Efficient packaging of HIV-1 genomes via recognition of its adenosine-rich content by a heterologous RNA-binding domain

Hung R. Vuong[1], Qianzi Zhou[1], Sydney Lesko[2], Kasyap Tenneti[1], Keanu Davis[1], Shanyqua Scott[2], Daphne Boodwa-Ko[1], Jenna E. Eschbach[1], Kamya Gopal[3], Jessica M. Porter[1], Yiqing Wang[1], Shawn Mohammed[1], Nakyung Lee[1], Alice Telesnitsky[3, 4], Nathan Sherer[2], Sebla B. Kutluay[1*]

[1]Department of Molecular Microbiology, Washington University School of Medicine, Saint Louis, Missouri, USA

[2]McArdle Laboratory for Cancer Research and Institute for Molecular Virology, University of Wisconsin, Madison, USA.

[3]Cell and Molecular Biology Program, University of Michigan Medical School, Ann Arbor, MI USA.

[4]Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI USA.

*Corresponding author: kutluay@wustl.edu

ABSTRACT

Selective packaging of a dimeric HIV-1 genome (gRNA) is thought to be driven by specific binding of the nucleocapsid (NC) domain of viral Gag protein to the packaging signal (Ψ) in the host cell cytosol. Through replacement of NC with heterologous RNA-binding domains (RBDs) with distinct RNA-binding properties, we show that the biased adenosine-rich nucleotide content of the gRNA facilitates its selective packaging. Despite disparate RNA binding specificities, all Gag-RBD chimeras successfully recruited the gRNA to the plasma membrane, but many were arrested at the assembly stage. Only the Gag-SRSF5 chimera, which multimerized efficiently on adenosine-rich sequences on the gRNA, packaged gRNA at near wild-type levels. Many Gag chimeras displayed potent dominant negative activities, highlighting NC functions as a targetable step in virus replication. Together, our findings reveal an unexpected aspect of the HIV-1 gRNA, its biased nucleotide content, as a key driver of selective genome packaging.

INTRODUCTION

Selective packaging of viral genomes is integral for viral replication and contributes to rapid viral evolution. For HIV-1, a non-covalent dimer of the unspliced viral mRNA serves as the packageable viral genome (gRNA)[1-3]. During reverse transcription in a target cell, the gRNA dimer allows the virus to undergo recombination at a high rate, which maintains viral genome integrity and helps HIV-1 overcome the selective pressures imposed by host immune responses and antiviral drugs[4-7]. Similar to HIV-1, RNA viruses with segmented genomes (e.g. members of *Orthomyxoviridae*, *Sedoreoviridae*) can acquire novel antigenic properties through genetic reassortment, where genome segments from co-infecting viruses are exchanged and packaged by viral progeny[8-12].

HIV-1 gRNA packaging is a highly efficient process with nearly all HIV-1 particles (>90%) containing a dimer of gRNA[1]. How such a high level of enrichment is achieved despite gRNA competing with cellular mRNAs in infected cells is a long-standing question. It is well established that the interaction of the nucleocapsid (NC) domain of the major HIV-1 structural protein, Gag, with gRNA is required for genome encapsidation[13,14]. Efficient gRNA packaging is widely thought to be initiated in the cytosol by high affinity binding of NC to distinct sites within the cis-acting packaging signal, Ψ, located near the 5' end of gRNA[15,16]. Ψ folds into three conserved stem loops known as SL1, SL2, and SL3[3]. Multiple *in vitro* studies have shown that unpaired guanosine residues within these stem loops are preferentially bound by Gag/NC and act synergistically to mediate gRNA packaging[17-19]. In agreement, crosslinking immunoprecipitation (CLIP) studies in cells have also shown a clear propensity of cytosolic Gag to bind to multiple regions within Ψ[20]. However, deletions of the Ψ stem loops do not result in a complete loss of packaging[21-23]. Furthermore, Gag binds with high affinity to both Ψ-containing and Ψ-null RNAs at physiological ionic strength *in vitro*[24] and packages cellular mRNAs into

virions, particularly in the absence of HIV-1 gRNA[25,26]. Hence, selective binding of Gag to Ψ appears to be insufficient to support selective gRNA packaging.

While Gag-Ψ interactions likely initiate in the cytosol[20,27], whether this early genome recognition event is the most critical step for selective gRNA packaging is unclear. Based on live-cell imaging studies, a small number of Gag molecules appear to be required for trafficking and stable anchoring of the gRNA at the plasma membrane (PM)[28-30]. This nucleation complex subsequently recruits additional Gag molecules from the cytosol to drive the formation of high-order Gag multimers, which ultimately results in encapsidation of the gRNA in budding virions[3,31-33]. In addition, RNA plays a structural component in virions[34] and capsid (CA) domain mutations that block Gag lattice assembly can impair specific binding to Ψ and genome packaging[35,36]. Based on these findings, gRNA was proposed to be selectively packaged due to having multiple Gag binding sites that help increase local Gag concentration and consequently making HIV-1 RNAs more efficient at promoting Gag-Gag interactions that drive virus assembly[3,16]. However, the specific features of HIV-1 gRNA that promote efficient virus assembly and the mechanism by which Gag multimerizes more readily on viral RNAs are not understood.

The HIV-1 gRNA has a highly biased nucleotide content: rich in adenosines (~36%) and poor in cytosines (~18%) with purines overrepresented in all codon positions of all HIV-1 genes[37,38]. It is plausible that the unique nucleotide content of the HIV-1 gRNA can facilitate high avidity Gag-gRNA interactions at the PM, which in turn enhances Gag multimerization on the HIV-1 gRNA during virion assembly. While prior CLIP studies have revealed the propensity of Gag to bind to A/G-rich motifs in cellular mRNAs recruited to the PM[20], direct evidence that supports the role of purine-rich elements in gRNA packaging is lacking.

In this study, we generated Gag-RBD chimeras with altered RNA binding specificities by

replacing its NC domain with RNA binding domains (RBDs) derived from the heterogeneous nuclear ribonucleoprotein (hnRNP) and serine-arginine-rich (SRSF) protein families with the hypothesis that like NC, Gag-RBD chimeras with purine-rich binding preference could package gRNA efficiently. Remarkably while all chimeras were capable of binding to Ψ in the cytosol and recruiting gRNA to the PM, we found that Gag-SRSF5, which preferentially binds to A/G-rich sequences, was the sole chimera that packaged gRNA at near WT level. Gag-SRSF5 chimera efficiently multimerized on the HIV-1 gRNA at the PM and displayed a remarkably similar purine-rich RNA binding specificity with WT Gag in immature virions. Notably, mutations that reduced the specificity of Gag-SRSF5 for A/G-rich motifs also reduced its gRNA packaging efficiency. While each of the Gag-RBD chimeras we generated yielded non-infectious virions and several displayed particle release defects, they were nonetheless able to co-assemble with WT Gag and potently inhibited virion infectivity. Taken together, our findings indicate that Gag binding to gRNA in the cytosol and trafficking it to PM assembly sites are insufficient to ensure selective genome packaging. Rather, our findings provide strong support for the notion that efficient Gag multimerization on the unusually A-rich HIV-1 gRNA at the PM favors virus assembly on—and thus selective encapsidation of—gRNA over competing cellular mRNAs.

RESULTS

Gag-SRSF5 is unique in its ability amongst Gag-RBD chimeras to efficiently package the HIV-1 gRNA

To alter the RNA binding specificity of Gag, we replaced its NC domain with tandem RBDs (e.g. RNA-recognition motifs (RRMs) and K homology domains (KH)) from the hnRNP and SRSF protein families within the context of a fully infectious HIV-1_{NL4-3} proviral plasmid bearing a 3xHA tag between capsid (CA) and matrix (MA) domains of Gag (**Fig. 1A**). We reasoned that RBDs that preferentially bind to and multimerize on the A/G-rich sequences of HIV-1 gRNA would be

able to replace the genome packaging function of NC. Internal controls included replacement of NC with leucine zipper domain from GCN4 (bZIP) that facilitates Gag multimerization in the absence of NC [39] and complementation by NC itself. The set of hnRNP and SRSF proteins (**Fig. 1A**) were chosen because their RNA binding properties are well defined and prior studies have shown that they directly bind to HIV-1 RNAs to regulate their alternative splicing [40-45]. For example, hnRNP A1 recognizes “AGG” motifs, hnRNP F binds to stretches of oligonucleotide G motifs, hnRNP K binds to short C-rich tracks, and SRSF1, 4 and 5 all recognize “AAG” motifs. Analogous to NC, which binds to cognate RNA via two zinc finger motifs, hnRNP and SR proteins possess tandem RBDs that create a larger binding interface enabling recognition of longer sequence motifs with enhanced affinity compared to individual RBDs.

As expected, replacing the NC domain, which has multiple functions in HIV-1 replication beyond packaging[46-49], rendered all chimeric viruses non-infectious **(Fig. 1B)**. Notwithstanding, all chimeras expressed similar levels of Gag and Gag-Pol in cells, irrespective of the presence or absence of a functional PR **(Fig. 1C, S1A).** The expression of Env and Nef proteins, generated through translation of partially and completely spliced HIV-1 transcripts, respectively, were similar across WT Gag and Gag-RBD chimeric viruses, suggesting that the isolated RBDs of hnRNPs and SR proteins did not have unintended consequences on viral mRNA splicing (**Fig. 1C, S1A**). In the case of PR^+ viruses, Gag processing intermediates and fully processed CA (p24) were present (**Fig. 1C**). Notably, particle release was largely blocked for Gag-hnRNP A1, -hnRNP F and hnRNP K chimeras **(Fig. 1C, S1B)**. Gag-SRSF1 and Gag-SRSF4 displayed a more modest release defect whereas Gag-SRSF5 release was largely unaffected **(Fig. 1C, S1B)**. Gag-SRSF5 particles contained equivalent levels of Gag-Pol and integrase, demonstrating that virion incorporation and processing of Pol was also unaffected **(Fig. 1C, S1B)**.

As intended, all Gag-RBD chimeras could bind RNAs in cells, with a much higher propensity to form Gag-RNA adducts than WT as revealed by autoradiography analysis of immunoprecipitated Gag-RNA complexes (**Fig. 1D, 1E**). However, only the Gag-SRSF5 packaged gRNA nearly as efficiently as WT Gag based on RT-qPCR analysis of purified mature and immature virions **(Fig. 1F, S1C, S1D)**. On the other hand, Gag-SRSF5 appeared to be less selective for gRNA compared to WT Gag and packaged a higher proportion of multiply spliced HIV-1 RNAs **(Fig. 1F, S1C, S1D)**. In addition to the viral gRNA, certain host RNAs including 7SL RNA are markedly enriched in HIV-1 virions[20,50-52]. Surprisingly, Gag-SRSF5 also encapsidated host 7SL RNA at levels comparable to that of WT Gag (**Fig. 1F, S1C, S1D**). Northern blotting in Gag-SRSF5 virions confirmed these results and additionally demonstrated that both gRNA and 7SL RNA in Gag-SRSF5 particles were intact (**Fig. S1E**).

To gain further insight into the extent with which Gag-SRSF5 selectively packages gRNA over cellular RNAs, we next conducted total RNA-seq experiments in purified virions. Expectedly, for WT Gag and ΔNC-NC virions HIV-1 gRNA mapping reads constituted ~80% of the total reads (**Fig. 1G**). Roughly half of the remaining reads were derived from rRNAs (**Fig. S1F**) and other host-derived RNAs primarily composed of mRNAs and 7SL RNA (**Fig.1H, S1G**). In comparison, only about 25% of RNA-seq-derived reads were of HIV-1 origin in Gag-SRSF5 virions, whereas the majority of the remaining reads mapped to host mRNAs as well as 7SL RNA (**Fig. 1G, 1H, S1F, S1G**), supporting the above finding that Gag-SRSF5 is less selective towards gRNAs. With the exception of hnRNP A1, viral RNAs constituted only a small proportion of RNA-seq derived reads for the closely related Gag-SRSF1 and Gag-SRSF4 as well as other Gag-RBD chimeras, and the remaining reads mapped to mRNAs and host rRNAs (**Fig. 1G, 1H, S1F)**. Note that the observed high abundance of rRNA reads in Gag-RBD chimeras with low virus yields as well as Gag-bZIP and Gag-ΔNC chimeras is likely due to PCR overamplification (per RNA-seq protocol) of otherwise low abundance RNAs in virus preps. Nevertheless, comparison of mRNA-mapping

reads revealed that the host mRNAs packaged by ΔNC-NC virions were most similar to those packaged by Gag-SRSF1, 4 and 5 as indicated by the highest Jaccard index values with > 0.2 generally considered significant correlation (**Fig. 1I**).

Single virion analysis of Gag-SRSF5 reveals its ability to package the HIV-1 gRNA at near WT Gag efficiency

We next sought to gain more granular insight into gRNA packaging and virion assembly properties of Gag-SRSF5 at a single virion level. To do so, we used a modified, visible virus encoding YFP tagged Gag (Gag-YFP) and gRNA bearing 24 copies of an RNA stem loop derived from the MS2 bacteriophage (MSL) that can be tracked in proxy by a virally-encoded MS2 coat protein fused to mCherry (**Fig. 2A**)[28,53-55]. No differences were observed in virion-incorporated Gag-YFP levels for WT compared to Gag-SRSF5 particles (**Fig. S2A**), suggesting that Gag incorporation per virion and particle size were not affected. In contrast, ΔNC virions incorporated considerably lower levels of Gag-YFP per virion (**Fig. 2B, S2A**). In agreement with previous reports[1], ~90% of WT and ΔNC-NC particles contained gRNA, whereas the majority of ΔNC virions lacked detectable genomes, also as expected (**Fig. 2B**). While ~70% of Gag-SRSF5 chimeras encapsidated gRNA **(Fig. 2B),** we noted a trend towards lower gRNA signal intensity per virion (**Fig. 2C, S2A**) suggesting that not all Gag-SRSF5 particles may contain a gRNA dimer.

To interrogate the virion morphology of Gag-SRSF5 particles further, we conducted thin section electron microscopy (TEM) analyses of purified virions. Similar to WT and ΔNC-NC particles, Gag-SRSF5 particles were ~100nm in diameter but contained eccentrically localized condensates and a higher fraction of particles with immature morphology **(Fig. 2D).** Biochemical fractionation of viral cores isolated from Gag-SRSF5 virions indicated the presence of unstable or improperly assembled cores similar to those observed for ΔNC and bZIP virions **(Fig. 2E, 2F)**

and possibly due to the presence of aberrant Gag processing intermediates (**Fig. S2B**). These aberrations combined with the inability to execute various NC functions likely explain the early reverse transcription defects of Gag-SRSF5 particles in target cells (**Fig. S2C**) and the accompanying decrease in virion infectivity (**Fig. 1B**).

Gag-RBD chimeras efficiently recruit HIV-1 gRNA to the PM but many display virion assembly defects

HIV-1 virion assembly involves stepwise localization of a gRNA dimer at the PM, Gag multimerization on the PM-associated gRNA forming an immature Gag lattice, and recruitment of host ESCRT-III machinery for budding[56]. The particle release defects of the several Gag-RBD chimeras prompted us to evaluate at what step of virion assembly they were blocked at, which we reasoned would help explain what makes WT Gag and Gag-SRSF5 successful at gRNA packaging.

First, we focused on Gag and gRNA localization at the PM, given that gRNA recruitment and its stable tethering at the PM are mediated by NC[28,29]. Like WT Gag, all Gag-RBD chimeras localized in the cytosol and at the PM at steady state, except for Gag myristoylation mutant (G2A) which is defective in PM anchoring and was enriched in the cytosol **(Fig. 3A, S3A, S3B)**. Surprisingly, even though only Gag-SRSF5 packaged gRNA, all Gag-RBD chimeras recruited gRNA to the PM as efficiently as WT Gag **(Fig. 3B)**. Of note, while Gag-hnRNP A1 chimera preferentially retained gRNA over multiply spliced viral RNAs at the PM, other Gag chimeras were comparably less selective **(Fig. 3B)**. In contrast to viral RNAs, the abundance of the control host GAPDH mRNA at the PM was not dependent on the presence of Gag with a functional NC domain or RBDs (**Fig. 3B**). To corroborate these findings and interrogate the influence of heterologous RRMs on the trafficking properties of Gag chimeras with gRNA at a single-cell level, we next conducted live-cell imaging studies using the dual labeling system

described in Fig. 2A. In agreement with biochemical experiments, all Gag chimeras were able to recruit the gRNA to the PM **(Fig. 3C, S3C)**. However, only Gag-SRSF5 formed WT-like foci with gRNA at the PM whereas the other Gag chimeras appeared to form sheet-like assemblies **(Fig. 3C, S3D-E)**.

These data suggested that the particle release defect of select Gag-RBD chimeras is not due to impaired trafficking with gRNA to the PM. To interrogate the next steps in virion assembly, namely whether Gag-RBD chimeras can multimerize at the PM and form membrane curvature necessary for recruitment of (ESCRT-III) machinery[57], we conducted TEM experiments in virus producer cells. In agreement with live-cell imaging experiments, we found that Gag-RBD chimeras with release defects formed extended electron dense sheet-like structures at the PM **(Fig. 3D)**, suggesting that they can all multimerize and form Gag lattices. However, unlike WT and ΔNC-NC which formed well-defined curved buds, the electron dense sheet-like structures from hnRNP A1, F, K and Gag-SRSF1 chimeras had little to no membrane curvature. Gag-SRSF4, and SRSF5, on the other hand, formed buds that were closer in shapes and sizes to WT and ΔNC-NC control **(Fig. 3D)**. Altogether, these data suggest that all Gag-RBD chimeras possessing diverse RNA-binding modules can find and retain gRNA at the PM. However, events post-Gag:gRNA localization, possibly the inability to multimerize on the gRNA at the PM, rendered most of the chimeras unable to bud and package gRNA efficiently.

Gag-RBD chimeras maintain the known RNA-binding specificities of the isolated RBDs but bind similarly to HIV-1 gRNA

To determine which feature of Gag-SRSF5 makes it uniquely capable of gRNA packaging, we interrogated the RNA-binding specificities of Gag-RBD chimeras using CLIP in cells, at the PM and in immature virions. In line with a prior report[20], a relatively small percentage of CLIP-derived reads mapped to viral RNAs for WT Gag and Gag-ΔNC-NC **(Fig. 4A)**. Similarly, despite

co-immunoprecipitating with higher levels of total RNA (**Fig. 1D, 1E**), a relatively small percentage of CLIP reads for Gag-RBD chimeras were derived from viral RNAs (**Fig. 4A**). Analysis of reads that map to cellular RNAs revealed that similar to WT Gag and Gag-ΔNC-NC, all Gag chimeras, with the exception of Gag-hnRNP F, bound most frequently to cellular tRNAs, mediated by the matrix domain[20], and mRNAs (**Fig. 4B**). mRNAs constituted comparably a greater fraction of cell-derived RNAs for Gag-hnRNP A1, -SRSF4 and -SRSF5 chimeras **(Fig. 4B)**. Of note, for Gag-hnRNP F, tRNAs made up a considerably smaller proportion of reads for compared to intron-mapping reads (**Fig. 4B**). This can be ascribed to the nuclear trafficking of this chimera **(Fig. 3C)**, due to the presence of glycine-tyrosine-arginine-rich (GYR) domain between RRM2 and RRM3 combined with its known preference to bind to intronic regions in cognate RNAs[43,45,58]. Motif analysis within the bound mRNAs revealed that sequence preferences of Gag-RBD chimeras are in alignment with the known RNA binding specificity of the hnRNP/SR proteins that the RBDs were derived from **(Fig. 4C)**[43,45,59-62]: WT Gag as well as Gag-SRSF1, 4 and 5 chimeras bound preferentially to sequences containing A/G-rich sequences (**Fig. 4C**), Gag-hnRNP K bound to C-rich oligonucleotides, Gag-hnRNP A1 and Gag-hnRNP F preferentially bound to sequences containing AGGGG motif and short G repeats, respectively (**Fig. 4C**).

On the HIV-1 genome, WT Gag and Gag-ΔNC-NC bound to discrete sites within the 5' leader, including SL1 and SL3 (**Fig. 4D, S4A**), in agreement with prior reports[20]. Surprisingly, all Gag-RBD chimeras also frequently bound to these sites despite their diverse RNA-binding specificities, though the extent with which they were bound to SL1 and SL3 varied **(Fig. 4D, S4A)**. Since Gag-SRSF5 also bound to the 5' leader, we tested whether progressive deletions of structured RNA elements within this region would affect gRNA packaging **(Fig. S4B)**. None of the 5'UTR deletion mutations impaired gRNA packaging for Gag-SRSF5 (**Fig. S4C, S4D**),

suggesting that sequences or features outside of the 5' leader contributes to its ability to mediate selective genome packaging.

Gag-SRSF5 multimerizes on the A-rich HIV-1 gRNA at the plasma membrane and in immature virions

Gag was proposed to undergo a multimerization-driven change in its RNA-binding specificity towards more frequent binding to A-rich motifs at the PM, which is further enhanced in immature virions[20]. We next examined whether a similar switch accounts for the unique ability of Gag-SRSF5 to package gRNA through CLIP experiments in PM fractions and immature virions. Comparison of membrane-associated Gag-RBD:RNA adducts revealed that unlike cytosolic Gag (**Fig. 1D, E**), WT Gag and Gag-RBD chimeras bound to comparable levels of RNA **(Fig. 5A, S5A).** The increased level of RNA binding for WT Gag demonstrates its ability to more extensively engage with cognate RNAs upon localizing to the PM through multimerization. While this pattern continued in virions for WT Gag, which had the highest level of immunoprecipitated Gag:RNA adducts, only the Gag-SRSF5 yielded comparable level of protein:RNA complexes whereas Gag-SRSF1 and Gag-SRSF4 were largely impaired in RNA-binding **(Fig. 5B, S5B)**. Expectedly, the percentages of viral RNA-derived CLIP reads were enriched in immature virions compared to cells for WT Gag and Gag-ΔNC-NC **(Fig. 5C)**. However, the level of enrichment was lower for Gag-SRSF5 **(Fig. 5C)**, in agreement with it packaging gRNA less selectively **(Fig. 1G)**.

Strikingly, in immature virions the binding profile of Gag-SRSF5 on gRNA mirrored that of WT Gag and Gag-ΔNC-NC (**Fig. 5D**) and exhibited a statistically significant correlation nearing that observed between WT Gag vs. Gag-ΔNC-NC (**Fig. 5E**). In fact, Gag-SRSF5, along with Gag-SRSF1 and Gag-SRSF4, which bind to A-rich motifs in cells **(Fig. 4C)**, exhibited a notable switch in their RNA binding profile upon association with the PM. Similar to WT Gag, each of

these chimeras bound more frequently to a larger number of sites throughout the HIV-1 gRNA at the PM and less prominently to the 5' UTR **(Fig. S5C)**. In contrast, Gag-CAΔCTD mutant that is significantly impaired in forming high-order multimers bound to more discrete sites on the gRNA (**Fig. S5C**). Gag-hnRNP A1 and Gag-hnRNP F, which bind to G-rich sequences in cells, showed little to no increase in binding sites, with their cytosol and PM RNA binding profiles resembling each other (**Fig. S5C vs. Fig. S4A**).

In addition to the marked similarity between WT Gag and Gag-SRSF5 binding to the HIV-1 gRNA in immature virions, Gag-SRSF5-bound host mRNAs also resembled those bound by WT Gag and Gag-ΔNC-NC, with mRNAs and 7SL RNA representing the most frequently bound RNAs in virions **(Fig. 5F)** Lastly, motif analyses on bound cellular mRNAs showed that virion-associated Gag-SRSF5, WT Gag, and Gag-ΔNC-NC were all preferentially bound to A/G-rich sequences **(Fig. 5G)**. Taken together, these results provide strong support for the notion that the ability of Gag to efficiently multimerize on the A-rich gRNA may underlie its capacity to efficiently package the HIV-1 genome.

Changing the RNA-binding specificity of Gag-SRSF5 specifically impairs gRNA packaging

SRSF1, SRSF4, and SRSF5 bear tandem RRMs bridged by a flexible linker. A previous structural study showed that the highly conserved S_{133}WQDLKD_{139} sequence residing within RRM2 of SRSF1 was responsible for its binding specificity for a minimal GGA motif [63]. Specifically, residues D136, K138 and D139 were engaged in an extensive hydrogen binding network within the purine-rich recognition site. Based on this structural prediction, we designed two Gag-SRSF5 mutants, mGGA-1 and mGGA-2, carrying substitutions in the homologous residues (e.g. D123, K125, D126) to weaken site specific RNA recognition (**Fig. S6A**). In

addition, we engineered two sets of mutations to shift the RNA-binding specificity of SRSF5 towards a CCG motif (mCCG-1 and mCCG-2) or a GGG motif (mGGG) (**Fig. S6B**).

Reassuringly, Gag-SRSF5 mutations did not impair its overall ability to bind RNA in cells, and all Gag-SRSF5 variants immunoprecipitated with RNA at higher than WT level **(Fig. 6A, 6C)**. In contrast, in virions, the parental Gag-SRSF5 variant bound to RNA at moderately higher levels than the mutants **(Fig. 6B, 6C)**. Gag-SRSF5 mGGA-1 and mGGA-2 mutants packaged gRNA at levels approximately two-fold lower than the parental Gag-SRSF5 variant **(Fig. 6D)**. The mCCG-1, mCCG-2 and mGGG mutants displayed stronger defects, with gRNA, multiply spliced viral RNAs, and 7SL RNA packaged at significantly lower levels **(Fig. 6D-F)**. Importantly, there was no significant difference for any of these mutants in the packaging efficiency of cellular RPL36A, RPS6 and RPS16 mRNAs that we and others have found to be abundantly incorporated into HIV-1 VLPs [25] **(Fig. 6G-I)**. These findings strongly suggest that the shifting the RNA-binding specificity of Gag-SRSF5 towards A-poor sequences specifically impair packaging of the A-rich gRNA without affecting its overall ability to bind and package non-viral RNAs.

Gag-RBD chimeras potently interfere with virion infectivity

Since Gag-RBD chimeras are non-infectious, with some having severe release defects, we next tested whether they can inhibit the release of WT Gag and/or co-assemble with it and subsequently interfere with HIV-1 replication. Presence of WT Gag in producer cells enhanced virion incorporation of Gag-hnRNP A1, F and K chimeras with negligible effects on Gag-SRSF1-5 chimeras **(Fig. 7A, 7B)**. On the other hand, Gag-hnRNP F and Gag-hnRNP K chimeras were present in virions at comparably lower levels than WT Gag and their presence didn't impair virion release by WT Gag levels **(Fig. 7A, 7B)**. On the other hand, co-transfection with Gag-hnRNP A1, Gag-hnRNP K, Gag-SRSF1, Gag-SRSF4, and Gag-SRSF5 substantially reduced viral titers (≥100 fold), again with negligible effects on particle release **(Fig. 7C, 7D)**. With the

exception of the Gag-hnRNPK chimera which reduced gRNA packaging by ~3-fold, other Gag-RBD chimeras did not interfere with gRNA packaging (**Fig. 7E**). Hence, the potent antiviral activities of these chimeras are unlikely due to interference with gRNA packaging by NC. It is noteworthy that Gag-hnRNP K exerted the most substantial dominant negative effect even though it was incorporated into particles at a much lower level than WT Gag **(Fig. 7A)**. In agreement, dose-response experiments corroborated these findings and demonstrated that Gag-hnRNP K is the most potent antiviral chimera **(Fig. 7F, 7G)**. Interestingly, co-transfection with Gag-ΔNC and Gag-bZIP also led to a significant decrease in virion infectivity (**Fig. 7C, 7D**), suggesting that the functional amount of NC copies in virions is finely gauged.

These findings prompted us to test whether the antiviral activity of Gag-RBD chimeras can be mapped to individual RBDs, and conversely, whether other hnRNP/SR proteins that naturally bear a single RBDs (e.g. hnRNP C, SRSF2, SRSF3, SRSF7) can also exert antiviral effects. All Gag chimeras bearing a single RBD were expressed in cells and virion release largely correlated with their cognate expression level in cells **(Fig. S7A, S7B)**. Similar to Gag chimeras with tandem RBDs, Gag-single RRM viruses were non-infectious **(Fig. S7C)**. While all of these Gag chimeras bound RNA in cells **(Fig. S7A, S7B),** none of them, including the Gag-SRSF5-RRM1 and Gag-SRSF5-RRM2 chimeras, were able to package gRNA **(Fig. S7D)**. Numerous Gag-RBD chimeras reduced virion infectivity, with unique patterns emerging for each chimera. For example, individual RRMs of hnRNP A1 each reduced virion infectivity by 10-fold as compared to 100-fold reduction with both RRMs, suggesting cooperativity between the RRMs (**Fig. 7C, 7H, 7I**). The potent antiviral activities of Gag-hnRNP K and Gag-SRSF4 largely mapped to the KH2 and RRM1 domains, respectively (**Fig. 7C, 7H, 7I**). Interestingly, while the Gag-hnRNP F chimera did not impair virion infectivity, Gag-hnRNP F-RRM1 chimera was potently antiviral (**Fig. 7C, 7H, 7I**). Of the proteins with single RBDs, Gag-SRSF2 and Gag-SRSF3 were the most potently antiviral, whereas Gag-SRSF7 did not inhibit virion infectivity and

Gag-hnRNP C1 modestly inhibited infectivity (**Fig. 7H, 7I**). Like with the full-length chimeras, despite their potent antiviral activities, none of the chimeras interfered with gRNA packaging by WT Gag (**Fig. 7J**).Lasty, dose-response experiments demonstrated that Gag-hnRNP K-KH2 has the most potent antiviral activity **(Fig. 7K-L)**.While it is unclear how the Gag-RBD chimeras inhibit virion infectivity and whether they all inhibit the same step in HIV-1 replication, these findings highlight that they can be repurposed as potent antiviral modules to target NC-dependent functions post gRNA packaging.

DISCUSSION

HIV-1 genome packaging is widely thought to depend on the selective binding of Gag to gRNA through recognition of a *cis*-acting packaging signal, Ψ, within the 5'-UTR. Gag-gRNA interaction initiates in the cytosol in the excess of cellular mRNAs and this initial "finding" step is thought to be the major driver of gRNA selection. Our findings challenge this notion and demonstrate that heterologous RBDs with diverse RNA-binding specificities, without exception, can successfully recruit gRNA to the host cell PM. This suggests that NC in and on its own is not required for gRNA selection in the cytosol and rather its properties post-PM recruitment is more critical.

Our findings highlight that the high adenosine content of HIV-1 gRNA serves as a key molecular signature that mediates selective gRNA packaging following its recruitment to the PM. Hence an important feature of NC that emerges from our findings is its ability to multimerize on A-rich sequences on the HIV-1 gRNA. This idea is strengthened by the key finding in our study that Gag-SRSF5 chimera efficiently packages gRNA through mimicking the ability of NC to multimerize on A-rich sequences on the viral gRNA both at the PM and in immature virions. Strong support for this notion is further provided by targeted mutations that altered the RNA-binding specificity of Gag-SRSF5 resulting in specific inhibition of gRNA packaging.

An unexpected finding from our study is that Gag-RBD chimeras with diverse RNA-binding specificities appeared to bind regions within Ψ in cells, while maintaining their known RNA-binding specificities for cellular RNA targets. Multiple studies have shown that WT Gag recognizes Ψ via high affinity interactions with unpaired guanosines within SL1 and SL3[17,64-66]. Since the Gag-RBD chimeras tested here, with the exception of Gag-hnRNP K, preferentially bind to single-stranded G-rich or A/G-rich sequences, it is possible that they bound to Ψ through similar interactions with unpaired guanosines. Moving forward, it will be important to determine whether RNA-binding modules that recognize dsRNAs can similarly recruit the HIV-1 gRNA to the PM and successfully replace NC. In addition, understanding whether Ψ and HIV-1 gRNA themselves bear unique structural features that destine the gRNA more readily for PM trafficking by diverse RBDs will provide crucial insight into selective gRNA packaging.

While our findings suggest that multimerization-driven A-rich sequence content recognition underlies the ability of Gag-SRSF5 to package gRNA, why the closely related Gag-SRSF1 and Gag-SRSF4 chimeras with similar RNA-binding properties were unable to package gRNA is unclear. Together, these findings suggest that simply switching to A/G-rich binding at the PM may be insufficient for virion genesis. It is possible that the higher RNA binding affinity of these Gag-RBD chimeras could lead to an excessive amount of RNA being brought to virus assembly sites, sterically hindering formation of Gag high-order multimers around a single gRNA dimer and instead forming multimers on other non-productive RNA molecules that is not conducive to forming membrane curvature. In line with this, Gag-hnRNP and Gag-SRSF1 chimeras exhibited defects in membrane curvature while effectively multimerizing on the PM. Notably, the release defects of these chimeras could be rescued at least in part by WT Gag, presumably by reducing the levels of RNA retained at the PM and suggesting that these chimeras are not inherently defective in making it into budding virions. Future studies in defined in vitro systems will be informative to more directly test these models.

The unusual nucleotide composition of HIV-1 gRNA has remarkably remained stable through multiple decades since the beginning of the HIV-1 pandemic despite the HIV-1 genome being highly variable. While recent evidence suggest that this property of HIV-1 gRNA is partly shaped by selective pressure imposed by host antiviral factors and higher G-to-A mutation rates during reverse transcription, our study raises the possibility that selective packaging of the HIV-1 genome also contribute to the maintenance of A-richness. Specifically, viral RNAs with higher adenosine content will be preferentially packaged because they are more efficient at driving Gag multimerization. Furthermore, because Gag binds to unpaired nucleotides, A-richness could also contribute to genome packaging by decreasing the prevalence of structured regions throughout the HIV-1 gRNA, owing to adenosine only forming a weak base pair with uracil. Indeed, an analysis of the secondary structure model of a complete HIV-1 genome found that single-stranded regions have a high degree of adenosine accumulation, but regions where secondary RNA structures are formed[67].

In addition to providing a unique tool for studying gRNA packaging, Gag-RBD chimeras will be instrumental in investigating functions of NC beyond gRNA packaging and a multitude of other RNA-binding proteins essential for virus replication. NC has been implicated in various steps of HIV-1 replication, including budding, reverse transcription and integration[47,49,68]. Deeper dissection of what makes a Gag-RBD chimera successful in executing these steps will provide key insights into NC functions critical for virus replication. Considering that a single dimeric gRNA in a virion is bound by a magnitude of viral (e.g. integrase and reverse transcriptase) and host RNA-binding proteins (e.g. APOBEC3 family), the ability to package gRNAs independent of NC by Gag-SRSF5 will also help determine the interdependency of how these diverse RNA-binding proteins interact with the same gRNA.

Given the potent antiviral activities of Gag-RBD chimeras, understanding how they inhibit HIV-1 replication can reveal new vulnerabilities of HIV-1 and inform the development of new virus-

targeted therapeutics. Another relevant application of our findings is the design of gene therapy vectors. While altering the nucleotide content of a transgene in this setting may enhance vector production, whether protein expression from mRNAs bearing suboptimal codons will suffice to obtain the desired phenotype remains to be determined. Hence, fine-tuning of the transgene codon usage may help enhance packaging while still allowing sufficient protein production.

Because HIV-1 has a high mutation rate, a less restrictive mechanism of selecting viral genomes for packaging — through efficient virion assembly on viral RNAs with high adenosine bias rather than specific binding of viral proteins to a packaging signal made up of conserved RNA sequences — could allow Gag to package the constantly diversifying genomes. Besides HIV-1, a large number of other viruses with RNA genomes (e.g. HTLV-1 and Rubella) also have a biased nucleotide composition including [69]. For many RNA viruses, in particular for those with segmented genomes, molecular features that allow genome selection are poorly defined. While our study suggests that nucleotide composition may serve as a molecular signature for HIV-1 genome packaging, similar principles can in principle be applicable to a broad range of RNA viruses.

FIGURE LEGENDS

Figure 1. Replacement of HIV-1 NC with heterologous RNA-binding modules reveals the unique ability of Gag-SRSF5 chimera to package gRNA. A. Cartoon diagram of HIV-1 Gag-RBD chimeras in which the NC domain was replaced by RNA-binding domains from heterologous RNA-binding proteins of host hnRNP and SRSF proteins. **B.** The indicated HIV-1 virions were titered on TZM-bl reporter cells, with 100 μg/mL dextran-sulfate added at 6-16 hpi to limit infection to a single cycle. The titers were subsequently normalized based on particle numbers as quantified by a reverse transcriptase (RT) activity assay and presented relative WT (set to 1). Titers are mean ± SEM from n=3 independent biological replicates. **C.** HEK293T cells were transfected with PR^+ proviral plasmids encoding the indicated Gag-RBD chimeric proteins. Cell lysates and purified virions were harvested 2 days post-transfection and analyzed with immunoblotting for Gag-Pol (GP), Gag, Envelope (gp120), Nef, Capsid (p24) and Integrase (IN). **D.** Representative autoradiogram of Gag:RNA adducts (top) and corresponding anti-HA immunoblot of Gag proteins (bottom) immunoprecipitated from cells that were transfected with PR^- proviral plasmids encoding the indicated Gag-RBD chimeras following PAR-CLIP protocol. **E.** Densitometry analysis of Gag:RNA adducts from autoradiograms (e.g. panel D) of independent biological replicates. Data are normalized relative to immunoprecipitated protein levels and graphed as fold differences relative to WT (set to 1). Graph shows the average of 3-5 independent experiments denoted by symbols, error bars show the SEM. **F.** RT-qPCR quantification of relative copies of gRNA, multiply spliced HIV-1 RNA (multiply spl.), and 7SL RNA (7SL) from purified PR^+ virions collected from HEK293T cells expressing the indicated Gag chimeras. Data show the average of 3 independent experiments and the error bars represent standard error of the mean (SEM). Indicated statistical analyses (gRNA, multiply spliced RNA and 7SL RNA) were done by one-way ANOVA with Dunnet's multiple comparison test for correction, whereby n.s. denotes non-significant comparisons. Other group comparisons are not

shown due to clarity. **G-H.** Proportion of reads that map to viral RNA (G) and cellular non-ribosomal RNAs (H) obtained from total RNA-seq experiments performed with purified virions. Data show the average of 2-3 independent experiments, error bars represent the SEM. **I.** Bar plot showing Jaccard index of reads that map to host mRNAs derived from virion total RNA-seq to quantify the similarity between the indicated Gag chimeras and WT Gag.

Figure 2. Single virion analysis of gRNA packaging by and virion morphology of the Gag-SRSF5 chimera. A. Cartoon depiction of strategy used for simultaneous fluorescence imaging of Gag protein and HIV-1 gRNA. YFP, yellow fluorescent protein; MSL, 24 copies of MS2 RNA stem-loop; MS2-mChe, bacteriophage MS2 coat protein fused with the fluorescent protein mCherry. **B.** Representative fluorescence microscopy images of HIV-1 particles harvested from 293T cells transfected with MS2-mCherry plasmid and MSL HIV-1 plasmids expressing the indicated Gag chimera-YFP fusion protein. Quantification of the percentage of total virions that contains gRNA ($mCherry^{+}YFP^{+}$) based on single-virion analysis. **C.** Violin plot showing MS2-mCherry/Gag-YFP mean fluorescence intensity (MFI) ratios measured for WT, Gag-ΔNC, and Gag-SRSF5 particles shown in panel B. 60 particles were quantified for each condition. **D.** Representative thin section electron microscopy (TEM) images showing WT Gag, Gag-ΔNC, Gag-ΔNC-NC,Gag-bZIP, and Gag-SRSF5 virions. **E-F.** The indicated virions were subjected to equilibrium density sedimentation following treatment with 0.5% Triton as detailed in Materials and Methods. Ten fractions harvested from the top of sucrose gradients were analyzed for RT activity via a qPCR assay (E) or CA content via immunoblotting using a mouse monoclonal anti-CA antibody (F). Results in E and F show the mean values ± SEM of three independent biological replicates.

Figure 3. Gag-RBD chimeras efficiently recruit and retain gRNA at the PM. A. Immunoblot analysis of Gag from membrane (M) and cytoplasmic (C) fractions isolated from 293T cells transfected with the indicated PR^{-} proviral plasmids using a mouse monoclonal anti-HA

antibody. **B.** RT-qPCR quantitation of the relative PM/cytosol ratios of gRNA, multiply spliced HIV-1 mRNA, and GAPDH mRNA copies. Data show the average of three independent experiments and the error bars represent SEM. Indicated statistical analyses for gRNA were done by one-way ANOVA with Dunnet's multiple comparison test for correction (**** $p<0.0001$). Non-significant comparisons are not shown for clarity. **C.** Fluorescence imaging of HeLa cells transfected with MS2-mCherry and HIV-1 constructs. Scale bars represent 20 µm in full images and 2.5 µm in regions of interest (ROI) at the PM. ROIs are indicated by yellow dotted boxes. Yellow arrowheads indicate where Gag and gRNA colocalize. **D.** Representative TEM images from two independent experiments showing virion assembly sites at the PM of 293T cells transfected with the indicated Gag plasmids. Scale bar represents 500 nm.

Figure 4. CLIP-seq analysis of Gag-RBD chimeras reveals their RNA-binding properties in cells. A. Proportion of reads that map to cellular and viral genomes obtained from Gag-CLIP experiments performed on cell lysates. Each point represents an independent experiment. **B.** Classification of read clusters generated from cell Gag-CLIP experiments that map to cellular RNAs; total number of reads is indicated below each pie chart. **C.** Representative sequence motifs most frequently occurring in Gag-bound cellular mRNA clusters generated from CLIP experiments. **D.** Read density distribution within the 5' leader (1-400 nt) from cell Gag-CLIP experiments.

Figure 5. Gag-SRSF5 mimics NC in its ability to multimerize on the gRNA in immature virions. A-B Densitometry analysis of Gag:RNA adducts from autoradiograms (See Figure S5A, S5B) of independent CLIP replicates done on PM fractions (A) or immature virions (B). Data are normalized relative to immunoprecipitated protein levels and graphed as fold differences relative to WT (set to 1). Graph shows the average of 3 independent experiments and the error bars represent SEM. **C.** Proportion of reads that map to cellular and viral genomes obtained from virion Gag-CLIP experiments. Each point represents an independent experiment.

D. Read density distribution on full-length HIV-1 gRNA from virion CLIP-seq experiments. **E.** Correlation analysis of WT, ΔNC-NC, and Gag-SRSF5 binding to viral RNA in virions from independent CLIP experiments. **F.** Classification of read clusters generated from virion Gag-CLIP experiments that map to cellular RNAs; total number of reads is indicated below each pie chart**.** Representative sequence motifs most frequently occurring in Gag-bound cellular mRNA clusters generated from virion Gag-CLIP experiments.

Figure 6. Altering the RNA-binding specificity of Gag-SRSF5 specifically impairs gRNA packaging. A-B. Representative autoradiograms of Gag-RNA complexes (top) and immunoblot analysis of Gag-SRSF5 mutants (bottom) immunoprecipitated from cells expressing the indicated PR$^-$ Gag chimeras (A) and progeny virions (B). (*) marks the IgG-heavy chain from the immunoprecipitating antibody. Eluates in virion IPs in (B) were normalized to load similar amounts of protein for Gag and Gag-SRSF5 variants to more clearly illustrate RNA binding ability per protein. **C.** Densitometry analysis of Gag:RNA adducts from autoradiograms (e.g. panels A and B) of independent replicates. Data are normalized relative to immunoprecipitated protein levels and graphed as fold differences relative to WT (set to 1). Graph shows the average of 3 technical replicates, representative of two independent experiments, and the error bars represent SEM. **D-I.** RT-qPCR quantification of relative copies of gRNA, multiply spliced HIV-1 RNA, 7SL, RPL36A, RPS6, and RPS16 RNA from HIV-1 particles collected from 293T cells expressing the indicated Gag. Data show the average of 3 independent experiments and the error bars represent SEM. Indicated statistical analyses (gRNA, multiply spliced RNA and 7SL RNA) were done by one-way ANOVA with Dunnet's multiple comparison test for correction, whereby n.s. denotes non-significant comparisons (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$). Other group comparisons are not shown due to clarity.

Figure 7. Gag-RBD chimeras are potently antiviral. A. Proviral plasmids encoding ΔGag or WT Gag was co-transfected with the indicated Gag-RBD proviral plasmids in 293T cells. 2 days

post-transfection, cell lysates and purified virions were harvested for immunoblot analysis of Gag using a mouse monoclonal antibody against the HA-tag. **B.** Quantitation of virion immunoblot results from panel A. Fold increase of the virion-incorporated Gag-RBD chimeric proteins in the presence of WTGag over ΔGag co-transfection control was graphed. Data show the mean fold increase from n=2 independent experiments. **C-L.** 293T cells were transfected with equal ratio of tandem RRM or single RRM Gag-RBD chimeric HIV-1 proviral plasmids and a WT Gag HIV-1 proviral plasmid. Progeny virions were titered on TZM-bl reporter cells, with dextran-sulfate added to limit infection to one cycle. The titers were presented as relative to WT which was set to 1 (C, H). Particle release from (C) and (H) was quantified by RT activity assays (D, I). Data are mean ± SEM from three-four independent experiments. RT-qPCR quantitation of relative HIV-1 gRNA and multiply spliced HIV-1 RNA copies from purified virions co-assembled from tandem RRM (E) or single RRM (J) with WT Gag. Purified virions harvested from 293T co-expressing tandem RRM (F) or single RRM (K) Gag chimera with WT Gag at the indicated ratios were titered on Tzm-bl reporter cells. RT activity from purified virions were quantified to assess particle release (G, L). Data are mean ± SEM of 2-3 independent experiments.

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Properties of Gag-RBD chimeras with heterologous RNA binding modules in place of NC. A-B. HEK293T cells were transfected with PR^- pNL4-3 proviral plasmids encoding the indicated Gag-RBD chimeras. Cell lysates (A) and purified virions (B) were harvested 2 days post-transfection and analyzed by immunoblotting for Gag, Gag-Pol (GP), Nef and Env (gp120). **C-D.** RT-qPCR quantification of relative copies of gRNA, multiply spliced viral RNAs, and 7SL RNA (7SL) from purified virions with (C, PR^-) or without (D, PR^+) an inactivation mutation in viral protease collected from 293T cells expressing the indicated Gag chimeras. For PR^+ virions, copies of packaged RNA was normalized relative to virus particle numbers quantified based on RT activity. The columns represent the average of 3 independent experiments and the error bars represent SEM. Indicated statistical analyses (gRNA) were done by one-way ANOVA with Dunnet's multiple comparison test for correction, whereby (*) denotes $p<0.05$ comparisons. **E.** Northern blot analysis of unspliced (US), partially spliced (PS), multiply spliced (MS) HIV-1 RNAs and host 7SL RNA isolated from transfected cells and purified HIV-1 particles. **F.** Proportion of cellular reads, following removal of viral RNA-derived reads, that map to rRNA from total RNA-seq experiments using purified virions. Data show the average of 2-3 independent experiments and the error bars represent SEM. **G.** Classification of non-rRNA cellular RNA reads from virion total RNA-seq experiments as 7SL, mRNA, and Other.

Figure S2. Assembly and gRNA packaging properties of Gag-SRSF5 virions. A. Violin plots of MFI of Gag-YFP (left) and MS2-mCherry (right) measured for 60 WT, Gag-ΔNC, Gag-SRSF5 particles imaged in Figure 2B. **B.** WT and Gag-SRSF5 virions were purified from culture supernatants of transfected HEK293T cells and subjected to equilibrium density sedimentation following treatment with 0.5% Triton as detailed in Materials and Methods. Ten fractions (top-to-bottom) were analyzed by immunoblotting using a mouse monoclonal antibody against CA. **C.** MT4-GFP cells were infected with VSV-G pseudotyped HIV-1 at MOI of 0.5 or 1 in the presence

and absence nevirapine (NVP, 25 μM). At 24 hpi, cells were collected for DNA extraction as detailed in Materials and Methods. Extracted DNA was analyzed by qPCR for accumulation of early RT products (ERT), late RT products (LRT), 2-LTR circles (2-LTR), and integrated viral DNA (Alu-PCR).

Figure S3. Trafficking and assembly properties of Gag-RBD chimeras. A. Representative (n=3) immunoblot analysis of Na^+K^+ ATPase and β-tubulin from membrane (M) and cytoplasmic (C) fractions isolated from 293T cells transfected with the indicated PR^- proviral plasmids. Samples are the same as those shown in Fig. 3A. **B.** Representative confocal microscopy images of HeLa cells transfected with PR^- proviral plasmids encoding the indicated Gag. Cells were stained for Gag (red) and nuclei by DAPI (blue). **C.** Quantification of Gag-gRNA colocalization rate from fluorescence imaging experiments shown in Figure 3C. **D.** Confocal imaging of the bottom surface of cells transfected with two-color Gag-YFP-SRSF5/MS2-mCherry or Gag-YFP-hnRNPA1/MS2-mCherry viruses. **E.** Images from live cell imaging of cells expressing two-color Gag-YFP-hnRNPA1/MS2-mCherry virus.

Figure S4. Gag-SRSF5 gRNA packaging is not affected by deletions of structured elements within the HIV-1 5' leader. A. Read density distribution on full-length HIV-1 gRNA (x-axis) from CLIP-seq experiments in which PR^- WT Gag and Gag-RBD chimeras were immunoprecipitated from transfected 293T cell lysates. **B.** Depiction of secondary structures within the 5'-UTR of HIV-1 RNA (top) and table (bottom) showing sequences deleted in ΔΨ1-6 mutants used for this study. TAR, trans-activation response element; PBS, primer-binding site; DIS, dimerization initiation signal; SD, major splice donor. **C-D.** RT-qPCR analysis of relative copies of virion-incorporated HIV-1 gRNA and multiply spliced HIV-1 mRNA (C) and immmunoblot analysis using a mouse monoclonal anti-HA antibody (D) of purified PR^- ΔΨ virions. The columns in panel C represent the average of two independent experiments (each

done in two technical replicates representing independent transfections), and the error bars represent SEM.

Figure S5. Analysis of the RNA binding properties of Gag-RBD chimeras at the PM and immature virions. A-B. Autoradiogram of Gag-RNA complexes (top) and immunoblot analysis of Gag proteins (bottom) immunoprecipitated from the PM fraction (A) or immature virions (B) in CLIP experiments. (*) marks the IgG-heavy chain from the immunoprecipitating anti-HA antibody. **C.** Frequency distribution of nucleotide occurrence (read density) in reads mapping to the HIV-1 genome derived from CLIP experiments done in PM fractions.

Figure S6. Rational design of Gag-SRSF5 variants with reduced affinity for GGA and enhanced affinity for CCG or GGG. A-B. Structural depiction of key residues involved in GGA motif recognition in Gag-SRSF5 variants with reduced GGA affinity (A, mGGA1- and mGGA-2). Residues of Gag-SRSF5 variants with enhanced affinity for CCG or GGG motif (mCCG-1, mCCG-2, and mGGG) are shown in panel B. The amino acid positions shown are based on full-length SRSF5.

Figure S7. Single RRM Gag chimeras bound RNA in cells, but do not package gRNA efficiently. A-B. 293T cells were transfected with PR^- pNL4-3 proviral plasmids encoding single RRM (e.g. M1, M2) Gag chimeras. Cell lysates and purified virions were harvested 2 days post-transfection and analyzed with immunoblot for Gag and Gag-Pol (GP). Single RRM Gag-RNA complexes were immunoprecipitated from cell lysates and analyzed using autoradiography and immunoblotting. **C.** The relative titers of purified virions were determined via single-cycle infection of TZM-bl reporter cells. **D.** RT-qPCR quantification of relative copies of gRNA, multiply spliced HIV-1 RNA, and 7SL RNA from purified virions. Data are mean ± SEM from 4 independent experiments.

MATERIALS AND METHODS

Plasmids and cloning.

pNL4-3-derived plasmids containing three copies of a HA-tag in the stalk region of MA [20] and with (PR-) or without (PR+) an inactivating mutation in the protease (PR) gene were modified to insert the heterologous RNA-binding domains as follows. First, the NC coding sequence was replaced by a NotI restriction endonuclease site through conventional cloning using outward BssHII and SphI sites (see list below). Sequences coding for the first four and last three amino acids of NC were left intact to facilitate the proper proteolytic cleavage of the inserted heterologous RNA-binding domains. PCR amplicons containing individual or the complete set of RNA-binding domains from hnRNP and SR proteins were inserted into the Not I site (see list below). Chimeras bearing a subset of RRMs were also subcloned into a previously described two-color virus backbone [70] using SphI and SacII cut sites for microscopy analyses.

Deletions within the 5' leader were generated by overlap-extension PCR mutagenesis. The overlap PCR fragments were restriction digested and subsequently cloned into pNL4-3/MA-3xHA backbone via conventional cloning. ΔΨ5 was built on the ΔΨ1 backbone. ΔΨ3 and ΔΨ6 was built on the ΔΨ2 backbone. AatII/SphI fragment derived from these ΔΨ plasmids were then cloned into Gag-SRSF5 backbone. Gag-SRSF5 mutants (mGGA-1, mGGA-2, mCCG-1, mCCG-2, and mGGG) were generated using overlap-extension PCR mutagenesis, with PCR amplicons being cloned into the Not I site via ligation or Gibson Assembly. mGGA-1 and mGGA-2 overlap PCR amplicons were generated with 2 fragments; mCCG-1 and mCCG-2 were generated with 4 fragments; mGGG was generated with 3 fragments.

Primer pairs used in cloning are listed in the below table:

650

Primer	Sequence
ΔNC-NotI-F	ACTGAGCGGCCGCAATTTTAGGAACCAAAGAAAG
ΔNC-NotI-R	ACTGAGCGGCCGCCTCTCTCAGTACAATCTTTCATTTG
BssHII-F	ATGGCGCCCGAACAGGGACT
AatII-F	CCTGACGTCTAAGAAACCATTATTATC
SphI-R	GCCCTGCATGCACTGGATGCAATCTATCCCA
hnRNP A1-RRM1-F	ACTGAGCGGCCGCCAGCTGAGGAAGCTCTTCATTG
hnRNP A1-RRM1-R	ACTGAGCGGCCGCCTCTGGAGACAGCTCTCTTTGG
hnRNP A1-RRM2-F	ACTGAGCGGCCGCAAAAAGATATTTGTTGGTGGC
hnRNP A1-RRM2-R	ACTGAGCGGCCGCCCAGGGCTTTTCTAACTTCACAG
hnRNP F-RRM1-F	ACTGAGCGGCCGCGGCTTTGTGGTCAAGCTCCG
hnRNP F-RRM1-R	ACTGAGCGGCCGCCGGACTTGAACACCTCAATG
hnRNP F-RRM2-F	ACTGAGCGGCCGCATGGATTGGGTGTTGAAGC
hnRNP F-RRM2-R	ACTGAGCGGCCGCCCTGGCTGCTCTTAAACACCTC
hnRNP F-RRM3-F	ACTGAGCGGCCGCCACTGTGTCCACATGAGGG
hnRNP F-RRM3-R	ACTGAGCGGCCGCCTGAATTCAAGAAGAGTTCTAT
hnRNP K-KH1-F	ACTGAGCGGCCGCATGGTTGAATTACGCATTCTG
hnRNP K-KH1-R	ACTGAGCGGCCGCCGATTTTCTTCAGAATTTCTC
hnRNP K-KH2-F	ACTGAGCGGCCGCGACTGCGAGTTGAGGCTGTTG
hnRNP K-KH2-R	ACTGAGCGGCCGCCGATGATCTTTATGCACTCTAC
hnRNP K-KH3-F	ACTGAGCGGCCGCATTATTACTACACAAGTAAC
hnRNP K-KH3-R	ACTGAGCGGCCGCCCTGCAGCAAATACTGTGCAT
SRSF1-RRM1-F	ACTGAGCGGCCGCTGCCGCATCTACGTGGGTA
SRSF1-RRM1-R	ACTGAGCGGCCGCCGCTTCGAGGAAACTCCACCC

SRSF1-RRM2-F	ACTGAGCGGCCGCAACAGAGTGGTTGTCTCTGG
SRSF1-RRM2-R	ACTGAGCGGCCGCCATCAACTTTAACCCGGATGTA
SRSF3-RRM1-F	ACTGAGCGGCCGCTGTAAGGTTTATGTAGGCAATC
SRSF3-RRM1-R	ACTGAGCGGCCGCCACCATTCGACAGTTCCACTCT
SRSF4-RRM1-F	ACTGAGCGGCCGCCCGCGGGTGTACATCGGCC
SRSF4-RRM1-R	ACTGAGCGGCCGCCGCCGCGGGCATGCTCAACAA
SRSF4-RRM2-F	ACTGAGCGGCCGCTACAGACTTATTGTGGAGAA
SRSF4-RRM2-R	ACTGAGCGGCCGCCTGGCTTGTCTTCAACTAATCT
SRSF5-RRM1-F	ACTGAGCGGCCGCTGTCGGGTATTCATCGGGA
SRSF5-RRM1-R	ACTGAGCGGCCGCCAGCCCTAGCATGTTCAATAGT
SRSF5-RRM2-F	ACTGAGCGGCCGCAATCGTCTTATAGTTGAGA
SRSF5-RRM2-R	ACTGAGCGGCCGCCTTTGCTGCCTTCAATTAATTTTA
Fragment 1-F (mGGA mGGA-2, mCCG-1, mCCG-2, mGGG)	TAATGATACAGAAAGGCAGCGGCCGCTGTCGGGTATTCATCGGG
Fragment 1-R (mGGA	GGCGGCGAGGGCCTGCCAGCTGACTCTTGAGGATAAATTCTC
Fragment 2-F (mGGA-1)	CAGCTGGCAGGCCCTCGCCGCCTTCATGAGACAAGCTGGGGAAG
Fragment 1-R (mGGA-2)	GTTCTGGAGGTTCTGCCAGCTGACTCTTGAGGATAAATTCTC
Fragment 2-F (mGGA-2)	CAGCTGGCAGAACCTCCAGAACTTCATGAGACAAGCTGGGGAAG
Fragment 1-R (mCCG-1, mCCG-2)	CTCTACAGGTGGAGCATTTCGTCTATCATTTCGAGGTCTG
Fragment 2-F (mCCG-1, mCCG-2)	ACGAAATGCTCCACCTGTAGAGAGAACAGAAAATCGTCTTATAG
Fragment 2- R (mCCG-1)	GGCGGCGAGCTTCTGCCAGCTGACTCTTGAGGATAAATTCTC
Fragment 3- F (mCCG-1)	CAGCTGGCAGAAGCTCGCCGCCTTCATGAGACAAGCTGGGGAAG

Fragment 3-R (mCCG-1, mCCG-2) & Fragment 2-R (mGGG)	CTCCTCCCCATTTATTTCCTTTCCAGAAAGTTTTTCAATAGC
Fragment 4-F (mCCG-1, mCCG-2) & Fragment 3-F (mGGG)	AAAGGAAATAAATGGGGAGGAGATAAAATTAATTGAAGGC
Fragment 2-R (mCCG-2) & Fragment 1-R(mGGG)	ATCTTTGAGCTTCTGCCAGCTGACTCTTGAGGATAAATTCTC
Fragment 3-F (mCGG-2) & Fragment 2-F (mGGG)	CAGCTGGCAGAAGCTCAAAGATTTCATGAGACAAGCTGGGGAAG
Fragment 2 (mGGA-1, mGGA-2) & Fragment 4 (mCCG-1, mCCG-2) & Fragment 3 (mGGG) (R)	CCCTAAAAAATTAGCCTGGCGGCCGCCTTTGCT

Cells, transfections, and viruses.

HEK293T (ATCC CRL-11268), HeLa (ATCC CCL-2) and HeLa-derived TZM-bL (NIH AIDS Reagent Program) cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) at 37°C. MT4-GFP indicator cells containing an HIV-1 LTR-driven GFP reporter gene were described before[71] and grown in RPMI-10% FBS supplemented with puromycin. For virus production, HEK293T cells were transfected with the indicated proviral plasmid using polyethylenimine (PEI) and viral particles were harvested at 48 hours post transfection from filtered cell culture supernatants. HeLa cell transfections were performed with either PEI or Lipofectamine 3000 (Thermo Fisher Scientific) reagent according to the manufacturer's instructions.

Immunoblotting.

Cells were lysed in 1X radioimmunoprecipitation assay (RIPA) buffer and cleared lysates were diluted in 1X final sodium dodecyl sulfate (SDS) sample buffer. Viruses were pelleted on a 20% sucrose cushion prepared in 1X PBS and similarly resuspended in 1X sodium dodecyl sulfate (SDS) sample buffer. Samples were separated by electrophoresis on Bolt 4–12% Bis-Tris Plus gels (Life Technologies), transferred to nitrocellulose membranes and probed with the following antibodies in Odyssey/Intercept Blocking Buffer (LI-COR): mouse monoclonal anti-CA (183-H12-5C, NIH AIDS reagents), mouse monoclonal anti-HA (Biolegend, clone 16B12), goat polyclonal anti-Env (ARP, 12-6205-1), rabbit polyclonal anti-Nef (NIH AIDS Reagents, ARP2949), mouse monoclonal anti-β-tubulin (Santa Cruz, SC-5274), mouse monoclonal anti-actin (Santa Cruz, SC-8432), mouse monoclonal anti-Na^+-K^+ ATPase (Santa Cruz, SC-21712). Membranes were probed with fluorophore-conjugated secondary antibodies and visualized using the LI-COR Odyssey system. Where indicated, Gag protein in cells and virions were quantified using the LI-COR Image Studio software to normalize for CLIP protein:RNA adduct signal.

RNA extraction and RT-qPCR.

Cell and virion associated RNA was harvested with TRIzol (Thermo Fisher Scientific) and precipitated overnight in isopropanol. RNA was resuspended and reverse transcribed with the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) according to the manufacturer's instructions. qPCR was performed using PowerUp 2X SYBR Green Master Mix (Applied Biosystems). Following primers were used to quantify their respective targets:

Primer	**Sequence**
Unspliced/Gag-Pol-F	TTCTTCAGAGCAGACCAGAGC

Unspliced/Gag-Pol-R	GCTGCCAAAGAGTGATCTGA
Multiply spliced-F	TCTATCAAAGCAACCCACCTC
Multiply spliced-R	CGTCCCAGATAAGTGCTAAGG
7SL RNA-F	CACCAGGTTGCCTAAGGAGGGG
7SL RNA-R	TCCCACTACTGATCAGCACGGG
GAPDH-F	AGGTGAAGGTCGGAGTCAACG
GAPDH-R	GGTCATTGATGGCAACAATATCCACTTTAC
RPL36A-F	GCACCAACCCCATAAAGTGAC
RPL36A-R	CTGGGCGTACAGAGAATCCT
RPS6-F	TGGACGATGAACGCAAACTTC
RPS6-R	TTCGGACCACATAACCCTTCC
RPS16-F	TCGGACGCAAGAAGACAGC
RPS16-R	AGCAGCTTGTACTGTAGCGTG

Quantification of reverse transcription and integration

MT4-GFP cells were grown in 24-well plates and infected with VSV-G pseudotyped WT or Gag chimera viruses at an MOI of 0.5 or 1 in the presence of polybrene. Cells were collected at 24 hpi, washed with PBS, and resuspended in 200 µL lysis buffer (100 mM NaCl, 10 mM Tris-HCl pH 8, 25 mM EDTA pH 8, 0.5% SDS, 0.1 mg/mL Proteinase K). Samples were incubated at 50°C by constant agitation (2,000 rpm) in a thermal mixer (Eppendorf) for 2 h. DNA was extracted by phenol-chloroform extraction. Extracted DNA was analyzed for accumulation of RT products, 2-LTR circles, and integration as described before [72,73].

Northern blotting

For northern blotting experiments, 293T cells were transiently transfected with 5 µg of each plasmid using a 4:1 ratio of PEI and 150 mM NaCl. Media was replaced 24- and 44-hours post-infection. 4 hours after the 44-hour media change, extracellular viral particles were harvested by

passing culture supernatants through a 0.22 um filter. Particles were then concentrated via ultracentrifugation through a 20% sucrose cushion. RNA was isolated from transfected cells and concentrated virions using Trizol Reagent. Northern blot analysis was performed using a modified version of a previously described protocol. Briefly, RNA samples were incubated in 1x MOPS electrophoresis buffer, 8% formaldehyde, 50% formamide, and gel loading buffer and then heated at 85°C for 10 minutes. The denatured samples were loaded into the wells of a 1% agarose gel containing 2.2M formaldehyde, and electrophoresis proceeded for approximately 7 hours at 50V. RNA species were then vertically transferred to a nylon membrane. Following transfer, the membrane was exposed to UV to induce crosslinking. The membrane was then incubated in prehybridization buffer (50 mM Tris, pH 7.5, 120 µg/mL salmon sperm DNA) at 55°C. After 4 hours, a ^{32}P-labelled probe complementary to the viral genome or 7SL RNA was added to the prehybridization buffer and allowed to hybridize with the membrane at 55°C overnight. The membrane was washed the next day and then exposed overnight in a phosphorimaging cassette. The screen was imaged using an Amersham™ Typhoon™ scanner.

Membrane Flotation Assays.

Membrane flotation assays were performed as previously described [33]. HEK293T cells grown in 10 cm plates were transfected with 10µg of the indicated proviral plasmids using PEI and cell culture media was replaced 6-8 hours post-transfection. 24 hours post transfection, cells were collected, washed once with 1X STE buffer (100 mM NaCl, 10 mM Tris-Cl pH 8.0, 1 mM EDTA). Afterwards, cells were resuspended in 500µl of hypotonic solution (10mM Tris [pH 7.4], 10mM KCl, 1 mM EDTA) supplemented with Complete Protease Inhibitor Cocktail (Roche), and placed on ice for 15 minutes. Cells were then lysed by dounce homogenization and centrifuged at 1000 x g for 5 minutes, 4°C. 350µl of post-nuclear supernatant was collected and mixed with 1.65 mL of 90% sucrose solution. 6.5 mL of 65% sucrose solution followed by 2.5 mL of 10% sucrose solution were layered on top of the mixture. Centrifugation was then performed with a Beckman

SW41 Ti rotor at 35,000 rpm for 4-18 hours, 4°C. 1 mL fractions were collected, with the 3rd fraction and 11th fraction from top corresponding to the membrane and cytosol, respectively. 100 µL of each fraction was kept for RNA analysis while the remaining 900 µL was used for protein precipitation. Aliquots for RNA analysis were treated with Proteinase K for 10 min at 37°C and RNA was extracted with an equal volume of phenol:chloroform:isoamyl alcohol (125:24:1) (Sigma-Aldrich) and precipitated with the addition of 1 µL Glycoblue (Ambion), 0.25 volume of 3M NaOAc pH 5.2, and 2.5 volumes of 1:1 ethanol:isopropanol. The resulting RNA samples were treated with DNase, reextracted using one more round of phenol:chloroform extraction and used for RT-qPCR. Protein was precipitated in 10% trichloroacetic acid (TCA) for 2-4 hours and pelleted by centrifugation. Protein pellets were washed twice with 10% TCA and once with ice cold acetone before being resuspended in SDS-PAGE sample buffer and used for western blot analyses.

Equilibrium density sedimentation of virion core components in vitro

Equilibrium density sedimentation of virion core components was performed as previously described [74]. Briefly, HEK293T cells grown on 10-cm dishes were transfected with the above-mentioned plasmids. 48 hours post-transfection, cell-free virions collected from cell culture supernatants were pelleted through a 20% sucrose cushion. Pelleted virions were resuspended in 1X PBS and treated with 0.5% Triton X-100 for 2 minutes at room temperature. Immediately after, samples were layered on top of a 30-70% linear sucrose gradients in 1X STE buffer and centrifuged for 16 hours at 4°C and 28,500 rpm using an SW55Ti rotor. 500 µL fractions collected from the top were analyzed by immunoblotting using a mouse monoclonal anti-HIV p24 antibody (183-H12-5C, NIH AIDS reagents) and Q-PCR-based assays for RT activity [75].

Transmission electron microscopy.

To visualize virus assembly sites in cells, HEK293T cells were grown in 10 cm plates and transfected with 10µg of proviral plasmid as previously described. 24 h post-transfection, cells were harvested and fixed in 2% paraformaldehyde/2.5% glutaraldehyde (Ted Pella Inc., Redding, CA) in 100mM cacodylate buffer, pH 7.2 for 2 h at room temperature. Samples were washed in cacodylate buffer and postfixed in 1% osmium tetroxide (Ted Pella Inc.)/1.5% potassium ferricyanide (Sigma, St. Louis, MO) for 1 h. Samples were then rinsed extensively in dH20 prior to en bloc staining with 1% aqueous uranyl acetate (Ted Pella Inc.) for 1 h. Following several rinses in water, samples were dehydrated in a graded series of ethanol and embedded in Eponate 12 resin (Ted Pella Inc.). Ultrathin sections of 95nm were cut with a Leica Ultracut UCT ultramicrotome (Leica Microsystems Inc., Bannockburn, IL), stained with uranyl acetate and lead citrate. To visualize viral particles, HEK293T cells grown in 15-cm plates were transfected with 30 µg of the indicated plasmid. Cell free virions were harvested from the supernatant 48 h post transfection, filtered through 0.22 µm filters and pelleted through a 20% sucrose cushion via centrifugation using a Beckman SW32-Ti rotor at 28,000 rpm for 1.5 h, 4°C. Virion pellets were fixed and processed for EM as described above. All samples were viewed on a JEOL 1200 EX transmission electron microscope (JEOL USA Inc., Peabody, MA) equipped with an AMT 8-megapixel digital camera and AMT Image Capture Engine V602 software (Advanced Microscopy Techniques, Woburn, MA).

Immunofluorescence microscopy.

HeLa cells were plated in 24-well glass-bottom dishes (Mattek Corporation, Ashland, MA, USA) coated with poly-L-lysine (50,000 cells/well) and transfected with 500ng of the indicated proviral plasmid using either PEI or Lipofectamine 3000 transfection reagent. At 24 h post-transfection cells were fixed with 4% paraformaldehyde in PBS, for 10 min and, for immunofluorescence-

based experiments, permeabilized with 0.1% Triton X-100 in PBS for 10 minutes. Permeabilized cells were blocked with 5% goat serum in PBS-T (PBS, 0.1% Tween 20) blocking solution for 30 min at room temperature. Cells were probed with a mouse monoclonal anti-HA (Biolegend, clone 16B12) in blocking solution for 1-2 h at room temperature, followed by goat anti-mouse Alexa Fluor 594 secondary antibody (1:1000, Thermo Fisher Scientific) for 1 h at room temperature. Nuclei were stained with DAPI for 10 min. Cells were visualized with a Zeiss LSM880 Confocal Microscope with Airyscan, oil immersion x63 or A1R Confocal Microscope, oil immersion x60.

Single virion analysis

To generate labeled, two-colored virus-like particles for single virion analysis, approximately 500,000 HEK293T cells were plated in each well of a six-well dish and transfected with plasmids encoding wild-type, SRSF5 RBD, or ΔNC two-color self-tagging viruses using PEI. The media was exchanged at 24 h post-transfection. Virus particle-containing supernatants were harvested at 48 h post-transfection, filtered through a 0.22-µm filter, and centrifuged through 20% sucrose for 2 h at 18,213 rpm. The medium was discarded after centrifugation, and concentrated viral particles were resuspended in 1× PBS, plated in a 24-well glass-bottom dish (Cellvis, Mountain View, CA), and left overnight at 4°C to allow virus particles to settle down on the glass wells. Microscopy was performed using a Nikon Ti-Eclipse inverted wide-field microscope (Nikon Corp, Minato, Tokyo, Japan) using a 100× Plan Apo oil objective lens (numerical aperture [NA] 1.45). Cell and virion images were captured using an ORCA—Flash4.0 CMOS camera (Hamamatsu Photonics, Skokie, IL, USA) and the following excitation/emission filter sets: 510/535 nm (YFP) and 585/610 nm (mCherry). All images were processed and analyzed using FIJI/ImageJ2. FIJI/ImageJ2's built-in tools were used to measure the mean fluorescent intensities (MFIs) of both the virions and the background by selecting regions of interest in the Gag-YFP channel and the corresponding areas in the MS2–mCherry (unspliced RNA) channel.

Virion background subtracted MFIs for MS2–mCherry and Gag–YFP channels were plotted using GraphPad Prism (version 10.1.2). The data were plotted for MS2–mCherry MFI, Gag–YFP MFI, and MS2–mCherry/Gag–YFP MFI.

RNA-seq.

HEK293T cells were grown in 24-well plates and transfected with 400ng of the indicated proviral plasmid using PEI. 48 h post transfection, cell culture media was collected and filtered through 0.22 µm filters, before being pelleted through a 20% sucrose cushion at 15,000 rpm for 1.5 h, 4°C. RNA was harvested from purified cell free virions with TRIzol as previously described. RNA samples were prepared for sequencing using the TruSeq Stranded Total RNA Sample Prep Kit (Illumina) according to the manufacturer's instructions but without rRNA depletion and as detailed previously [76]. Barcoded samples were pooled and sequenced using an Illumina NextSeq platform (1x75) at the Center for Genome Sciences at Washington University.

Photoactivatable ribonucleoside-enhanced cross-linking and immunoprecipitation (PAR-CLIP).

PAR-CLIP experiments were conducted as previously described with some modifications [20,77,78]. HEK293T cells were grown in 10-cm plates and transfected with 10 µg of the indicated proviral plasmid. Cell culture media was replaced 6-8 h post-transfection, and 4-thiouridine (4SU) was added to the cell culture medium at a concentration of 100 µM 16 h before harvesting cells and virions. 48 hours post-transfection, virions and cells were collected. Virions in the cell culture supernatant were filtered through 0.22 µm filters and pelleted through a 20% sucrose cushion at 28,000 rpm for 1.5 hr, 4°C. Transfected cells were harvested with a cell scraper. Both cells and purified virions were resuspended in PBS and protein-RNA complexes UV-crosslinked (365nm) before being lysed in 1X RIPA lysis buffer. Cell and virus lysates were then treated with RNase A (Thermo Fisher Scientific) and DNase I (Roche) at a final concentration of 20and 60 U/mL,

respectively. HA-tagged Gag proteins were immunoprecipitated using mouse monoclonal anti-HA antibody conjugated Protein G-coupled Dynabeads (Thermo Fisher Scientific), as previously described. Gag associated RNAs were radiolabeled with γ-^{32}P-ATP using T4 Polynucleotide Kinase (New England Biolabs). The protein-RNA adducts were separated via SDS-PAGE and transferred onto nitrocellulose membranes before being visualized with autoradiography film. Protein-RNA adducts were excised from nitrocellulose membranes and digested with Proteinase K, and RNA was extracted with phenol:chloroform:isoamyl alcohol (125:24:1) (Sigma-Aldrich). Resultant RNA was pelleted with the addition of 1μL Glycoblue (Ambion), 0.25 volume of 3M NaOAc pH 5.2, and 2.5 volumes of 1:1 ethanol:isopropanol. The RNA pellet was washed with 80% ethanol, dried and dissolved in 13 μL RNase-free H2O. RNA was ligated with barcoded adapters and ligation products were separated and purified on Novex 15% TBE-Urea gel (Invitrogen) as previously described with some exceptions [77]. Here we used barcoded 3' adapters and thus samples were pooled into groups of two prior to 5' adapter ligation. Furthermore, the amount of 3' adapter added to each reaction was reduced to 5pmole. An end-labeled 15-nucleotide RNA oligonucleotide was prepared in parallel as positive control for adapter ligations. The resulting ligation products were reverse transcribed using Superscript III First Strand Synthesis Kit (Invitrogen), and cDNA was amplified using Phusion Polymerase (New England Biolabs). PCR products were gel purified and sequenced on a NextSeq platform (1x75bp) at the Center for Genome Sciences at Washington University.

3' barcoded adapters	**Sequence**
CLIP1	5′adenylated/NNTGACTG TGG AAT TCT CGG GTG CCA AGG-3′dideoxyC
CLIP2	5′adenylated/NNACACTCTGG AAT TCT CGG GTG CCA AGG-3′dideoxyC
CLIP3	5′adenylated/NNACAGAG TGG AAT TCT CGG GTG CCA AGG-3′dideoxyC
CLIP4	5′adenylated/NNGCGATA TGG AAT TCT CGG GTG CCA AGG-3′dideoxyC

CLIP31	5′adenylated/NNTAGCGA TGG AAT TCT CGG GTG CCA AGG-3′dideoxyC
CLIP32	5′adenylated/NNCTGTAG TGG AAT TCT CGG GTG CCA AGG-3′dideoxyC
CLIP33	5′adenylated/NNTAGTCG TGG AAT TCT CGG GTG CCA AGG-3′dideoxyC
CLIP34	5′adenylated/NNAGTGTC TGG AAT TCT CGG GTG CCA AGG-3′dideoxyC
CLIP35	5′adenylated/NNATCGAC TGG AAT TCT CGG GTG CCA AGG-3′dideoxyC
CLIP36	5′adenylated/NNGACATG TGG AAT TCT CGG GTG CCA AGG-3′dideoxyC
CLIP37	5′adenylated/NNGACTAC TGG AAT TCT CGG GTG CCA AGG-3′dideoxyC
CLIP38	5′adenylated/NNATCTCG TGG AAT TCT CGG GTG CCA AGG-3′dideoxyC
5’ adapter	rGUU CAG AGU UCU ACA GUC CGA CGA UC
RT primer	GCCTTGGCACCCGAGAATTCCA
cDNA PCR (F)	AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGA
cDNA PCR (R)	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
Control oligo	rAUAGCUACGAUUGCA

Bioinformatic analysis of NGS data

Clip-seq:

Mapping:

Adapter trimming was performed using BBDuk from BBTools version 39.06. The 3′ adapter sequences were removed with parameters k=8 and mink=7 to define the k-mer size and minimum k-mer size. Additional parameters included ml=18 and maxlength=52 to filter reads based on length. The rcomp function was disabled to only consider kmers in the forward orientation.

Trimmed reads were further processed with BBDuk for separation based on their 3′ barcode sequences. Parameters included a k-mer size of 8, maskmiddle=f to ensure only exact matches were trimmed, and copyundefined was set to handle undefined bases. The resulting FASTQ files are the trimmed reads specific to each barcode.

Next, the trimmed reads were mapped to their corresponding Gag chimera viral genomes and WT viral genome (HIV-1_{NL4-3}/PR^{-}) with Bowtie version 1.3.1. Uniquely mapped reads were reported, allowing one mismatch (-v 1 -m 1). Mapped reads were processed with the mpileup function form SAMtools to generate pileups for HIV-1_{NL4-3}/PR^{-}, and these pileups were then analyzed using customized in-house scripts and visualized as viral density maps with GraphPad Prism (v6). Reads that failed to map to their corresponding viral genomes were extracted using SAMtools (-b -f 4) and then aligned to the hg19 human genome with Bowtie, allowing up to 10 multi-mapped reads with one mismatch. Various file formats of mapped reads were generated, including BED and SAM files, for further analysis.

Cluster Generation by PARalyzer and Cluster Annotations

PARalyzer was used for peak calling using SAM or BOWTIE outputs to generate clusters, which represent binding sites derived from overlapping mapped reads. Clusters were annotated using publicly available databases, ENSEMBL v72 and tRNASCANse [79].

For a cluster or a read that mapped to a location with multiple gene features (i.e., intron-exon or coding-sequence-3′ UTR junctions), we required greater than half the length of the read/cluster to map to the assigned class. In other cases where a cluster/read could be assigned to multiple gene features due to the presence of alternatively spliced isoforms in the ENSEMBL database, the following order of preference was chosen: coding sequence, 5′UTR, 3′UTR, intron. In some experiments, the number of reads associated with each cluster was taken as a measurement of how frequently a given cluster (i.e., binding site) was bound by Gag. To identify sequence motifs

within mRNAs, PARalyzer parameters were adjusted to generate clusters with a mean length of ~10 nucleotides surrounding the crosslinking site and which overlapped > 95% with the clusters generated above. Perl scripts were written for comparison of clusters in two separate data sets, determining the nucleotide composition of clusters and identifying features of mRNA molecules (i.e., 5′UTR, 3′UTR, CDS length) bound by Gag.

Correlation Analysis

Cross-correlation analysis was performed to measure the similarity of Gag binding frequencies between two datasets (Oppenheim & Schafer, 1975). The correlation function (CF(s)) compared the signals by shifting one dataset relative to the other and calculating their normalized correlation. The search window for correlation analysis was set to 400 nucleotides.

To assess the statistical significance of observed correlations, Monte Carlo simulations were used to generate 10,000 randomized datasets with the same structure as the original (i.e., identical numbers of binding sites and reads per site). Confidence intervals were calculated for p-values of 0.05 based on the cross-correlation results of these randomized samples.

Motif analysis

The bed files were annotated and reads in the mRNA regions were extracted using in-house scripts. These extracted reads were then converted to FASTA format using BEDTools for the subsequent analysis.

Motif analysis was conducted using MEME Suite with the following parameters: -nmotifs 10 -minw 5 -maxw 20 -dna. The parameters specified the identification of up to 10 motifs with a minimum width of 5 and a maximum width of 20 base pairs, applied to DNA sequences.

RNA-seq:

The STAR aligner was used to map the RNA-seq data to human genome indices (hg19) with a mismatch rate threshold of 4%. Mapped reads were than annotated with featureCounts.

All CLIP-seq and RNA-seq data were analyzed using publicly available programs (BBDuk toolkit, Bowtie, and SAMtools) and in-house scripts [20,77]. Briefly, reads that contained ambiguous nucleotides, did not contain the 3′ adapter sequence or were shorter than 10nt were discarded. The remaining reads were then separated based on their 3′ barcode sequences and collapsed to generate a set of unique sequences. Reads were then aligned to the human genome (hg19) or to corresponding viral genomes, with up to 1 mismatch allowed using the Bowtie algorithm [80]. In the case of reads that mapped to the human genome, locations were only reported for those with the minimum number of observed mismatches for each read (Bowtie criteria: -v 1 -m 10 for mapping to hg19, and - v 1 -m 1 for mapping to the viral genomes). SAMtools was used to generate pileups of mapped reads, which were subsequently processed using in-house scripts and visualized by Prism GraphPad software (v6) [81].

ACKNOWLEDGEMENTS

We thank Dr. Wandy Beatty and the Molecular Microbiology Imaging Facility at the Washington University School of Medicine for their assistance with confocal microscopy and transmission electron microscopy experiments. H.R.V. was supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1745038, DGE-2139839 and the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Number T32AI007172. This work was supported by NIH grants U54 AI170660 (A.T., S.B.K., N.S.), R01 AI179691 (S.B.K., N.S.) and R56 AI110221 (N.S.).

CLIP-seq and RNA-seq data were deposited to the Gene Expression Omnibus Database with accession numbers GSE281422, GSE281423, and GSE281424.

REFERENCES

913 1. Chen, J., Nikolaitchik, O., Singh, J., Wright, A., Bencsics, C.E., Coffin, J.M., Ni, N., Lockett,
914 S., Pathak, V.K., and Hu, W.S. (2009). High efficiency of HIV-1 genomic RNA packaging
915 and heterozygote formation revealed by single virion analysis. Proc Natl Acad Sci U S A
916 *106*, 13535-13540. 10.1073/pnas.0906822106.
917 2. Moore, M.D., and Hu, W.S. (2009). HIV-1 RNA dimerization: It takes two to tango. AIDS
918 Rev *11*, 91-102.
919 3. Duchon, A., and Hu, W.S. (2024). HIV-1 RNA genome packaging: it's G-rated. mBio *15*,
920 e0086123. 10.1128/mbio.00861-23.
921 4. Hu, W.S., and Temin, H.M. (1990). Genetic consequences of packaging two RNA
922 genomes in one retroviral particle: pseudodiploidy and high rate of genetic
923 recombination. Proc Natl Acad Sci U S A *87*, 1556-1560. 10.1073/pnas.87.4.1556.
924 5. Rawson, J.M.O., Nikolaitchik, O.A., Keele, B.F., Pathak, V.K., and Hu, W.S. (2018).
925 Recombination is required for efficient HIV-1 replication and the maintenance of viral
926 genome integrity. Nucleic Acids Res *46*, 10535-10545. 10.1093/nar/gky910.
927 6. Streeck, H., Li, B., Poon, A.F., Schneidewind, A., Gladden, A.D., Power, K.A., Daskalakis,
928 D., Bazner, S., Zuniga, R., Brander, C., et al. (2008). Immune-driven recombination and
929 loss of control after HIV superinfection. J Exp Med *205*, 1789-1796.
930 10.1084/jem.20080281.
931 7. Kellam, P., and Larder, B.A. (1995). Retroviral recombination can lead to linkage of
932 reverse transcriptase mutations that confer increased zidovudine resistance. J Virol *69*,
933 669-674. 10.1128/jvi.69.2.669-674.1995.
934 8. Vijaykrishna, D., Mukerji, R., and Smith, G.J. (2015). RNA Virus Reassortment: An
935 Evolutionary Mechanism for Host Jumps and Immune Evasion. PLoS Pathog *11*,
936 e1004902. 10.1371/journal.ppat.1004902.
937 9. McDonald, S.M., Matthijnssens, J., McAllen, J.K., Hine, E., Overton, L., Wang, S., Lemey,
938 P., Zeller, M., Van Ranst, M., Spiro, D.J., and Patton, J.T. (2009). Evolutionary dynamics
939 of human rotaviruses: balancing reassortment with preferred genome constellations.
940 PLoS Pathog *5*, e1000634. 10.1371/journal.ppat.1000634.
941 10. Nelson, M.I., Viboud, C., Simonsen, L., Bennett, R.T., Griesemer, S.B., St George, K.,
942 Taylor, J., Spiro, D.J., Sengamalay, N.A., Ghedin, E., et al. (2008). Multiple reassortment
943 events in the evolutionary history of H1N1 influenza A virus since 1918. PLoS Pathog *4*,
944 e1000012. 10.1371/journal.ppat.1000012.
945 11. Ghedin, E., Fitch, A., Boyne, A., Griesemer, S., DePasse, J., Bera, J., Zhang, X., Halpin,
946 R.A., Smit, M., Jennings, L., et al. (2009). Mixed infection and the genesis of influenza
947 virus diversity. J Virol *83*, 8832-8841. 10.1128/jvi.00773-09.
948 12. Komoto, S., Tacharoenmuang, R., Guntapong, R., Ide, T., Tsuji, T., Yoshikawa, T.,
949 Tharmaphornpilas, P., Sangkitporn, S., and Taniguchi, K. (2016). Reassortment of Human
950 and Animal Rotavirus Gene Segments in Emerging DS-1-Like G1P[8] Rotavirus Strains.
951 PLOS ONE *11*, e0148416. 10.1371/journal.pone.0148416.
952 13. Aldovini, A., and Young, R.A. (1990). Mutations of RNA and protein sequences involved
953 in human immunodeficiency virus type 1 packaging result in production of noninfectious
954 virus. J Virol *64*, 1920-1926. 10.1128/jvi.64.5.1920-1926.1990.

955 14. Gorelick, R.J., Nigida, S.M., Jr., Bess, J.W., Jr., Arthur, L.O., Henderson, L.E., and Rein, A.
956 (1990). Noninfectious human immunodeficiency virus type 1 mutants deficient in
957 genomic RNA. J Virol *64*, 3207-3211. 10.1128/jvi.64.7.3207-3211.1990.
958 15. Kuzembayeva, M., Dilley, K., Sardo, L., and Hu, W.S. (2014). Life of psi: how full-length
959 HIV-1 RNAs become packaged genomes in the viral particles. Virology *454-455*, 362-370.
960 10.1016/j.virol.2014.01.019.
961 16. Rein, A. (2019). RNA Packaging in HIV. Trends Microbiol *27*, 715-723.
962 10.1016/j.tim.2019.04.003.
963 17. Nikolaitchik, O.A., Somoulay, X., Rawson, J.M.O., Yoo, J.A., Pathak, V.K., and Hu, W.S.
964 (2020). Unpaired Guanosines in the 5' Untranslated Region of HIV-1 RNA Act
965 Synergistically To Mediate Genome Packaging. J Virol *94*. 10.1128/jvi.00439-20.
966 18. De Guzman, R.N., Wu, Z.R., Stalling, C.C., Pappalardo, L., Borer, P.N., and Summers, M.F.
967 (1998). Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA
968 recognition element. Science *279*, 384-388. 10.1126/science.279.5349.384.
969 19. Ding, P., Kharytonchyk, S., Waller, A., Mbaekwe, U., Basappa, S., Kuo, N., Frank, H.M.,
970 Quasney, C., Kidane, A., Swanson, C., et al. (2020). Identification of the initial
971 nucleocapsid recognition element in the HIV-1 RNA packaging signal. Proc Natl Acad Sci
972 U S A *117*, 17737-17746. 10.1073/pnas.2008519117.
973 20. Kutluay, S.B., Zang, T., Blanco-Melo, D., Powell, C., Jannain, D., Errando, M., and
974 Bieniasz, P.D. (2014). Global changes in the RNA binding specificity of HIV-1 gag regulate
975 virion genesis. Cell *159*, 1096-1109. 10.1016/j.cell.2014.09.057.
976 21. Clever, J.L., and Parslow, T.G. (1997). Mutant human immunodeficiency virus type 1
977 genomes with defects in RNA dimerization or encapsidation. J Virol *71*, 3407-3414.
978 10.1128/JVI.71.5.3407-3414.1997.
979 22. Laham-Karam, N., and Bacharach, E. (2007). Transduction of human immunodeficiency
980 virus type 1 vectors lacking encapsidation and dimerization signals. J Virol *81*, 10687-
981 10698. 10.1128/JVI.00653-07.
982 23. McBride, M.S., and Panganiban, A.T. (1997). Position dependence of functional hairpins
983 important for human immunodeficiency virus type 1 RNA encapsidation in vivo. J Virol
984 *71*, 2050-2058. 10.1128/JVI.71.3.2050-2058.1997.
985 24. Comas-Garcia, M., Datta, S.A., Baker, L., Varma, R., Gudla, P.R., and Rein, A. (2017).
986 Dissection of specific binding of HIV-1 Gag to the 'packaging signal' in viral RNA. Elife *6*.
987 10.7554/eLife.27055.
988 25. Rulli, S.J., Hibbert, C.S., Mirro, J., Pederson, T., Biswal, S., and Rein, A. (2007). Selective
989 and Nonselective Packaging of Cellular RNAs in Retrovirus Particles. Journal of Virology
990 *81*, 6623-6631. doi:10.1128/jvi.02833-06.
991 26. Muriaux, D., Mirro, J., Harvin, D., and Rein, A. (2001). RNA is a structural element in
992 retrovirus particles. Proc Natl Acad Sci U S A *98*, 5246-5251. 10.1073/pnas.091000398.
993 27. Kutluay, S.B., and Bieniasz, P.D. (2010). Analysis of the initiating events in HIV-1 particle
994 assembly and genome packaging. PLoS Pathog *6*, e1001200.
995 10.1371/journal.ppat.1001200.
996 28. Jouvenet, N., Simon, S.M., and Bieniasz, P.D. (2009). Imaging the interaction of HIV-1
997 genomes and Gag during assembly of individual viral particles. Proc Natl Acad Sci U S A
998 *106*, 19114-19119. 10.1073/pnas.0907364106.

999 29. Sardo, L., Hatch, S.C., Chen, J., Nikolaitchik, O., Burdick, R.C., Chen, D., Westlake, C.J.,
1000 Lockett, S., Pathak, V.K., and Hu, W.S. (2015). Dynamics of HIV-1 RNA Near the Plasma
1001 Membrane during Virus Assembly. J Virol *89*, 10832-10840. 10.1128/jvi.01146-15.
1002 30. Chen, J., Rahman, S.A., Nikolaitchik, O.A., Grunwald, D., Sardo, L., Burdick, R.C., Plisov,
1003 S., Liang, E., Tai, S., Pathak, V.K., and Hu, W.S. (2016). HIV-1 RNA genome dimerizes on
1004 the plasma membrane in the presence of Gag protein. Proc Natl Acad Sci U S A *113*,
1005 E201-208. 10.1073/pnas.1518572113.
1006 31. Freed, E.O. (2015). HIV-1 assembly, release and maturation. Nat Rev Microbiol *13*, 484-
1007 496. 10.1038/nrmicro3490.
1008 32. Ivanchenko, S., Godinez, W.J., Lampe, M., Kräusslich, H.-G., Eils, R., Rohr, K., Bräuchle,
1009 C., Müller, B., and Lamb, D.C. (2009). Dynamics of HIV-1 Assembly and Release. PLOS
1010 Pathogens *5*, e1000652. 10.1371/journal.ppat.1000652.
1011 33. Kutluay, S.B., and Bieniasz, P.D. (2010). Analysis of the Initiating Events in HIV-1 Particle
1012 Assembly and Genome Packaging. PLOS Pathogens *6*, e1001200.
1013 10.1371/journal.ppat.1001200.
1014 34. Muriaux, D., Mirro, J., Harvin, D., and Rein, A. (2001). RNA is a structural element in
1015 retrovirus particles. Proceedings of the National Academy of Sciences *98*, 5246-5251.
1016 doi:10.1073/pnas.091000398.
1017 35. Duchon, A., Santos, S., Chen, J., Brown, M., Nikolaitchik, O.A., Tai, S., Chao, J.A., Freed,
1018 E.O., Pathak, V.K., and Hu, W.-S. (2021). Plasma Membrane Anchoring and Gag:Gag
1019 Multimerization on Viral RNA Are Critical Properties of HIV-1 Gag Required To Mediate
1020 Efficient Genome Packaging. mBio *12*, e03254-03221. doi:10.1128/mbio.03254-21.
1021 36. Lei, X., Gonçalves-Carneiro, D., Zang, T.M., and Bieniasz, P.D. (2023). Initiation of HIV-1
1022 Gag lattice assembly is required for recognition of the viral genome packaging signal.
1023 eLife *12*, e83548. 10.7554/eLife.83548.
1024 37. Kypr, J., and Mrazek, J. (1987). Unusual codon usage of HIV. Nature *327*, 20.
1025 10.1038/327020a0.
1026 38. van der Kuyl, A.C., and Berkhout, B. (2012). The biased nucleotide composition of the
1027 HIV genome: a constant factor in a highly variable virus. Retrovirology *9*, 92.
1028 10.1186/1742-4690-9-92.
1029 39. Accola, M.A., Strack, B., and Göttlinger, H.G. (2000). Efficient particle production by
1030 minimal Gag constructs which retain the carboxy-terminal domain of human
1031 immunodeficiency virus type 1 capsid-p2 and a late assembly domain. J Virol *74*, 5395-
1032 5402. 10.1128/jvi.74.12.5395-5402.2000.
1033 40. Han, Siew P., Tang, Yue H., and Smith, R. (2010). Functional diversity of the hnRNPs:
1034 past, present and perspectives. Biochemical Journal *430*, 379-392. 10.1042/bj20100396.
1035 41. Jeong, S. (2017). SR Proteins: Binders, Regulators, and Connectors of RNA. Mol Cells *40*,
1036 1-9. 10.14348/molcells.2017.2319.
1037 42. Sertznig, H., Hillebrand, F., Erkelenz, S., Schaal, H., and Widera, M. (2018). Behind the
1038 scenes of HIV-1 replication: Alternative splicing as the dependency factor on the quiet.
1039 Virology *516*, 176-188. 10.1016/j.virol.2018.01.011.
1040 43. Huelga, S.C., Vu, A.Q., Arnold, J.D., Liang, T.Y., Liu, P.P., Yan, B.Y., Donohue, J.P., Shiue,
1041 L., Hoon, S., Brenner, S., et al. (2012). Integrative genome-wide analysis reveals

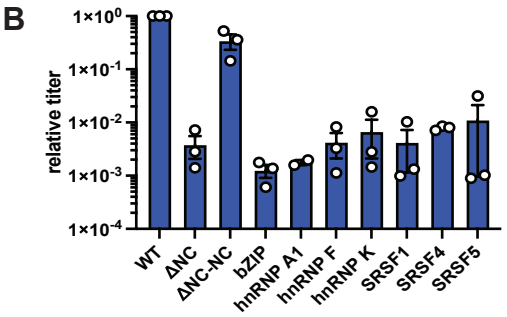
1042 cooperative regulation of alternative splicing by hnRNP proteins. Cell Rep *1*, 167-178.
1043 10.1016/j.celrep.2012.02.001.
1044 44. Pandit, S., Zhou, Y., Shiue, L., Coutinho-Mansfield, G., Li, H., Qiu, J., Huang, J., Yeo, G.W.,
1045 Ares, M., Jr., and Fu, X.D. (2013). Genome-wide analysis reveals SR protein cooperation
1046 and competition in regulated splicing. Mol Cell *50*, 223-235.
1047 10.1016/j.molcel.2013.03.001.
1048 45. Kutluay, S.B., Emery, A., Penumutchu, S.R., Townsend, D., Tenneti, K., Madison, M.K.,
1049 Stukenbroeker, A.M., Powell, C., Jannain, D., Tolbert, B.S., et al. (2019). Genome-Wide
1050 Analysis of Heterogeneous Nuclear Ribonucleoprotein (hnRNP) Binding to HIV-1 RNA
1051 Reveals a Key Role for hnRNP H1 in Alternative Viral mRNA Splicing. J Virol *93*.
1052 10.1128/jvi.01048-19.
1053 46. Derdowski, A., Ding, L., and Spearman, P. (2004). A novel fluorescence resonance energy
1054 transfer assay demonstrates that the human immunodeficiency virus type 1 Pr55Gag I
1055 domain mediates Gag-Gag interactions. J Virol *78*, 1230-1242. 10.1128/jvi.78.3.1230-
1056 1242.2004.
1057 47. Levin, J.G., Mitra, M., Mascarenhas, A., and Musier-Forsyth, K. (2010). Role of HIV-1
1058 nucleocapsid protein in HIV-1 reverse transcription. RNA Biol *7*, 754-774.
1059 10.4161/rna.7.6.14115.
1060 48. Darlix, J.L., Gabus, C., Nugeyre, M.T., Clavel, F., and Barré-Sinoussi, F. (1990). Cis
1061 elements and trans-acting factors involved in the RNA dimerization of the human
1062 immunodeficiency virus HIV-1. J Mol Biol *216*, 689-699. 10.1016/0022-2836(90)90392-y.
1063 49. Gao, K., Gorelick, R.J., Johnson, D.G., and Bushman, F. (2003). Cofactors for human
1064 immunodeficiency virus type 1 cDNA integration in vitro. J Virol *77*, 1598-1603.
1065 10.1128/jvi.77.2.1598-1603.2003.
1066 50. Onafuwa-Nuga, A.A., Telesnitsky, A., and King, S.R. (2006). 7SL RNA, but not the 54-kd
1067 signal recognition particle protein, is an abundant component of both infectious HIV-1
1068 and minimal virus-like particles. Rna *12*, 542-546. 10.1261/rna.2306306.
1069 51. Itano, M.S., Arnion, H., Wolin, S.L., and Simon, S.M. (2018). Recruitment of 7SL RNA to
1070 assembling HIV-1 virus-like particles. Traffic *19*, 36-43. 10.1111/tra.12536.
1071 52. Eckwahl, M.J., Arnion, H., Kharytonchyk, S., Zang, T., Bieniasz, P.D., Telesnitsky, A., and
1072 Wolin, S.L. (2016). Analysis of the human immunodeficiency virus-1 RNA packageome.
1073 Rna *22*, 1228-1238. 10.1261/rna.057299.116.
1074 53. Becker, J.T., and Sherer, N.M. (2017). Subcellular Localization of HIV-1 gag-pol mRNAs
1075 Regulates Sites of Virion Assembly. J Virol *91*. 10.1128/jvi.02315-16.
1076 54. Guerrero, J.F., Lesko, S.L., Evans, E.L., and Sherer, N.M. (2024). Studying Retroviral Life
1077 Cycles Using Visible Viruses and Live Cell Imaging. Annu Rev Virol *11*, 125-146.
1078 10.1146/annurev-virology-100422-012608.
1079 55. GC, K., Lesko, S., Emery, A., Burnett, C., Gopal, K., Clark, S., Swanstrom, R., Sherer, N.M.,
1080 Telesnitsky, A., and Kharytonchyk, S. (2025). HIV-1 single-transcription start-site mutants
1081 display complementary replication functions that are restored by reversion. Journal of
1082 Virology *0*, e02139-02124. doi:10.1128/jvi.02139-24.
1083 56. Sundquist, W.I., and Kräusslich, H.G. (2012). HIV-1 assembly, budding, and maturation.
1084 Cold Spring Harb Perspect Med *2*, a006924. 10.1101/cshperspect.a006924.

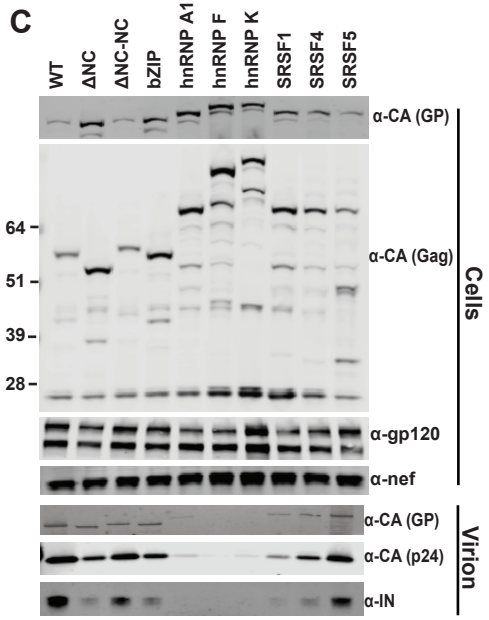
1085 57. Lee, I.H., Kai, H., Carlson, L.A., Groves, J.T., and Hurley, J.H. (2015). Negative membrane
1086 curvature catalyzes nucleation of endosomal sorting complex required for transport
1087 (ESCRT)-III assembly. Proc Natl Acad Sci U S A *112*, 15892-15897.
1088 10.1073/pnas.1518765113.
1089 58. Van Dusen, C.M., Yee, L., McNally, L.M., and McNally, M.T. (2010). A glycine-rich domain
1090 of hnRNP H/F promotes nucleocytoplasmic shuttling and nuclear import through an
1091 interaction with transportin 1. Mol Cell Biol *30*, 2552-2562. 10.1128/mcb.00230-09.
1092 59. Sanford, J.R., Wang, X., Mort, M., Vanduyn, N., Cooper, D.N., Mooney, S.D., Edenberg,
1093 H.J., and Liu, Y. (2009). Splicing factor SFRS1 recognizes a functionally diverse landscape
1094 of RNA transcripts. Genome Res *19*, 381-394. 10.1101/gr.082503.108.
1095 60. Änkö, M.L., Müller-McNicoll, M., Brandl, H., Curk, T., Gorup, C., Henry, I., Ule, J., and
1096 Neugebauer, K.M. (2012). The RNA-binding landscapes of two SR proteins reveal unique
1097 functions and binding to diverse RNA classes. Genome Biol *13*, R17. 10.1186/gb-2012-
1098 13-3-r17.
1099 61. Li, J., Li, G., Qi, Y., Lu, Y., Wang, H., Shi, K., Li, D., Shi, J., Stovall, D.B., and Sui, G. (2021).
1100 SRSF5 regulates alternative splicing of DMTF1 pre-mRNA through modulating SF1
1101 binding. RNA Biol *18*, 318-336. 10.1080/15476286.2021.1947644.
1102 62. Buratti, E., Muro, A.F., Giombi, M., Gherbassi, D., Iaconcig, A., and Baralle, F.E. (2004).
1103 RNA folding affects the recruitment of SR proteins by mouse and human polypurinic
1104 enhancer elements in the fibronectin EDA exon. Mol Cell Biol *24*, 1387-1400.
1105 10.1128/mcb.24.3.1387-1400.2004.
1106 63. Cléry, A., Sinha, R., Anczuków, O., Corrionero, A., Moursy, A., Daubner, G.M., Valcárcel,
1107 J., Krainer, A.R., and Allain, F.H.-T. (2013). Isolated pseudo–RNA-recognition motifs of SR
1108 proteins can regulate splicing using a noncanonical mode of RNA recognition.
1109 Proceedings of the National Academy of Sciences *110*, E2802-E2811.
1110 doi:10.1073/pnas.1303445110.
1111 64. Keane, S.C., Heng, X., Lu, K., Kharytonchyk, S., Ramakrishnan, V., Carter, G., Barton, S.,
1112 Hosic, A., Florwick, A., Santos, J., et al. (2015). Structure of the HIV-1 RNA packaging
1113 signal. Science *348*, 917-921. 10.1126/science.aaa9266.
1114 65. Wilkinson, K.A., Gorelick, R.J., Vasa, S.M., Guex, N., Rein, A., Mathews, D.H., Giddings,
1115 M.C., and Weeks, K.M. (2008). High-Throughput SHAPE Analysis Reveals Structures in
1116 HIV-1 Genomic RNA Strongly Conserved across Distinct Biological States. PLOS Biology *6*,
1117 e96. 10.1371/journal.pbio.0060096.
1118 66. Abd El-Wahab, E.W., Smyth, R.P., Mailler, E., Bernacchi, S., Vivet-Boudou, V., Hijnen, M.,
1119 Jossinet, F., Mak, J., Paillart, J.-C., and Marquet, R. (2014). Specific recognition of the
1120 HIV-1 genomic RNA by the Gag precursor. Nature Communications *5*, 4304.
1121 10.1038/ncomms5304.
1122 67. van Hemert, F.J., van der Kuyl, A.C., and Berkhout, B. (2013). The A-nucleotide
1123 preference of HIV-1 in the context of its structured RNA genome. RNA Biol *10*, 211-215.
1124 10.4161/rna.22896.
1125 68. Sette, P., O'Connor, S.K., Yerramilli, V.S., Dussupt, V., Nagashima, K., Chutiraka, K.,
1126 Lingappa, J., Scarlata, S., and Bouamr, F. (2016). HIV-1 Nucleocapsid Mimics the
1127 Membrane Adaptor Syntenin PDZ to Gain Access to ESCRTs and Promote Virus Budding.
1128 Cell Host Microbe *19*, 336-348. 10.1016/j.chom.2016.02.004.

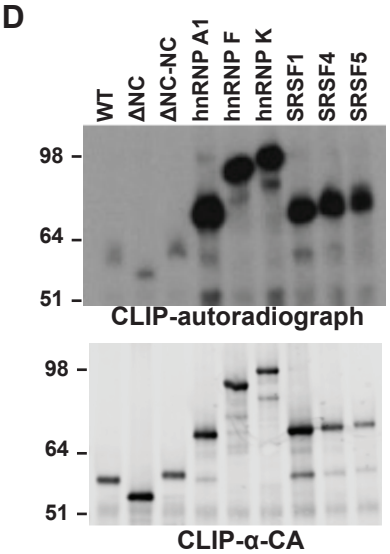
1129 69. Gaunt, E.R., and Digard, P. (2022). Compositional biases in RNA viruses: Causes,
1130 consequences and applications. Wiley Interdiscip Rev RNA *13*, e1679.
1131 10.1002/wrna.1679.
1132 70. Gc, K., Lesko, S., Emery, A., Burnett, C., Gopal, K., Clark, S., Swanstrom, R., Sherer, N.M.,
1133 Telesnitsky, A., and Kharytonchyk, S. (2024). HIV-1 single transcription start site mutants
1134 display complementary replication functions that are restored by reversion. bioRxiv.
1135 10.1101/2024.12.04.626847.
1136 71. Kane, M., Zang, T.M., Rihn, S.J., Zhang, F., Kueck, T., Alim, M., Schoggins, J., Rice, C.M.,
1137 Wilson, S.J., and Bieniasz, P.D. (2016). Identification of Interferon-Stimulated Genes with
1138 Antiretroviral Activity. Cell Host Microbe *20*, 392-405. 10.1016/j.chom.2016.08.005.
1139 72. Mohammed, K.D., Topper, M.B., and Muesing, M.A. (2011). Sequential deletion of the
1140 integrase (Gag-Pol) carboxyl terminus reveals distinct phenotypic classes of defective
1141 HIV-1. J Virol *85*, 4654-4666. 10.1128/jvi.02374-10.
1142 73. Shema Mugisha, C., Dinh, T., Kumar, A., Tenneti, K., Eschbach, J.E., Davis, K., Gifford, R.,
1143 Kvaratskhelia, M., and Kutluay, S.B. (2022). Emergence of Compensatory Mutations
1144 Reveals the Importance of Electrostatic Interactions between HIV-1 Integrase and
1145 Genomic RNA. mBio *13*, e0043122. 10.1128/mbio.00431-22.
1146 74. Madison, M.K., Lawson, D.Q., Elliott, J., Ozanturk, A.N., Koneru, P.C., Townsend, D.,
1147 Errando, M., Kvaratskhelia, M., and Kutluay, S.B. (2017). Allosteric HIV-1 Integrase
1148 Inhibitors Lead to Premature Degradation of the Viral RNA Genome and Integrase in
1149 Target Cells. J Virol *91*. 10.1128/JVI.00821-17.
1150 75. Pizzato, M., Erlwein, O., Bonsall, D., Kaye, S., Muir, D., and McClure, M.O. (2009). A one-
1151 step SYBR Green I-based product-enhanced reverse transcriptase assay for the
1152 quantitation of retroviruses in cell culture supernatants. J Virol Methods *156*, 1-7.
1153 10.1016/j.jviromet.2008.10.012.
1154 76. Mugisha, C.S., Dinh, T., Kumar, A., Tenneti, K., Eschbach, J.E., Davis, K., Gifford, R.,
1155 Kvaratskhelia, M., and Kutluay, S.B. (2022). Emergence of Compensatory Mutations
1156 Reveals the Importance of Electrostatic Interactions between HIV-1 Integrase and
1157 Genomic RNA. mBio *13*, e00431-00422. doi:10.1128/mbio.00431-22.
1158 77. Shema Mugisha, C., Tenneti, K., and Kutluay, S.B. (2020). Clip for studying protein-RNA
1159 interactions that regulate virus replication. Methods *183*, 84-92.
1160 https://doi.org/10.1016/j.ymeth.2019.11.011.
1161 78. Kutluay, S.B., and Bieniasz, P.D. (2016). Analysis of HIV-1 Gag-RNA Interactions in Cells
1162 and Virions by CLIP-seq. Methods Mol Biol *1354*, 119-131. 10.1007/978-1-4939-3046-
1163 3_8.
1164 79. Chan, Patricia P., Lin, Brian Y., Mak, Allysia J., and Lowe, Todd M. (2021). tRNAscan-SE
1165 2.0: improved detection and functional classification of transfer RNA genes. Nucleic
1166 Acids Research *49*, 9077-9096. 10.1093/nar/gkab688.
1167 80. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-
1168 efficient alignment of short DNA sequences to the human genome. Genome Biol *10*,
1169 R25. 10.1186/gb-2009-10-3-r25.
1170 81. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A.,
1171 Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and
1172 BCFtools. GigaScience *10*. 10.1093/gigascience/giab008.

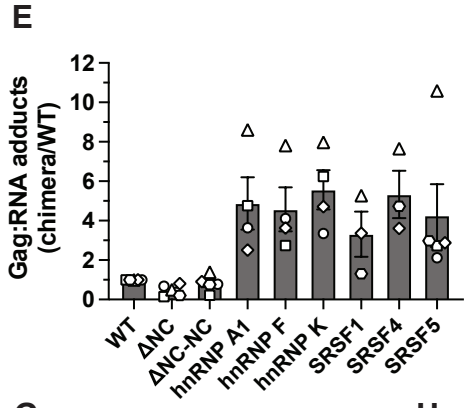
1173

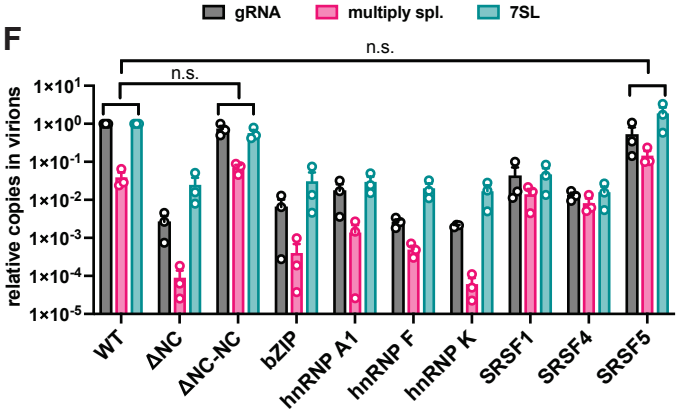
A
WT
ΔNC
ΔNC-NC
bZIP
hnRNP A1
hnRNP F
hnRNP K
SRSF1
SRSF4
SRSF5
MA
CA
NC
NotI

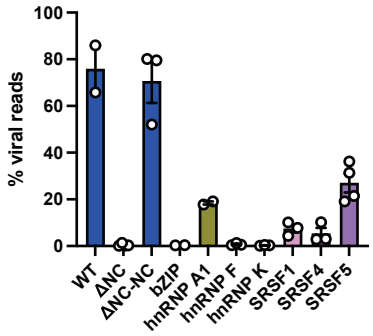

B
relative titer
1×10^0
1×10^-1
1×10^-2
1×10^-3
1×10^-4
WT
ΔNC
ΔNC-NC
bZIP
hnRNP A1
hnRNP F
hnRNP K
SRSF1
SRSF4
SRSF5

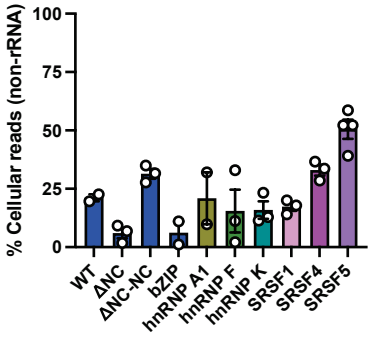

C
WT
ΔNC
ΔNC-NC
bZIP
hnRNP A1
hnRNP F
hnRNP K
SRSF1
SRSF4
SRSF5
α-CA (GP)
α-CA (Gag)
64
51
39
28
α-gp120
α-nef
Cells
α-CA (GP)
α-CA (p24)
α-IN
Virion

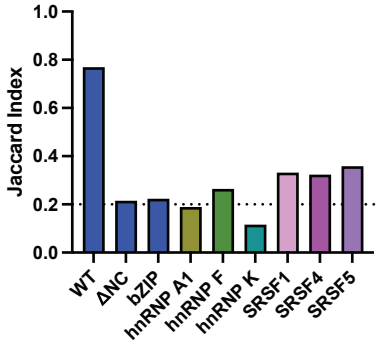

D
WT
ΔNC
ΔNC-NC
hnRNP A1
hnRNP F
hnRNP K
SRSF1
SRSF4
SRSF5
98
64
51
CLIP-autoradiograph
98
64
51
CLIP-α-CA

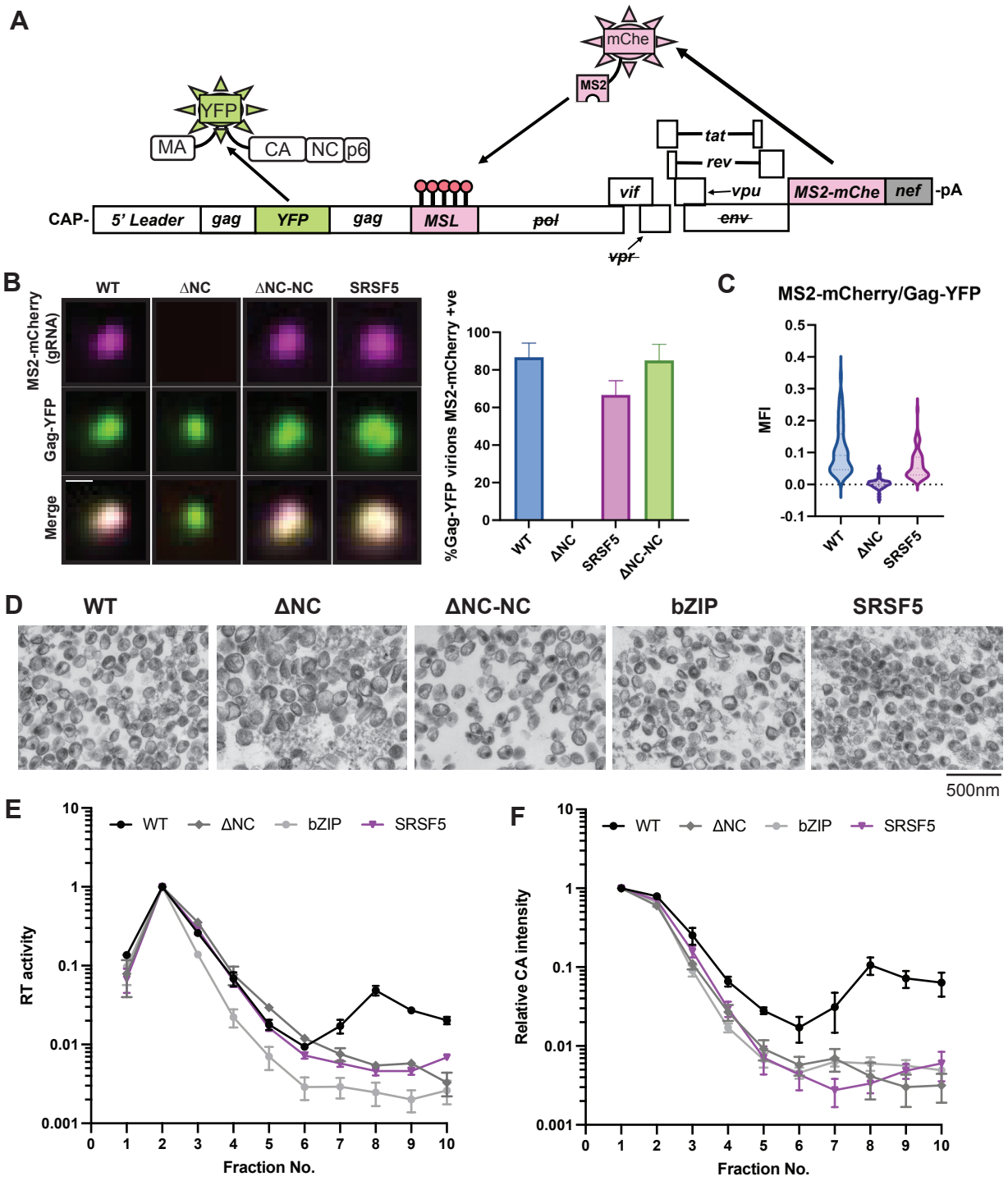

E
Gag:RNA adducts (chimera/WT)
12
10
8
6
4
2
0
WT
ΔNC
ΔNC-NC
hnRNP A1
hnRNP F
hnRNP K
SRSF1
SRSF4
SRSF5

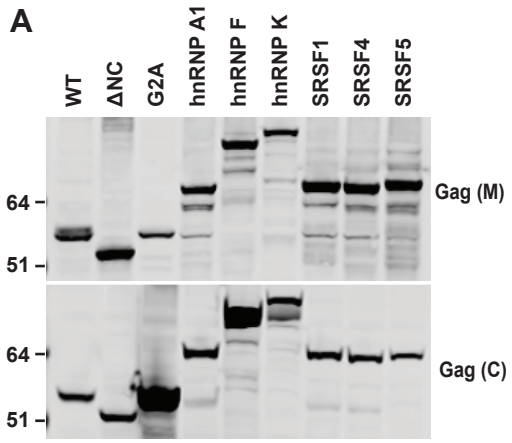

gRNA
multiply spl.
7SL
F
n.s.
n.s.
relative copies in virions
1×10^1
1×10^0
1×10^-1
1×10^-2
1×10^-3
1×10^-4
1×10^-5
WT
ΔNC
ΔNC-NC
bZIP
hnRNP A1
hnRNP F
hnRNP K
SRSF1
SRSF4
SRSF5

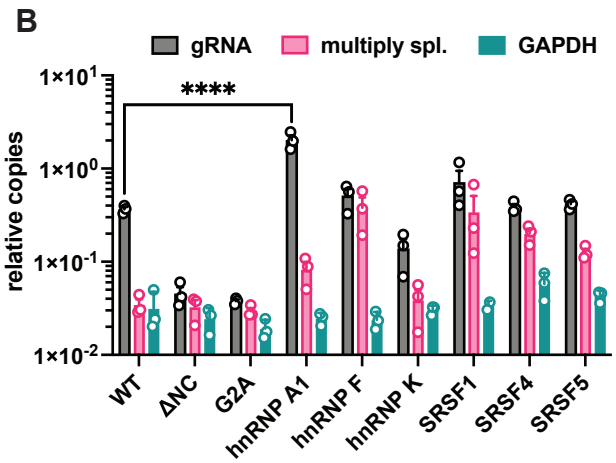

G
% viral reads
100
80
60
40
20
0
WT
ΔNC
ΔNC-NC
bZIP
hnRNP A1
hnRNP F
hnRNP K
SRSF1
SRSF4
SRSF5

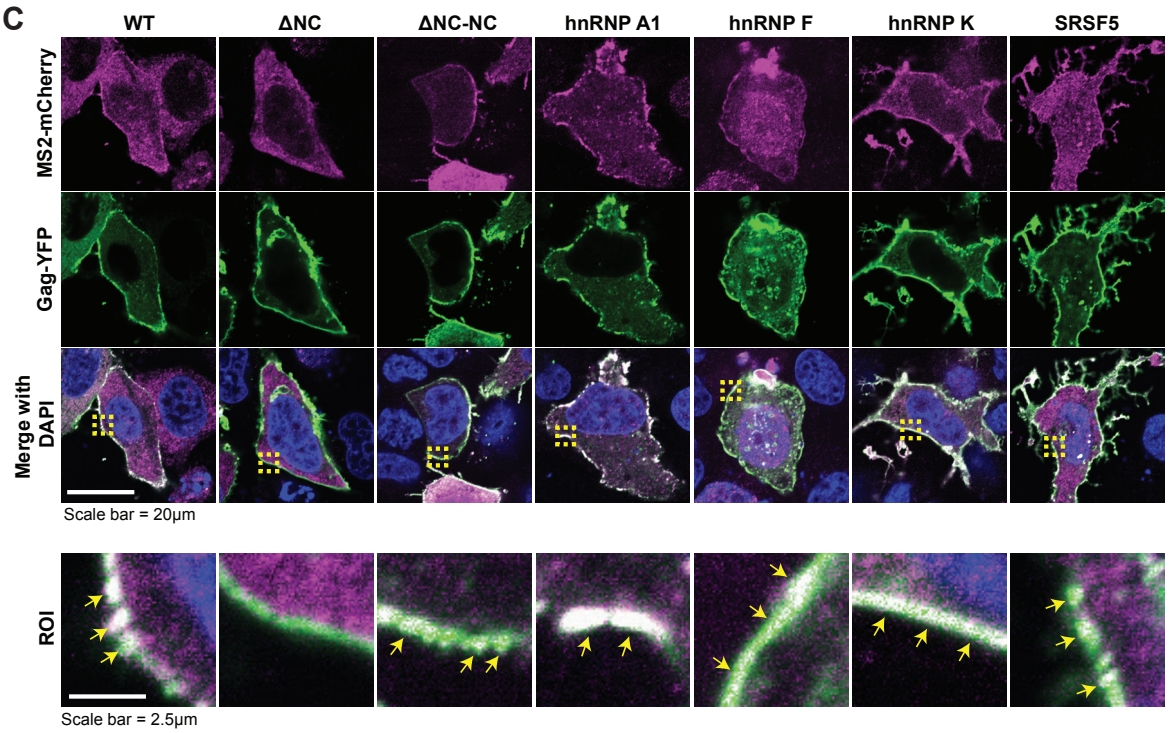

H
% Cellular reads (non-rRNA)
100
75
50
25
0
WT
ΔNC
ΔNC-NC
bZIP
hnRNP A1
hnRNP F
hnRNP K
SRSF1
SRSF4
SRSF5

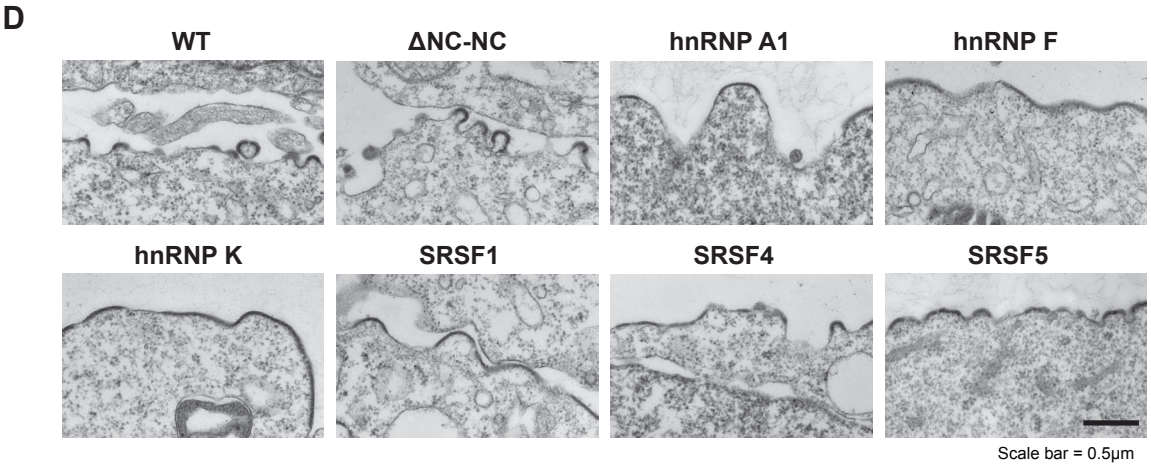

I
Jaccard Index
1.0
0.8
0.6
0.4
0.2
0.0
WT
ΔNC
bZIP
hnRNP A1
hnRNP F
hnRNP K
SRSF1
SRSF4
SRSF5

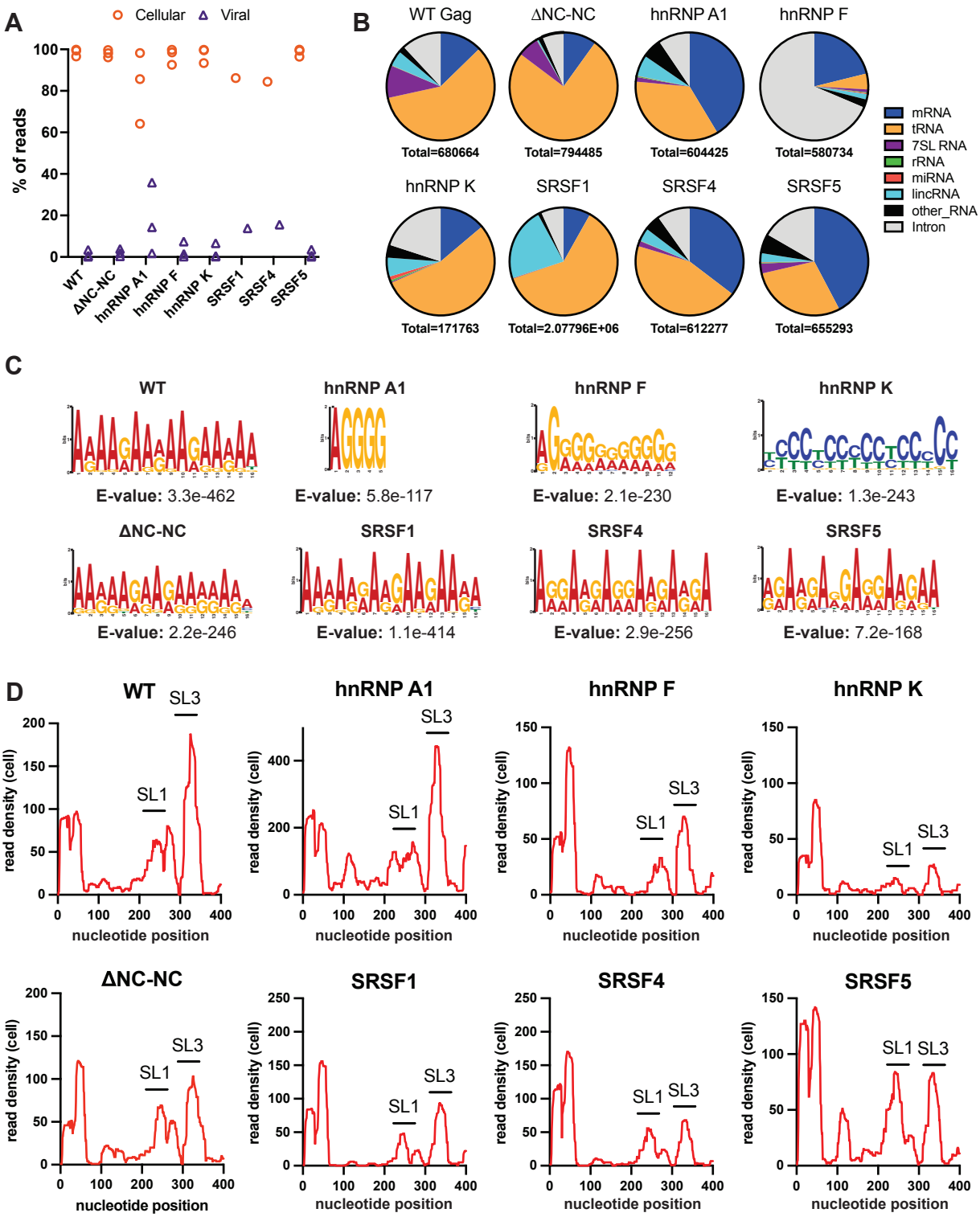

A
mChe
MS2
YFP
MA
CA
NC
p6
tat
rev
vpu
vif
MS2-mChe
nef
-pA
CAP-
5' Leader
gag
YFP
gag
MSL
pol
env
vpr
B
WT
ΔNC
ΔNC-NC
SRSF5
MS2-mCherry (gRNA)
Gag-YFP
Merge
%Gag-YFP virions MS2-mCherry +ve
100
80
60
40
20
0
WT
ΔNC
SRSF5
ΔNC-NC
C
MS2-mCherry/Gag-YFP
MFI
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
WT
ΔNC
SRSF5
D
WT
ΔNC
ΔNC-NC
bZIP
SRSF5
500nm
E
WT
ΔNC
bZIP
SRSF5
RT activity
10
1
0.1
0.01
0.001
0
1
2
3
4
5
6
7
8
9
10
Fraction No.
F
WT
ΔNC
bZIP
SRSF5
Relative CA intensity
10
1
0.1
0.01
0.001
0
1
2
3
4
5
6
7
8
9
10
Fraction No.

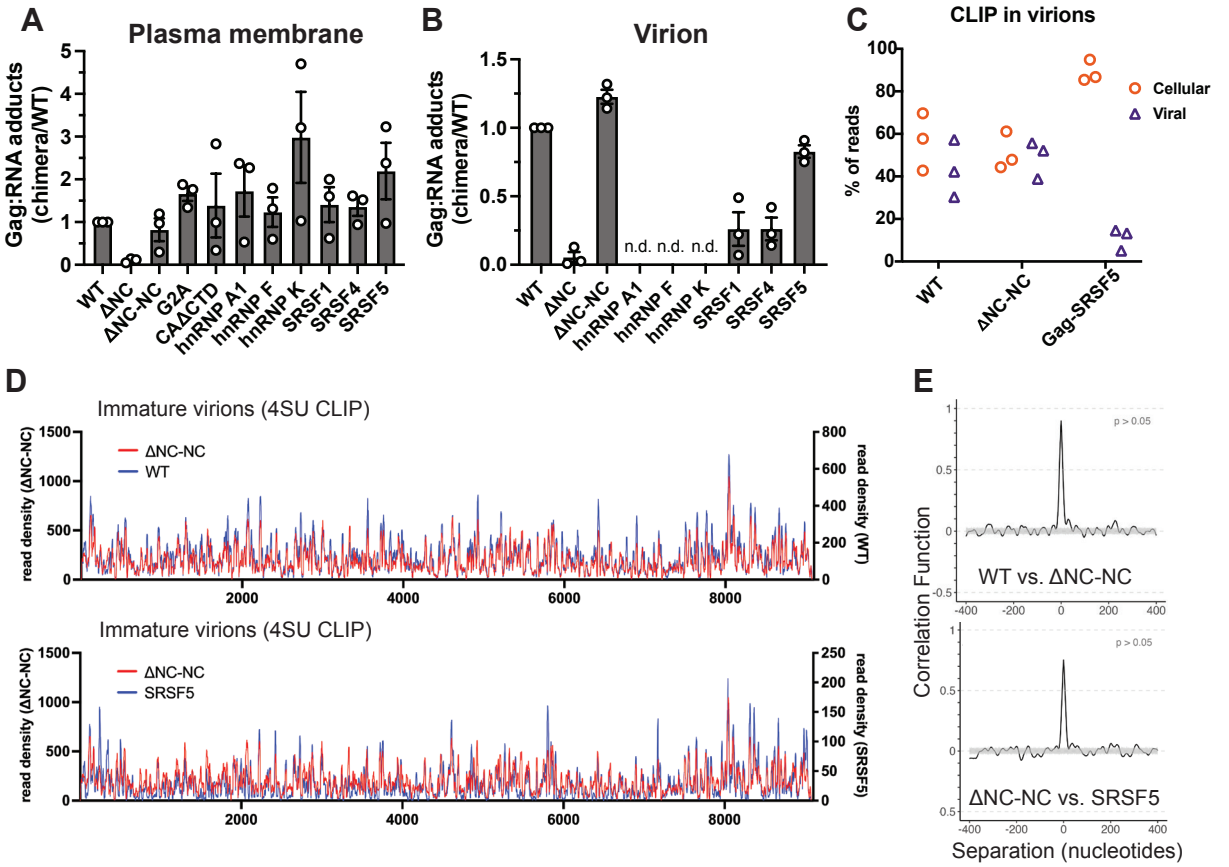

A
WT
ΔNC
G2A
hnRNP A1
hnRNP F
hnRNP K
SRSF1
SRSF4
SRSF5
64
51
Gag (M)
64
51
Gag (C)

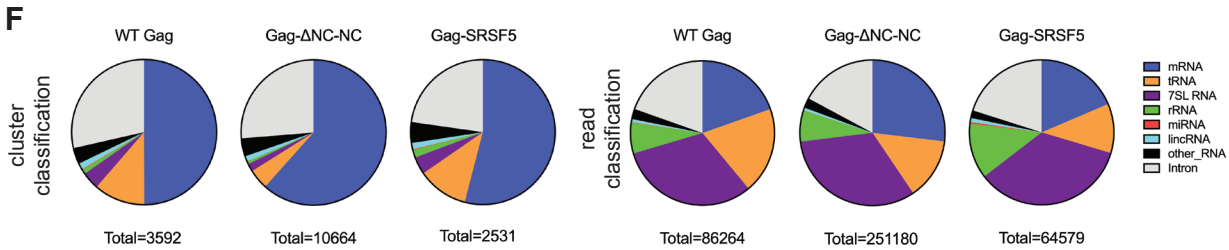

B
gRNA
multiply spl.
GAPDH
****
relative copies
1×10^1
1×10^0
1×10^-1
1×10^-2
WT
ΔNC
G2A
hnRNP A1
hnRNP F
hnRNP K
SRSF1
SRSF4
SRSF5

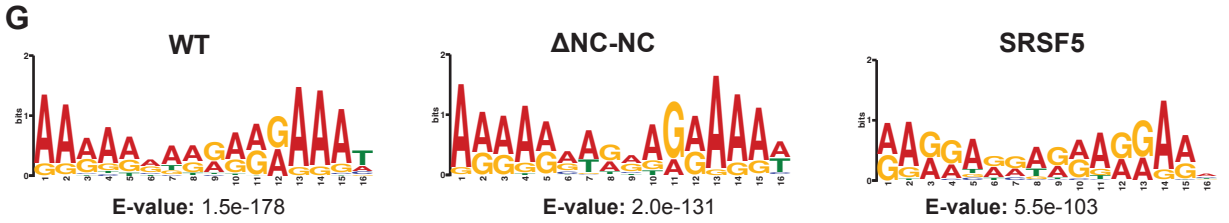

C
WT
ΔNC
ΔNC-NC
hnRNP A1
hnRNP F
hnRNP K
SRSF5
MS2-mCherry
Gag-YFP
Merge with DAPI
Scale bar = 20μm
ROI
Scale bar = 2.5μm

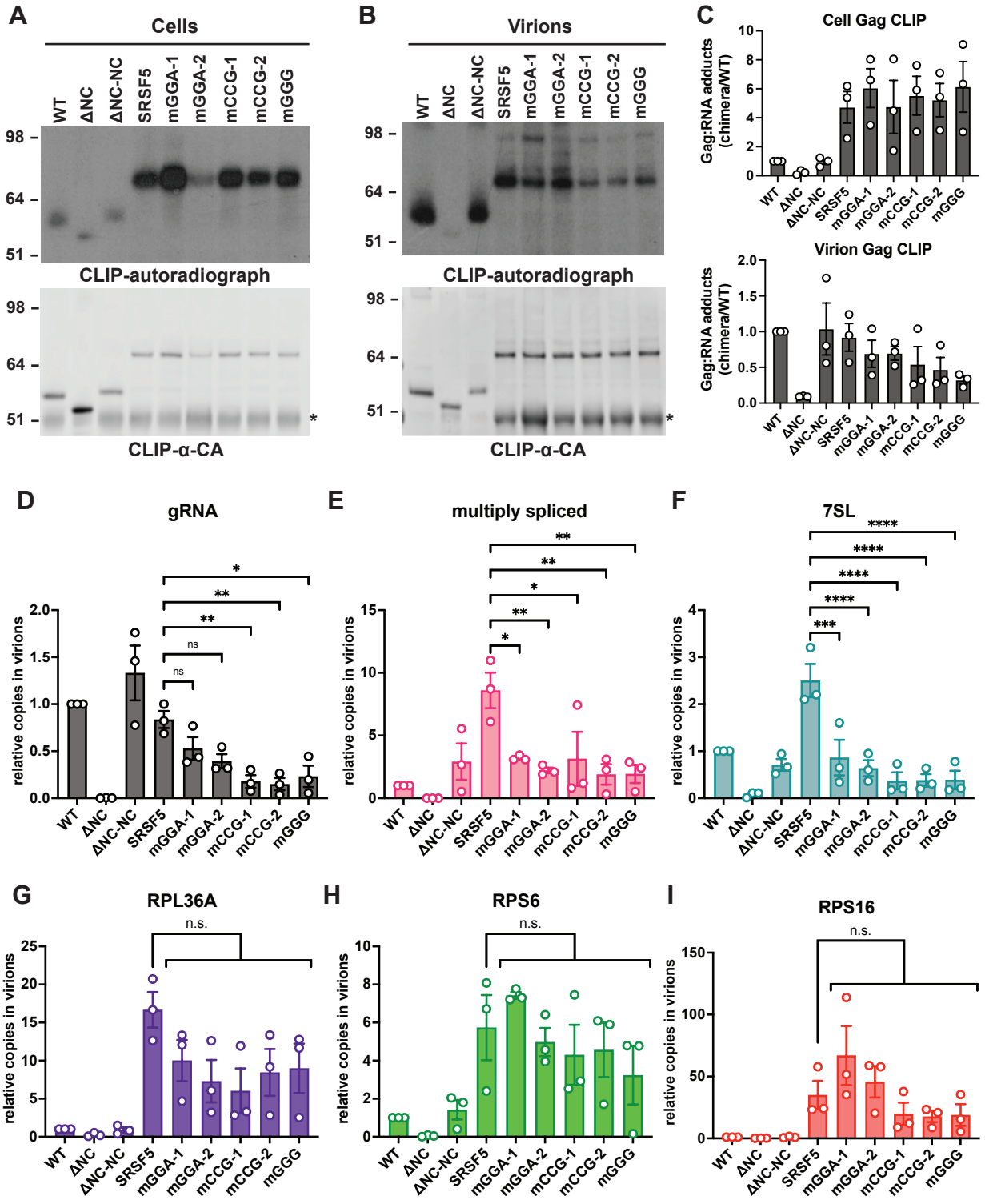

D
WT
ΔNC-NC
hnRNP A1
hnRNP F
hnRNP K
SRSF1
SRSF4
SRSF5
Scale bar = 0.5μm

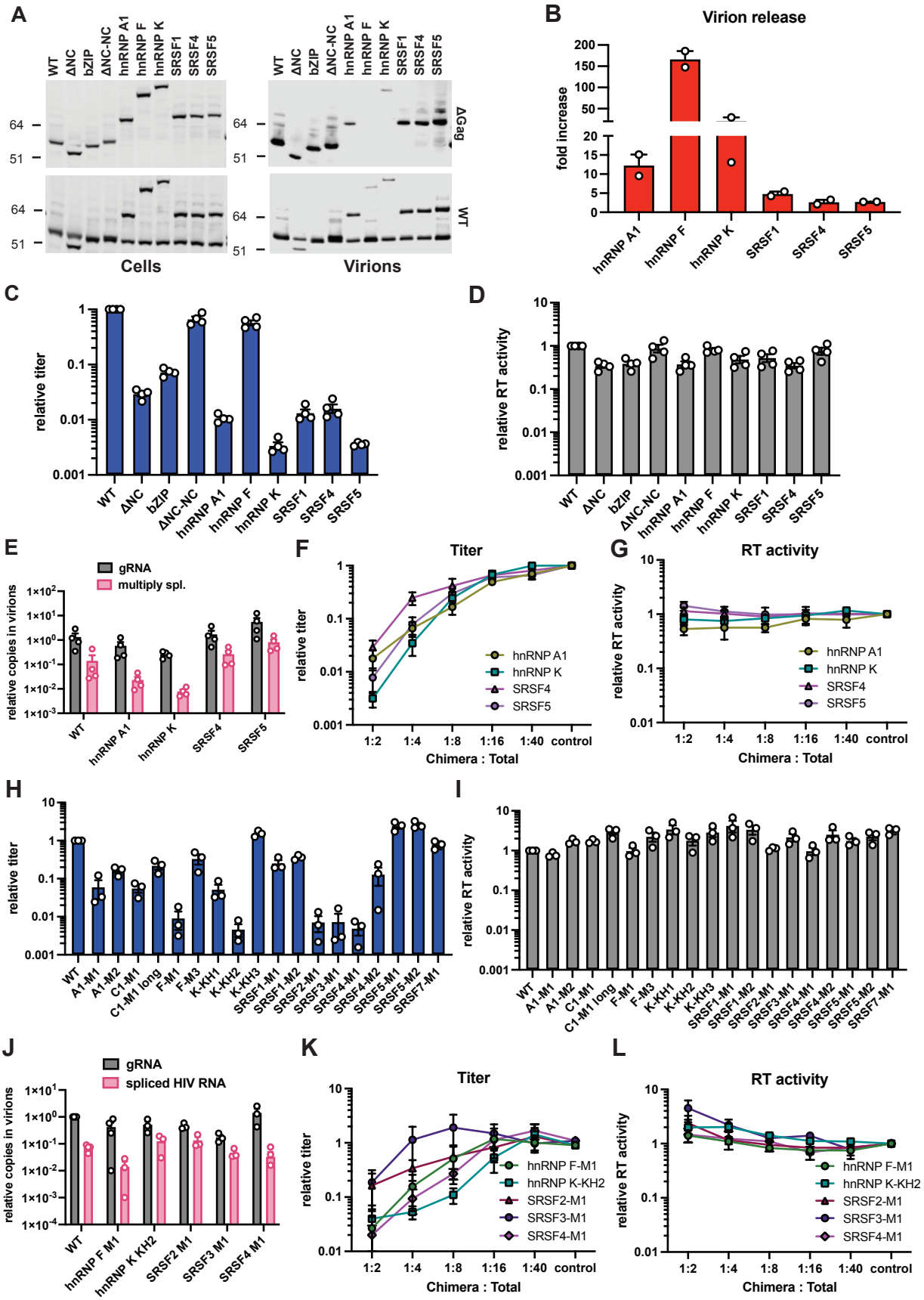

A
Cellular
Viral
% of reads
100
80
60
40
20
0
WT
ΔNC-NC
hnRNP A1
hnRNP F
hnRNP K
SRSF1
SRSF4
SRSF5
B
WT Gag
ΔNC-NC
hnRNP A1
hnRNP F
Total=680664
Total=794485
Total=604425
Total=580734
hnRNP K
SRSF1
SRSF4
SRSF5
Total=171763
Total=2.07796E+06
Total=612277
Total=655293
mRNA
tRNA
7SL RNA
rRNA
miRNA
lincRNA
other_RNA
Intron
C
WT
E-value: 3.3e-462
hnRNP A1
E-value: 5.8e-117
hnRNP F
E-value: 2.1e-230
hnRNP K
E-value: 1.3e-243
ΔNC-NC
E-value: 2.2e-246
SRSF1
E-value: 1.1e-414
SRSF4
E-value: 2.9e-256
SRSF5
E-value: 7.2e-168
D
WT
hnRNP A1
hnRNP F
hnRNP K
ΔNC-NC
SRSF1
SRSF4
SRSF5
SL1
SL3
read density (cell)
nucleotide position


A
Plasma membrane
Gag:RNA adducts (chimera/WT)
WT
ΔNC
ΔNC-NC
G2A
CAΔCTD
hnRNP A1
hnRNP F
hnRNP K
SRSF1
SRSF4
SRSF5
B
Virion
n.d. n.d. n.d.
C
CLIP in virions
% of reads
Cellular
Viral
D
Immature virions (4SU CLIP)
ΔNC-NC
WT
read density (ΔNC-NC)
read density (WT)
SRSF5
read density (SRSF5)
E
p > 0.05
WT vs. ΔNC-NC
ΔNC-NC vs. SRSF5
Correlation Function
Separation (nucleotides)


F
WT Gag
Gag-ΔNC-NC
Gag-SRSF5
cluster classification
read classification
mRNA
tRNA
7SL RNA
rRNA
miRNA
lincRNA
other_RNA
Intron
Total=3592
Total=10664
Total=2531
Total=86264
Total=251180
Total=64579


G
WT
ΔNC-NC
SRSF5
bits
E-value: 1.5e-178
E-value: 2.0e-131
E-value: 5.5e-103


A
Cells
WT
ΔNC
ΔNC-NC
SRSF5
mGGA-1
mGGA-2
mCCG-1
mCCG-2
mGGG
98
64
51
CLIP-autoradiograph
CLIP-α-CA
B
Virions
C
Cell Gag CLIP
Virion Gag CLIP
Gag:RNA adducts (chimera/WT)
D
gRNA
E
multiply spliced
F
7SL
relative copies in virions
G
RPL36A
H
RPS6
I
RPS16
n.s.


A
WT
ΔNC
bZIP
ΔNC-NC
hnRNP A1
hnRNP F
hnRNP K
SRSF1
SRSF4
SRSF5
64
51
ΔGag
WT
Cells
Virions
B
Virion release
fold increase
200
150
100
50
20
15
10
5
0
hnRNP A1
hnRNP F
hnRNP K
SRSF1
SRSF4
SRSF5
C
relative titer
1
0.1
0.01
0.001
WT
ΔNC
bZIP
ΔNC-NC
hnRNP A1
hnRNP F
hnRNP K
SRSF1
SRSF4
SRSF5
D
relative RT activity
10
1
0.1
0.01
0.001
E
gRNA
multiply spl.
relative copies in virions
1×10^2
1×10^1
1×10^0
1×10^-1
1×10^-2
1×10^-3
WT
hnRNP A1
hnRNP K
SRSF4
SRSF5
F
Titer
relative titer
hnRNP A1
hnRNP K
SRSF4
SRSF5
1:2
1:4
1:8
1:16
1:40
control
Chimera : Total
G
RT activity
relative RT activity
hnRNP A1
hnRNP K
SRSF4
SRSF5
1:2
1:4
1:8
1:16
1:40
control
Chimera : Total
H
relative titer
WT
A1-M1
A1-M2
C1-M1
C1-M1 long
F-M1
F-M3
K-KH1
K-KH2
K-KH3
SRSF1-M1
SRSF1-M2
SRSF2-M1
SRSF3-M1
SRSF4-M1
SRSF4-M2
SRSF5-M1
SRSF5-M2
SRSF7-M1
I
relative RT activity
WT
A1-M1
A1-M2
C1-M1
C1-M1 long
F-M1
F-M3
K-KH1
K-KH2
K-KH3
SRSF1-M1
SRSF1-M2
SRSF2-M1
SRSF3-M1
SRSF4-M1
SRSF4-M2
SRSF5-M1
SRSF5-M2
SRSF7-M1
J
gRNA
spliced HIV RNA
relative copies in virions
1×10^1
1×10^0
1×10^-1
1×10^-2
1×10^-3
1×10^-4
WT
hnRNP F M1
hnRNP K KH2
SRSF2 M1
SRSF3 M1
SRSF4 M1
K
Titer
relative titer
hnRNP F-M1
hnRNP K-KH2
SRSF2-M1
SRSF3-M1
SRSF4-M1
1:2
1:4
1:8
1:16
1:40
control
Chimera : Total
L
RT activity
relative RT activity
hnRNP F-M1
hnRNP K-KH2
SRSF2-M1
SRSF3-M1
SRSF4-M1
1:2
1:4
1:8
1:16
1:40
control
Chimera : Total