5' cap sequestration is required for sensing of unspliced HIV-1 RNA by MDA5

Ivy K. Hughes[1], Siarhei Kharytonchyk[2], Sita Ramaswamy[1], Sallieu Jalloh[1], James B. Hood[1], Xianbao He[3], Andrew J. Henderson[1 3], Hisashi Akiyama[1], Alice Telesnitsky[2], Suryaram Gummuluru[1]*

[1]Department of Virology, Immunology, & Microbiology, Boston University Chobanian & Avedisian School of Medicine

[2]Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor

[3]Department of Medicine, Section of Infectious Diseases, Boston Medical Center

Short title: MDA5 senses ^{cap}1G HIV-1 usRNA

*Corresponding author

Suryaram Gummuluru, Ph.D.

Department of Virology, Immunology & Microbiology

Boston University Chobanian & Avedisian School of Medicine

650 Albany St., X343C

Boston, MA 02118

Ph: (617) 358-1774

Fax: (617) 638-4286

Email: rgummulu@bu.edu

ABSTRACT

Heterogenous transcription start site (TSS) usage dictates the structure and function of unspliced HIV-1 RNAs (usRNA). We and others have previously reported that expression and Rev/CRM1-mediated nuclear export of HIV-1 usRNA in macrophages activates MDA5, MAVS, and innate immune signaling cascades. In this study, we reveal that MDA5 sensing of viral usRNA is strictly determined by TSS, 5' leader structure, and RNA function. We show that cap-sequestered HIV-1 usRNAs (^{cap}1G) destined for viral genome packaging are specifically targeted by MDA5, while translation-destined (^{cap}3G) usRNAs are remarkably immunologically silent. Using mutant viruses which express usRNA with altered 5' cap-exposed leader structure, or inclusion of a retroviral constitutive transport element which drives mRNA-like NXF1-dependent nuclear export of viral usRNA, we show that cap exposure and nuclear export pathway choice are major determinants of both lentiviral RNA immunogenicity and function. In total, we identify innate immune system evasion as a possible rationale for the universal conservation of heterogenous TSS usage among ancestral and extant HIV-1 isolates and shed light on how MDA5 fundamentally discriminates between self and non-self RNAs.

18 **SIGNIFICANCE**

19 Innate immune activation is critical to both the process of initial infection
20 establishment and ongoing chronic inflammation in HIV-1 infection. While MDA5 has
21 been identified as the sensor which detects unspliced HIV-1 RNA produced in infected
22 cells, it remains unclear how HIV-1 unspliced RNAs, which are generated by cellular
23 transcriptional processes, are recognized as non-self. Here, we reveal that HIV-1 RNA
24 function determines MDA5-driven immunogenicity. We show that only unspliced RNAs
25 which traffic to membrane-associated viral assembly sites are immunogenic, while
26 unspliced RNAs which are ribosomally translated to produce viral proteins are
27 immunologically silent. These findings not only advance our knowledge of how the
28 human innate immune system recognizes HIV-1 unspliced RNAs as foreign but also
29 provide a rationale for the selective advantage to generate two pools of unspliced RNAs
30 during HIV-1 replication.

31

32 **INTRODUCTION**

33 The HIV-1 genome encodes a variety of viral transcripts across its ~9 kb length
34 whose expression is reliant on the promoter activity located at the 5' long terminal
35 repeat (LTR).(1) To accomplish this, HIV-1 employs a complex splicing strategy which
36 allows for stoichiometric and temporal control over viral transcription.(2, 3) The first viral
37 RNAs generated, encoding for Tat, Rev, and Nef proteins, are short, multiply spliced
38 transcripts (msRNA) which are exported from the nucleus using canonical host mRNA
39 export pathways dependent on NXF1.(4) Nuclear import of the *de novo*-translated Rev
40 protein and subsequent engagement of the *cis*-acting Rev-response element (RRE) is
41 associated with the generation of singly spliced (encoding *vif, vpr, vpu* and *env*) and full-
42 length unspliced (usRNA, encoding *gag/pol* or acting as the viral genome) viral
43 transcripts, which Rev links to the host factor CRM1 for nuclear export.(5, 6)

44 The relatively small HIV-1 single-stranded RNA genome encodes a remarkable
45 amount of information not only in coding sequences across all three reading frames, but
46 in structure as well. Highly structured regions in the full-length unspliced HIV-1 transcript
47 such as those surrounding the *gag/pol* frameshift and RRE play documented roles in
48 viral replication.(7-9) Among these regions is the 5' untranslated region (5' UTR), a
49 structure-dense region of the HIV-1 genome which contains the *trans*-activation
50 response element (TAR), dimerization initiation sequence (DIS), and major splice donor
51 (SD) sequence, amongst others. Recent research has also revealed that the density of

52 structural information encoded in the 5' UTR is further enhanced by the transcription
53 initiation promiscuity of host RNA polymerase II, which initiates differentially on a
54 conserved GGG- tract at the 5' end of the HIV-1 provirus and produces HIV-1 RNAs with
55 either G- (^{cap}1G) or triple GGG- (^{cap}3G) 5' end sequences.(10, 11) Remarkably, this 2-
56 nucleotide difference at the 5' end of usRNA leads to major changes in 5' UTR structure
57 and fate of HIV-1 usRNA.(12-14) In the ^{cap}1G usRNA, the hairpin containing the polyA
58 signal is intact, which leads to a 5' UTR structure in which the 5' cap structure is buried
59 within the 5'UTR due to co-axial stacking of the TAR and polyA hairpins. In this
60 conformation, the 5' UTR DIS motif is exposed and available for dimerization with a
61 second gRNA for packaging into nascent particles.(15) This contrasts with the ^{cap}3G
62 usRNA, in which the polyA hairpin is unstable and instead forms a more continuous
63 central hairpin-like structure which includes DIS, sequestering the signals needed for
64 packaging and instead exposing the 5' cap needed for interactions with translation
65 initiation factors and ribosomal translation.(16, 17)

66 We and others have previously reported that cytoplasmic expression of intron-
67 containing HIV-1 RNAs exported from the nucleus via Rev/CRM1 activate innate
68 immune responses in macrophages and dendritic cells in an MDA5-dependent
69 manner.(18-21) MDA5, a RIG-I like receptor (RLR) family member, surveys the
70 cytoplasm for viral or host-derived immunostimulatory RNAs, binding of which induces
71 oligomerization of MDA5 through N-terminal caspase activation and recruitment
72 domains (CARD) and subsequent recruitment and activation of MAVS through CARD-
73 CARD interactions. We have recently demonstrated that activation of downstream
74 TRAF6 E3 ubiquitin ligase, IKK-β kinase, and IRF5 are required for expression of type I

interferons (type I IFN) and proinflammatory cytokines in HIV-1-infected macrophages.(21) While the host nucleic acid sensing and signal transduction machinery that detects HIV-1 intron-containing RNAs and activates type I IFN responses has been elucidated, sequence and structural requirements of the viral transcripts that define their immunostimulatory potential have remained unclear. In this report, using primary human monocyte-derived macrophages as a model of myeloid cell infection, we show that innate immune responses to intron-containing HIV-1 RNA initiated by MDA5 are restricted to ^{cap}1G usRNA specifically. Further, we show that MDA5-mediated immunogenicity is linked to nuclear export pathway, cytoplasmic RNA fate, and 5' cap exposure, implying that HIV-1 ^{cap}3G usRNAs destined for translation make use of host RNA metabolism pathways to dampen immunogenicity while facilitating viral protein expression. In summary, this work provides a new model to understand the ability of MDA5 to distinguish 5' UTR structures of HIV-1 usRNAs based on 5' cap exposure and the functional importance of TSS heterogeneity in the context of innate immune activation by HIV-1.

RESULTS

De novo-transcribed ^{cap}1G HIV-1 unspliced RNAs stimulate innate immunity in macrophages

In the HIV-1 LTR, a highly conserved GGG motif encompasses the transcription initiation site and is located approximately 30 bp downstream of the viral core promoter CATA box element.(22) The LTR sequence in the HIV-1/NL4-3 lab-adapted strain is sufficient to drive transcription of both ^{cap}1G and ^{cap}3G usRNAs (**Fig. 1A**/**B**).(10, 12) Mutation of the first two G residues in the NL4-3 LTR to TC instead drives transcription

98 only from the final G in the initiation tract, producing a virus in which only ^{cap}1G RNA is
99 transcribed (1G virus, where 1G refers to the mutant virus and ^{cap}1G refers to the RNA
100 transcripts it produces). Similarly, if the first two G residues of the GGG tract are not
101 replaced but instead shifted 2 bp downstream by insertion of the TC motif, transcription
102 results in only ^{cap}3G RNA expression (3G virus) (**Fig. 1B**). To determine the
103 immunological outcome of ^{cap}1G and ^{cap}3G HIV-1 usRNA expression in macrophages,
104 we infected primary human monocyte-derived macrophages (MDMs) with VSV-G-
105 pseudotyped replication-competent viruses (containing both G and Env surface
106 expression). To overcome SAMHD1-mediated inhibition of reverse transcription in
107 primary MDMs, we co-infected with SIVmac Vpx-containing VLPs (SIV3+ VLPs) as
108 previously described.(23) We first ensured that 1G and 3G infection of MDMs resulted in
109 the specific transcription of only ^{cap}1G and ^{cap}3G RNAs by harvesting RNA from infected
110 macrophages and performing a cap-dependent adaptor ligation/PCR assay (CaDAL) for
111 viral usRNA as previously described.(22) Indeed, we observed highly specific
112 expression of unspliced ^{cap}1G/3G RNAs upon infection with 1G/3G mutant viruses,
113 respectively (**Fig. 1C**). Further, staining for intracellular p24^{Gag} by flow cytometry and
114 measuring p24^{Gag} release in culture supernatants by ELISA at three days post infection
115 confirmed that the efficiency was equivalent between WT, 1G, and 3G virus infections of
116 MDMs (**Fig. 1D**/**E**).

117 To measure type I IFN secretion, we employed a bioassay to measure total
118 bioactive type I IFN levels. Though infection establishment was equivalent across all
119 three viruses, we observed type I IFN secretion only from NL4-3- and 1G-infected
120 MDMs, but not from 3G-infected MDMs (**Fig. 1F**). IFN-I secretion from 1G-infected

121 MDMs was also significantly greater than that observed with wild-type NL4-3-infected
122 MDMs, consistent with specific sensing of ^{cap}1G RNAs. We also observed significantly
123 less pro-inflammatory IP-10 secretion from 3G-infected cultures compared to 1G or wild
124 type virus infected MDMs (**Fig. 1G**). We confirmed that innate immune activation was
125 limited to *de novo*-transcribed ^{cap}1G usRNA rather than incoming virus-associated
126 gRNAs since inhibition of infection establishment (lenacapavir, efavirenz, and
127 raltegravir) or Rev/CRM1 usRNA export (KPT-330), but not protease activity (lopinavir)
128 blocked induction of type I IFN secretion (**Fig. 1H-J**). These findings are consistent with
129 the established model that innate immune activation is induced late in the virus lifecycle
130 after successful Rev/CRM1-mediated export of HIV-1 usRNA.(18-21) Additionally, we
131 infected MDMs with NL4-3, 1G, and 3G viruses without addition of Vpx-containing
132 SIV3+ VLPs, and observed type I IFN secretion peaking at 6 days post-infection in wild-
133 type and 1G infections but not 3G infection despite identical levels of *de novo* p55^{Gag}
134 expression and p24^{Gag} secretion (**Fig. S1**). Together, these findings establish that
135 expression of unspliced ^{cap}1G RNAs alone stimulates proinflammatory type I IFN
136 responses in macrophages.

137 **MDA5 detects ^{cap}1G unspliced HIV-1 RNAs and activates type I interferon**
138 **expression through MAVS**

139 We and others have previously reported that MDA5 senses HIV-1 usRNAs and
140 activates MAVS and downstream innate immune responses.(20, 21) To determine
141 whether MDA5 detects ^{cap}1G usRNA, we transfected MDMs with siRNAs targeting
142 MDA5, RIG-I, or MAVS and infected cells with wild-type (1G/3G) or 1G viruses and
143 measured induction of innate immune responses. We ensured that targeted siRNA

144 transfection resulted in efficient knockdown of MDA5, RIG-I, and MAVS expression (**Fig.**
145 **2A**-**F**). We also ensured that siRNA transfection did not impact infection efficiency or
146 expression of usRNA, evidenced by equivalent levels of $p24^{Gag}$ secretion and usRNA
147 quantification via RT-qPCR in siRNA-transfected MDMs (**Fig. 2G**/**H**). We did not
148 observe any significant diminution of cytokine secretion in siRIG-I-transfected wild-type
149 or 1G virus infected MDMs (**Fig. 2I**/**J**). Consistent with our hypothesis, type I IFN and
150 IP-10 release were abrogated in siMDA5- and siMAVS-transfected MDMs infected with
151 wild-type or 1G virus.

152 To confirm that MDA5 was responsible for specifically engaging viral ^{cap}1G
153 usRNA transcripts, we analyzed the *in situ* interaction of MDA5 with unspliced viral
154 transcripts by performing RNA immunoprecipitation in virus-infected HEK293T cells
155 overexpressing MDA5-FLAG (**Fig. 3A**).(21) Following UV crosslinking and anti-FLAG
156 immunoprecipitation, we observed significant infection-dependent anti-FLAG pulldown
157 of unspliced HIV-1 transcripts in NL4-3 and 1G virus infection conditions relative to
158 control IgG pulldown, but no usRNA pulldown in 3G-infected MDA5-FLAG-expressing
159 HEK293T cells (**Fig. 3B**). Consistent with our hypothesis, we did not observe pulldown
160 of multiply-spliced viral transcripts or host actin mRNA in any virus infection condition
161 (**Fig. 3C**/**D**). Taken together, these studies indicate that MDA5 senses productive HIV-1
162 infection by specific interaction with newly-transcribed ^{cap}1G unspliced transcripts,
163 leading to MAVS activation and induction of innate immune responses in macrophages.

164 **^{cap}1G/3G RNA splicing and usRNA localization in primary macrophages**

165 To better understand the dynamics of MDA5-driven innate responses in HIV-
166 infected cells, we characterized ^{cap}1G RNA metabolism and fate in primary

167 macrophages. We analyzed ^{cap}1G and ^{cap}3G unspliced RNA abundance in wild type
168 (^{cap}1G/3G) virus-infected syngeneic CD4+ T cells and macrophages using CaDAL
169 assay. In CD4+ T cells, we found a roughly 1:1 ratio of *de novo*-transcribed ^{cap}1G/^{cap}3G
170 unspliced RNA species in infected cell lysates, which was absent in efavirenz-treated
171 cell lysates (**Fig. 4A**). In MDMs, we observed a distinct shift towards ^{cap}1G unspliced
172 RNA abundance which was again dependent on establishment of productive infection
173 (**Fig. 4A**). We next analyzed usRNA levels in the wild-type, 1G, or 3G virus infections of
174 MDMs by RT-qPCR, and again observed a decrease in unspliced HIV-1 RNA signal in
175 3G infection compared to 1G and wild-type NL4-3 infection conditions (**Fig. 4B**). We
176 hypothesized that the low usRNA abundance in 3G-infected MDMs was either due to
177 increased splicing of ^{cap}3G RNA products, or a shorter cellular half-life of ^{cap}3G usRNA,
178 or both. To test these hypotheses, we analyzed the viral RNA transcriptome in MDMs
179 using a multiplex digital-droplet PCR (ddPCR)-based assay, simultaneously analyzing
180 the presence of unspliced (LTR/gag/env/nef), singly-spliced (LTR/env/nef), or multiply-
181 spliced (LTR/nef) transcript sequences using previously defined primer/probe sets.(24)
182 We observed a significant shift towards unspliced RNA fate (positive for all four probes
183 simultaneously) in 1G infection compared to NL4-3 and 3G infection of MDMs (**Fig. 4C**,
184 **Fig. S2**). This effect was accompanied by a decrease in singly-spliced RNA in 1G
185 infection, and a roughly equivalent abundance of multiply-spliced RNA in all three
186 infections, similar to findings in the previous report.(25)

187 To characterize the localization of ^{cap}1G/3G usRNA expression in primary
188 macrophages, we employed single-molecule RNA-FISH (smFISH) with co-staining for
189 plasma membrane (wheat germ agglutinin, WGA) or p24^{Gag} expression (**Fig. 4D**/**E**, **Fig.**

190 **S3**). For these experiments, we employed single-round (Δenv, GFP in place of *nef*)
191 viruses harboring wild-type (^{cap}1G/3G, WT-LTR) or ^{cap}1G- or ^{cap}3G-only LTRs (1G-LTR,
192 3G-LTR). We also ensured that cytoplasmic usRNA signal was dependent on productive
193 infection using reverse transcriptase inhibition (**Fig. S3**), and Rev/CRM1 nuclear export
194 using a virus with an inactivating mutation in Rev (WT-LTR/M10, **Fig. S3**).(18, 26)
195 usRNA expressed in 1G infection of MDMs preferentially co-localized with both WGA-
196 positive plasma membrane and sites of Gag accumulation (**Fig. 4F**/**G**), consistent with
197 the established function of ^{cap}1G usRNA transcripts as packaging-fated virus
198 genomes.(12, 27) Taken together, these experiments suggest that HIV-1 ^{cap}1G usRNA
199 accumulate to higher levels at membrane-associated sites of viral assembly in primary
200 macrophages, and that preferential accumulation at these peripheral membrane sites
201 may contribute to specific surveillance of ^{cap}1G usRNA by MDA5 and induction of innate
202 immune responses.

203 **RNA fate links expression of ^{cap}1G unspliced HIV-1 RNA to innate immune**
204 **activation**

205 While both unspliced ^{cap}1G and ^{cap}3G contain highly structured 5' UTRs, ^{cap}1G
206 unspliced RNAs predominantly dimerize to serve as viral genomes, which could
207 theoretically explain their susceptibility to MDA5 sensing. To test if dimerization is the
208 trigger of innate immune responses in macrophages, we infected MDMs with an
209 additional mutant virus which results in the expression of usRNA with a novel 5' leader
210 structure. This virus construct, known as no bulge (NB), encodes a polyA hairpin with
211 greater internal base-pairing and stability (**Fig. 5A**), which when expressed from a
212 ^{cap}3G-only LTR (3G-NB virus) produces a single 'hybrid' unspliced RNA (in addition to

213 typical spliced transcripts) in which the 5' cap is exposed as in typical ^{cap}3G usRNA, but
214 which contains an overall ^{cap}1G-like fold characterized by a double three-way junction
215 structure and exposure of DIS.(28) Previous work has established that while ^{cap}3G-NB
216 usRNAs dimerize *in vitro*, they are efficiently translated in cells and not packaged in
217 competition assays with ^{cap}1G usRNA, suggesting that 5' cap exposure is the major
218 determinant of whether an usRNA will adopt a packaging or translation fate.(27, 28)
219 When we infected primary MDMs with WT, 3G, and 3G-NB viruses, we observed similar
220 levels of infection efficiency and p24^{Gag} release for all viruses (**Fig. 5B**/**C**, **Fig. S4**).
221 When we measured bioactive type I IFN from these cultures, we found that 3G-NB virus
222 infection resulted in minimal IFN secretion, similar to that observed upon 3G infection
223 (**Fig. 5D**). Interestingly, relatively low abundance of usRNA transcripts in 3G-NB
224 infection was observed, similar to 3G infection of MDMs (**Fig. 5E**). However, despite
225 these low usRNA levels, we observed high levels of Gag expression in 3G- and 3G-NB-
226 infected MDMs (**Fig. 5F**, **Fig. S4**), suggesting that 5' cap exposed 3G and 3G-NB
227 usRNAs are targeted primarily for translation. These experiments demonstrate that 5'
228 cap sequestration, and not DIS exposure or 5' leader structure itself *per se*, is the major
229 determinant of RNA fate and MDA5-mediated innate immune activation induced by HIV-
230 1 usRNA expression.

231 We next determined if Rev/CRM1 export was strictly necessary for ^{cap}1G usRNA-
232 driven interferon responses, as we and others have previously shown that substitution
233 of Rev/CRM1 RNA export with a retroviral constitutive transport element (CTE) from
234 Mason-Pfizer Monkey Virus (MPMV) fails to induce innate immune responses.(18, 19)
235 To address this, we generated single-round reporter viruses utilizing various RNA export

236 mechanisms with ^{cap}1G-only LTRs (**Fig. 6A**). We added either the inactivating M10
237 mutations in Rev which block Rev/CRM1 export of usRNA (1G-M10) or added M10 in
238 combination with four repeats of CTE (1G-M10/4XCTE) which rescues NXF1-
239 dependent nuclear export of usRNA independent of Rev/CRM1.(29) Upon infection of
240 MDMs, we observed robust infection with all three reporter viruses (**Fig. 6B**). We found
241 that Rev/CRM1 export was strictly required for type I IFN secretion and that 1G-
242 M10/4XCTE infection resulted in only minimal increases in IP-10 secretion relative to
243 1G-M10 infection (**Fig. 6C**/**D**). Interestingly, when we probed infected MDM lysates for
244 usRNA abundance via RT-qPCR, we observed low abundance of usRNA in 1G-
245 M10/4XCTE-infected cells, but no defects in p55^{Gag} expression (**Fig. 6E**/**F**). Altogether,
246 these results support a model where CTE/NXF1-exported ^{cap}1G usRNAs are efficiently
247 translated but not sensed by MDA5, similarly to Rev/CRM1-exported ^{cap}3G usRNA
248 transcripts (**Fig. 6G**). Collectively, these findings suggest that the immunostimulatory
249 potential of HIV-1 usRNA is modulated by both cap sequestration and nuclear export
250 trafficking itinerary, thus determining its susceptibility to surveillance by MDA5 in the
251 cytoplasm of infected cells.

252 **DISCUSSION**

253 The cell-intrinsic innate immune system manages the task of detecting foreign
254 RNAs among a massive background of abundant and diverse host RNA species.(30,
255 31) In this report, we establish that ^{cap}3G unspliced HIV-1 transcripts, which contain an
256 exposed 5' cap and are efficiently translated utilizing host translational machinery, are
257 remarkably non-immunostimulatory. In contrast, ^{cap}1G unspliced transcripts, which
258 contain a sequestered 5' cap, are recognized by MDA5 to activate innate immune

259 signaling cascades. Since nuclear export and cytoplasmic packaging of ^{cap}1G-RNA is an
260 immutable step in the virus life cycle, we hypothesize that TSS heterogeneity to
261 generate ^{cap}3G usRNA expression by HIV-1 is a viral innovation that represents viral co-
262 option of immunologically silent mRNA translation pathways to minimize induction of
263 type I IFN responses during viral replication. Importantly, cap sequestration is an
264 essential determinant of efficient viral genome packaging, suggesting that innate
265 immune activation is the putative fitness cost to effectively generate high titers of
266 infectious virus particles.(27) Further, preservation of ^{cap}3G unspliced HIV-1 transcript
267 expression in macrophages to facilitate innate immune evasion may at least partially
268 explain the striking conservation of heterogeneous TSS usage among the HIV-
269 1/SIVcpz/SIVgor lineages, as our previous reports have observed only minor replicative
270 defects in ^{cap}1G-only virus-infected CD4+ T cells.(22, 32)

271 We conclude that the innate immune responses observed here are most likely
272 due to usRNA and not singly-spliced transcripts, as splicing of the *gag*/*pol* intron at SD
273 would inhibit the native folded ^{cap}1G double three-way junction structure, causing
274 unfolding of the polyA hairpin and making cap sequestration, which triggers detection by
275 MDA5, unlikely. This would agree with a previous report indicating that MDA5 does not
276 engage singly-spliced HIV-1 RNAs.(20) In characterizing viral RNA expression in
277 primary macrophages, we found a greater accumulation of ^{cap}1G usRNA levels
278 compared to ^{cap}3G in both wild-type and 1G/3G mutant virus infection. In macrophages,
279 because virus particles are frequently retained post-budding in CD169+ virus-containing
280 compartments (VCCs)(33, 34) and particles produced during wild-type or 1G infection
281 are presumably more likely to contain packaged unspliced RNA transcripts than those

282 produced during 3G infection, a higher abundance of ^{cap}1G usRNA may not necessarily
283 reflect direct transcriptional differences. However, in line with previous studies, we also
284 found that ^{cap}1G RNA expression was biased towards unspliced RNA production.(25)
285 Further, we additionally speculate that cellular turnover of actively translated ^{cap}3G
286 transcripts may be faster than ^{cap}1G transcripts destined for packaging in macrophages,
287 though the cellular mechanisms postulated to account for differential lifespan and decay
288 of ^{cap}3G usRNA remain to be definitively determined. Notably, previous studies have
289 identified translation-dependent destabilization of mRNAs in human cells.(35) While
290 ^{cap}3G usRNA abundance was lower than ^{cap}1G, we repeatedly did not observe defects
291 in Gag production or particle release in ^{cap}3G-only infection, implying that ^{cap}3G
292 expression alone is sufficient to drive efficient Gag and Gag-Pol translation while
293 minimizing cytoplasmic usRNA abundance and innate immune activation. Further, we
294 consistently observed a lower ratio of type I IFN production to usRNA expression in 3G
295 infection compared to wild-type or 1G infection, indicating that either the per-transcript
296 immunogenicity of ^{cap}3G unspliced transcripts is lower than ^{cap}1G, or MDA5 sensing of
297 cytoplasmic HIV-1 usRNA requires a threshold level that ^{cap}3G expression does not
298 achieve, or both.

299 In infections with ^{cap}1G-M10/4XCTE virus, where CTE-containing usRNA
300 transcripts access NXF1-mediated rather than CRM1-dependent usRNA nuclear export,
301 we observed an absence of type I IFN production and low usRNA abundance, but
302 robust rescue of *de novo* p55^{Gag} production. We therefore speculate that the relatively
303 small and/or rapidly turned-over pool of cytoplasmic ^{cap}1G-M10/4XCTE usRNA
304 transcripts, though etiologically distinct, are functionally similar to ^{cap}3G usRNA

305 transcripts and preferentially translated, leading to co-translational decay-dependent
306 usRNA turnover and dampening of innate immune responses. This model would explain
307 previous observations by our group and others that CTE-exported HIV-1 usRNA is not
308 immunostimulatory.(18, 19) In addition, remarkably, this model is consistent with
309 previous reports demonstrating that unspliced murine leukemia virus (MLV) RNA
310 transcripts utilize both CRM1- and NXF1-dependent pathways for nuclear export, and
311 only CRM1-exported MLV-usRNA is effectively packaged into nascent virions, while
312 NXF1-exported MLV-usRNA is instead translated to produce Gag and Gag-Pol
313 polyproteins.(36) Our previous work has also established that CTE-exported HIV-1
314 usRNA does not effectively compete with Rev/CRM1-exported HIV-1 usRNA for
315 packaging, again implying that NXF1-exported usRNAs have a packaging defect,
316 similar to Rev/CRM1-exported ^{cap}3G usRNAs.(37) In summary, the findings in this report
317 and others point to the selection of unspliced retroviral RNA fate and function occurring
318 in the nucleus, which in HIV-1 infection is controlled by transcription start site and
319 determines the immunological outcome of HIV-1 usRNA expression in human
320 macrophages.(36, 38, 39) While HIV-1 exclusively uses Rev/CRM1-mediated nuclear
321 export for unspliced transcripts, ^{cap}3G usRNA effectively mimics host mRNA and
322 facilitates robust viral structural/enzymatic polyprotein production while likely minimizing
323 the innate immune cost of expression of ^{cap}1G unspliced transcripts required for efficient
324 genome packaging.

325 Our data additionally point to the sequestration/exposure of the 5' usRNA cap
326 playing a major role in both the function and immunogenicity of HIV-1 usRNA. A large
327 diversity of positive and negative strand RNA viruses encode mechanisms to both

328 acquire 5' caps and achieve a host-like methylation pattern on 5' cap structures on *de*
329 *novo*-transcribed RNAs, both to achieve efficient translation and to neutralize innate
330 immune responses.(40, 41) Notably, in coronavirus infection, virus-encoded cap 2'-O-
331 methylation activity prevents activation of MDA5 and influences coronavirus
332 pathogenesis in mice.(42, 43) In HIV-1 infection, immunogenicity of cap-sequestered
333 ^{cap}1G usRNAs could directly reflect MDA5's biochemical affinity for uncapped RNAs.
334 Given our previous finding that HIV-1 usRNA-induced innate immune activation requires
335 membrane association,(18) MDA5's ability to detect cap-sequestered ^{cap}1G HIV-1
336 usRNAs associated with the plasma membrane or coronavirus RNAs lacking cap-
337 methylation associated with double membrane vesicle assembly sites may highlight the
338 proximity of an uncapped or insufficiently-methylated capped RNA to a membrane site
339 as a distinct pathogen-associated molecular pattern (PAMP).(44, 45) Additional
340 variables could theoretically include regulation of MDA5 by the cap-exposed or cap-
341 sequestered usRNA-associated proteome or post-transcriptional RNA modifications.
342 Alternatively, driven by 5' cap exposure in ^{cap}3G HIV-1 usRNA transcripts, translation
343 itself may repress MDA5 sensing by promoting co-translational decay of usRNAs or
344 obfuscating access to physical RNA-protein interactions with MDA5. Protection of
345 efficiently ribosome-translated RNAs from MDA5 sensing could theoretically decrease
346 the propensity of the innate immune system from spontaneously recognizing host
347 mRNAs that contain significant secondary structure. Further, efficiently-translated
348 Rev/CRM1-exported ^{cap}3G usRNAs may enter a pool of cytoplasmic mRNAs not
349 sampled by MDA5, similarly to Rev/CRM1-exported ^{cap}3G-NB usRNA transcripts and
350 NXF1-exported transcripts (including multiply-spliced retroviral transcripts and CTE-

351 exported usRNA transcripts), highlighting a probable intracellular localization
352 requirement for MDA5 sensing. In total, our data indicate that HIV-1 usRNA leader
353 structure, function, and intracellular localization are intricately linked to susceptibility to
354 innate immune sensing. Further, these results provide a rationale for HIV-1 dependence
355 on TSS heterogeneity for generating at least two usRNA populations (^{cap}1G and ^{cap}3G)
356 to dampen MDA5 sensing and overcome restrictions to viral propagation in humans.

357 **MATERIALS AND METHODS**

358 **Viruses, plasmids, and cloning**

359 The NL4-3, NL4-3 1G, and NL4-3 3G proviral plasmids have been previously
360 described.(22) The NL4-3 3G-NB proviral plasmid has also been previously
361 described.(28) The SIV3+ plasmid used for producing SIV3+ VLPs has been
362 described.(23) The plasmid encoding the G protein of vesicular stomatitis virus (HCMV-
363 G) and the HIV-1 packaging plasmid (psPAX2) used for production of infectious viruses
364 have been described.(18) To generate ^{cap}1G-only or ^{cap}3G-only single-round GFP-
365 expressing viruses, the proviral genomes of single-round (Δenv, GFP in place of *nef*)
366 Lai-based plasmids (LaiΔenvGFP, LaiΔenvGFP-M10, and LaiΔenvGFP-M10-4XCTE)
367 used in previous studies were excised via *BssHII/XhoI* restriction sites and ligated into
368 the NL4-3 1G or 3G backbones.(18) LaiΔenvGFP and LaiΔenvGFP-M10 themselves
369 were additionally used for smFISH (WT-LTR, WT-LTR/M10). The MDA5-FLAG plasmid
370 used for RNA immunoprecipitation studies was a generous gift from Dr. Kate Fitzgerald
371 (University of Massachusetts Chan Medical School).

372 All viruses used were pseudotyped with VSV-G envelope protein (regardless of
373 intact or deleted *env*). Viruses were produced via calcium phosphate-mediated
374 transfection of HEK293T cells as described previously.(18) Briefly, HEK293T seeded in
375 10 cm dishes were transfected with 10-12 μg of plasmids using calcium
376 phosphate/BES-buffered saline. Virus-containing cell supernatants were harvested 48

377 hours post-transfection, 0.45 μm syringe-filtered and then ultracentrifuged at 100,000 x
378 g in an SW28 rotor for 2 hours at 4ºC on a 20% sucrose cushion. The viral pellet was
379 resuspended in DPBS (Gibco, cat. # 14190-144), aliquoted, stored at -80°C, and
380 subsequently titered for infectivity using the TZM-bl assay, as previously described.(18)
381 To produce VSV-G-pseudotyped replication-competent NL4-3-viruses, a plasmid ratio of
382 10:1 provirus/HCMV-G was utilized. To produce VSV-G-pseudotyped single-round
383 Δenv/GFP viruses, a plasmid ratio of 6:6:1 provirus/psPAX2/HCMV-G was transfected in
384 HEK293T cells. To produce Vpx-containing SIV3+ VLPs, a plasmid ratio of 10:1
385 SIV3+/HCMV-G was used, and SIV3+ VLPs were titered for p27 content before use with
386 a commercially available p27 ELISA kit (XpressBio, cat. # SK845).

387 **Cell lines and cell culture**

388 HEK293T cells were acquired from ATCC (cat. # CRL-3216). TZM-bl cells used
389 for titering viruses were acquired from the NIH HIV Reagent Program (cat. # HRP-
390 8129). 293 ISRE-luc cells used for type I IFN bioassay were a generous gift from Drs.
391 Junzhi Wang (National Institute for the Control of Pharmaceutical and Biological
392 Products, China) and Xuguang Li (University of Ottawa). HEK293T, TZM-bl, and 293
393 ISRE-luc cells were cultured in D10 medium, composed of DMEM (Gibco, cat. # 11965-
394 092) supplemented with 1% Pen-Strep (Gibco, cat. # 15-140-122) and 10% heat-
395 inactivated fetal bovine serum (FBS, Gibco, cat. # A52568-01). MDMs were cultured in
396 R10 medium, made of RPMI1640 medium (Gibco, cat. # 11875-093) supplemented with
397 1% Pen-Strep and 10% FBS. All cell lines used in these studies were routinely tested
398 for mycoplasma contamination.

399 **Isolation and differentiation of primary cells**

400 Human peripheral blood mononuclear cells (PBMCs) were isolated from de-
401 identified leukopaks obtained from NY Biologics (Long Island, NY) as described
402 previously.(18) CD14+ cells were isolated from PBMCs using anti-CD14 magnetic
403 beads (Miltenyi Biotech, cat. # 130-050-201) according to the manufacturer's protocol.
404 To differentiate CD14+ primary monocytes to macrophages, CD14+ cells were plated on
405 low-attachment six well plates (Corning, cat. # 3471) in RPMI media supplemented with
406 10% heat-inactivated human AB serum (Sigma, cat. # H4522) and 20 ng/mL human M-
407 CSF (Peprotech, cat. # 300-25). After 5-6 days of culture, cells were lifted, counted, and
408 re-plated on normal-attachment plates in R10 medium (or on low-attachment plates if
409 cells were to be analyzed via flow cytometry). CD4+ T cells were isolated from PBMCs
410 using anti-CD4 magnetic beads (Miltenyi Biotech, cat. # 130-045-101) and activated
411 with 2% phytohemagglutinin (PHA, Gibco, cat. # 10576015) and 50 U/mL IL-2
412 (NIH/NIAID HIV Reference Reagent Program, cat. # 136) for two days before washing
413 in DPBS and resuspension in R10 medium supplemented with 50 U/mL IL-2.

414 **Infections**

415 MDMs and activated CD4+ T cells were infected with viruses at MOI 1 based on
416 TZM-bl infectious titer. Infections were carried out in the presence of polybrene
417 (Millipore-Sigma, cat. # TR1003G, 10 µg/mL) with or without addition of Vpx-containing
418 SIV3+ VLPs. Cells were spinoculated at 2300 rpm at room temperature for 1 h,
419 incubated at 37ºC for 2-3 h, followed by washing with DPBS and culture in fresh R10
420 medium (with 50 U/mL IL-2 for CD4+ T cells). Cells were harvested at 3 dpi unless
421 otherwise noted. For replication-competent virus infections (MOI 1), $5x10^5$ MDMs were
422 plated in 24-well plates and co-infected with SIV3+ VLPs (5 ng $p27^{Gag}$). For GFP-

423 expressing single-round virus infections, $1x10^6$ MDMs were plated in 12-well plates and
424 co-infected with SIV3+ VLPs (15 ng $p27^{Gag}$).

425 **Antibodies and drugs**

426 Antibodies used included anti-$p24^{Gag}$ (KC57-RD1, Beckman Coulter cat. #
427 6604667, mouse monoclonal, used at 1:50 for flow cytometry), anti-$p24^{Gag}$ (NIH/NIAID
428 HIV Reagent Program cat. # HRP-20068, and mouse monoclonal clone # AG3.0, used
429 at 1:500 for co-staining with smFISH). The following antibodies were used for western
430 blot analysis: anti-β-actin (Invitrogen cat. # AM4302, mouse monoclonal, used at
431 1:5000), anti-β-actin (Sigma-Aldrich cat. # A2066, rabbit polyclonal, used at 1:5000),
432 anti-MDA5 (Proteintech cat. # 21775-1-AP, rabbit polyclonal used at 1:1000), anti-RIG-I
433 (AdipoGen cat. # AG-20B-0009-C100, mouse monoclonal, used at 1:1000), anti-MAVS
434 (Thermo Fisher cat. # PA5-17256, rabbit polyclonal, used at 1:1000), anti-$p24^{Gag}$
435 (NIH/NIAID HIV Reagent Program, cat. # ARP-6457, mouse monoclonal clone # 24-2,
436 used at 1:1000). For immunoprecipitation, anti-FLAG (Sigma cat. # F3165), and mouse
437 IgG (Novus cat. # NBP1-97019) were used. Secondary antibodies used were goat anti-
438 mouse DyLight 680 (for western blot analysis, Invitrogen cat. # 35518, used at
439 1:10,000) goat anti-rabbit DyLight 800 (for western blot analysis, Invitrogen cat. # SA5-
440 35571, used at 1:10,000), goat anti-mouse AlexaFluor647 (for anti-$p24^{Gag}$ co-staining
441 with smFISH, Thermo Fisher, cat. # A-21237, used at 1:200).

442 Antiretrovirals and drugs used included lenacapavir (1 μM, NIH HIV Reagent
443 Program, cat. # HRP-20266), efavirenz (1 μM, NIH HIV Reagent Program, cat. # HRP-
444 4624), raltegravir (30 μM, Selleckchem, cat. # S2005), KPT-330 (1 μM, Selleckchem,
445 cat. # S7252), lopinavir (1 μM, NIH HIV Reagent Program, cat. # HRP-9481).

446 **siRNA transfection**

447 Pooled siRNA stocks were obtained from Horizon Discovery (cat. # MDA5: L-
448 013041-00-0005, RIG-I: L-012511-00-0005, MAVS: L-024237-00-0005, control: D-
449 001810-10-05). siRNA was transfected into MDMs two days prior to infection using
450 TransIT-X2 (Mirus, cat. # 6000) at 50 nM final siRNA concentration. To assay for
451 knockdown efficiency, transfected MDMs were washed and either lysed for Western blot
452 analysis on the day of infection in RIPA buffer or lysed for RT-qPCR in Qiagen RLT
453 buffer on the day of infection or at 3 dpi.

454 **CaDAL assay**

455 5' ends of unspliced HIV-1 RNAs were analyzed by cap-dependent ligation/PCR
456 assay as previously described.(22) Total RNA from infected MDMs was analyzed via
457 cDNA synthesis from an RT primer downstream of the major splice donor site to restrict
458 analysis to solely unspliced RNAs. ^{cap}1G/^{cap}3G ratio within a given sample is
459 quantitative, while comparisons between samples or cell types are not quantitative.

460 **RT-qPCR**

461 RT-qPCR was performed as previously described.(18) Briefly, total RNA was
462 isolated using a Qiagen RNEasy kit (cat. # 74106). cDNA was generated from
463 equivalent amounts of total RNA using oligo-dT primers (Invitrogen, cat. # 18418020)
464 and SuperScript III RT (Invitrogen, cat. # 18080051). qPCR was performed with 5-20 ng
465 cDNA using 2X Maxima SYBR Green master mix (Thermo Scientific, cat. # FERK0242)
466 and primer sets outlined in **Table S1**. Quantification was performed using the ΔΔCt
467 method with GAPDH used as an internal control.

468 **Digital droplet PCR**

469 To analyze splicing patterns between NL4-3/1G/3G viruses, RT-ddPCR was
470 performed similarly to previous descriptions with modifications.(24) cDNA was
471 generated from 50 ng total RNA from infected MDMs as described above, using random
472 hexamers as primers. cDNA was diluted 1:1500 (or no-RT control samples were diluted
473 1:150) and mixed with Supermix for Probes (no dUTP) (Bio-Rad, cat. # 186-3023) and
474 primer/probe sets targeting LTR, *nef*, *env*, and *gag* outlined in **Table S1**. Droplets were
475 generated using a Bio-Rad QX200 ddPCR droplet generator. LTR and *nef* probes were
476 read in the FAM channel (with different probe concentrations to distinguish the signals),
477 while *env* and *gag* were read in the HEX channel (with different probe concentrations to
478 distinguish the signals). The final concentration of primers/probes in droplets was: 900
479 nM (all primers), 250 nM (LTR-FAM and *gag*-HEX probes), and 150 nM (*nef*-FAM and
480 *env*-HEX probes). PCR was performed and droplets were read using a Bio-Rad QX200
481 droplet reader. Data was analyzed in QuantaSoft software, and gates were drawn to
482 distinguish the LTR, *nef*, *env*, and *gag* signals. To analyze splicing patterns, only
483 LTR/nef, LTR/env/nef, and LTR/gag/env/nef droplets were compared (all other droplet
484 populations were discarded) and relative transcript abundance was calculated for each
485 primary cell donor.

486 **smRNA-FISH**

487 MDMs were cultured on gelatin-coated glass coverslips at $3x10^5$ cells/well in 24-
488 well plates before infection with LaiΔenvGFP (WT-LTR) or LaiΔenvGFP with NL4-3
489 LTR/backbone sequences (1G-LTR or 3G-LTR) at MOI 0.5 with SIV3+ VLP co-infection
490 (1 ng p27 per well). Infected MDMs were cultured for 3 days before fixation with

491 formaldehyde and permeabilization with DPBS/Tween-20 (0.1%, nuclease-free,
492 Promega, cat. # PRH5152). Coverslips were then blocked with 1% bovine serum
493 albumin (nuclease-free, Sigma-Aldrich, cat. # 126615-25ML) and stained for $p24^{Gag}$.
494 Alternatively, coverslips were stained with wheat germ agglutinin conjugated to
495 AlexaFluorPlus568 (WGA, 1 µg/mL for 10 min on ice, Invitrogen, cat. # W56133) to
496 label plasma membrane before fixation and blocking. After anti-$p24^{Gag}$ immunostaining,
497 coverslips were fixed again with formaldehyde before smFISH.

498 To perform smFISH, stained coverslips were equilibrated with wash buffer A
499 (Stellaris, cat. # SMF-WA1-60) before hybridization in hybridization buffer (Stellaris, cat.
500 # SMF-HB1-10) containing 125 nM probes. De-ionized formamide (Millipore Sigma, cat.
501 # 4610-100ML) and nuclease-free water were added to wash and hybridization buffers
502 before use per the manufacturer's instructions. For $p24^{Gag}$ staining, probes consisted of
503 48 CALFluor590-labeled oligos designed against HIV-1 Lai *gag/pol* as outlined in **Table**
504 **S2** (Stellaris). For WGA-568 staining, probes consisted of 48 Quasar670-labeled oligos
505 also outlined in **Table S2**. Coverslips were hybridized overnight at 37ºC in a humidified
506 chamber. Coverslips were then washed with wash buffer A at 37ºC for 30 min, followed
507 by DAPI staining at 37ºC or 30 min (5 ng/mL in wash buffer A), and then equilibrated
508 with wash buffer B (Stellaris, cat. # SMF-WB1-20) for 5 mins at room temperature
509 before mounting onto a slide using VECTASHIELD mounting medium (Vector
510 Laboratories, cat. # H-1900-10).

511 Images were taken using a Leica SP5 confocal microscope. For colocalization
512 analysis, regions of interest were drawn manually around cells and fluorescent co-
513 localization was quantified using the JACoP plugin for ImageJ/Fiji.

514 **RNA immunoprecipitation**

515 RNA immunoprecipitation in infected HEK293T cells transfected with MDA5-
516 FLAG was performed as described previously.(21) After UV-crosslinking and cytosolic
517 fractionation followed by immunoprecipitation, RNA was extracted from antibody-coated
518 beads using TRIzol/chloroform extraction before cDNA synthesis, DNase treatment, and
519 qPCR. Primers used for qPCR are listed in **Table S1**. Beads were washed in RIPA low-
520 salt (150 mM NaCl) followed by RIPA high-salt buffer (1M NaCl) before RNA extraction
521 or extraction for Western blot analysis. For Western extraction, beads were boiled with
522 neat 6X loading buffer (containing 4% SDS and 2% β-mercaptoethanol) before bead
523 removal and gel loading. Data is reported as fold enrichment of RT-qPCR signal
524 between anti-FLAG and IgG control immunoprecipitation conditions.

525 **Type I interferon bioassay**

526 Type I interferon bioactivity was measured using 293 ISRE-luc cells similar to
527 previous descriptions.(18) $4x10^4$ cells were plated with 100 µL of culture supernatant or
528 IFNα2 standard diluted in R10 medium (1.6-200 U/mL, PBL Assay Science cat. # 11100-
529 1) in 96-well plates and incubated at 37ºC for 24h. Cells were washed in DPBS, lysed in
530 Glo Lysis buffer (Promega, cat. # E266A), and luciferase activity in cell lysates was
531 determined using Bright-Glo reagent (Promega, cat. # E2610) using a LUMIstar Omega
532 reader (BMG LabTech). For quantification, the lowest standard (1.6 U/mL) was
533 considered the limit of detection and samples with lower luminescence than the limit of
534 detection were reported as 0 U/mL.

535 **ELISA**

536 IP-10 secreted into culture supernatants was quantified with ELISA (BD, cat. #
537 550926) after inactivation in DPBS/10% NCS/0.5%TX-100. $p24^{Gag}$ in culture
538 supernatants was detected as previously described using an in-house ELISA.(18)

539 **Western blot analysis**

540 To perform Western blot analysis, 30 µg of protein (quantified with Bradford
541 assay) was diluted to equivalent volume in DPBS/loading buffer (containing SDS and β-
542 mercaptoethanol), boiled, and loaded into Mini-PROTEAN TGX Precast Protein Gels
543 (BioRad, 4-20% gradient, cat. # 4561094) along with a protein ladder (PageRuler
544 Prestained NIR Protein Ladder, Thermo Scientific, cat. # 26635). Blots were blocked
545 with LI-COR Intercept blocking buffer (LI-COR, cat. # NC1660556) before staining and
546 imaging on a LI-COR Odyssey CLx imager.

547 **Flow cytometry**

548 To assess infection efficiency with replication-competent viruses, MDMs were
549 seeded on low-attachment plates and lifted using Cellstripper (Corning, cat. # 25-056-
550 CI). Cells were moved to a U- or V-bottom plate, fixed in 4% paraformaldehyde for 15-
551 30 min at 4ºC, permeabilized with Perm/Wash buffer (BD, cat. # 554723), and stained
552 with anti-$p24^{Gag}$ antibody for 30 min at room temperature. Flow cytometry was
553 performed immediately using a Cytek Aurora spectral analyzer. Data was analyzed and
554 spectral plots were generated in FlowJo 10.8.1.

555 **Statistics**

556 Statistics were performed in GraphPad Prism 10.2.3. Two-sample comparisons
557 were analyzed by t-test. For multi-sample comparisons, one-way or two-way ANOVA

558 were performed with Dunnett's or Tukey's post-test. For MDMs, each data point
559 represents an individual primary cell donor, while for cell lines, each data point
560 represents an independent experiment. Plots represent the mean +/- SEM of the data,
561 ns: $p \geq 0.05$; *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$; ****: $p<0.0001$.

562 **Data Availability**

563 The authors declare that the data that support the findings of this study are
564 available from the corresponding author upon reasonable request.

565 **Acknowledgements**

566 The authors thank the BU-CAMED Flow Cytometry Core Facility for technical
567 assistance and the NIH/NIAID HIV Reagent Program and associated investigators for
568 reagents used in this study. The authors thank NIH for funding used to perform this
569 study. This work was supported by grants R01DA059952 (S.G. and A.H.), R01AI187175
570 (S.G. and A.H.), R01DA055488 (S.G. and A.H.), R01DA051889 (S.G.), and
571 5P30AI042853 (S.G., A.H., and H.A.).

572 **Author contributions**

573 I.H., S.K., A.H., H.A., A.T., and S.G. conceived of and designed the experiments.
574 I.H. and S.G. wrote the manuscript. I.H., H.A., A.H., S.K., A.T. and S.G. edited the
575 manuscript. I.H., S.K., S.R., S.J., J.H., and X.H. performed the experiments.

576 **FIGURE LEGENDS**

577 **Fig. 1: *De novo*-transcribed ^{cap}1G HIV-1 unspliced RNAs stimulate innate immune**
578 **responses in macrophages.** (**A**) Schematic of 2D RNA structures demonstrating the
579 predominant folded 5' leader conformations of ^{cap}1G and ^{cap}3G unspliced HIV-1 RNAs.
580 (**B**) 5' proviral sequences of wild-type NL4-3, ^{cap}1G-only-expressing NL4-3 (1G), and
581 ^{cap}3G-only-expressing NL4-3 (3G) constructs. The GGG transcription initiation tract is
582 highlighted in pink, with ^{cap}1G and ^{cap}3G initiation sites labeled with arrows. In 1G virus,
583 GG is mutated to TC (in red). In 3G virus, TC is inserted (in blue) which pushes the
584 GGG initiation tract 2 nt downstream. (**C**) **Ca**p-**d**ependent **a**daptor **l**igation PCR assay
585 (**CaDAL**) analyzing 5' ends of unspliced HIV-1 RNA expressed in virus-infected MDMs.
586 +EFV refers to NL4-3 infection with efavirenz treatment. (**D**-**I**) Primary human MDMs
587 were co-infected with Vpx-containing SIV3+ VLPs and VSV-G-pseudotyped HIV-1/NL4-
588 3. Plots show (**D**) infection efficiency evidenced by p24^{Gag} immunostaining analyzed by
589 flow cytometry, (**E**) p24^{Gag} secretion into the culture supernatant assayed by ELISA, (**F**)
590 bioactive type I IFN secretion assayed by bioassay, (**G**) IP-10 secretion assayed by
591 ELISA, all at 3 dpi. (**H**-**J**) MDMs were co-infected with SIV3+ VLPs and 1G virus in the
592 presence of antiretrovirals lenacapavir (capsid, 1 μM), efavirenz (reverse transcriptase,
593 1 μM), raltegravir (integrase, 30 μM), KPT-330 (Rev/CRM1, 1 μM), or lopinavir
594 (protease, 1 μM). Plots show relative p24^{Gag} release (**H**, specific for cleaved p24^{Gag}),
595 HIV-1 usRNA expression by RT-qPCR (**I**), and IFN-I release (**J**) at 3 dpi. (**D**-**J**) Data was
596 analyzed by one-way ANOVA with Tukey's post-test (**D**-**G**) or Dunnett's post-test with
597 statistics referring to comparison with DMSO condition (**H**-**J**). (**D**/**E**) All comparisons
598 within virus infection conditions are nonsignificant. (**D**-**J**) Each data point is a primary

599 cell donor with plots showing mean +/- SEM, ns: p≥0.05; *: p<0.05; **: p<0.01; ***:
600 p<0.001; ****: p<0.0001.

601 **Fig. 2: MDA5/MAVS expression is required for ^{cap}1G HIV-1 usRNA-induced innate**
602 **immune responses in macrophages.** MDMs were transfected with siRNA targeting
603 the indicated sensors/adaptor and two days later co-infected with SIV3+ VLPs and VSV-
604 G-pseudotyped HIV-1/NL4-3. (**A**-**C**) RT-qPCR of RNA extracted from uninfected siRNA-
605 transfected MDMs normalized to GAPDH expression. (**D**-**F**) Western blots of uninfected
606 siRNA-transfected MDMs probed with the indicated antibodies. (**G**/**H**) p24^{Gag} release (**G**)
607 and RT-qPCR for HIV-1 usRNA normalized to GAPDH (**H**) at 3 dpi. (**I**/**J**) type I IFN
608 release (**I**) and IP-10 release (**J**) normalized to siCtrl-transfected infected MDMs. (**D**-**F**)
609 Western blots are of one representative primary cell donor. (**A**-**C**,**G**-**J**) Data points are
610 individual primary cell donors. Plots show mean +/- SEM analyzed by paired t-test (**A**-**C**)
611 or two-way ANOVA with Tukey's post-test (**G**/**H**) or Dunnett's post-test with statistics
612 referring to siCtrl-transfected condition (**I**/**J**), ns: p≥0.05; **: p<0.01; ***: p<0.001; ****:
613 p<0.0001. (**G**/**H**) all comparisons within the indicated virus infection condition are
614 nonsignificant.

615 **Fig. 3: MDA5 specifically engages ^{cap}1G HIV-1 usRNA.** (**A**) HEK293T cells were
616 infected with VSV-G-pseudotyped replication-competent viruses in the presence or
617 absence of 1 μM efavirenz. Infected cells were transfected with MDA5-FLAG plasmid.
618 Cells were UV-crosslinked and cytosolic fractions underwent RNA immunoprecipitation
619 at 2 dpi using anti-FLAG or IgG control antibodies. (**B**-**D**) Plots show RT-qPCR fold
620 enrichment of MDA5-FLAG signal over IgG control antibody conditions for (**B**) HIV-1
621 unspliced RNA, (**C**) HIV-1 multiply spliced RNA, and (**D**) actin mRNA. Plots show mean

622 +/- SEM and each data point is an independent experiment. (**B**) One-way ANOVA with
623 Dunnett's post-test was used with statistics referring to comparison with NL4-3
624 condition, ns: p≥0.05; **: p<0.01. (**C**/**D**) One-way ANOVA with Tukey's post-test was
625 used, all comparisons nonsignificant.

626 **Fig. 4: Transcription and intracellular localization of ^{cap}1G and ^{cap}3G usRNA in**
627 **macrophages.** (**A**) CaDAL assay analyzing unspliced HIV-1 transcripts using RNA
628 isolated from syngeneic activated primary CD4+ T cells and MDMs infected with
629 replication-competent VSV-G-pseudotyped NL4-3 (with SIV3+ VLP co-infection for
630 MDMs) with or without efavirenz pre-treatment. Comparisons within samples are
631 quantitative, comparisons between samples are not quantitative. (**B**) Relative RT-qPCR
632 for unspliced HIV-1 transcripts using RNA isolated from MDMs co-infected with HIV-
633 1/SIV3+ VLPs. (**C**) ddPCR transcriptomic analysis of RNA used in (**B**). cDNA was
634 generated using random hexamers and analyzed with multiplex ddPCR analyzing LTR,
635 *gag*, *env*, and *nef* expression. Droplets containing single cDNA molecules were classed
636 as multiply-spliced (LTR/nef), singly-spliced (LTR/env/nef), or unspliced
637 (LTR/gag/env/nef) or discarded (other combinations). Transcript abundance was
638 normalized for each donor (all three populations totaling 100% for each donor). (**D**-**F**)
639 MDMs were co-infected with SIV3+ VLPs and single-round (Δenv/GFP) viruses
640 harboring wild-type, 1G, or 3G LTRs. smRNA-FISH was performed targeting unspliced
641 *gag* RNA with WGA (**D**, wheat germ agglutinin, plasma membrane stain) or anti-p24^{Gag}
642 co-stain (**E**). (**F**/**G**) Plots show % colocalization of *gag*RNA signal with WGA or p24^{Gag}.
643 (**A**) CaDAL was performed with two primary cell donors. (**B**/**C**) Data was analyzed with
644 one-way or two-way ANOVA with Tukey's post-test. Data is shown as mean +/- SEM, all

645 data points are individual primary cell donors, ns: p≥0.05; *: p<0.05; **: p<0.01; ***:
646 p<0.001. (**D**/**E**) Images show one representative field. (**F**/**G**) Data points are individual
647 fields of view within one experiment, representative of over three independent
648 experiments. Data was analyzed using one-way ANOVA with Tukey's post-test, ns:
649 p≥0.05; **: p<0.01; ****: p<0.0001.

650 **Fig. 5: 5' cap sequestration on HIV-1 usRNA determines MDA5-driven**
651 **immunogenicity.** (**A**) Schematic of the predominant folded 5' leader structure of ^{cap}3G
652 unspliced HIV-1 RNA with 'no bulge' (NB) contact-stabilizing mutations in the polyA
653 hairpin. (**B**-**F**) MDMs were co-infected with replication-competent VSV-G-pseudotyped
654 HIV-1/SIV3+ VLPs. Plots show (at 3 dpi) (**B**) infection efficiency analyzed by flow
655 cytometry after anti-p24^{Gag} antibody staining, (**C**) p24^{Gag} release into the culture
656 medium, (**D**) relative bioactive type I IFN secretion into the culture medium, (**E**) relative
657 RT-qPCR for HIV-1 usRNA normalized to GAPDH expression, and (**F**) p24^{Gag} flow
658 cytometry mean fluorescence intensity (MFI) of p24^{Gag}-positive cells. Plots show mean
659 +/- SEM and each data point is an individual primary cell donor. Data was analyzed by
660 one-way ANOVA with Tukey's post-test, ns: p≥0.05; *: p<0.05; **: p<0.01; ***: p<0.001;
661 ****: p<0.0001.

662 **Fig. 6: MDA5 sensing of ^{cap}1G transcripts requires Rev/CRM1 nuclear export.** (**A**)
663 Schematic of HIV-1Δenv/GFP reporter virus genomes cloned into NL4-3 ^{cap}1G-only LTR
664 proviral backbone. (**B**-**F**) MDMs were co-infected with VSV-G-pseudotyped reporter
665 viruses and SIV3+ VLPs with or without efavirenz (EFV) treatment. Plots show GFP
666 positivity (**B**), bioactive type I IFN release (**C**), IP-10 release (**D**), relative usRNA
667 abundance (**E**), and Gag expression analyzed by Western blot (**F**) at 3 dpi. (**G**) Model of

668 the findings of this and previous studies, highlighting cytoplasmic pools of HIV-1 usRNA
669 subject to different fates and immunological outcomes. Plots show mean +/- SEM and
670 each data point is an individual primary cell donor. Data was analyzed by one-way
671 ANOVA with Tukey's post-test, ns: $p \geq 0.05$; **: $p < 0.01$; ***: $p < 0.001$; ****: $p < 0.0001$.

672 **SUPPLEMENTAL FIGURE LEGENDS**

673 **Fig. S1: ^{cap}1G unspliced HIV-1 RNA expression stimulates type I interferon release**
674 **in macrophages.** MDMs were infected with HIV-1 alone (no SIV3+ VLP addition) and
675 culture supernatants were analyzed at 3, 6, or 9 days post-infection. Plots show $p24^{Gag}$
676 release (**A**) and IFN-I release (**B**) at the indicated timepoint. (**C**) Western blot analyzing
677 Gag expression, harvested at 9 dpi from cells used in (**A**/**B**). Data points are individual
678 primary cell donors, plots show mean +/- SEM. Data was analyzed by two-way ANOVA
679 with Tukey's post-test (**A**) or Dunnett's post-test with statistics referring to comparison
680 with mock infection (**B**). (**A**) All virus infection comparisons nonsignificant, comparing
681 the same timepoint across virus infections. (**B**) **: $p < 0.01$ comparing NL4-3 and 1G
682 infection to mock infection at 6 dpi, all other comparisons nonsignificant ($p \geq 0.05$). (**C**)
683 Western blot was performed with two individual primary cell donors.

684 **Fig. S2: ddPCR for viral transcriptomic analysis in primary macrophages.**
685 Additional data from experiment described in **Fig. 4C**. (**A**) Raw transcript abundance
686 values obtained from ddPCR, NRT: no-RT control. (**B**) Sample gating strategy to
687 distinguish viral RNA populations, for +RT samples diluted 1:1500 or NRT controls
688 diluted 1:150. (**A**) Plot is shown as mean +/- SEM, each data point is an individual
689 primary cell donor. (**B**) Data is from one primary cell donor.

690 **Fig. S3: smFISH of ^{cap}1G/3G usRNA expression in primary macrophages.**
691 Additional data from experiment described in **Fig. 4D**-**G**. Additional magnified images
692 and colocalization quantification showing smRNA-FISH analysis merged with WGA (**A**,
693 wheat germ agglutinin, plasma membrane stain) or anti-p24^{Gag} co-stain (**C**). In (**B**), % of
694 *gag* RNA signal colocalizing with WGA is shown. EFV: reverse transcriptase inhibitor
695 efavirenz treatment. WT-LTR/M10: wild-type ^{cap}1G/3G-expressing virus with inactivating
696 M10 mutation in Rev. Means +/- SEM of individual fields of view from one representative
697 experiment are shown, with data analyzed by one-way ANOVA with Tukey's post-test,
698 ****: $p<0.0001$.

699 **Fig. S4: p24^{Gag} production efficiency in 3G and 3G-NB infection.** Additional data
700 from experiment described in **Fig. 5B**. (**A**) Plots show log p24^{Gag} flow cytometry signal
701 as a histogram with positive population quantification for four primary cell donors.

REFERENCES

1. W. Li *et al.*, Characterizing transcripts of HIV-1 different substrains using direct RNA sequencing. *Heliyon* **10**, e39474 (2024).
2. K. E. Ocwieja *et al.*, Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. *Nucleic Acids Res* **40**, 10345-10355 (2012).
3. A. Emery, R. Swanstrom, HIV-1: To Splice or Not to Splice, That Is the Question. *Viruses* **13** (2021).
4. J. Katahira, Nuclear export of messenger RNA. *Genes (Basel)* **6**, 163-184 (2015).
5. M. Hadzopoulou-Cladaras *et al.*, The rev (trs/art) protein of human immunodeficiency virus type 1 affects viral mRNA and protein expression via a cis-acting sequence in the env region. *J Virol* **63**, 1265-1274 (1989).
6. V. W. Pollard, M. H. Malim, The HIV-1 Rev protein. *Annu Rev Microbiol* **52**, 491-532 (1998).
7. J. M. Watts *et al.*, Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**, 711-716 (2009).
8. T. Jacks *et al.*, Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature* **331**, 280-283 (1988).
9. J. Fernandes, B. Jayaraman, A. Frankel, The HIV-1 Rev response element: an RNA scaffold that directs the cooperative assembly of a homo-oligomeric ribonucleoprotein complex. *RNA Biol* **9**, 6-11 (2012).
10. S. Kharytonchyk *et al.*, Transcriptional start site heterogeneity modulates the structure and function of the HIV-1 genome. *Proc Natl Acad Sci U S A* **113**, 13378-13383 (2016).
11. T. Masuda *et al.*, Fate of HIV-1 cDNA intermediates during reverse transcription is dictated by transcription initiation site of virus genomic RNA. *Sci Rep* **5**, 17680 (2015).
12. J. D. Brown *et al.*, Structural basis for transcriptional start site control of HIV-1 RNA fate. *Science* **368**, 413-417 (2020).
13. O. A. Nikolaitchik *et al.*, Selective packaging of HIV-1 RNA genome is guided by the stability of 5' untranslated region polyA stem. *Proc Natl Acad Sci U S A* **118** (2021).
14. L. Ye *et al.*, Short- and long-range interactions in the HIV-1 5' UTR regulate genome dimerization and packaging. *Nat Struct Mol Biol* **29**, 306-319 (2022).
15. A. Rein, RNA Packaging in HIV. *Trends Microbiol* **27**, 715-723 (2019).
16. Z. Cheng *et al.*, Translation of HIV-1 unspliced RNA is regulated by 5' untranslated region structure. *J Virol* **98**, e0116024 (2024).
17. O. A. Nikolaitchik *et al.*, HIV-1 usurps transcription start site heterogeneity of host RNA polymerase II to maximize replication fitness. *Proc Natl Acad Sci U S A* **120**, e2305103120 (2023).
18. H. Akiyama *et al.*, HIV-1 intron-containing RNA expression induces innate immune activation and T cell dysfunction. *Nat Commun* **9**, 3450 (2018).
19. S. M. McCauley *et al.*, Intron-containing RNA from the HIV-1 provirus activates type I interferon and inflammatory cytokines. *Nat Commun* **9**, 5305 (2018).

747 20. M. H. Guney *et al.*, IFIH1 (MDA5) is required for innate immune detection of
748 intron-containing RNA expressed from the HIV-1 provirus. *Proc Natl Acad Sci U S*
749 *A* **121**, e2404349121 (2024).
750 21. S. Ramaswamy *et al.*, The macrophage-intrinsic MDA5-IRF5 axis drives HIV-1
751 intron-containing RNA-induced inflammatory responses. *J Clin Invest*
752 10.1172/JCI187663 (2025).
753 22. S. Kharytonchyk, C. Burnett, K. Gc, A. Telesnitsky, Transcription start site
754 heterogeneity and its role in RNA fate determination distinguish HIV-1 from other
755 retroviruses and are mediated by core promoter elements. *J Virol* **97**, e0081823
756 (2023).
757 23. K. Hrecka *et al.*, Vpx relieves inhibition of HIV-1 infection of macrophages
758 mediated by the SAMHD1 protein. *Nature* **474**, 658-661 (2011).
759 24. N. A. J. Cassidy *et al.*, HIV reservoir quantification using cross-subtype multiplex
760 ddPCR. *iScience* **25**, 103615 (2022).
761 25. J. M. Esquiaqui, S. Kharytonchyk, D. Drucker, A. Telesnitsky, HIV-1 spliced RNAs
762 display transcription start site bias. *RNA* **26**, 708-714 (2020).
763 26. M. H. Malim, S. Bohnlein, J. Hauber, B. R. Cullen, Functional dissection of the
764 HIV-1 Rev trans-activator--derivation of a trans-dominant repressor of Rev
765 function. *Cell* **58**, 205-214 (1989).
766 27. P. Ding *et al.*, 5'-Cap sequestration is an essential determinant of HIV-1 genome
767 packaging. *Proc Natl Acad Sci U S A* **118** (2021).
768 28. S. Yasin *et al.*, Role of RNA structural plasticity in modulating HIV-1 genome
769 packaging and translation. *Proc Natl Acad Sci U S A* **121**, e2407400121 (2024).
770 29. A. E. Pasquinelli *et al.*, The constitutive transport element (CTE) of Mason-Pfizer
771 monkey virus (MPMV) accesses a cellular mRNA export pathway. *EMBO J* **16**,
772 7500-7510 (1997).
773 30. J. Rehwinkel, M. U. Gack, RIG-I-like receptors: their regulation and roles in RNA
774 sensing. *Nat Rev Immunol* **20**, 537-551 (2020).
775 31. K. A. Cottrell, R. J. Andrews, B. L. Bass, The competitive landscape of the dsRNA
776 world. *Mol Cell* **84**, 107-119 (2024).
777 32. K. Gc *et al.*, HIV-1 single-transcription start-site mutants display complementary
778 replication functions that are restored by reversion. *J Virol* **99**, e0213924 (2025).
779 33. W. B. Puryear *et al.*, Interferon-inducible mechanism of dendritic cell-mediated
780 HIV-1 dissemination is dependent on Siglec-1/CD169. *PLoS Pathog* **9**, e1003291
781 (2013).
782 34. J. E. Hammonds *et al.*, Siglec-1 initiates formation of the virus-containing
783 compartment and enhances macrophage-to-T cell transmission of HIV-1. *PLoS*
784 *Pathog* **13**, e1006181 (2017).
785 35. P. Dave *et al.*, Single-molecule imaging reveals translation-dependent
786 destabilization of mRNAs. *Mol Cell* **83**, 589-606 e586 (2023).
787 36. M. Mougel *et al.*, NXF1 and CRM1 nuclear export pathways orchestrate nuclear
788 export, translation and packaging of murine leukaemia retrovirus unspliced RNA.
789 *RNA Biol* **17**, 528-538 (2020).
790 37. S. Kharytonchyk *et al.*, Influence of gag and RRE Sequences on HIV-1 RNA
791 Packaging Signal Structure and Function. *J Mol Biol* **430**, 2066-2079 (2018).

792 38. G. M. Pocock, J. T. Becker, C. M. Swanson, P. Ahlquist, N. M. Sherer, HIV-1 and
793 M-PMV RNA Nuclear Export Elements Program Viral Genomes for Distinct
794 Cytoplasmic Trafficking Behaviors. *PLoS Pathog* **12**, e1005565 (2016).
795 39. J. Chen *et al.*, Impact of Nuclear Export Pathway on Cytoplasmic HIV-1 RNA
796 Transport Mechanism and Distribution. *mBio* **11** (2020).
797 40. E. Decroly, F. Ferron, J. Lescar, B. Canard, Conventional and unconventional
798 mechanisms for capping viral mRNA. *Nat Rev Microbiol* **10**, 51-65 (2011).
799 41. J. L. Hyde, M. S. Diamond, Innate immune restriction and antagonism of viral
800 RNA lacking 2′-O methylation. *Virology* **479-480**, 66-74 (2015).
801 42. R. Zust *et al.*, Ribose 2'-O-methylation provides a molecular signature for the
802 distinction of self and non-self mRNA dependent on the RNA sensor Mda5. *Nat*
803 *Immunol* **12**, 137-143 (2011).
804 43. G. Mandilara *et al.*, The Role of Coronavirus RNA-Processing Enzymes in Innate
805 Immune Evasion. *Life (Basel)* **11** (2021).
806 44. G. Wolff, C. E. Melia, E. J. Snijder, M. Barcena, Double-Membrane Vesicles as
807 Platforms for Viral Replication. *Trends Microbiol* **28**, 1022-1033 (2020).
808 45. J. Saraste, K. Prydz, Assembly and Cellular Exit of Coronaviruses: Hijacking an
809 Unconventional Secretory Pathway from the Pre-Golgi Intermediate
810 Compartment via the Golgi Ribbon to the Extracellular Space. *Cells* **10** (2021).

811

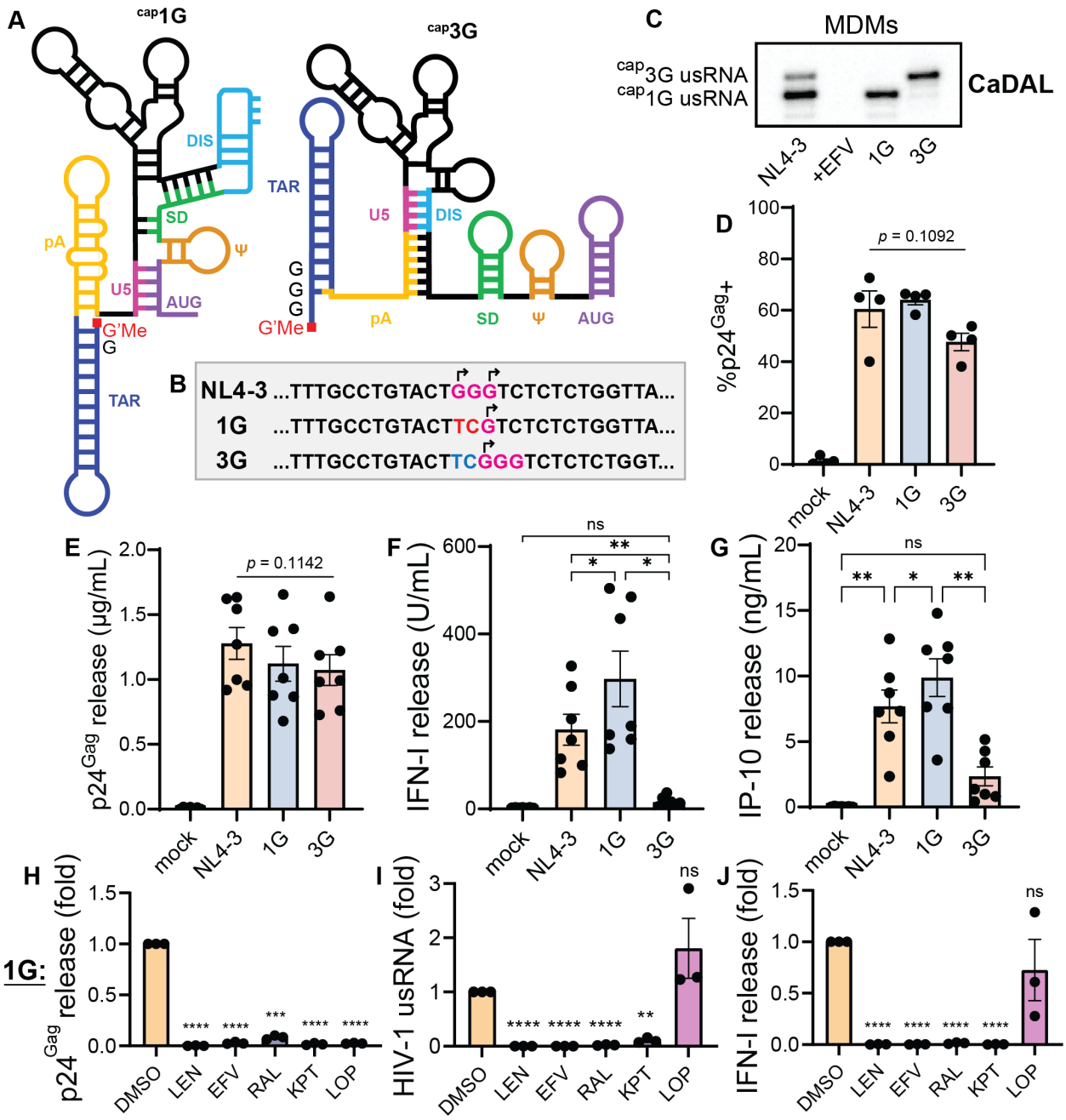

Fig. 1: *De novo*-transcribed capG1G HIV-1 unspliced RNAs stimulate innate immune responses in macrophages. (**A**) Schematic of 2D RNA structures demonstrating the predominant folded 5' leader conformations of capG1G and capG3G unspliced HIV-1 RNAs. (**B**) 5' proviral sequences of wild-type NL4-3, capG1G-only-expressing NL4-3 (1G), and capG3G-only-expressing NL4-3 (3G) constructs. The GGG transcription initiation tract is highlighted in pink, with capG1G and capG3G initiation sites labeled with arrows. In 1G virus, GG is mutated to TC (in red). In 3G virus, TC is inserted (in blue) which pushes the GGG initiation tract 2 nt downstream. (**C**) **Ca**p-**d**ependent **a**daptor **l**igation PCR assay (**CaDAL**) analyzing 5' ends of unspliced HIV-1 RNA expressed in virus-infected MDMs.

+EFV refers to NL4-3 infection with efavirenz treatment. (**D**-**I**) Primary human MDMs were co-infected with Vpx-containing SIV3+ VLPs and VSV-G-pseudotyped HIV-1/NL4-3. Plots show (**D**) infection efficiency evidenced by $p24^{Gag}$ immunostaining analyzed by flow cytometry, (**E**) $p24^{Gag}$ secretion into the culture supernatant assayed by ELISA, (**F**) bioactive type I IFN secretion assayed by bioassay, (**G**) IP-10 secretion assayed by ELISA, all at 3 dpi. (**H**-**J**) MDMs were co-infected with SIV3+ VLPs and 1G virus in the presence of antiretrovirals lenacapavir (capsid, 1 μM), efavirenz (reverse transcriptase, 1 μM), raltegravir (integrase, 30 μM), KPT-330 (Rev/CRM1, 1 μM), or lopinavir (protease, 1 μM). Plots show relative $p24^{Gag}$ release (**H**, specific for cleaved $p24^{Gag}$), HIV-1 usRNA expression by RT-qPCR (**I**), and IFN-I release (**J**) at 3 dpi. (**D**-**J**) Data was analyzed by one-way ANOVA with Tukey's post-test (**D**-**G**) or Dunnett's post-test with statistics referring to comparison with DMSO condition (**H**-**J**). (**D**/**E**) All comparisons within virus infection conditions are nonsignificant. (**D**-**J**) Each data point is a primary cell donor with plots showing mean +/- SEM, ns: $p \geq 0.05$; *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$; ****: $p<0.0001$.

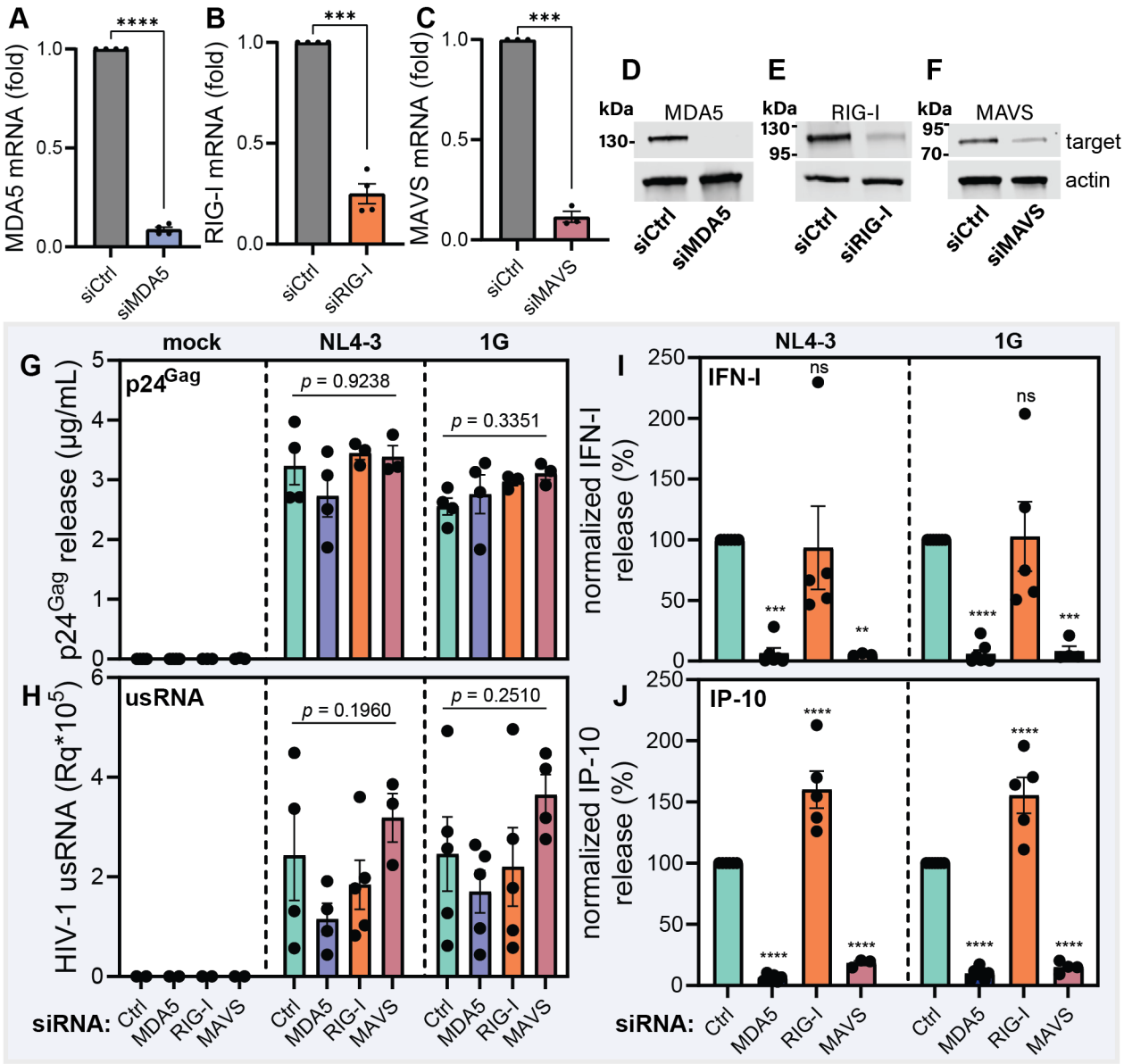

Fig. 2: MDA5/MAVS expression is required for ^{cap}1G HIV-1 usRNA-induced innate immune responses in macrophages. MDMs were transfected with siRNA targeting the indicated sensors/adaptor and two days later co-infected with SIV3+ VLPs and VSV-G-pseudotyped HIV-1/NL4-3. (**A**-**C**) RT-qPCR of RNA extracted from uninfected siRNA-transfected MDMs normalized to GAPDH expression. (**D**-**F**) Western blots of uninfected siRNA-transfected MDMs probed with the indicated antibodies. (**G**/**H**) p24^{Gag} release (**G**) and RT-qPCR for HIV-1 usRNA normalized to GAPDH (**H**) at 3 dpi. (**I**/**J**) type I IFN release (**I**) and IP-10 release (**J**) normalized to siCtrl-transfected infected MDMs. (**D**-**F**) Western blots are of one representative primary cell donor. (**A**-**C**,**G**-**J**) Data points are individual primary cell donors. Plots show mean +/- SEM analyzed by paired t-test (**A**-**C**)

or two-way ANOVA with Tukey's post-test (**G**/**H**) or Dunnett's post-test with statistics referring to siCtrl-transfected condition (**I**/**J**), ns: $p \geq 0.05$; **: $p < 0.01$; ***: $p < 0.001$; ****: $p < 0.0001$. (**G**/**H**) all comparisons within the indicated virus infection condition are nonsignificant.

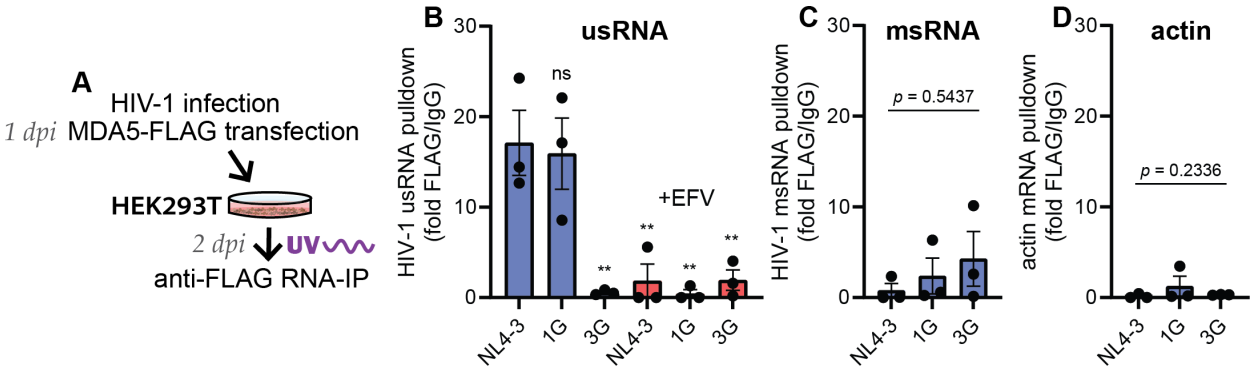

Fig. 3: MDA5 specifically engages ^{cap}1G HIV-1 usRNA. (**A**) HEK293T cells were infected with VSV-G-pseudotyped replication-competent viruses in the presence or absence of 1 μM efavirenz. Infected cells were transfected with MDA5-FLAG plasmid. Cells were UV-crosslinked and cytosolic fractions underwent RNA immunoprecipitation at 2 dpi using anti-FLAG or IgG control antibodies. (**B**-**D**) Plots show RT-qPCR fold enrichment of MDA5-FLAG signal over IgG control antibody conditions for (**B**) HIV-1 unspliced RNA, (**C**) HIV-1 multiply spliced RNA, and (**D**) actin mRNA. Plots show mean +/- SEM and each data point is an independent experiment. (**B**) One-way ANOVA with Dunnett's post-test was used with statistics referring to comparison with NL4-3 condition, ns: $p \geq 0.05$; **: $p < 0.01$. (**C**/**D**) One-way ANOVA with Tukey's post-test was used, all comparisons nonsignificant.

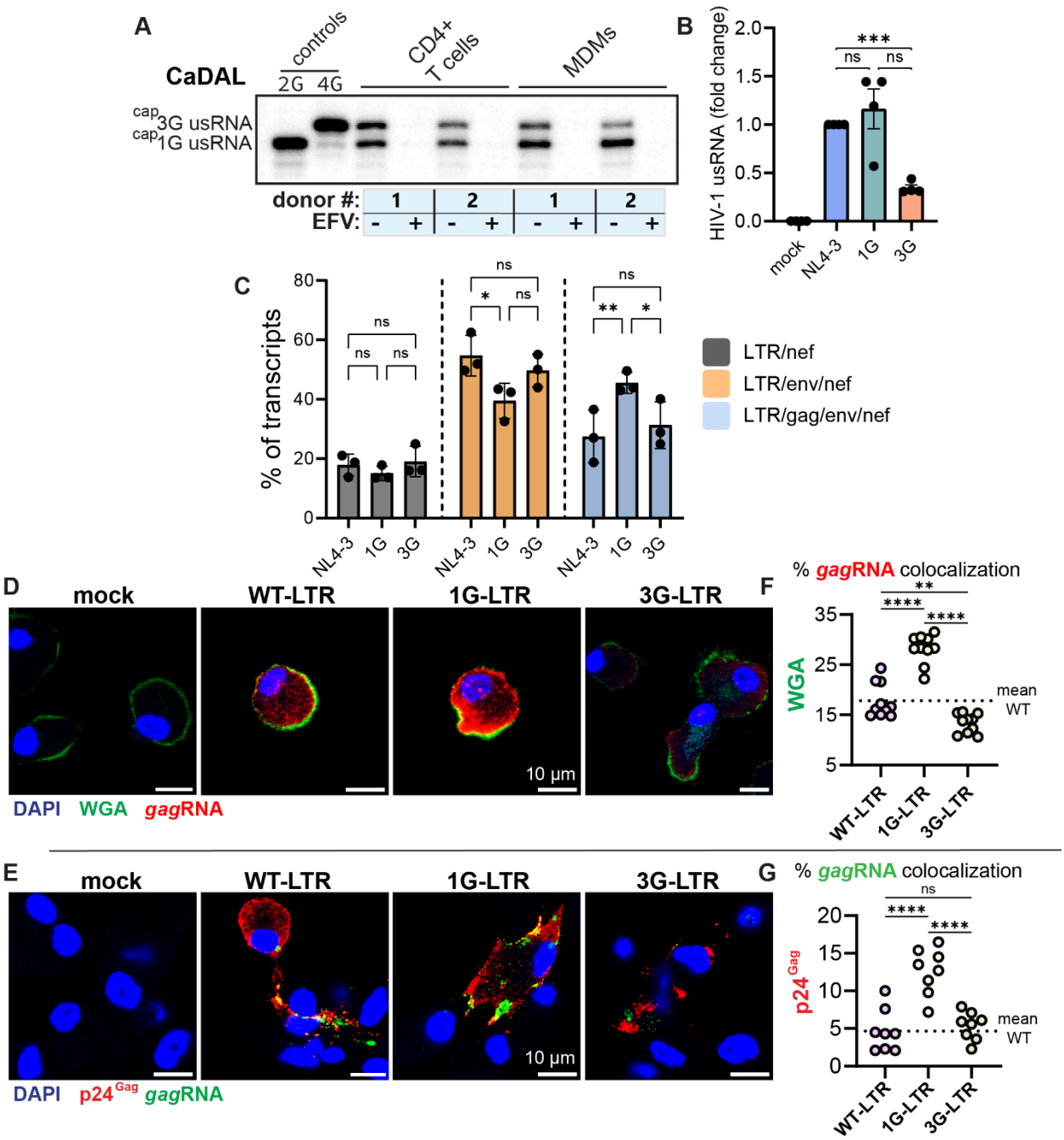

Fig. 4: Transcription and intracellular localization of ^{cap}1G and ^{cap}3G usRNA in macrophages. (**A**) CaDAL assay analyzing unspliced HIV-1 transcripts using RNA isolated from syngeneic activated primary CD4+ T cells and MDMs infected with replication-competent VSV-G-pseudotyped NL4-3 (with SIV3+ VLP co-infection for MDMs) with or without efavirenz pre-treatment. Comparisons within samples are quantitative, comparisons between samples are not quantitative. (**B**) Relative RT-qPCR for unspliced HIV-1 transcripts using RNA isolated from MDMs co-infected with HIV-1/SIV3+ VLPs. (**C**) ddPCR transcriptomic analysis of RNA used in (**B**). cDNA was

generated using random hexamers and analyzed with multiplex ddPCR analyzing LTR, *gag*, *env*, and *nef* expression. Droplets containing single cDNA molecules were classed as multiply-spliced (LTR/nef), singly-spliced (LTR/env/nef), or unspliced (LTR/gag/env/nef) or discarded (other combinations). Transcript abundance was normalized for each donor (all three populations totaling 100% for each donor). (**D**-**F**) MDMs were co-infected with SIV3+ VLPs and single-round (Δenv/GFP) viruses harboring wild-type, 1G, or 3G LTRs. smRNA-FISH was performed targeting unspliced *gag* RNA with WGA (**D**, wheat germ agglutinin, plasma membrane stain) or anti-$p24^{Gag}$ co-stain (**E**). (**F**/**G**) Plots show % colocalization of *gag*RNA signal with WGA or $p24^{Gag}$. (**A**) CaDAL was performed with two primary cell donors. (**B**/**C**) Data was analyzed with one-way or two-way ANOVA with Tukey's post-test. Data is shown as mean +/- SEM, all data points are individual primary cell donors, ns: $p \geq 0.05$; *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$. (**D**/**E**) Images show one representative field. (**F**/**G**) Data points are individual fields of view within one experiment, representative of over three independent experiments. Data was analyzed using one-way ANOVA with Tukey's post-test, ns: $p \geq 0.05$; **: $p<0.01$; ****: $p<0.0001$.

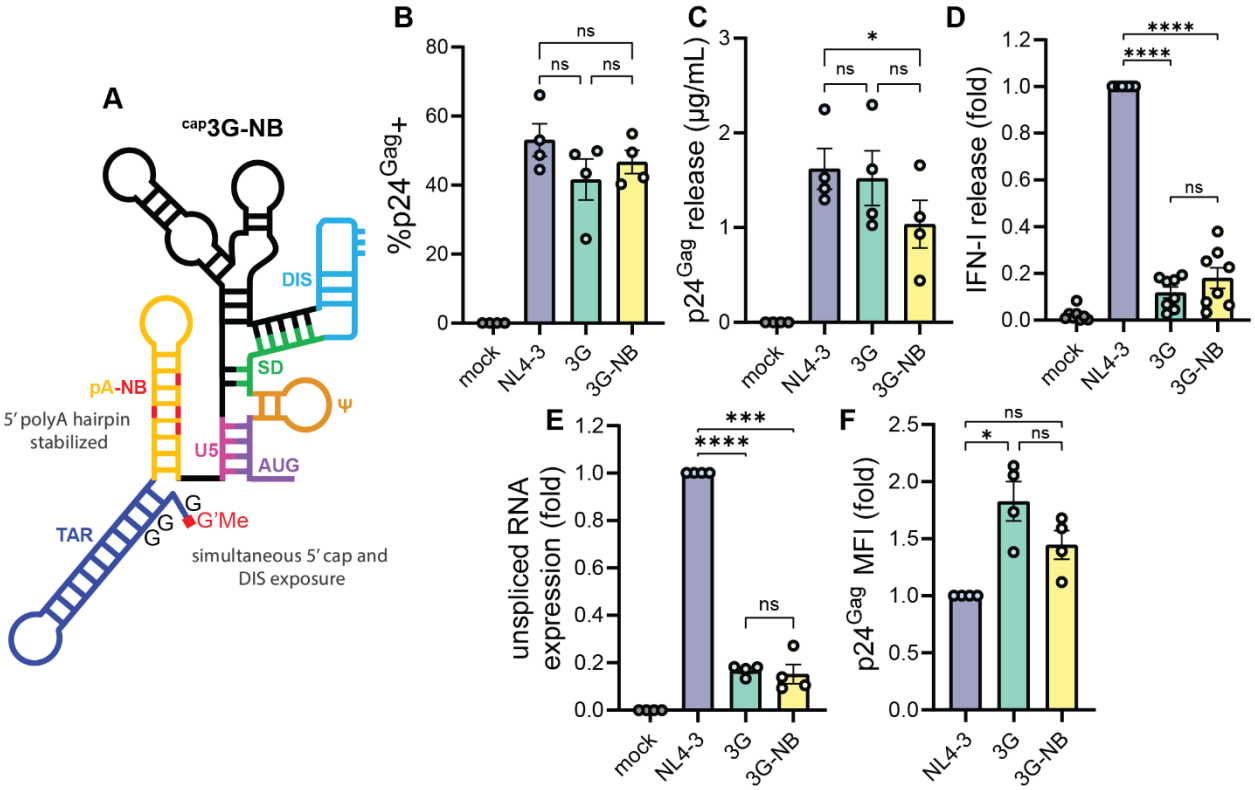

Fig. 5: 5' cap sequestration on HIV-1 usRNA determines MDA5-driven immunogenicity. (**A**) Schematic of the predominant folded 5' leader structure of ^{cap}3G unspliced HIV-1 RNA with 'no bulge' (NB) contact-stabilizing mutations in the polyA hairpin. (**B-F**) MDMs were co-infected with replication-competent VSV-G-pseudotyped HIV-1/SIV3+ VLPs. Plots show (at 3 dpi) (**B**) infection efficiency analyzed by flow cytometry after anti-p24^{Gag} antibody staining, (**C**) p24^{Gag} release into the culture medium, (**D**) relative bioactive type I IFN secretion into the culture medium, (**E**) relative RT-qPCR for HIV-1 usRNA normalized to GAPDH expression, and (**F**) p24^{Gag} flow cytometry mean fluorescence intensity (MFI) of p24^{Gag}-positive cells. Plots show mean +/- SEM and each data point is an individual primary cell donor. Data was analyzed by one-way ANOVA with Tukey's post-test, ns: p≥0.05; *: p<0.05; **: p<0.01; ***: p<0.001; ****: p<0.0001.

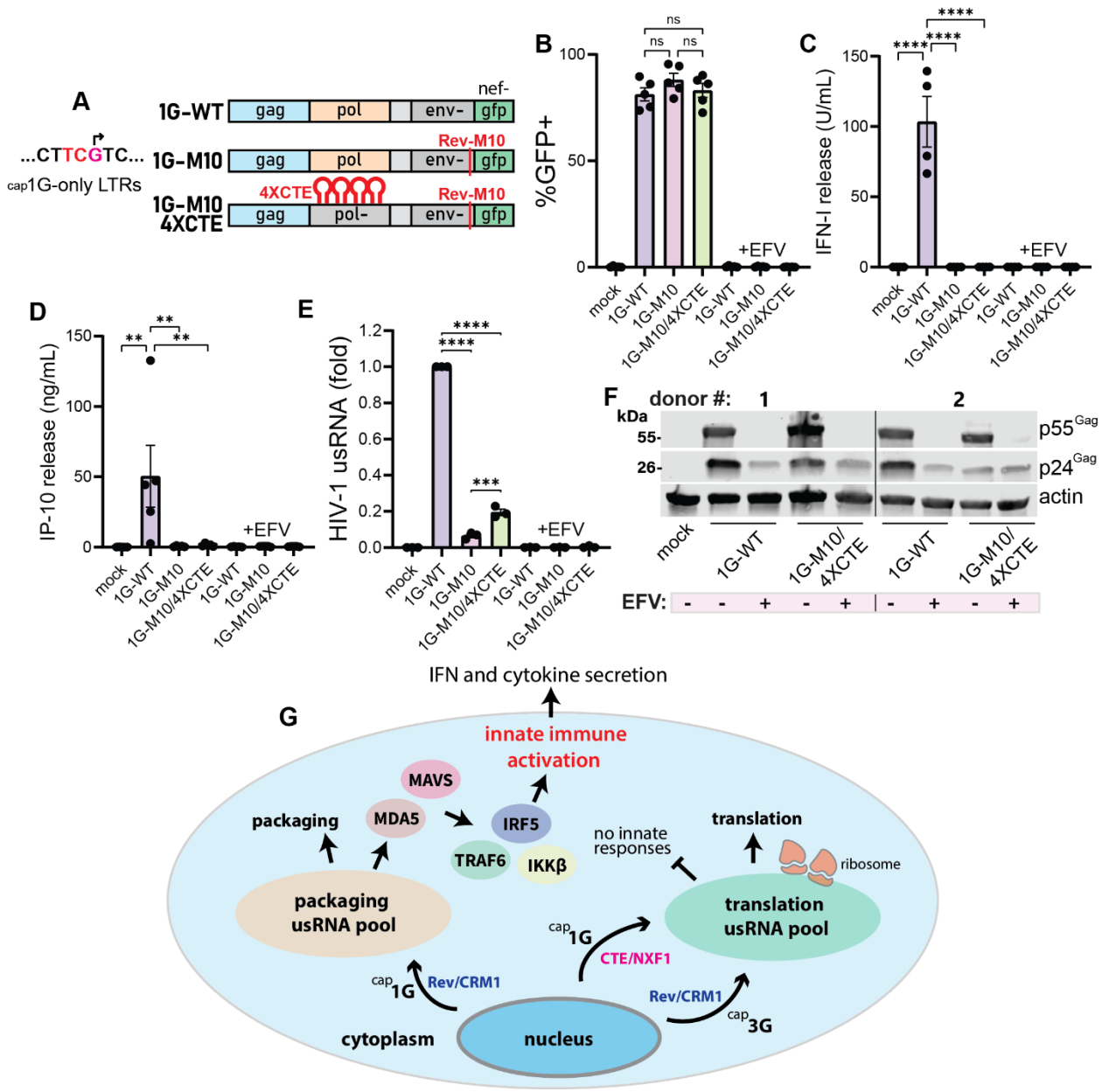

Fig. 6: MDA5 sensing of capG transcripts requires Rev/CRM1 nuclear export. (**A**) Schematic of HIV-1Δenv/GFP reporter virus genomes cloned into NL4-3 capG-only LTR proviral backbone. (**B**-**F**) MDMs were co-infected with VSV-G-pseudotyped reporter viruses and SIV3+ VLPs with or without efavirenz (EFV) treatment. Plots show GFP positivity (**B**), bioactive type I IFN release (**C**), IP-10 release (**D**), relative usRNA abundance (**E**), and Gag expression analyzed by Western blot (**F**) at 3 dpi. (**G**) Model of the findings of this and previous studies, highlighting cytoplasmic pools of HIV-1 usRNA subject to different fates and immunological outcomes. Plots show mean +/- SEM and each data point is an individual primary cell donor. Data was analyzed by one-way ANOVA with Tukey's post-test, ns: p≥0.05; **: p<0.01; ***: p<0.001; ****: p<0.0001.

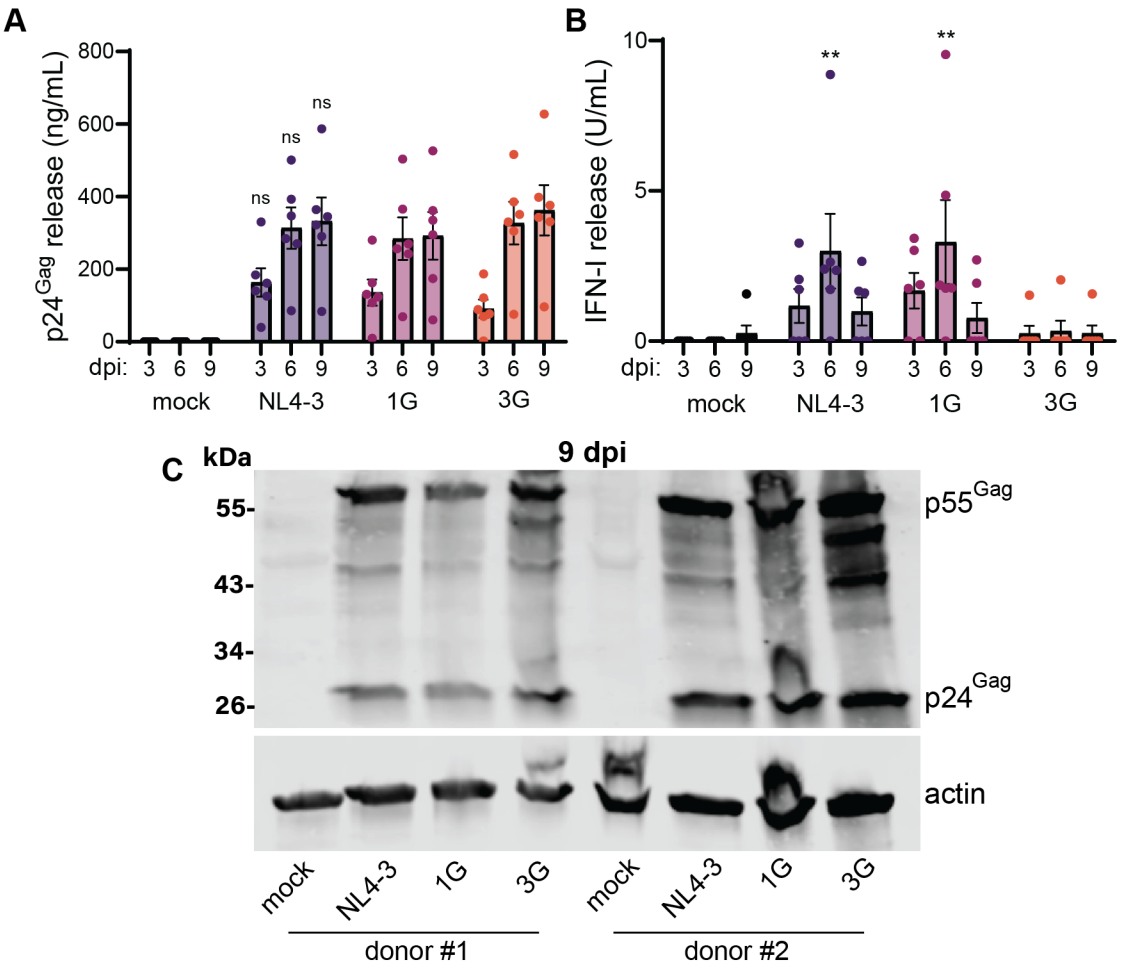

Fig. S1: ^{cap}1G unspliced HIV-1 RNA expression stimulates type I interferon release in macrophages. MDMs were infected with HIV-1 alone (no SIV3+ VLP addition) and culture supernatants were analyzed at 3, 6, or 9 days post-infection. Plots show p24^{Gag} release (**A**) and IFN-I release (**B**) at the indicated timepoint. (**C**) Western blot analyzing Gag expression, harvested at 9 dpi from cells used in (**A**/**B**). Data points are individual primary cell donors, plots show mean +/- SEM. Data was analyzed by two-way ANOVA with Tukey's post-test (**A**) or Dunnett's post-test with statistics referring to comparison with mock infection (**B**). (**A**) All virus infection comparisons nonsignificant, comparing the same timepoint across virus infections. (**B**) **: $p<0.01$ comparing NL4-3 and 1G infection to mock infection at 6 dpi, all other comparisons nonsignificant ($p\geq0.05$). (**C**) Western blot was performed with two individual primary cell donors.

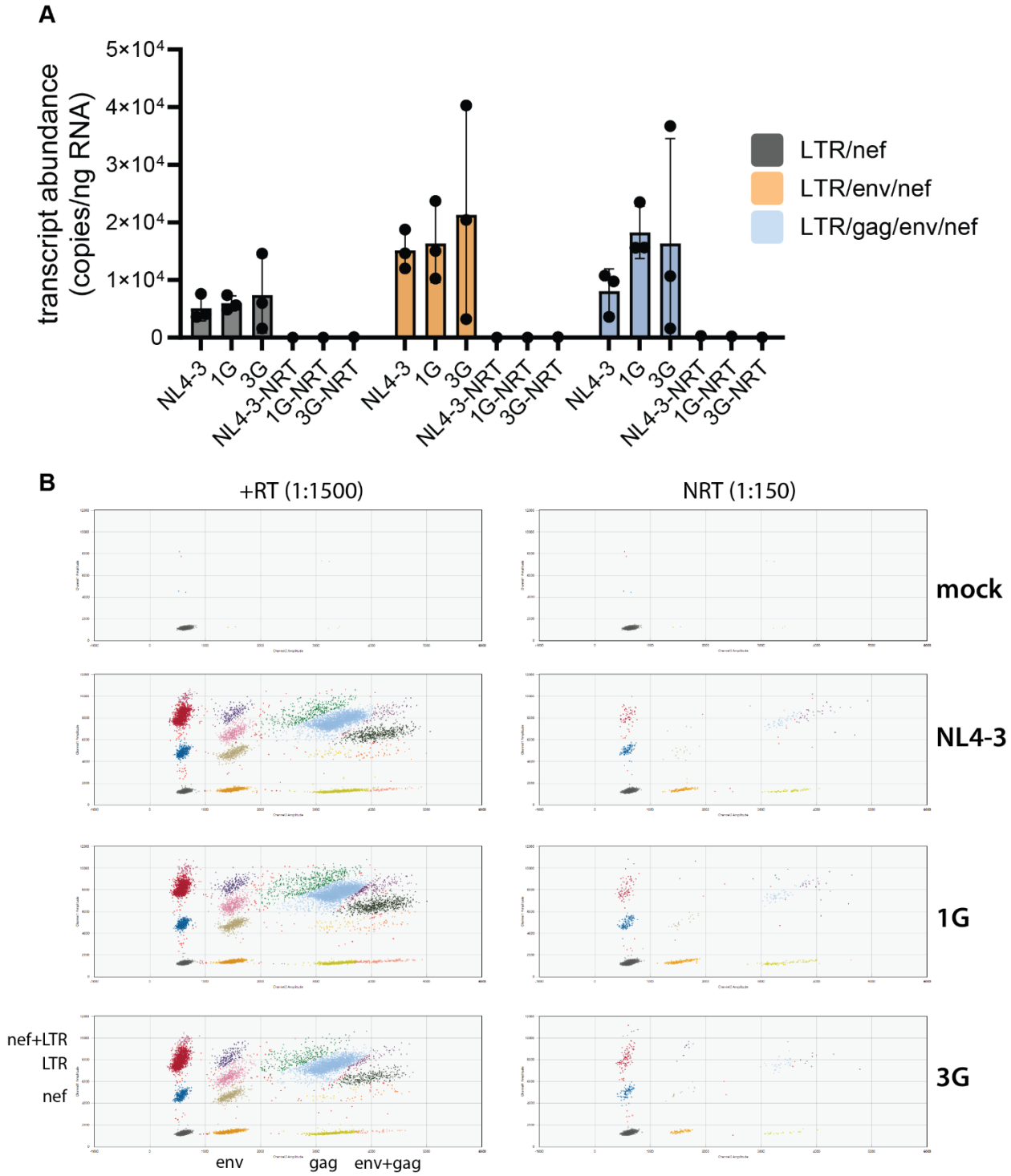

Fig. S2: ddPCR for viral transcriptomic analysis in primary macrophages. Additional data from experiment described in **Fig. 4C**. (**A**) Raw transcript abundance values obtained from ddPCR, NRT: no-RT control. (**B**) Sample gating strategy to distinguish viral RNA populations, for +RT samples diluted 1:1500 or NRT controls diluted 1:150. (**A**) Plot is shown as mean +/- SEM, each data point is an individual primary cell donor. (**B**) Data is from one primary cell donor.

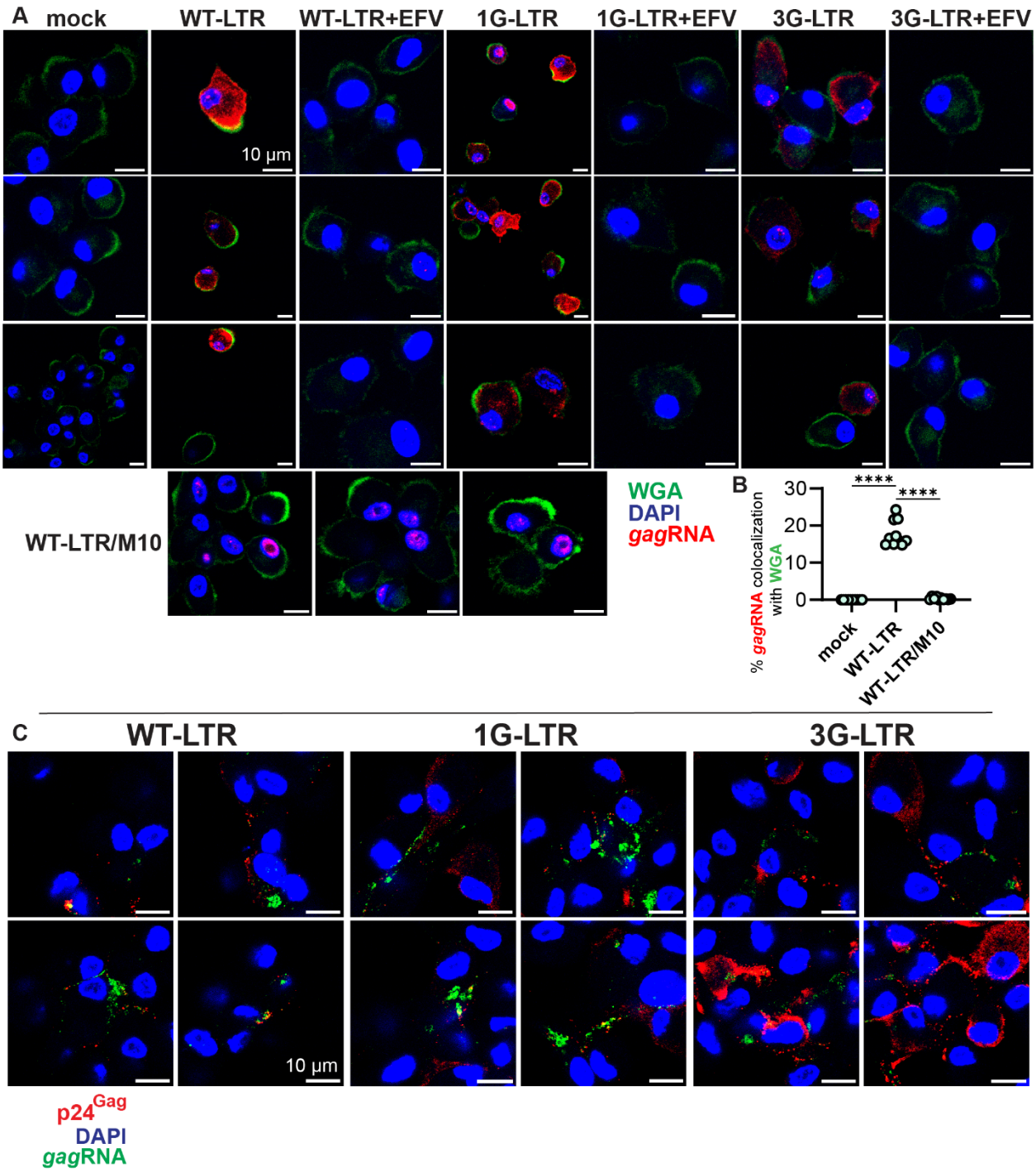

Fig. S3: smFISH of ^{cap}1G/3G usRNA expression in primary macrophages. Additional data from experiment described in **Fig. 4D-G**. Additional magnified images and colocalization quantification showing smRNA-FISH analysis merged with WGA (**A**, wheat germ agglutinin, plasma membrane stain) or anti-p24^{Gag} co-stain (**C**). In (**B**), % of *gag* RNA signal colocalizing with WGA is shown. EFV: reverse transcriptase inhibitor efavirenz treatment. WT-LTR/M10: wild-type ^{cap}1G/3G-expressing virus with inactivating M10 mutation in Rev. Means +/- SEM of individual fields of view from one representative

experiment are shown, with data analyzed by one-way ANOVA with Tukey's post-test, ****: $p<0.0001$.

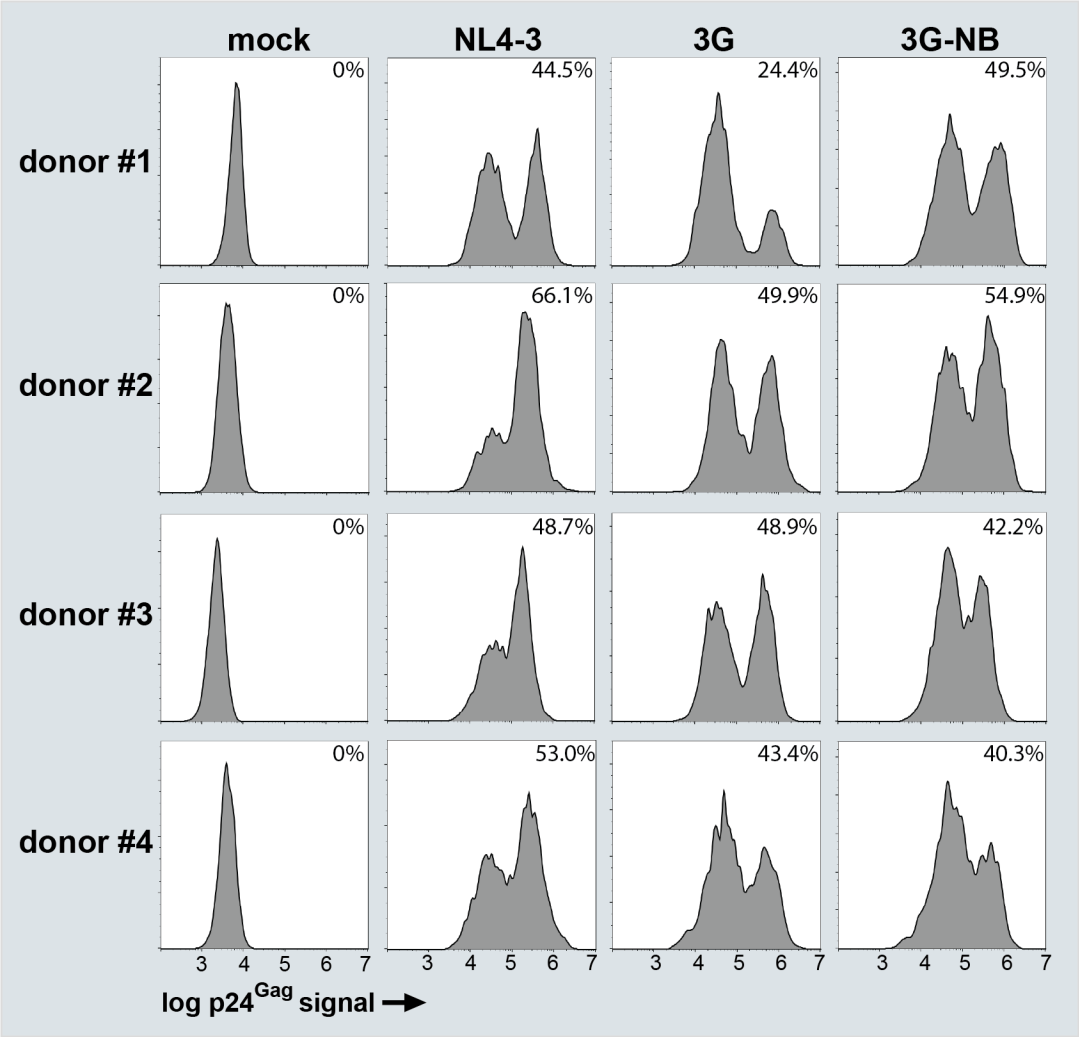

Fig. S4: p24^{Gag} production efficiency in 3G and 3G-NB infection. Additional data from experiment described in **Fig. 5B**. (**A**) Plots show log p24^{Gag} flow cytometry signal as a histogram with positive population quantification for four primary cell donors.

Primer	Sequence	Application
usRNA_Fwd	tgtgtgcccgtctgttgtgt	RT-qPCR, RNA-IP
usRNA_Rev	ctctccttctagcctccgct	RT-qPCR, RNA-IP
msRNA_Fwd	ccaggaagtcagcctaaaactgct	RNA-IP
msRNA_Rev	gggaggtgggttgtcttggatagag	RNA-IP
actin_Fwd	tggtgggcatgggtcagaag	RNA-IP
actin_Rev	gtcccggccagccaggtccag	RNA-IP
GAPDH_Fwd	caagatcatcagcaatgcct	RT-qPCR
GAPDH_Rev	agggatgatgttctggagag	RT-qPCR
LTR_Fwd	gcctcaataaagcttgccttga	ddPCR
LTR_Rev	gggcgccactgctagaga	ddPCR
LTR_Prb	FAM-ccagagtcacacaacagacgggcaca	ddPCR
Gag_Fwd	gactagcggaggctagaaggagaga	ddPCR
Gag_Rev	ctaattttccsccdcttaataytgacg	ddPCR
Gag_Prb	HEX-atgggtgcgaga	ddPCR
Env_Fwd	tvttcmttgggttcttrggagcagcagg	ddPCR
Env_Rev	gcactatrccagacaatavytgtctggcctgtacc	ddPCR
Env_Prb	HEX-agcackatggg	ddPCR
Nef_Fwd	ggtgggagcagyatctcgaga	ddPCR
Nef_Rev	tgtaagtcattggtcttaaaggtacctgagg	ddPCR
Nef_Prb	FAM-ccaggcacaakcagcatt	ddPCR

NOTE: Some ddPCR probe sets are oligonucleotide pools originally designed to analyze primary virus isolates.

Table S1: Primers used for PCR-based assays.

#	Probe sequence (CALFluor590)	Probe sequence (Quasar670)
1	AACTGCGAATCGTTCTAGCT	TTAATACTGACGCTCTCGCA
2	ATGTTTCTAACAGGCCAGGA	TTGCCCATACTATATGTTTT
3	TGAAGGGATGGTTGTAGCTG	ATTAACTGCGAATCGTTCTA
4	TCTTGTCTAAAGCTTCCTTG	TGTCCCAGTATTTGTCTACA
5	TACTACTTTTACCCATGCAT	TTCTTCTGATCCTGTCTGAA
6	ACATGGGTATCACTTCTGGG	TGTATTATATAATGATCTAA
7	CATGCACTGGATGCACTCTA	TTTGATGCACACAATAGAGG
8	ATCCTATTTGTTCCTGAAGG	TTTTTCTTACTTTTGTTTTG
9	GGGATAGGTGGATTATTTGT	TATAGGGTAATTTTGGCTGA
10	TGGTAGGGCTATACATTCTT	TATGGCCTGATGTACCATTT
11	CTCGGCTCTTAGAGTTTTAT	ATGCATTTAAAGTTCTAGGT
12	GGTTTCTGTCATCCAATTTT	TTCTCTTCTACTACTTTTAC
13	CAATCTGGGTTCGCATTTTG	TCTGATAATGCTGAAAACAT
14	TTCAGCCAAAACTCTTGCCT	TAGCATGGTGTTTAAATCTT
15	AATCTTTCTTTGGTTCCTAA	TTTGCATGGCTGCTTGATGT
16	CAGACCTGAAGCTCTCTTCT	ATTGATGGTCTCTTTTAACA
17	GAGTGATCTGAGGGAAGTTA	TAGTTCCTGCTATGTCACTT
18	TCCACAGATTTCTATGAGTA	TTTGTTCCTGAAGGGTACTA
19	GTCCTACTAATACTGTACCT	ATTATTTGTCATCCATCCTA
20	CCATTGTTTAACTTTTGGGC	TTAATCCCAGGATTATCCAT
21	CCCAGAAGTCTTGAGTTCTC	TTATGTCCAGAATGCTGGTA
22	GCAGTATACTTCCTGAAGTC	TACATAGTCTCTAAAGGGTT
23	TGGAATATTGCTGGTGATCC	TCTTAGAGTTTTATAGAACC
24	ATGTTTTTTGTCTGGTGTGG	AATTTTTTACCTCCTGTGAA
25	GATGGAGTTCATAACCCATC	ACCAACAAGGTTTCTGTCAT
26	CCCTGGGTAAATCTGACTTG	ACAATCTGGGTTCGCATTTT
27	GTATTACTTCTGTTAGTGCT	AATGCTTTTAAAATAGTCTT
28	TAGAATCTCTCTGTTTTCTG	TCATCATTTCTTCTAGTGTA
29	TAATACACTCCATGTACTGG	TGTTACTTGGCTCATTGCTT
30	TGCTTCTGTATTTCTGCTAT	ATCATTATGGTAGCTGAATT
31	CGTTCTTGCATATTTTCCTG	TTTCTTTGGTTCCTAAAATT
32	TCTTTCCCCATATTACTATG	AATTGAAACACTTAACAATC
33	CATGTTTCCTTTTGTATGGG	ATTTCCAACAGCCCTTTTTC
34	CTAAAGGAGGGGTATTGACA	TACAATCTTTCATTTGGTGT
35	AGTTAGGGTGACAACTTTTT	TGCTCTGAAGAAAATTCCCT
36	TGCTTGTAACTCAGTCTTCT	AAAGAGTGATCTGAGGGAAG
37	TTACTTCTAATCCCGAATCC	AGTTAGGGTGACAACTTTTT
38	GTGCTTGAATGATTCCTAAT	TGCTTGTAACTCAGTCTTCT
39	TTGTTCATTTCCTCCAATTC	TTACTTCTAATCCCGAATCC
40	ATAGTACTTTCCTGATTCCA	GTGCTTGAATGATTCCTAAT
41	GCCTTATCTATTCCATCTAA	TTGTTCATTTCCTCCAATTC
42	AAATCACTAGCCATTGCTCT	ATAGTACTTTCCTGATTCCA
43	TGGCTACTATTTCTTTTGCT	GCCTTATCTATTCCATCTAA
44	TCCTTTTAGCTGACATTTAT	AAATCACTAGCCATTGCTCT
45	TACATGAACTGCTACCAGGA	TGGCTACTATTTCTTTTGCT
46	TGGTGAAATTGCTGCCATTG	TCCTTTTAGCTGACATTTAT
47	ACTTTGGGGATTGTAGGGAA	TACATGAACTGCTACCAGGA
48	TCTGCTGTCCCTGTAATAAA	TGGTGAAATTGCTGCCATTG

Table S2: Fluorescent smRNA-FISH probe sequences.