 uM) for 2 h. After having changed the medium and labeled with HaloTag®TMR Ligand (5 µM) (see previous “HaloTag labelling” section) and Hoechst 333342 diluted 1:80000 in complete DMEM medium, 4D movies were acquired.

3D movies were acquired using a Nikon Ti2-E Confocal Inverted Spinning Disk microscope, using a 63x objective (Plan Apochromat, oil immersion, NA = 1.4) and a sCMOS Hamamatsu camera, Orca Flash 4. Pixel size 6.5 µm, 2048 x 2044 pixel, QE 82%. For the Z stuck a Z

piezo stage was used, with 0.33 µm interval. For live imaging acquisitions, cells were placed in an environmental chamber with 37°C temperature, 5% CO_2 and 21% O_2.

Protein expression and purification.

pET-11a vectors were used to express the HIV-1 capsid protein. Point mutations, A14C and E45C were introduced using the QuikChange II site-directed mutagenesis kit (Stratagene) according to the manufacturer's instructions. All proteins were expressed in *Escherichia coli* one-shoot BL21star™ (DE3) cells (Invitrogen). Briefly, LB medium was inoculated with overnight cultures, which were grown at 30°C until mid log-phase (A_{600}, 0.6-0.8). Protein expression was induced with 1 mM isopropyl-β-d-thiogalactopyranoside (IPTG) overnight at 18°C. Cells were harvested by centrifugation at 5,000 x g for 10 min at 4°C, and pellets were stored at −80°C until purification. Purification of capsid was carried out as follows. Pellets from 2 L of bacteria were lysed by sonication (Qsonica microtip: 4420; A=45; 2 minutes; 2 seconds on; 2 seconds off for 12 cycles), in 40 ml of lysis buffer (50 mM Tris pH=8, 50mM NaCl, 100mM β-mercaptoethanol and Complete EDTA-free protease inhibitor tablets). Cell debris were removed by centrifugation at 40,000g for 20 min at 4°C. Proteins from the supernatant were precipitated by incubation with 1/3 of volume of saturated ammonium sulfate containing 100 mM β-mercaptoethanol for 20 min at 4°C and centrifugation at 8,000g for 20 min at 4°C. Precipitated proteins were resuspended in 30 ml of buffer A (25mM MES pH6.5, 100 mM β-mercaptoethanol) and sonicated 2-3 times (Qsonica microtip: 4420; A=45; 2 minutes; 1 second on; 2 seconds off). Sample was dialyzed 3 times in buffer A (2 h, overnight, 2 h). The sample was sonicated and diluted in 500 ml of buffer A and was chromatographed sequentially on a 5 ml HiTrap™ Q HP column and on a 5 ml HiTrap™ SP FF column (GE Healthcare), both pre-equilibrated with buffer A. The capsid protein was eluted from HiTrap™ SP FF column using a linear gradient from 0-2 M of NaCl. Absorbance at 280 nm was checked to take the eluted fraction that had higher protein levels. Pooled fractions were dialyzed 3 times (2 h, overnight, 2 h) in storage buffer (25mM MES, 2M NaCl, 20 mM β-mercaptoethanol). Sample was concentrated using centricons to a concentration of 20 mg/ml and stored at -80°C.

Assembly of stabilized HIV-1 capsid tubes

1 mL of monomeric capsid (3 mg/mL or 1 mg/mL) was dialyzed in SnakeSkin dialysis tubing 10,000 MWCO (Thermo Scientific) against a buffer that is high in salt and contains a reducing agent (buffer 1: 50 mM Tris, pH 8, 1 M NaCl, 100 mM β-mercaptoethanol) at 4°C for 8 hours. Subsequently the protein was dialyzed against the same buffer without the reducing agent β-mercaptoethanol (buffer 2: 50 mM Tris, pH 8, 1 M NaCl) at 4°C for 8 hours. The

absence of β-mercaptoethanol in the second dialysis allows formation of disulfide bonds between Cysteine 14 and 43 inter-capsid monomers in the hexamer. Finally the protein is dialyzed against buffer 3 (20 mM Tris, pH 8,0, 40 mM NaCl) at 4°C for 8 hours. Assembled complexes were kept at 4°C up to 1 month.

Capsid binding assay protocol

Human HEK293T cells were transfected for 24 h with a plasmid expressing the specified CPSF6 variant tagged with mNeonGreen. Cell media was completely removed and cells were lysed in 300 µL of capsid binding buffer (CBB: 10 mM Tris, pH 8,0, 1,5 mM MgCl2, 10 mM KCl) by scrapping off the plate. Cells were rotated at 4°C for 15 min and then centrifuged to remove cellular debris (21,000 *x g* ,15 minutes, 4 °C). Cell lysates were incubated with stabilized HIV-1 capsid tubes for 1 h at 25 °C. Subsequently, stabilized HIV-1 capsid tubes were washed by pelleting the complexes by centrifugation at 21,000 x g for 2 min. Pellets were washed using by resuspension in CBB or PBS. Pellets were resuspended in Laemmli buffer 1X and analyzed by Western blotting using anti-p24 or anti-mNeonGreen antibodies.

Bioinformatics analysis of CPSF6

Intrinsic disorder propensity of CPSF6 was evaluated using the Rapid Intrinsic Disorder Analysis Online platform (RIDAO) (https://ridao.app/) designed to predict disordered residues and regions in a query protein based on its amino acid sequence (Dayhoff and Uversky, 2022). RIDAO yields results by combining the outputs of several commonly used per-residue disorder predictors, such as PONDR® VLXT (Romero et al., 2001), PONDR® VL3(Peng et al., 2006), PONDR® VLS2B(Peng et al., 2005), PONDR® FIT (Xue et al., 2010), as well as IUPred2 (Short) and IUPred2 (Long)(Dosztanyi et al., 2005a, b). RIDAO also computes a mean disorder score for each residue based on these. In the resulting intrinsic disorder profile, the disorder score of 0.5 is the threshold between order and disorder, where residues/regions above 0.5 are disordered, and residues/regions below 0.5 are ordered. The disorder score of 0.15 is the threshold between order and flexibility, where residues/regions with the disorder scores above 0.15 but below 0.5 are flexible, and residues/regions below 0.15 are highly ordered.

Amino acid compositions of the intrinsically disordered C-terminal domain (residues 261-358) of human CPSF6 and its different variants (CPSF6 ΔFG ΔLCR, CPSF6 ΔLCR, CPSF6 ΔFG, and CPSF6 ADD2 ΔLCR) were analyzed to evaluate the relative abundance of prion-like low complexity region (LCR) defining uncharged, charged, and Pro residues in these protein regions. The corresponding values of the relative abundance of these residue groups were

calculated by dividing numbers of prion-like LCR defining uncharged (Ala, Gly, Val, Phe, Tyr, Leu, Ile, Ser, Thr, Pro, Asn, Gln), charged (Asp, Glu, Lys, Arg), and Pro residues by total number of amino acids in the corresponding protein fragments. As references, we used the corresponding data for protein sequences deposited to the UniProtKB/Swiss-Prot database that provides information on the overall distribution of amino acids in nature (Bairoch et al., 2005); PDB Select 25 (Berman et al., 2000), which is a subset of structures from the Protein Data Bank with less than 25% sequence identity, biased towards the composition of proteins amenable to crystallization studies; and DisProt 3.4 that is comprised of a set of consensus sequences of experimentally determined disordered regions (Sickmeier et al., 2007). Per-residue intrinsic disorder propensities of the LCR-FG and ADD2-FG sequences were evaluated by PONDR® VLXT (Romero et al., 2001), which is sensitive to local peculiarities of the amino acid sequences freely available at http://www.pondr.com/ (accessed on August 03, 2024). Linear distribution of the net charge per residue (NCPR) within the LCR-FG and ADD2-FG sequences were evaluated by CIDER (Holehouse et al., 2017), which is a webserver for the analysis of a wide range of the physicochemical properties encoded by IDP sequences freely available at http://pappulab.wustl.edu/CIDER (accessed on August 03, 2024). Secondary structure propensities of the LCR-FG and ADD2-FG sequences were evaluated by PSIPRED (McGuffin et al., 2000), which is a highly accurate secondary structure prediction method freely available to non-commercial users at http://globin.bio.warwick.ac.uk/psipred/ (accessed on August 03, 2024).

IMAGING AND STATISTICAL ANALYSIS

All images were analysed using Fiji Software. More in detail, for the count of the nuclear speckles a macro was computed to segment cellular nuclei and to select and count the nuclear speckles. For the nuclei segmentation a Gaussian Blur with sigma=2 was used and for the channel related to the nuclear speckles the set threshold was at 7000-46012 (min and max). For the counting, a “size =10 – Infinity summarize” was used for the nuclei, whereas for the nuclear speckles the size was reduced to “size=0-Infinity summarize”.

For live imaging analysis, Arivis software was used to reconstruct the 3D movies.

All data were statistically analyzed with GraphPad Prism 9 (GraphPad Software, La Jolla California USA, www.graphpad.com). Calculations were performed and figures were drawn using Excel 365 or GraphPad Prism 8.0. Statistical analysis was performed, with Wilcoxon matched paired t-tests or Mann–Whitney unpaired t-tests. Spearman correlation coefficients (r) were calculated using GraphPad Prism.

Figure legends

Figure 1. Role of HIV-induced CPSF6 puncta in the nuclear RT upon removal of NEV. A) A) THP-1 cells, infected with VSV-G ∆env HIV-1 (NL4.3) ∆R LUC (MOI 10) in presence or not of Nevirapine (10 µM) for 5 days, or in presence of Nevirapine (10 µM) for 2 days and then the remaining 3 days without drug or in presence of Nevirapine (10 µM) for 2 days then in presence of PF74 (25 µM). Confocal microscopy images, to verify the presence of CPSF6 clusters, the cells are stained with anti-CPSF6 antibody (green). Nuclei are stained with Hoechst (blue). Scale bar 10 µm. **B)** Luciferase Assay, to verify luciferase expression in the aforementioned samples. Luciferase values were normalized by total proteins revealed with the Bradford kit. One-way ANOVA statistical test with multiple comparison was performed (****=p<0.0001; *=p<0.05; ns=p>0.05). Data are representative of two independent experiments.

Figure 2. Role of CPSF6 Domains in HIV-Induced CPSF6 Puncta. A) Western blots demonstrate CPSF6 depletion using a specific antibody against CPSF6 in THP-1 cells subjected to CRISPR Cas9 methods: CRISPR Cas9 bulk (left), and CRISPR Cas9 clones selected by limiting dilution (right). Each condition is normalized for actin labeling. The ratio between the intensity signal of CPSF6 and actin was analyzed via ImageJ and is plotted below each western blot. **B)** Confocal microscopy images of THP-1 ctrl CRISPR clone 2 cells (Ctrl 2) and THP-1 duplex1-2-3 CRISPR clone 4 cells (CPSF6 KO 4) infected with HIV-1ΔEnvINHAΔNef Vpx (LAI) Bru (MOI 10) in the presence of Nevirapine (10 µM). The cells are stained 30 h p.i. with anti-CPSF6 antibody and anti-HA antibodies to detect HA tagged integrase (IN). **C)** Schema of CPSF6 isoform 588aa deletion mutants. **D)** Confocal microscopy images of THP-1 CPSF6 KO cells, transduced with different mutants of CPSF6, infected with HIV-1ΔEnvINHAΔNef Vpx (LAI) Bru (MOI 10) in presence of Nevirapine (10 µM). The cells are stained with CPSF6 and HA antibody 30 h p.i.. Scale bar 5µm. **E)** Analysis of the number of CPSF6 clusters in THP-1 CPSF6 KO cells transduced with different mutants of CPSF6, not infected or infected in presence of Nevirapine (10 µM) (the number of analyzed cells is shown under the x-axis). Statistical test: ordinary one-way ANOVA (****=p<0.0001; ***=p<0.001; *=p<0.05; ns=p>0.05). **F)** The plot compares the number of CPSF6 clusters per cell in THP-1 CPSF6 KO cells transduced with different mutants of CPSF6, infected with HIV-1 in the presence of Nevirapine (10 µM). Statistical test: ordinary one-way ANOVA (****=p<0.0001; ns=p>0.05). **G)** Confocal microscopy images of THP-1

CPSF6 KO clone 4, non-transduced and non-infected or transduced with WT CPSF6 and CPSF6 3xNLSΔMCD and infected with HIV-1ΔEnvINHAΔNef Vpx (LAI) Bru (MOI 10) in presence of Nevirapine (10 µM). Immuno-RNA FISH: the cells are stained with CPSF6 (green) antibody and with 24 probes against HIV-1 Pol sequence (gray)(RNA-FISH) 25 h p.i. Nuclei are stained with Hoechst (blue). Scale bar 10 µm. Violin plot presenting the percentage of CPSF6 clusters colocalizing with the viral RNA in THP-1 CPSF6 KO clone 4 cells transduced with LVs expressing CPSF6 WT or CPSF6 3xNLSΔMCD (respectively n=73 and n=103) and infected with HIV-1ΔEnvIN$_{HA}$ΔNef Vpx (LAI) Bru (MOI 10) in presence of Nevirapine (10 µM). A total of 198 CPSF6 WT clusters and 264 CPSF6 3xNLSΔMCD clusters were counted. Statistical test: unpaired t-test, ns=p>0.05. **H)** Confocal microscopy images of THP-1 KO CPSF6 cells transduced with WT CPSF6 and CPSF6 ΔMCD without NLS, with 3xNLS or with PY NLS, respectively. Cells were differentiated for 3 days, transduced with CPSF6 lentiviral vectors (MOI 1) for 3 days and infected for 24 h with HIV-1ΔEnvINHAΔNef Vpx (LAI) Bru (MOI 10) in presence of Nevirapine (10 uM) (left panels). The panels on the right show transduced and uninfected cells. CPSF6 and the IN tagged with the HA are labeled with anti-CPSF6 (green) and anti-HA (white) antibodies, respectively. Nuclei are stained with Hoechst (blue). The arrows show CPSF6 puncta in colocalization with IN-HA. Scale bar 10µm.

Figure 3. Evaluation of CPSF6 Deletion Mutants' Binding Capacity to the Viral Core. A) Ability of wild type and mutant CPSF6 proteins to bind to the HIV-1 core. Cellular extracts derived from human 293T cells expressing similar levels of the indicated CPSF6 proteins (INPUT) were incubated with HIV-1 capsid stabilized tubes for 1 hour at room temperatures in the presence and absence of 10 µM PF74, as described in materials and methods. As a carrier control, we utilized DMSO. Subsequently, HIV-1 capsid stabilized tubes were washed, and the bound proteins eluted 1X Laemmli buffer 1X. The BOUND fractions were analysed by Western blotting using antibodies against neon-GFP and the HIV-1 capsid. **B)** Experiments were repeated at least three times and the average BOUND fraction relative to the INPUT fraction normalized to wild type binding are shown for the different CPSF6 mutants. *** indicates a p-value < 0.001; **** indicates a p-value < 0.0001; and ns indicates no significant difference as determined by unpaired t-tests. **C)** Physicochemical characteristics of the LCR-FG and ADD2-FG sequences. Intrinsic disorder predispositions evaluated by PONDR® VLXT. Position of the FR segment within the LCR-FG and ADD2-FG sequences is shown as gray shaded area. **D)** Linear distribution of the net charge per residue (NCPR) within the LCR-FG sequence evaluated by CIDER. **E)** Linear distribution of the net charge per residue (NCPR) within the ADD2-FG sequence evaluated by CIDER. **F)**

Secondary structure propensity of the LCR-FG sequence evaluated by PSIPRED. **G)** Secondary structure propensity of the ADD2-FG sequence evaluated by PSIPRED. **H)** Analysis of the peculiarities of the amino acid compositions of the intrinsically disordered C-terminal domain (residues 261-358) of human CPSF6 and its different mutants. Relative abundance of prion-like LCR defining uncharged residues in analyzed protein segments. **I)** Relative abundance of proline residues in analyzed protein segments. **L)** Relative abundance of charged residues in analyzed protein segments. The values were calculated by dividing numbers of prion-like LCR defining uncharged (Ala, Gly, Val, Phe, Tyr, Leu, Ile, Ser, Thr, Pro, Asn, Gln), Pro, and charged (Asp, Glu, Lys, Arg) residues by the total number of amino acids in the respective protein fragments. Corresponding values for all protein sequences deposited in the UniProtKB/Swiss-Prot database, PDB Select25, and DisProt are shown for comparison.

Figure 4. A) Epifluorescence microscopy images of both infected and non-infected 293T cells showing the presence of CPSF6 clusters only in the infected condition. CPSF6 and SC35 are labeled with anti-CPSF6 (red) and anti-SC35 (gray) antibodies, respectively. Nuclei are stained with Hoechst (blue). Scale bar 5µm. **B)** Confocal microscopy images of THP-1 KO CPSF6 cells, differentiated for 3 days, transduced with CPSF6 lentiviral vector (MOI 1) (specifically WT CPSF6, CPSF6 ∆LCRs, CPSF6 ∆MCD with 3xNLS, without NLS, or with PY NLS) for 3 days and infected for 24 h with HIV-1ΔEnvINHAΔNef Vpx (LAI) Bru (MOI 10) in presence of Nevirapine (10 uM). CPSF6 and nuclear speckles were labeled with anti-CPSF6 (green) and anti-SC35 (red) antibodies, respectively. Nuclei are stained with Hoechst (blue). Scale bar 10µm. The percentage of CPSF6 puncta associated with SC35 per field of view is shown in the graph. N cells were counted in each condition and a one-way ANOVA statistical test with multiple comparison was performed; ns= p value > 0.05.

Figure 5. Dynamics of the HIV-induced CPSF6 puncta formation and their fusion with NSs. A) Time course of infection of THP-1, 6h.p.i., 9h.p.i., 12h.p.i., 30h.p.i. or non infected. Cells were stained with antibodies against CPSF6 (green) and SC35 (red). **B)** The graph shows the percentage of CPSF6 puncta associated with NS or adjacent to NS or isolated from NS at different time post-infection. **C)** The graph shows the progression of CPSF6 puncta associated to NS during the time post-infection. N indicates the number of cells analyzed. One-way ANOVA statistical test with multiple comparison was performed; ns= p value > 0.05; **** indicates a p-value < 0.0001.

Figure 6. Role of SRRM2 and SON in the Formation of HIV-Induced CPSF6 puncta. A) Depletion of SON and SRRM2 in THP-1 cells using AUM*silence*™ ASO technology. The degree of depletion is quantified by mean intensity through immunofluorescence using antibodies against SON and SRRM2, respectively. Scale bar 5µm. **B)** The percentage of CPSF6 puncta formation is quantified by IF in THP-1 cells knocked down for SON, SRRM2, and control (Ctrl) infected with HIV-1 (MOI 25) for 48 h. CPSF6 is stained with an antibody against CPSF6 (green), and nuclei are stained with Hoechst (blue). The graph on the right reports the percentage of CPSF6 clusters calculated from more than 100 cells. Scale bar 10µm. Experiments were performed at least twice. **C)** Confocal microscopy images of ∆IDR HaloTag SRRM2 HEK293 and HaloTag SRRM2 HEK293 cells stained with the halo tag ligand (red), and nuclei (blue). Scale bar 10µm. **D)** Confocal microscopy images of HaloTag SRRM2 HEK293 and ∆IDR HaloTag SRRM2 HEK293 cells, both labeled with anti-SRRM2 (red) and anti-SON (gray) antibodies. Nuclei are stained with Hoechst (blue). Scale bar 10µm. Statistical studies are summarized in the violin plot which displays the distribution of the number of SON puncta per cell in the two conditions. N cells were counted and Kolmogorov Smirnov test was performed, ns= $p>0.05$. **E)** Confocal microscopy images of HaloTag SRRM2 HEK293 and ∆IDR HaloTag SRRM2 HEK293 cells, either non-infected or infected for 24 h with ΔEnvHIV-1 LAI (BRU) (MOI 10) in the presence of Nevirapine (10 µM). CPSF6 and SC35 are labeled with anti-CPSF6 (red) and anti-SC35 (gray) antibodies, respectively. Nuclei are stained with Hoechst (blue). The plot shows the mean ± SD of the percentage of cells with CPSF6 clusters calculated in n fields of view (n=24, 29, 32); N is the number of cells analyzed for each of the three different cell lines; an unpaired t-test was performed, ****=$p<0.0001$, ns=$p>0.05$. Scale bar 10µm. Experiments were performed at list twice.

Movies 1 A and B. CPSF6 MLOs can form independently from SRRM2 MLOs in HEK293 SRRM2 HaloTag cells. Spinning Disk confocal images of HEK293 SRRM2 HaloTag cells acquired 9 hours after ΔEnvHIV-1 LAI (BRU) Vpx (MOI 10) infection in presence of Nevirapine (10 µM). Cells were previously transduced for 24h with CPSF6-mNeonGreen lentiviral vector (MOI 0.5) to detect CPSF6 and SRRM2 was visualized incubating cells with the TMR-halo tag ligand. The nuclei were stained with Hoechst. Acquisitions were performed continuously for 27 minutes.

Supplementary Figures

Suppl. Figure 1. Multiple examples (**A-E**) of confocal microscopy images of HaloTag SRRM2 HEK293 and ∆IDR HaloTag SRRM2 HEK293 cells, either non-infected or infected for 24 h with ΔEnvHIV-1 LAI (BRU) (MOI 10) in the presence of Nevirapine (10 µM). CPSF6 and SC35 are labeled with anti-CPSF6 (red) and anti-SC35 (gray) antibodies, respectively. Nuclei are stained with Hoechst (blue). Scale bar 10µm. **F)** HEK 293 HaloTag SRRM2 or ∆IDR were labelled with Halo ligand (red) and an anti-SON antibody (gray), nuclei are stained by Hoechst (blue). Scale bar 10µm. **G)** THP-1 infected with ΔEnvHIV-1 LAI (BRU) Vpx were labelled with specific antibodies against CPSF6 (green) and SC35 (red), nuclei are stained with Hoechst (blue). Scale bar 10µm.

Suppl. Figure 2. A) Per-residue intrinsic disorder propensity of the CPSF6 isoform 588 aa evaluated by the Rapid Intrinsic Disorder Analysis Online platform (RIDAO) (Dayhoff and Uversky, 2022) that yields results for IU-Pred_short (yellow line), IUPred_long (blue line), PONDR® VL3 (green line), PONDR® VLXT (black line), PONDR® VSL2 (red line), and PONDR® FIT (pink line) and computes a mean disorder score for each residue based on these predictors (MDP, thick, dark pink, dashed line). Light pink shadow represents MDP error distribution. The thin black line at the disorder score of 0.5 is the threshold between order and disorder, where residues/regions with disorder scores above 0.5 are disordered, and residues/regions below 0.5 are ordered. The dashed line at the disorder score of 0.15 is the threshold between order and flexibility, where residues/regions above 0.15 are flexible, and residues/regions below 0.15 are highly ordered (upper). Schema of the deletion mutants of CPSF6 (bottom). **B)** Lentiviral Vector Transduction of PMA-Differentiated THP-1 Cells expressing CPSF6 ∆MCD fused to mNeonGreen (left), CPSF6 NLS∆MCD fused to mNeonGreen (center), CPSF6 3xNLS∆MCD fused to mNeonGreen (right). CPSF6 is represented in green, and nuclei are stained in blue. Scale bar 5µm. **C)** Lentiviral Vector Transduction of PMA-Differentiated THP-1 cells expressing CPSF6 deletion mutants fused to mNeonGreen. CPSF6 is represented in green, and nuclei are stained in blue. Scale bar 5µm.

Suppl. Figure 3. A) Multiple examples of confocal microscopy images of THP-1 ctrl CRISPR clone 2 cells and THP-1 CPSF6 KO 4 cells infected with HIV-1ΔEnvINHAΔNef Vpx (LAI) Bru (MOI 10) in the presence of Nevirapine (10 µM). The cells are stained 30 h p.i. with CPSF6 (green) and HA (red) antibodies to detect integrase (IN). **B)** Multiple examples of THP-1 CPSF6 KO clone 4 cells transduced with different LVs carrying CPSF6 WT or mutants and stained with CPSF6 and HA antibody 30 h p.i.. Scale bar 5µm.

Suppl. Figure 4. Sequences of FG and LCRs or substituted amino acids sequences analyzed in figure 5.

Suppl. Figure 5. SRRM2 depletion is confirmed using antibodies against both SRRM2 and SC35 by IF. Statistical analysis: One-Way ANOVA (****=p<0.0001; *=p<0.05; ns=p>0.05).

ACKNOWLEDGMENTS

F.D.N. is supported by the Institut Pasteur and ANRS grants (ECTZ192036, ECTZ137593), ANR-PRCI grant, Sidaction grant. C.T. is supported by fellowships Sidaction, ANR-PRCI. S.A. is supported by ANRS fellowship ECTZ204694. F.D.-G., B.C., M.R., C.L., C.B. are supported by NIH Grants R01AI087390 and R01AI150455. We gratefully acknowledge the UtechS Photonic BioImaging platform (Imagopole), C2RT at Institut Pasteur. We thank the NIH AIDS Reagents program to support us with precious reagents and Addgene.

Bibliography

Ay, S., and Di Nunzio, F. (2023). HIV-Induced CPSF6 Condensates. J Mol Biol *435*, 168094.
Bairoch, A., Apweiler, R., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M.*, et al.* (2005). The Universal Protein Resource (UniProt). Nucleic Acids Res *33*, D154-159.
Barre-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C.*, et al.* (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science *220*, 868-871.
Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res *28*, 235-242.
Blanco-Rodriguez, G., and Di Nunzio, F. (2021). The Viral Capsid: A Master Key to Access the Host Nucleus. Viruses *13*.
Blanco-Rodriguez, G., Gazi, A., Monel, B., Frabetti, S., Scoca, V., Mueller, F., Schwartz, O., Krijnse-Locker, J., Charneau, P., and Di Nunzio, F. (2020). Remodeling of the Core Leads HIV-1 Preintegration Complex into the Nucleus of Human Lymphocytes. J Virol *94*.
Buffone, C., Martinez-Lopez, A., Fricke, T., Opp, S., Severgnini, M., Cifola, I., Petiti, L., Frabetti, S., Skorupka, K., Zadrozny, K.K.*, et al.* (2018). Nup153 Unlocks the Nuclear Pore Complex for HIV-1 Nuclear Translocation in Nondividing Cells. J Virol *92*.
Burdick, R.C., Li, C., Munshi, M., Rawson, J.M.O., Nagashima, K., Hu, W.S., and Pathak, V.K. (2020). HIV-1 uncoats in the nucleus near sites of integration. Proc Natl Acad Sci U S A *117*, 5486-5493.
Chen, N.Y., Zhou, L., Gane, P.J., Opp, S., Ball, N.J., Nicastro, G., Zufferey, M., Buffone, C., Luban, J., Selwood, D.*, et al*. (2016). HIV-1 capsid is involved in post-nuclear entry steps. Retrovirology *13*, 28.
Chen, Y., Zhang, Y., Wang, Y., Zhang, L., Brinkman, E.K., Adam, S.A., Goldman, R., van Steensel, B., Ma, J., and Belmont, A.S. (2018). Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. J Cell Biol *217*, 4025-4048.
Dayhoff, G.W., 2nd, and Uversky, V.N. (2022). Rapid prediction and analysis of protein intrinsic disorder. Protein Sci *31*, e4496.

921 Dharan, A., Bachmann, N., Talley, S., Zwikelmaier, V., and Campbell, E.M. (2020). Nuclear pore
922 blockade reveals that HIV-1 completes reverse transcription and uncoating in the nucleus. Nature
923 Microbiology *5*, 1088-1095.
924 Di Nunzio, F. (2013). New insights in the role of nucleoporins: a bridge leading to concerted steps
925 from HIV-1 nuclear entry until integration. Virus Res *178*, 187-196.
926 Di Nunzio, F., Danckaert, A., Fricke, T., Perez, P., Fernandez, J., Perret, E., Roux, P., Shorte, S., Charneau,
927 P., Diaz-Griffero, F.*, et al.* (2012). Human nucleoporins promote HIV-1 docking at the nuclear pore,
928 nuclear import and integration. PLoS One *7*, e46037.
929 Di Nunzio, F., Uversky, V.N., and Mouland, A.J. (2023). Biomolecular condensates: insights into early
930 and late steps of the HIV-1 replication cycle. Retrovirology *20*, 4.
931 Dickson, C.F., Hertel, S., Tuckwell, A.J., Li, N., Ruan, J., Al-Izzi, S.C., Ariotti, N., Sierecki, E., Gambin, Y.,
932 Morris, R.G.*, et al.* (2024). The HIV capsid mimics karyopherin engagement of FG-nucleoporins.
933 Nature.
934 Dosztanyi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005a). IUPred: web server for the prediction of
935 intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics *21*,
936 3433-3434.
937 Dosztanyi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005b). The pairwise energy content estimated
938 from amino acid composition discriminates between folded and intrinsically unstructured proteins. J
939 Mol Biol *347*, 827-839.
940 Durand, S., Nguyen, X.N., Turpin, J., Cordeil, S., Nazaret, N., Croze, S., Mahieux, R., Lachuer, J., Legras-
941 Lachuer, C., and Cimarelli, A. (2013). Tailored HIV-1 vectors for genetic modification of primary human
942 dendritic cells and monocytes. J Virol *87*, 234-242.
943 Francis, A.C., Marin, M., Singh, P.K., Achuthan, V., Prellberg, M.J., Palermino-Rowland, K., Lan, S.,
944 Tedbury, P.R., Sarafianos, S.G., Engelman, A.N.*, et al.* (2020). Publisher Correction: HIV-1 replication
945 complexes accumulate in nuclear speckles and integrate into speckle-associated genomic domains.
946 Nat Commun *11*, 6165.
947 Fu, L., Weiskopf, E.N., Akkermans, O., Swanson, N.A., Cheng, S., Schwartz, T.U., and Görlich, D. (2024).
948 HIV-1 capsids enter the FG phase of nuclear pores like a transport receptor. Nature *626*, 843-851.
949 Fu, X.D., and Maniatis, T. (1990). Factor required for mammalian spliceosome assembly is localized to
950 discrete regions in the nucleus. Nature *343*, 437-441.
951 Gao, L., Zhang, W., Zhang, L., Gromova, B., Chen, G., Csizmadia, E., Cagle, C., Nastasio, S., Ma, Y.,
952 Bonder, A.*, et al.* (2024). Silencing of aryl hydrocarbon receptor repressor restrains Th17 cell
953 immunity in autoimmune hepatitis. Journal of Autoimmunity *143*, 103162.
954 Greig, J.A., Nguyen, T.A., Lee, M., Holehouse, A.S., Posey, A.E., Pappu, R.V., and Jedd, G. (2020).
955 Arginine-Enriched Mixed-Charge Domains Provide Cohesion for Nuclear Speckle Condensation. Mol
956 Cell *77*, 1237-1250.e1234.
957 Holehouse, A.S., Das, R.K., Ahad, J.N., Richardson, M.O., and Pappu, R.V. (2017). CIDER: Resources to
958 Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. Biophys J *112*, 16-21.
959 Hülsmann, B.B., Labokha, A.A., and Görlich, D. (2012). The permeability of reconstituted nuclear
960 pores provides direct evidence for the selective phase model. Cell *150*, 738-751.
961 Ilik, I.A., Malszycki, M., Lubke, A.K., Schade, C., Meierhofer, D., and Aktas, T. (2020). SON and SRRM2
962 are essential for nuclear speckle formation. Elife *9*.
963 Karlsson, E., Schnatwinkel, J., Paissoni, C., Andersson, E., Herrmann, C., Camilloni, C., and Jemth, P.
964 (2022). Disordered Regions Flanking the Binding Interface Modulate Affinity between CBP and NCOA.
965 J Mol Biol *434*, 167643.
966 Lee, K., Ambrose, Z., Martin, T.D., Oztop, I., Mulky, A., Julias, J.G., Vandegraaff, N., Baumann, J.G.,
967 Wang, R., Yuen, W.*, et al.* (2010). Flexible use of nuclear import pathways by HIV-1. Cell Host Microbe
968 *7*, 221-233.
969 Lelek, M., Casartelli, N., Pellin, D., Rizzi, E., Souque, P., Severgnini, M., Di Serio, C., Fricke, T., Diaz-
970 Griffero, F., Zimmer, C.*, et al.* (2015). Chromatin organization at the nuclear pore favours HIV
971 replication. Nat Commun *6*, 6483.

Lester, E., Ooi, F.K., Bakkar, N., Ayers, J., Woerman, A.L., Wheeler, J., Bowser, R., Carlson, G.A., Prusiner, S.B., and Parker, R. (2021). Tau aggregates are RNA-protein assemblies that mislocalize multiple nuclear speckle components. Neuron *109*, 1675-1691 e1679.

Li, C., Burdick, R.C., Nagashima, K., Hu, W.S., and Pathak, V.K. (2021). HIV-1 cores retain their integrity until minutes before uncoating in the nucleus. Proc Natl Acad Sci U S A *118*.

Link, J.O., Rhee, M.S., Tse, W.C., Zheng, J., Somoza, J.R., Rowe, W., Begley, R., Chiu, A., Mulato, A., Hansen, D.*, et al.* (2020). Clinical targeting of HIV capsid protein with a long-acting small molecule. Nature *584*, 614-618.

Luchsinger, C., Lee, K., Mardones, G.A., KewalRamani, V.N., and Diaz-Griffero, F. (2023). Formation of nuclear CPSF6/CPSF5 biomolecular condensates upon HIV-1 entry into the nucleus is important for productive infection. Sci Rep *13*, 10974.

Marasca, F., Sinha, S., Vadalà, R., Polimeni, B., Ranzani, V., Paraboschi, E.M., Burattin, F.V., Ghilotti, M., Crosti, M., Negri, M.L.*, et al.* (2022). LINE1 are spliced in non-canonical transcript variants to regulate T cell quiescence and exhaustion. Nature Genetics *54*, 180-193.

Matreyek, K.A., Yucel, S.S., Li, X., and Engelman, A. (2013). Nucleoporin NUP153 phenylalanine-glycine motifs engage a common binding pocket within the HIV-1 capsid protein to mediate lentiviral infectivity. PLoS Pathog *9*, e1003693.

Mazzeo, L., Ghosh, S., Di Cicco, E., Isma, J., Tavernari, D., Samarkina, A., Ostano, P., Youssef, M.K., Simon, C., and Dotto, G.P. (2024). ANKRD1 is a mesenchymal-specific driver of cancer-associated fibroblast activation bridging androgen receptor loss to AP-1 activation. Nature Communications *15*, 1038.

McGuffin, L.J., Bryson, K., and Jones, D.T. (2000). The PSIPRED protein structure prediction server. Bioinformatics *16*, 404-405.

Muller, T.G., Zila, V., Peters, K., Schifferdecker, S., Stanic, M., Lucic, B., Laketa, V., Lusic, M., Muller, B., and Krausslich, H.G. (2021). HIV-1 uncoating by release of viral cDNA from capsid-like structures in the nucleus of infected cells. Elife *10*.

Peng, K., Radivojac, P., Vucetic, S., Dunker, A.K., and Obradovic, Z. (2006). Length-dependent prediction of protein intrinsic disorder. BMC Bioinformatics *7*, 208.

Peng, K., Vucetic, S., Radivojac, P., Brown, C.J., Dunker, A.K., and Obradovic, Z. (2005). Optimizing long intrinsic disorder predictors with protein evolutionary information. J Bioinform Comput Biol *3*, 35-60.

Pornillos, O., Ganser-Pornillos, B.K., Kelly, B.N., Hua, Y., Whitby, F.G., Stout, C.D., Sundquist, W.I., Hill, C.P., and Yeager, M. (2009). X-ray structures of the hexameric building block of the HIV capsid. Cell *137*, 1282-1292.

Price, A.J., Jacques, D.A., McEwan, W.A., Fletcher, A.J., Essig, S., Chin, J.W., Halambage, U.D., Aiken, C., and James, L.C. (2014). Host cofactors and pharmacologic ligands share an essential interface in HIV-1 capsid that is lost upon disassembly. PLoS Pathog *10*, e1004459.

Rai, A.K., Chen, J.X., Selbach, M., and Pelkmans, L. (2018). Kinase-controlled phase transition of membraneless organelles in mitosis. Nature *559*, 211-216.

Rasaiyaah, J., Tan, C.P., Fletcher, A.J., Price, A.J., Blondeau, C., Hilditch, L., Jacques, D.A., Selwood, D.L., James, L.C., Noursadeghi, M.*, et al.* (2013). HIV-1 evades innate immune recognition through specific cofactor recruitment. Nature *503*, 402-405.

Rensen, E., Mueller, F., Scoca, V., Parmar, J.J., Souque, P., Zimmer, C., and Di Nunzio, F. (2021). Clustering and reverse transcription of HIV-1 genomes in nuclear niches of macrophages. EMBO J *40*, e105247.

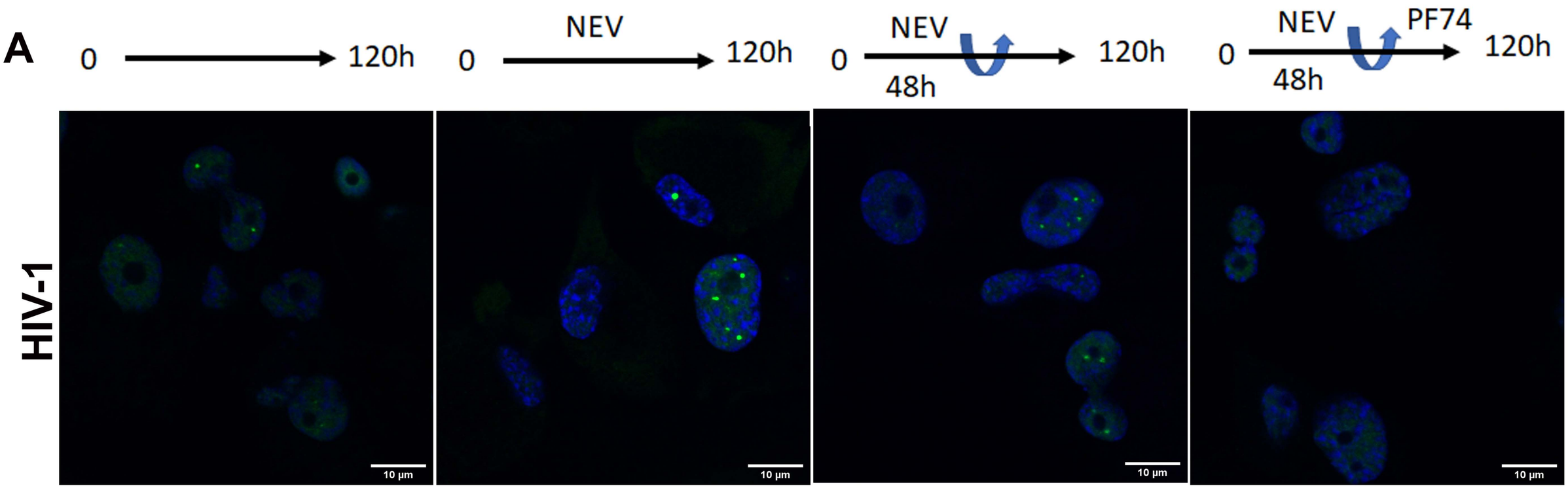
Romero, P., Obradovic, Z., Li, X., Garner, E.C., Brown, C.J., and Dunker, A.K. (2001). Sequence complexity of disordered protein. Proteins *42*, 38-48.

Scoca, V., Morin, R., Collard, M., Tinevez, J.Y., and Di Nunzio, F. (2023). HIV-induced membraneless organelles orchestrate post-nuclear entry steps. J Mol Cell Biol *14*.

Segal-Maurer, S., DeJesus, E., Stellbrink, H.J., Castagna, A., Richmond, G.J., Sinclair, G.I., Siripassorn, K., Ruane, P.J., Berhe, M., Wang, H.*, et al.* (2022). Capsid Inhibition with Lenacapavir in Multidrug-Resistant HIV-1 Infection. N Engl J Med *386*, 1793-1803.

1023 Selyutina, A., Bulnes-Ramos, A., and Diaz-Griffero, F. (2018). Binding of host factors to stabilized HIV-1
1024 capsid tubes. Virology *523*, 1-5.
1025 Selyutina, A., Persaud, M., Lee, K., KewalRamani, V., and Diaz-Griffero, F. (2020). Nuclear Import of
1026 the HIV-1 Core Precedes Reverse Transcription and Uncoating. Cell Rep *32*, 108201.
1027 Sickmeier, M., Hamilton, J.A., LeGall, T., Vacic, V., Cortese, M.S., Tantos, A., Szabo, B., Tompa, P., Chen,
1028 J., Uversky, V.N.*, et al.* (2007). DisProt: the Database of Disordered Proteins. Nucleic Acids Res *35*,
1029 D786-793.
1030 Suzuki, Y., and Craigie, R. (2007). The road to chromatin — nuclear entry of retroviruses. Nature
1031 Reviews Microbiology *5*, 187-196.
1032 Taylor, I.A., and Fassati, A. (2024). The capsid revolution. J Mol Cell Biol *15*.
1033 Tsanov, N., Samacoits, A., Chouaib, R., Traboulsi, A.M., Gostan, T., Weber, C., Zimmer, C., Zibara, K.,
1034 Walter, T., Peter, M.*, et al.* (2016). smiFISH and FISH-quant - a flexible single RNA detection approach
1035 with super-resolution capability. Nucleic Acids Res *44*, e165.
1036 Wei, G., Iqbal, N., Courouble, V.V., Francis, A.C., Singh, P.K., Hudait, A., Annamalai, A.S., Bester, S.,
1037 Huang, S.W., Shkriabai, N.*, et al.* (2022). Prion-like low complexity regions enable avid virus-host
1038 interactions during HIV-1 infection. Nat Commun *13*, 5879.
1039 Xue, B., Dunbrack, R.L., Williams, R.W., Dunker, A.K., and Uversky, V.N. (2010). PONDR-FIT: a meta-
1040 predictor of intrinsically disordered amino acids. Biochim Biophys Acta *1804*, 996-1010.
1041 Yamashita, M., and Emerman, M. (2004). Capsid is a dominant determinant of retrovirus infectivity in
1042 nondividing cells. J Virol *78*, 5670-5678.
1043 Zhang, L., Zhang, Y., Chen, Y., Gholamalamdari, O., Wang, Y., Ma, J., and Belmont, A.S. (2021). TSA-seq
1044 reveals a largely conserved genome organization relative to nuclear speckles with small position
1045 changes tightly correlated with gene expression changes. Genome Res *31*, 251-264.
1046 Zhang, Z.-Y., Harischandra, D.S., Wang, R., Ghaisas, S., Zhao, J.Y., McMonagle, T.P., Zhu, G., Lacuarta,
1047 K.D., Song, J., Trojanowski, J.Q.*, et al.* TRIM11 protects against tauopathies and is down-regulated in
1048 Alzheimer's disease. Science *381*, eadd6696.
1049 Zila, V., Margiotta, E., Turonova, B., Muller, T.G., Zimmerli, C.E., Mattei, S., Allegretti, M., Borner, K.,
1050 Rada, J., Muller, B.*, et al.* (2021). Cone-shaped HIV-1 capsids are transported through intact nuclear
1051 pores. Cell *184*, 1032-1046 e1018.

1052


A
0
120h
0
NEV
120h
0
NEV
48h
120h
0
NEV
PF74
48h
120h
HIV-1
10 µm
10 µm
10 µm
10 µm

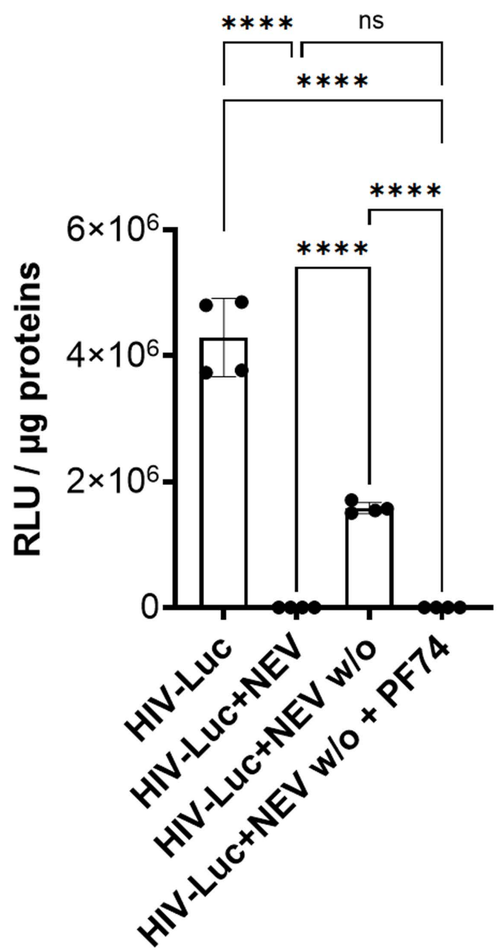

B
****
ns
****
****
6×10^6
4×10^6
2×10^6
0
RLU / µg proteins
HIV-Luc
HIV-Luc+NEV
HIV-Luc+NEV w/o
HIV-Luc+NEV w/o + PF74

Figure 1

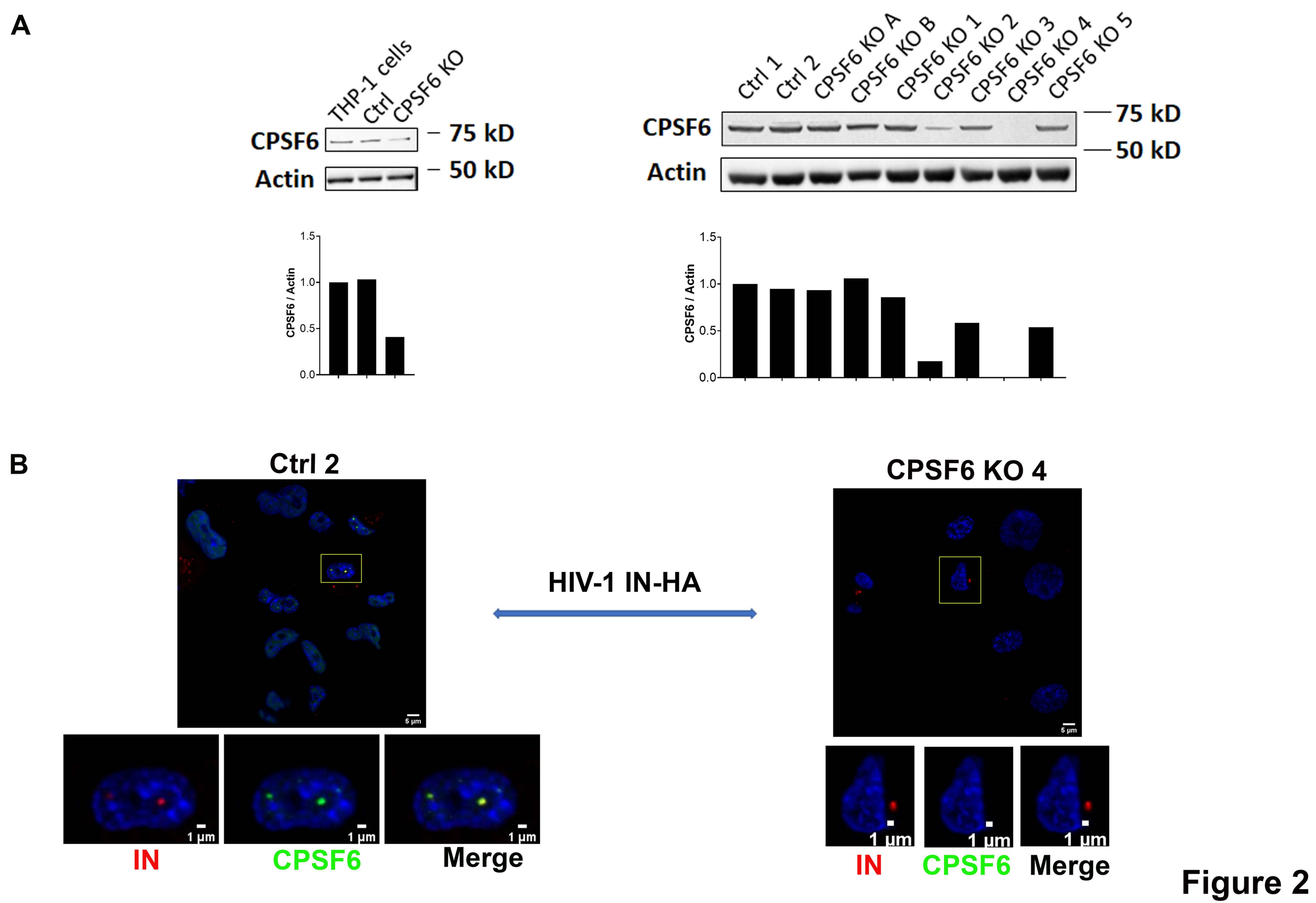

A
THP-1 cells
Ctrl
CPSF6 KO
CPSF6
Actin
75 kD
50 kD
CPSF6 / Actin
1.5
1.0
0.5
0.0
Ctrl 1
Ctrl 2
CPSF6 KO A
CPSF6 KO B
CPSF6 KO 1
CPSF6 KO 2
CPSF6 KO 3
CPSF6 KO 4
CPSF6 KO 5
B
Ctrl 2
CPSF6 KO 4
HIV-1 IN-HA
5 µm
1 µm
IN
CPSF6
Merge

Figure 2

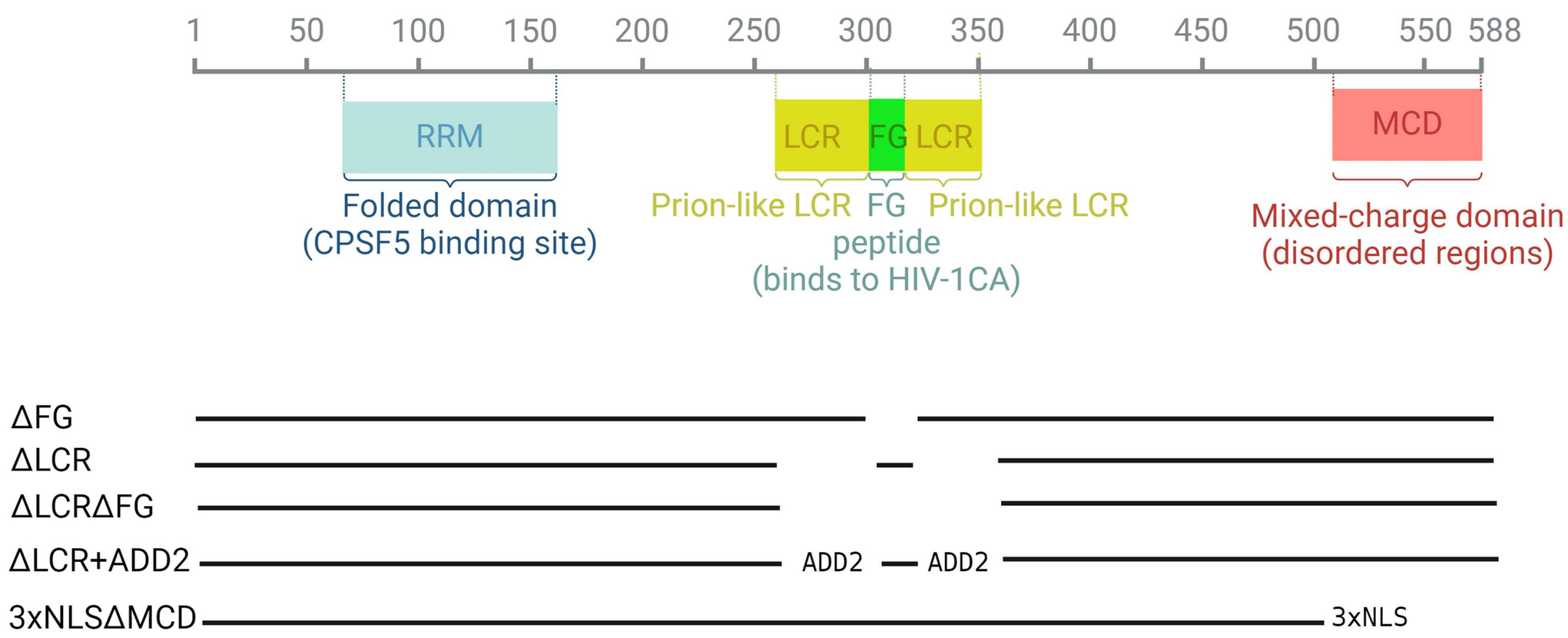

C
1
50
100
150
200
250
300
350
400
450
500
550
588
RRM
LCR
FG
LCR
MCD
Folded domain
(CPSF5 binding site)
Prion-like LCR
FG
peptide
(binds to HIV-1CA)
Prion-like LCR
Mixed-charge domain
(disordered regions)
ΔFG
ΔLCR
ΔLCRΔFG
ΔLCR+ADD2
ADD2
ADD2
3xNLSΔMCD
3xNLS

Figure 2

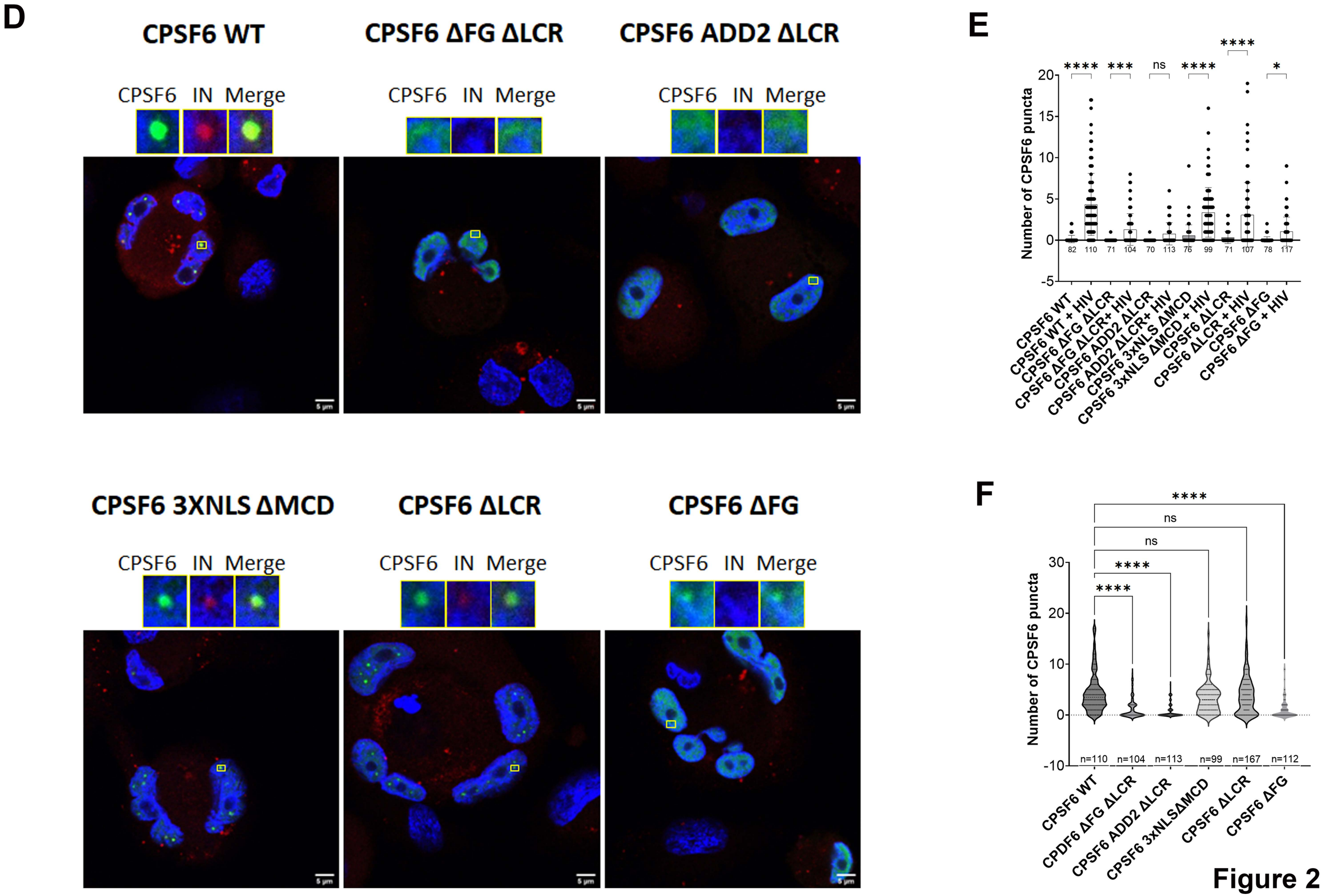

D
CPSF6 WT
CPSF6 ΔFG ΔLCR
CPSF6 ADD2 ΔLCR
CPSF6 IN Merge
CPSF6 3XNLS ΔMCD
CPSF6 ΔLCR
CPSF6 ΔFG
5 μm
E
Number of CPSF6 puncta
20
15
10
5
0
-5
82 110 71 104 70 113 76 99 71 107 78 117
**** *** ns **** **** *
CPSF6 WT
CPSF6 WT + HIV
CPSF6 ΔFG ΔLCR
CPSF6 ΔFG ΔLCR+ HIV
CPSF6 ADD2 ΔLCR
CPSF6 ADD2 ΔLCR+ HIV
CPSF6 3xNLS ΔMCD
CPSF6 3xNLS ΔMCD + HIV
CPSF6 ΔLCR
CPSF6 ΔLCR + HIV
CPSF6 ΔFG
CPSF6 ΔFG + HIV
F
Number of CPSF6 puncta
30
20
10
0
-10
n=110 n=104 n=113 n=99 n=167 n=112
**** ns ns **** ****
CPSF6 WT
CPDF6 ΔFG ΔLCR
CPSF6 ADD2 ΔLCR
CPSF6 3xNLSΔMCD
CPSF6 ΔLCR
CPSF6 ΔFG

Figure 2

G

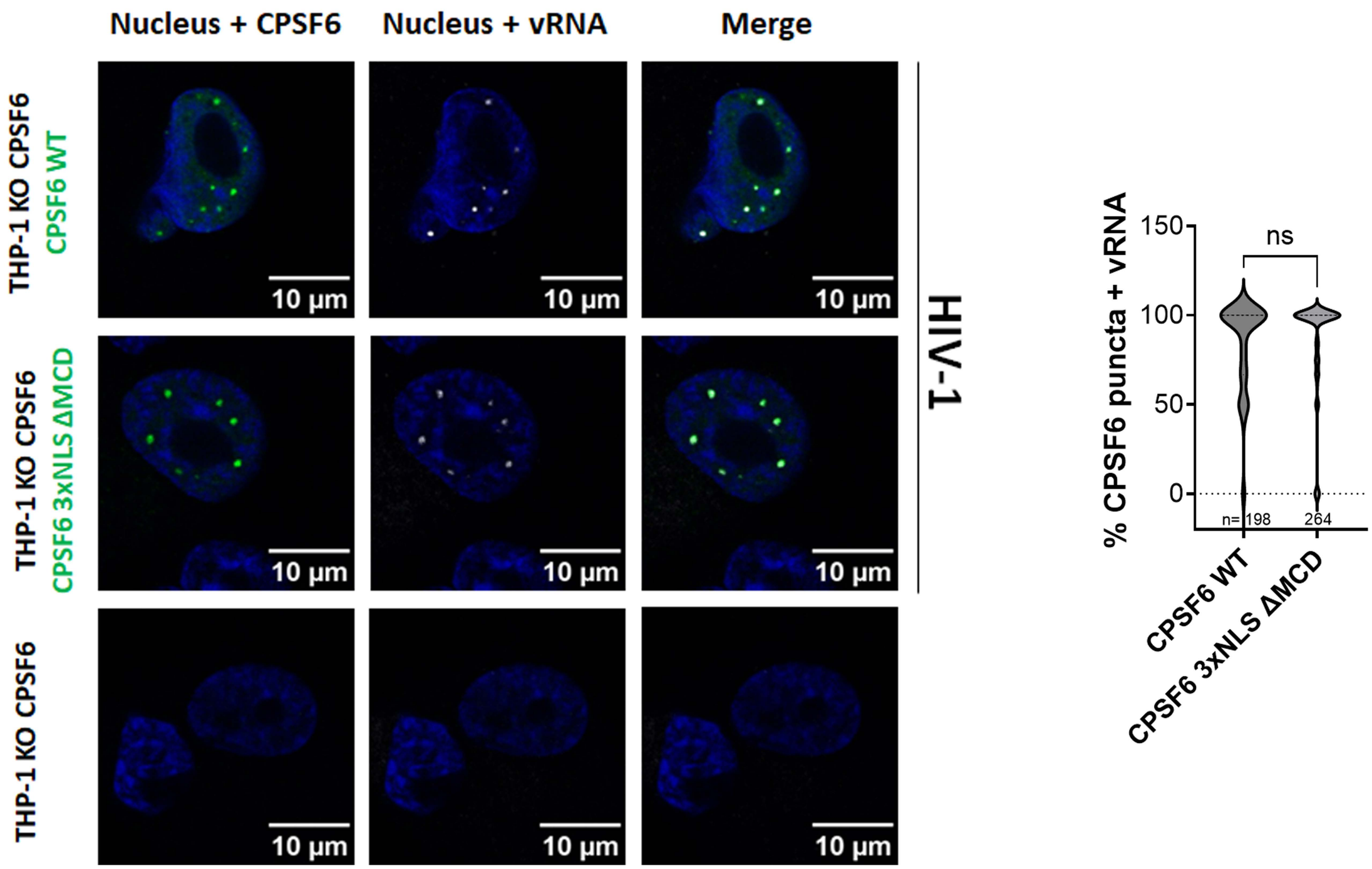

Nucleus + CPSF6
Nucleus + vRNA
Merge
THP-1 KO CPSF6
CPSF6 WT
THP-1 KO CPSF6
CPSF6 3xNLS ΔMCD
THP-1 KO CPSF6
10 µm
HIV-1
% CPSF6 puncta + vRNA
150
100
50
0
ns
n= 198
264
CPSF6 WT
CPSF6 3xNLS ΔMCD

Figure 2

H

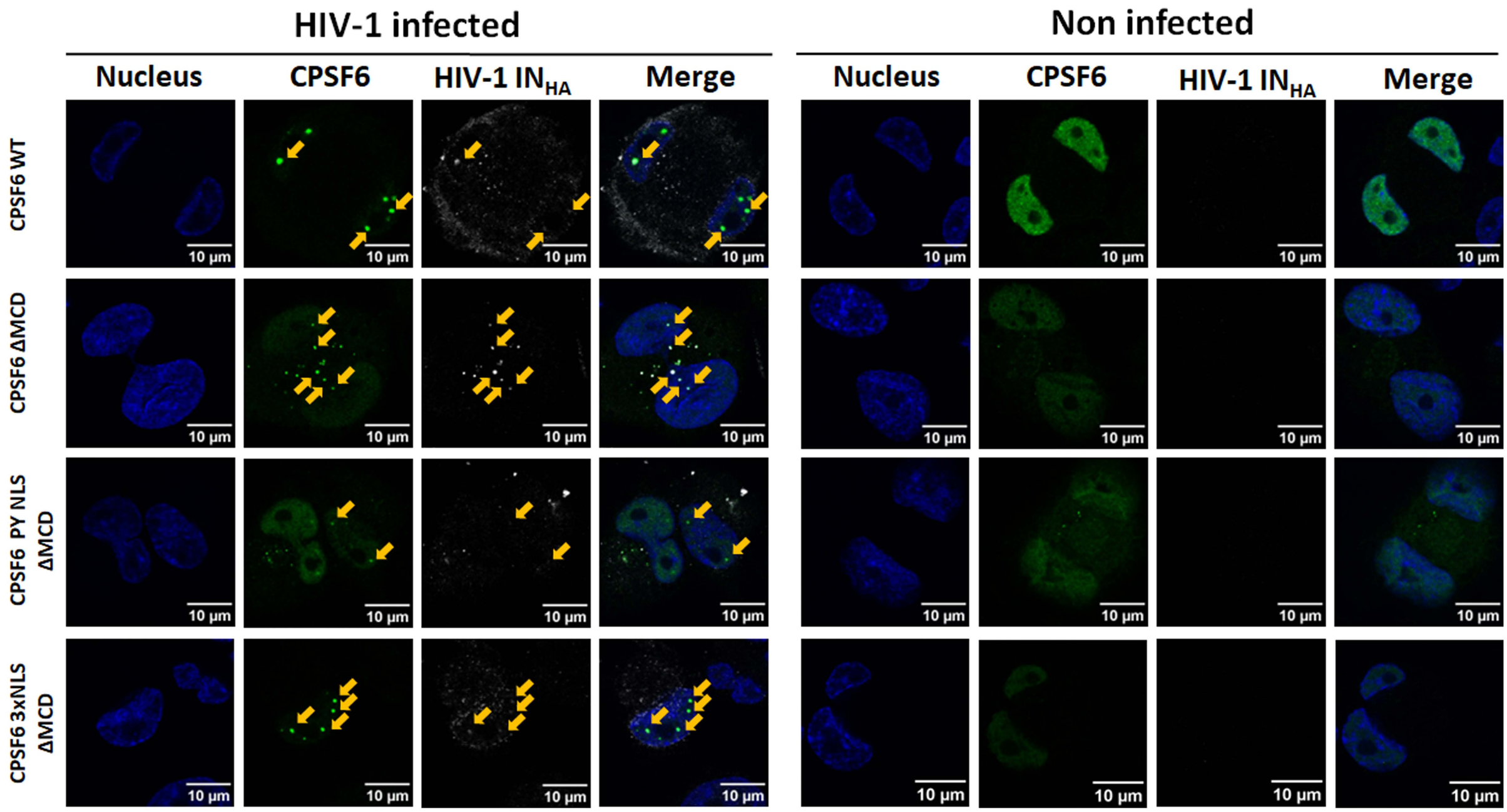

HIV-1 infected
Non infected
Nucleus
CPSF6
HIV-1 IN$_{HA}$
Merge
CPSF6 WT
CPSF6 ΔMCD
CPSF6 PY NLS ΔMCD
CPSF6 3xNLS ΔMCD
10 µm

Figure 2

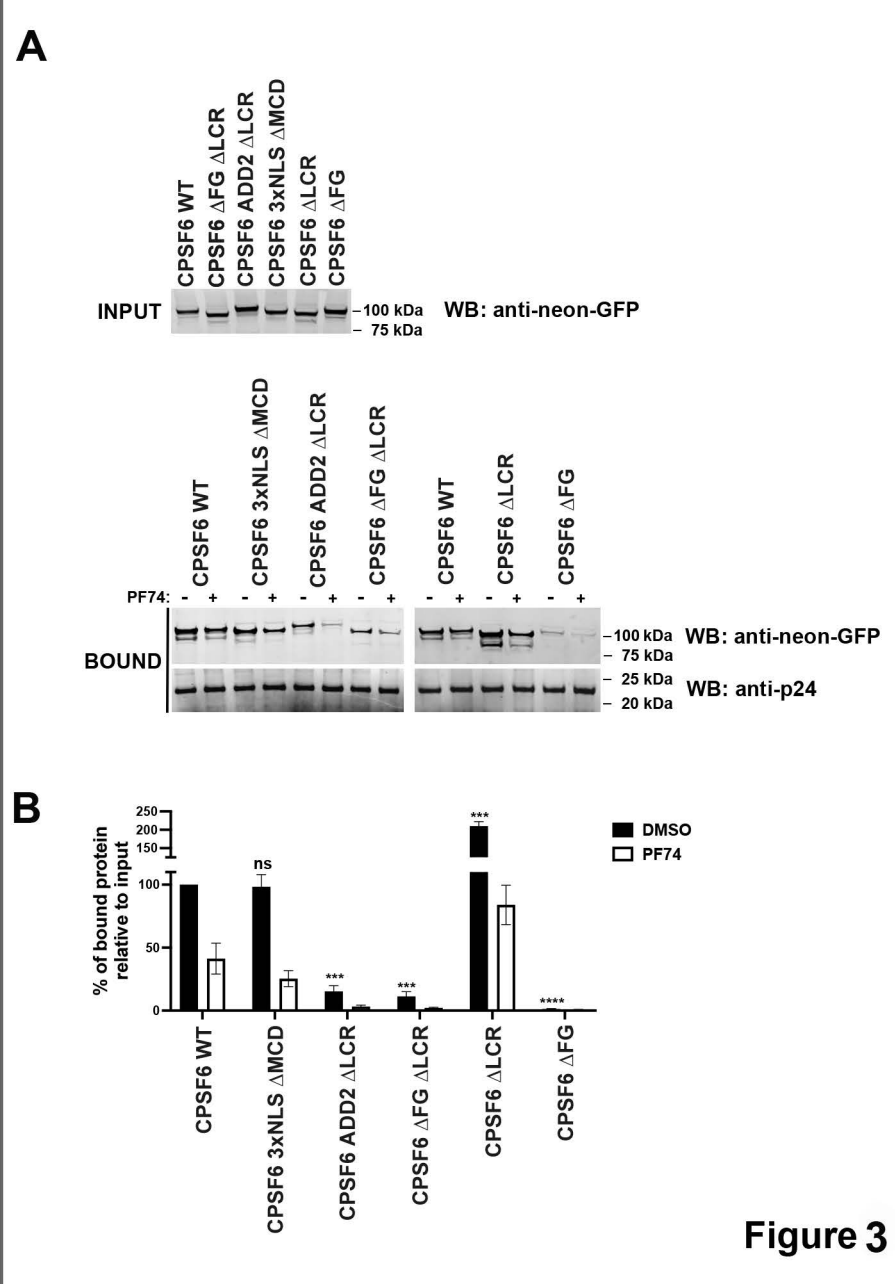

A
CPSF6 WT
CPSF6 ΔFG ΔLCR
CPSF6 ADD2 ΔLCR
CPSF6 3xNLS ΔMCD
CPSF6 ΔLCR
CPSF6 ΔFG
INPUT
100 kDa
75 kDa
WB: anti-neon-GFP
CPSF6 WT
CPSF6 3xNLS ΔMCD
CPSF6 ADD2 ΔLCR
CPSF6 ΔFG ΔLCR
CPSF6 WT
CPSF6 ΔLCR
CPSF6 ΔFG
PF74:
BOUND
100 kDa
75 kDa
WB: anti-neon-GFP
25 kDa
20 kDa
WB: anti-p24
B
% of bound protein relative to input
250
200
150
100
50
0
ns
***
***
***
****
DMSO
PF74
CPSF6 WT
CPSF6 3xNLS ΔMCD
CPSF6 ADD2 ΔLCR
CPSF6 ΔFG ΔLCR
CPSF6 ΔLCR
CPSF6 ΔFG

Figure 3

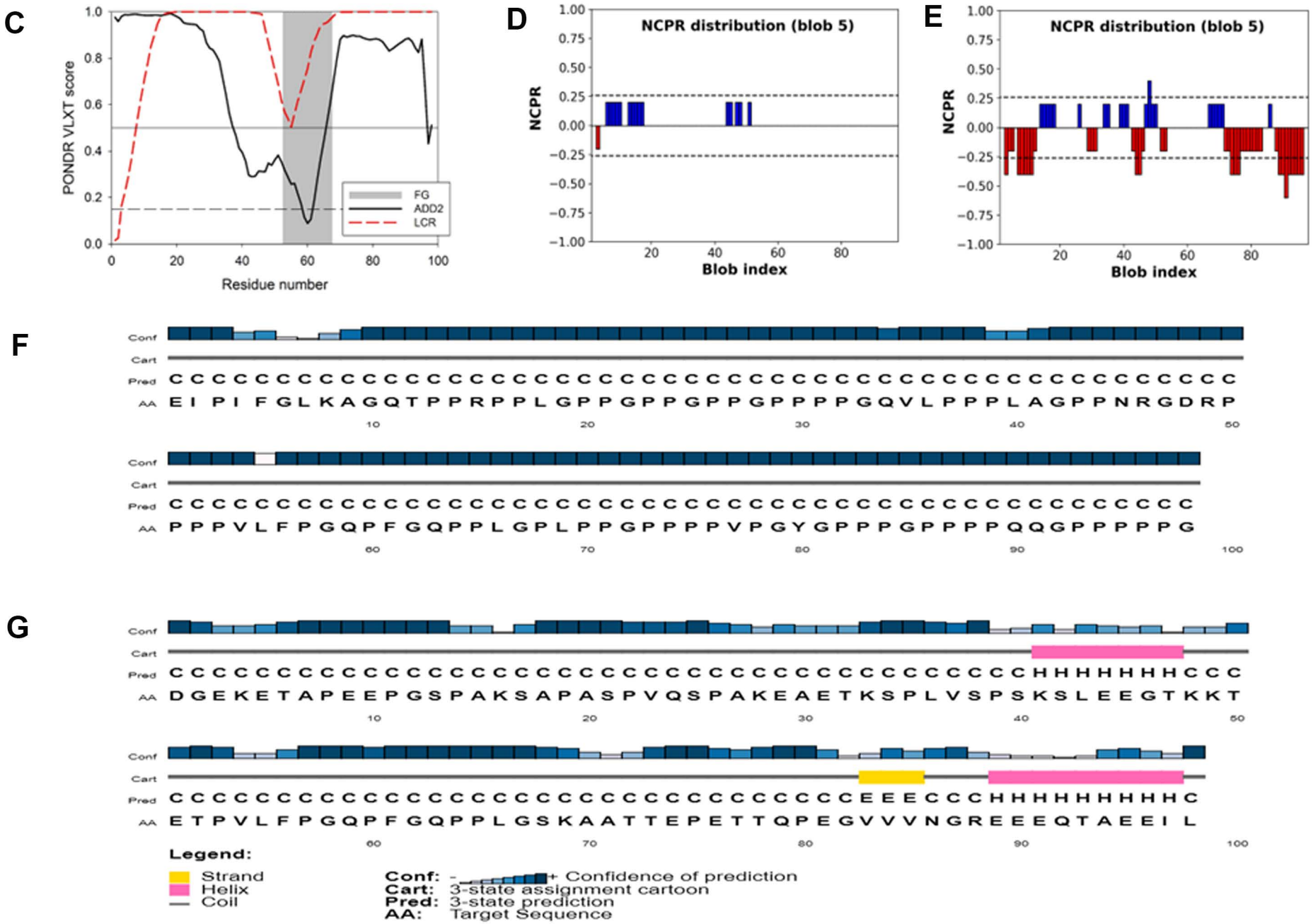
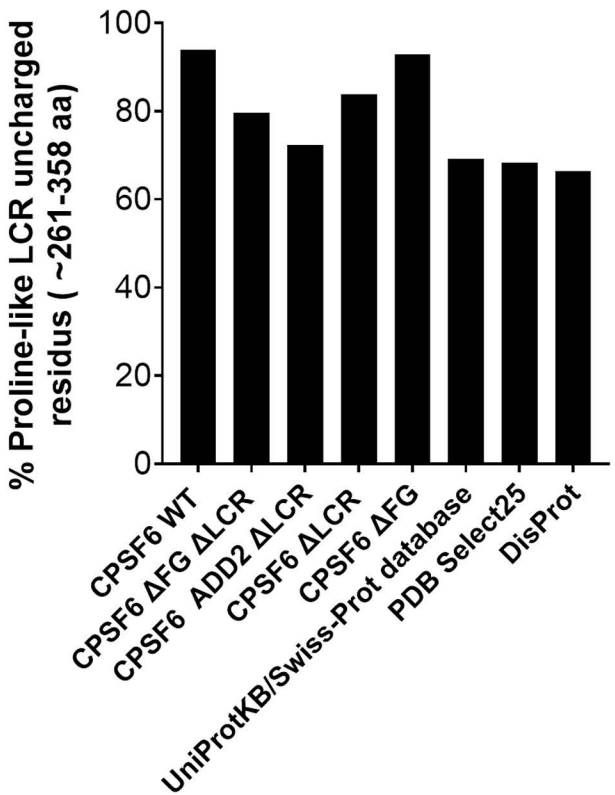
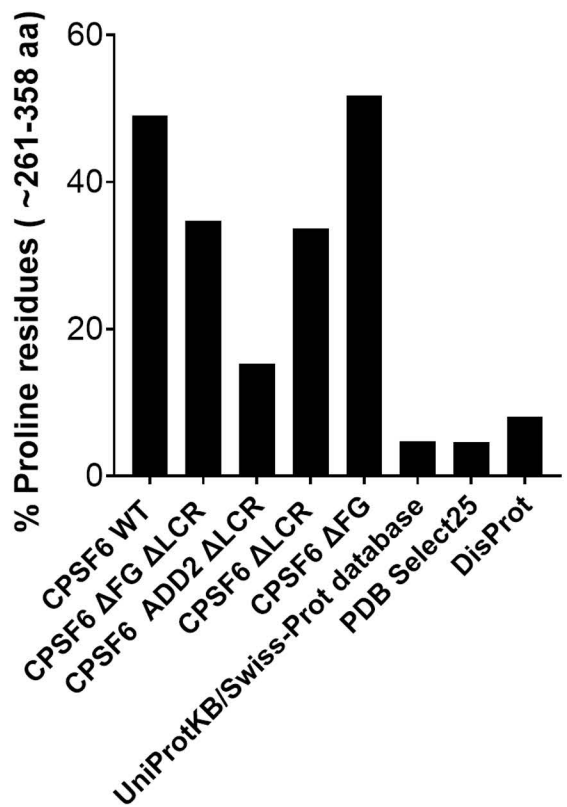

C
PONDR VLXT score
Residue number
FG
ADD2
LCR
D
NCPR distribution (blob 5)
NCPR
Blob index
E
NCPR distribution (blob 5)
NCPR
Blob index
F
Conf
Cart
Pred
AA
G
Conf
Cart
Pred
AA
Legend:
Strand
Helix
Coil
Conf: Confidence of prediction
Cart: 3-state assignment cartoon
Pred: 3-state prediction
AA: Target Sequence

Figure 3

H


% Proline-like LCR uncharged residus (~261-358 aa)
100
80
60
40
20
0
CPSF6 WT
CPSF6 ΔFG ΔLCR
CPSF6 ADD2 ΔLCR
CPSF6 ΔLCR
CPSF6 ΔFG
UniProtKB/Swiss-Prot database
PDB Select25
DisProt

I


% Proline residues (~261-358 aa)
60
40
20
0
CPSF6 WT
CPSF6 ΔFG ΔLCR
CPSF6 ADD2 ΔLCR
CPSF6 ΔLCR
CPSF6 ΔFG
UniProtKB/Swiss-Prot database
PDB Select25
DisProt

L

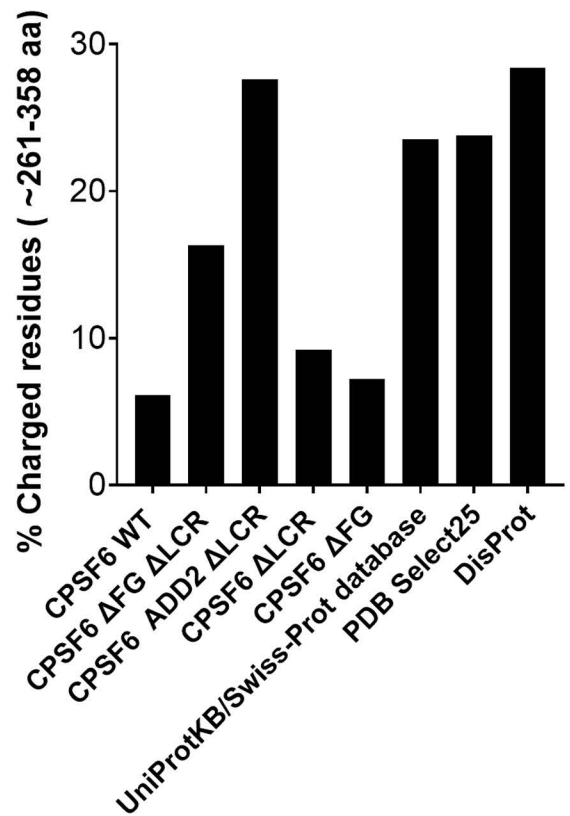
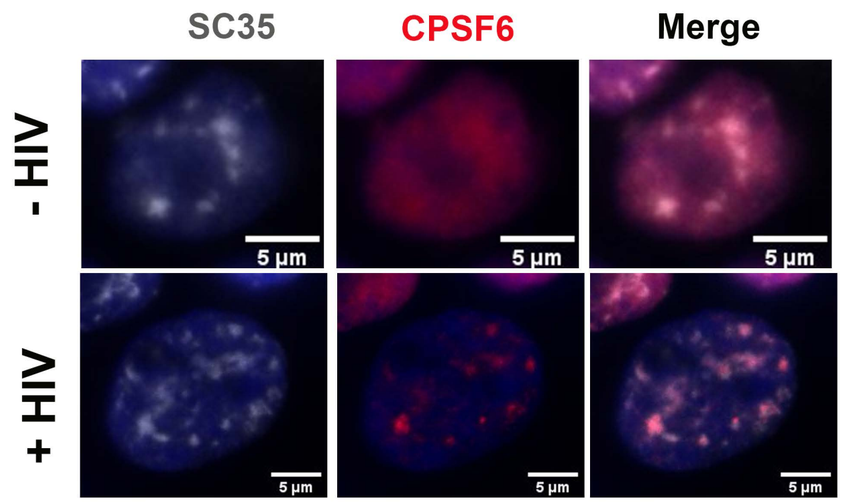
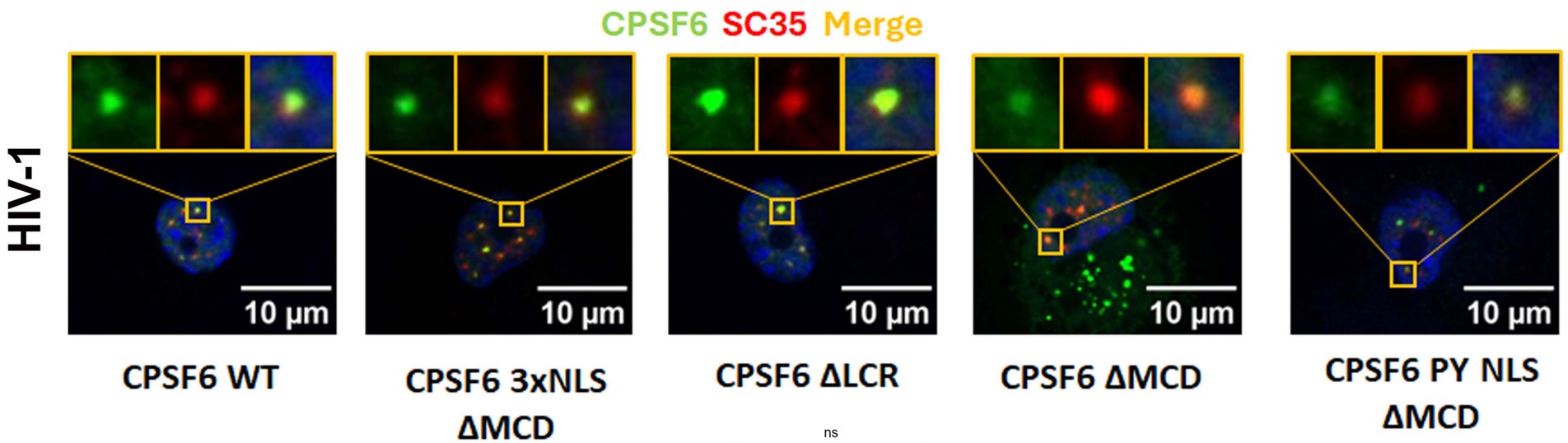

% Charged residues (~261-358 aa)
30
20
10
0
CPSF6 WT
CPSF6 ΔFG ΔLCR
CPSF6 ADD2 ΔLCR
CPSF6 ΔLCR
CPSF6 ΔFG
UniProtKB/Swiss-Prot database
PDB Select25
DisProt

Figure 3

A


SC35
CPSF6
Merge
- HIV
+ HIV
5 µm

B


CPSF6 SC35 Merge
HIV-1
10 µm
CPSF6 WT
CPSF6 3xNLS ΔMCD
CPSF6 ΔLCR
CPSF6 ΔMCD
CPSF6 PY NLS ΔMCD

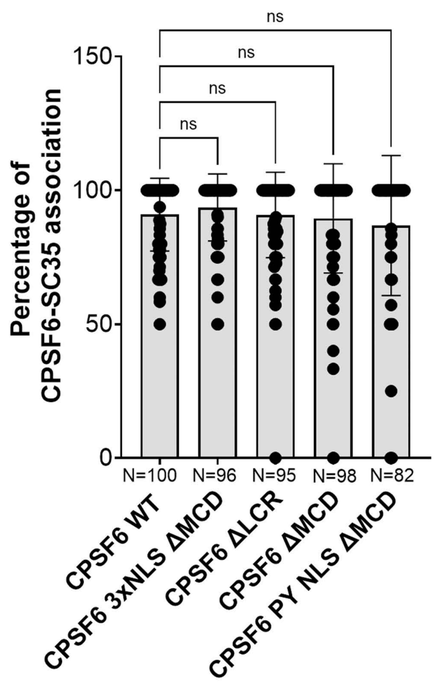

Percentage of CPSF6-SC35 association
150
100
50
0
ns
N=100
N=96
N=95
N=98
N=82
CPSF6 WT
CPSF6 3xNLS ΔMCD
CPSF6 ΔLCR
CPSF6 ΔMCD
CPSF6 PY NLS ΔMCD

Figure 4

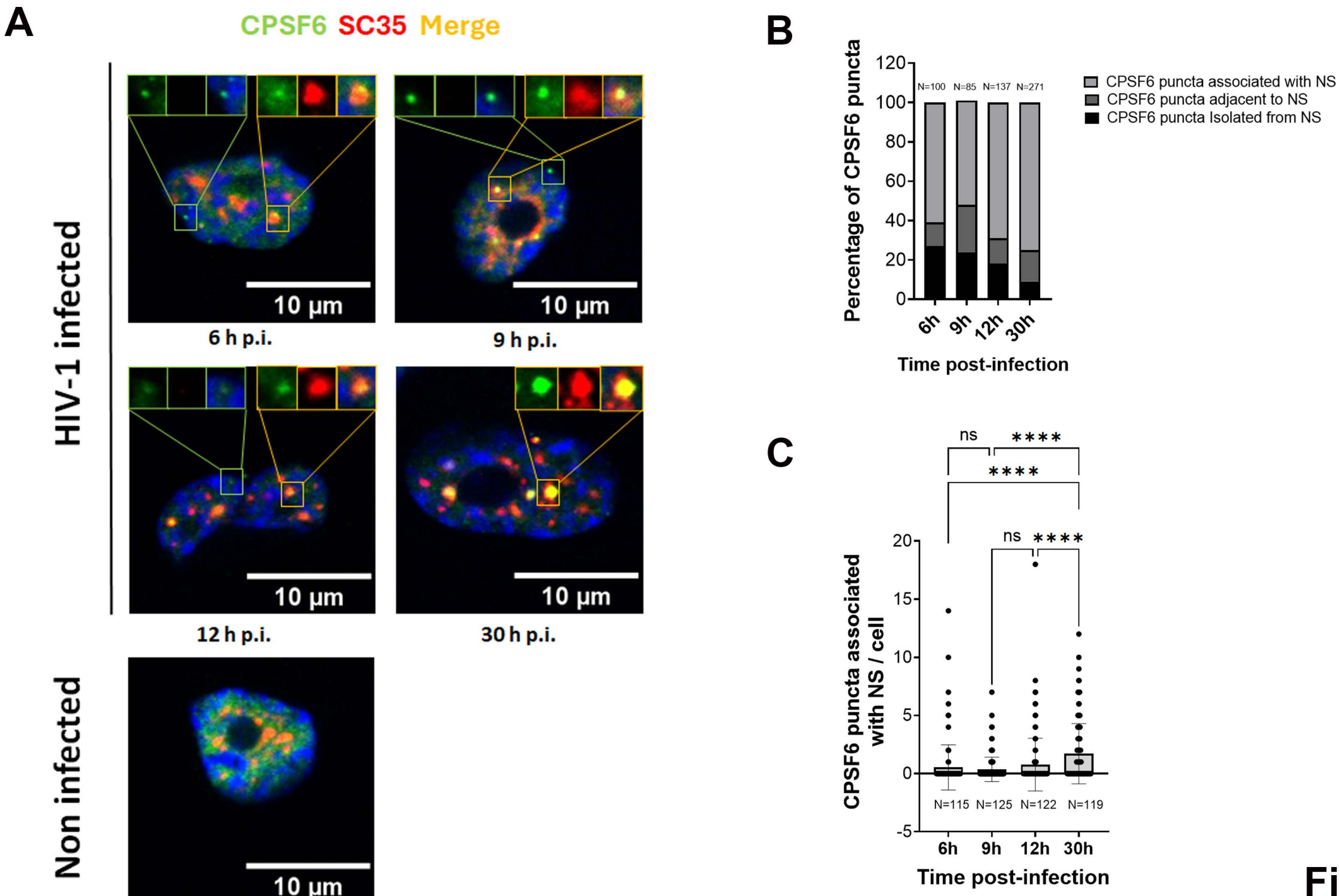

A
CPSF6 SC35 Merge
HIV-1 infected
Non infected
10 µm
6 h p.i.
9 h p.i.
12 h p.i.
30 h p.i.
B
Percentage of CPSF6 puncta
N=100 N=85 N=137 N=271
CPSF6 puncta associated with NS
CPSF6 puncta adjacent to NS
CPSF6 puncta Isolated from NS
6h 9h 12h 30h
Time post-infection
C
ns
****
CPSF6 puncta associated with NS / cell
N=115 N=125 N=122 N=119
6h 9h 12h 30h
Time post-infection

Figure 5

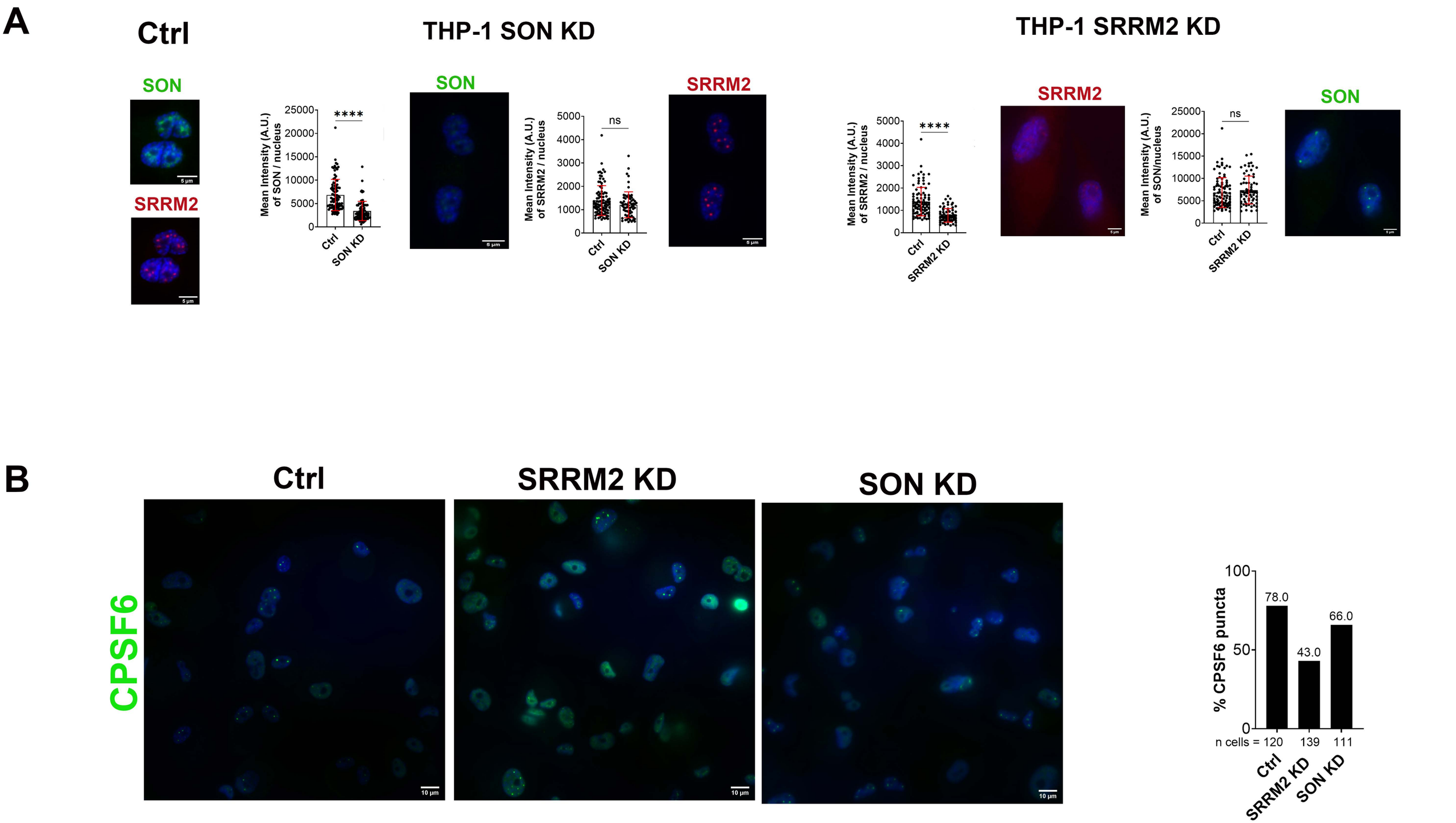
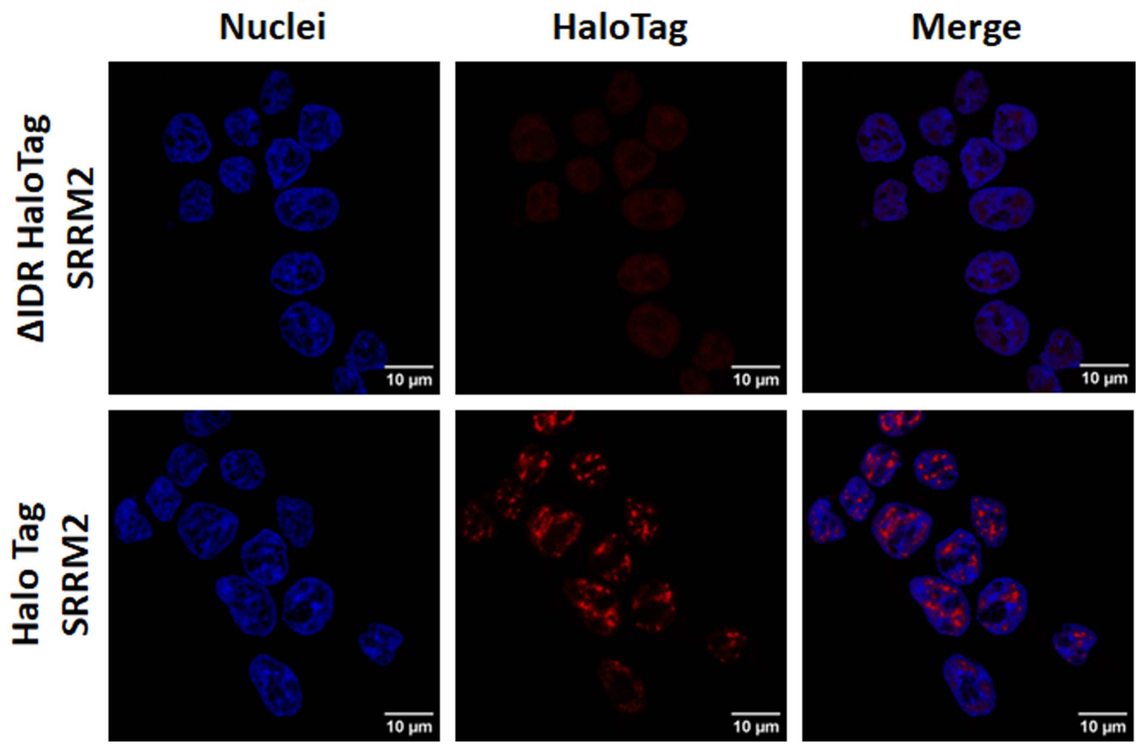
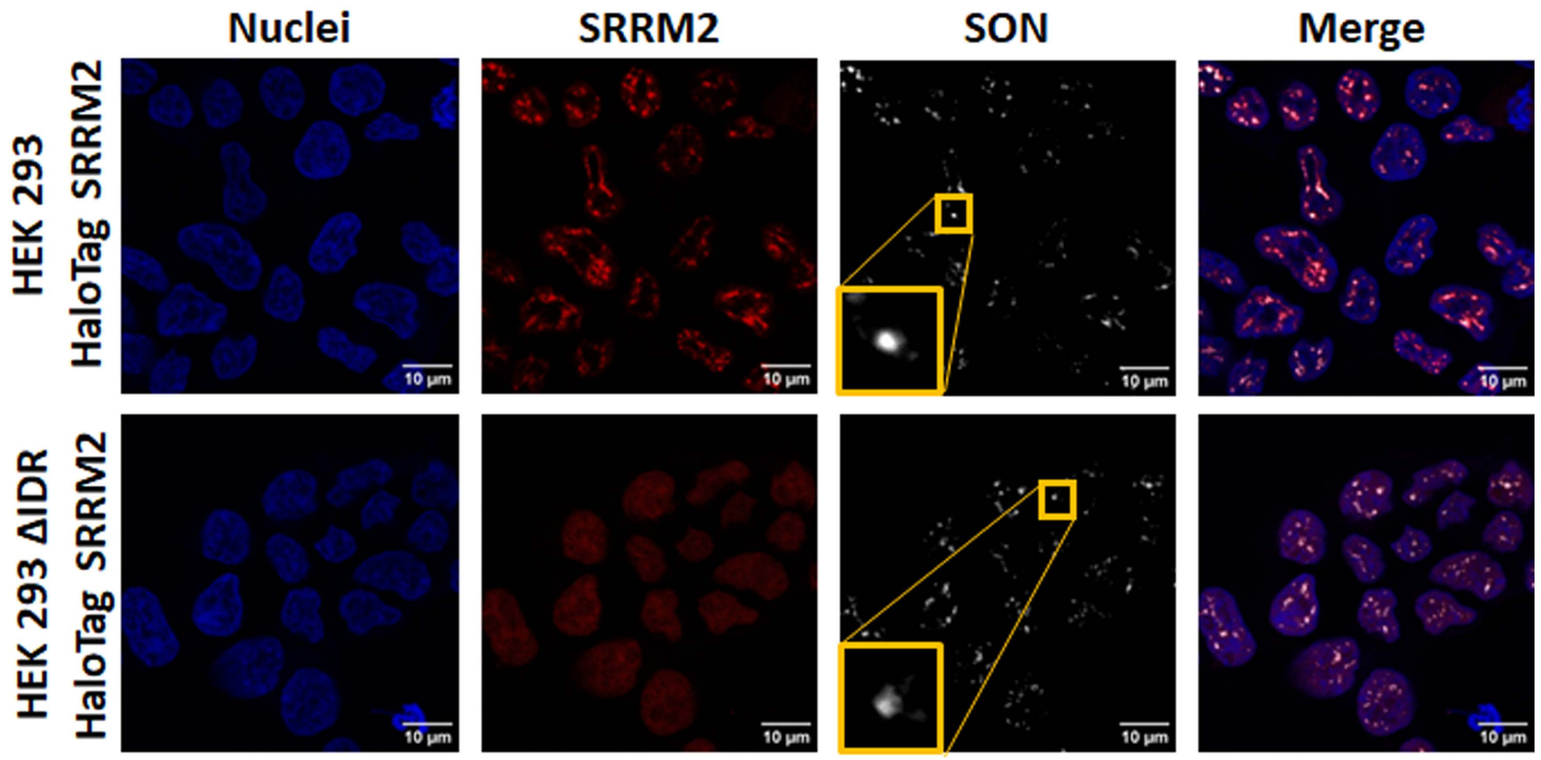

A
Ctrl
THP-1 SON KD
THP-1 SRRM2 KD
SON
SRRM2
Mean Intensity (A.U.) of SON / nucleus
****
Ctrl
SON KD
Mean Intensity (A.U.) of SRRM2 / nucleus
ns
Mean Intensity (A.U.) of SRRM2 / nucleus
****
Ctrl
SRRM2 KD
Mean Intensity (A.U.) of SON/nucleus
ns
B
Ctrl
SRRM2 KD
SON KD
CPSF6
% CPSF6 puncta
78.0
43.0
66.0
n cells = 120
139
111
Ctrl
SRRM2 KD
SON KD

Figure 6

C


Nuclei
HaloTag
Merge
ΔIDR HaloTag SRRM2
Halo Tag SRRM2
10 µm

D


Nuclei
SRRM2
SON
Merge
HEK 293 HaloTag SRRM2
HEK 293 ΔIDR HaloTag SRRM2
10 µm

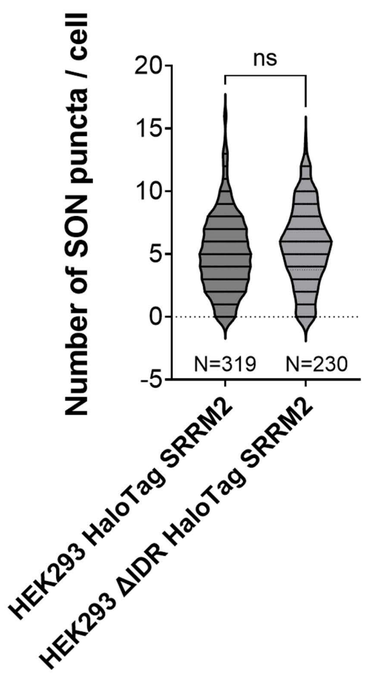
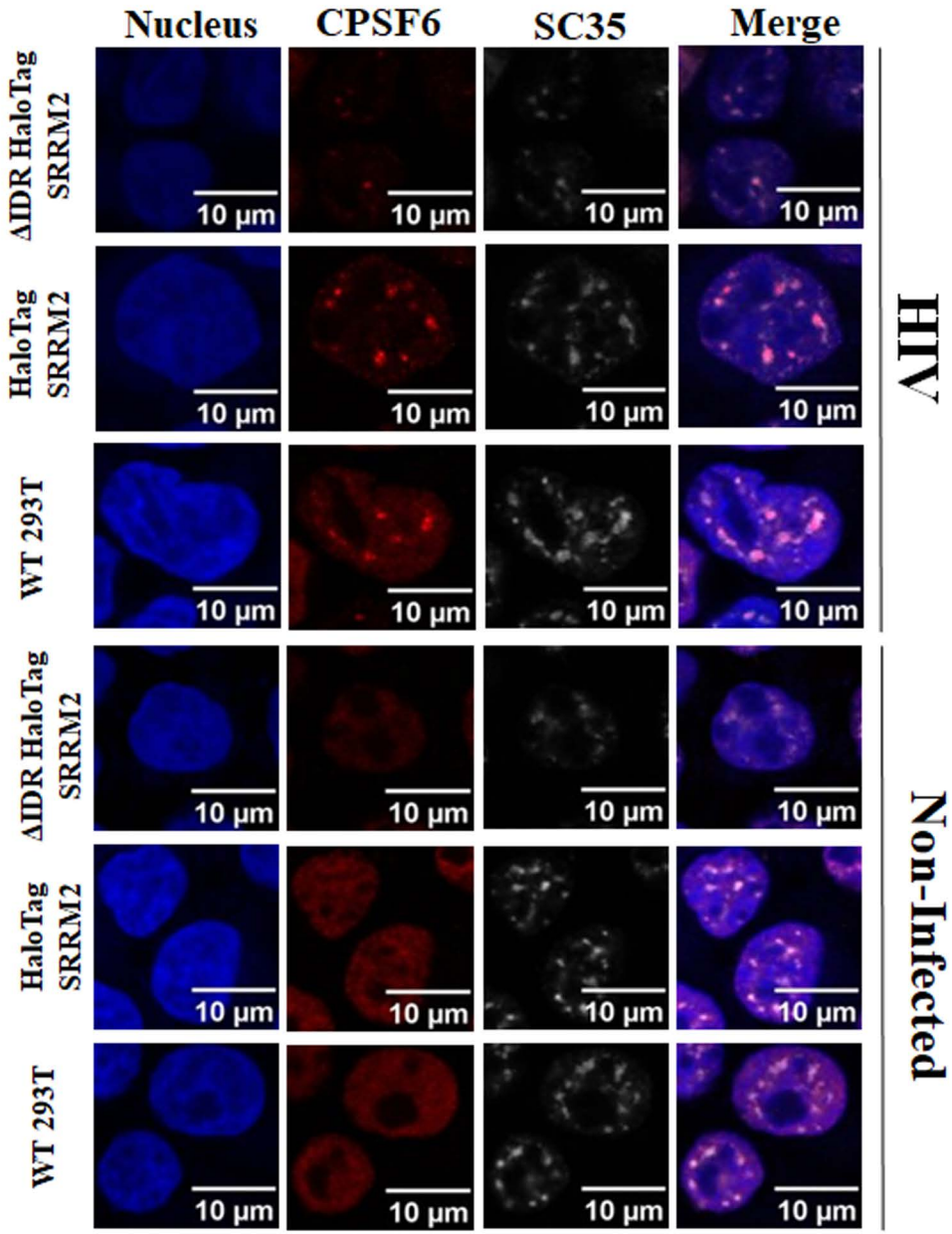

Number of SON puncta / cell
20
15
10
5
0
-5
ns
N=319
N=230
HEK293 HaloTag SRRM2
HEK293 ΔIDR HaloTag SRRM2

Figure 6

E


Nucleus
CPSF6
SC35
Merge
ΔIDR HaloTag SRRM2
HaloTag SRRM2
WT 293T
ΔIDR HaloTag SRRM2
HaloTag SRRM2
WT 293T
10 µm
HIV
Non-Infected

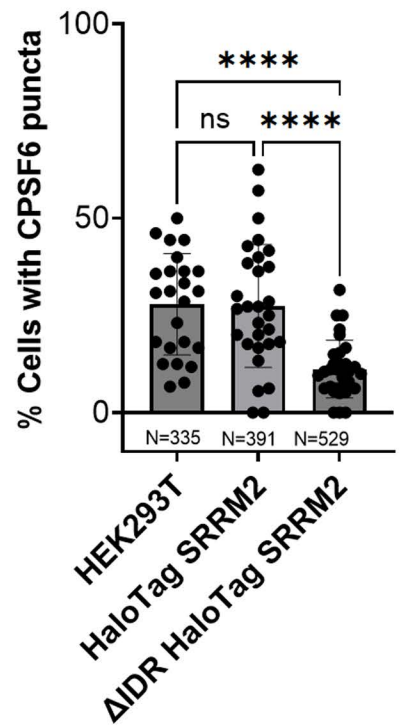

% Cells with CPSF6 puncta
100
50
0
****
ns
****
N=335
N=391
N=529
HEK293T
HaloTag SRRM2
ΔIDR HaloTag SRRM2

Figure 6