Defining the ordered pathway for ZAP-mediated RNA decay

Clément R. Bouton[1,7], Grega Gimpelj Domjanič[2,3,7], María José Lista[1,7], Rui Pedro Galão[1], Thomas Courty[1], Harry D. Wilson[1], Peter W. S. Hill[1], Hannah E. Mischo[1], Anob M. Chakrabarti[4], Mario Poljak[3], Jernej Ule[2,5,6,*], Stuart J. D. Neil[1,*], Chad M. Swanson[1,*]

[1]Department of Infectious Diseases, King's College London, London, UK

[2]National Institute of Chemistry, Ljubljana, Slovenia

[3]Institute of Microbiology and Immunology, University of Ljubljana, Ljubljana, Slovenia

[4]UCL Respiratory, University College London, London, UK

[5]UK Dementia Research Institute at King's College London, London, UK

[6]Department of Basic and Clinical Neuroscience, Institute of Psychiatry Psychology & Neuroscience, King's College London, London, United Kingdom.

[7]These authors contributed equally

*Correspondence: Jernej Ule (jernej.ule@kcl.ac.uk), Stuart Neil (stuart.neil@kcl.ac.uk), Chad Swanson (chad.swanson@kcl.ac.uk)

ABSTRACT

Zinc-finger Antiviral Protein (ZAP)-mediated RNA decay (ZMD) restricts replication of viruses containing CpG dinucleotide clusters. However, why ZAP isoforms differ in antiviral activity and how they recruit cofactors to mediate RNA decay is unclear. Therefore, we determined the ordered events of the ZMD pathway. The long ZAP isoform preferentially binds viral RNA, which is promoted by TRIM25. The endoribonuclease KHNYN then cleaves viral RNA at positions of ZAP binding. The 5' cleavage fragment undergoes TUT4/TUT7-mediated 3' uridylation and degradation by DIS3L2. The 3' cleavage fragment is degraded by XRN1. ZAP and TRIM25 interact with KHNYN, TUT7, DIS3L2 and XRN1 in a RNase-resistant manner. Viral infection promotes the interaction between ZAP and TRIM25 with these enzymes, leading to viral RNA degradation while also decreasing the abundance of many cellular transcripts. Overall, the long isoform of ZAP recruits key enzymes to assemble an RNA decay complex on viral RNA.

36 **INTRODUCTION**

37 A critical component of the innate immune response is the ability to distinguish
38 between self and non-self nucleic acids. The best characterised non-self RNA
39 detection systems include RIGI, IFIH1, EIF2AK2, OAS1-3 and ZBP1, which promote
40 signalling that leads to antiviral gene expression, cap-dependent translation inhibition,
41 global RNA degradation and cell death[1]. These pattern recognition receptors (PRRs)
42 bind double-stranded RNA (dsRNA), which acts as the pathogen-associated
43 molecular pattern (PAMP). In contrast, antiviral proteins that bind single-stranded RNA
44 (ssRNA) often directly restrict viral replication by inhibiting viral RNA translation or
45 targeting it for degradation[2]. Zinc-finger Antiviral Protein (ZAP, also known as PARP13
46 or ZC3HAV1) is a paradigmatic example of a ssRNA-binding protein critical for the cell
47 autonomous innate immune response[3].

48

49 ZAP is encoded by an interferon-stimulated gene that restricts diverse viruses
50 including retroviruses, alphaviruses, Ebola virus and human cytomegalovirus[4]. It has
51 two primary mechanisms for inhibiting viral infection: promoting viral RNA decay[5] and
52 inhibiting viral mRNA translation[6-8]. The RNA binding domain (RBD) in ZAP contains
53 four zinc finger motifs (ZnF1-4) and directly binds a single CpG dinucleotide in pocket
54 formed by ZnF2[5,9-12]. ZAP-mediated restriction requires the presence of multiple CpGs
55 in a target RNA molecule[13]. To form a potent ZAP-response element (ZRE) in viral
56 RNA, CpGs are optimally separated by ~15-30 nucleotides[14], indicating that
57 multivalency is required for substrate recognition. Many mammalian RNA viruses have
58 low CpG abundance, likely due to evolutionary pressure from ZAP[13,15]. ZAP also
59 regulates cellular transcript expression[16-19] and this, in addition to CpG DNA

methylation[20], has been proposed to contribute to the selective pressure against CpG dinucleotide abundance in cellular genes[18,21,22].

There are several ZAP isoforms, with ZAP-long (ZAP-L) and ZAP-short (ZAP-S) being the most abundant[5,23,24]. ZAP-S expression is induced by type I interferon while ZAP-L is constitutively expressed[24,25]. Both ZAP-S and ZAP-L contain the RBD and a central domain that binds ADP-ribose[23,26,27]. ZAP-L also contains a catalytically inactive ADP-ribosyl transferase-like (ART-like) domain and a C-terminal CaaX box that mediates S-farnesylation[23,28]. This post-translational modification localises ZAP-L to the cytoplasmic endomembrane system[17,28-30].

ZAP is the core of a multicomponent antiviral system. It has no enzymatic activity and acts as a scaffold that recognises target transcripts to recruit cofactors that inhibit translation or mediate RNA decay[3]. It has been reported to interact with a large number of proteins, though it remains unclear which are relevant for ZAP-mediated RNA decay (ZMD)[3]. ZAP was originally proposed to interact with the exosome to mediate ZMD[5,31]. It was also reported to recruit the decapping complex DCP1A-DCP2 through the RNA helicase DDX17 as well as the deadenylase PARN[32]. Subsequently, ZAP was shown to bind the endoribonuclease KHNYN through its RBD, which was proposed to mediate endonucleolytic cleavage of viral RNA[33-35]. The KHNYN paralog N4BP1 has also been implicated in ZMD and these proteins have at least partially redundant activity for some transcripts[34,36,37]. ZAP also interacts with TRIM25[38,39], though it is unclear how this E3 ubiquitin ligase regulates ZAP antiviral activity.

84 There are several therapeutic applications arising from understanding the targeting
85 specificity of the ZAP antiviral system and how it mediates ZMD. TRIM25, ZAP,
86 KHNYN and N4BP1 target mRNAs delivered by lipid nanoparticles when they do not
87 contain N^1-methylpseudouridine modifications, highlighting its importance for
88 therapeutic RNA design[40]. Synonymous CpG-recoding of ZAP-resistant viral
89 genomes has been shown to be a novel way to develop live-attenuated viral
90 vaccines[14,41,42]. ZAP expression is downregulated in some types of cancer and,
91 because ZAP-sensitive viruses would preferentially replicate in cancers with reduced
92 ZAP expression compared to healthy tissue, they have been proposed to be used as
93 oncolytic viruses for cancers with low ZAP expression[43].
94
95 However, several critical mechanistic questions regarding ZMD remain unanswered.
96 It is unclear whether ZAP targeting specificity is dictated solely by the RBD or whether
97 other factors modulate target recognition. ZAP isoforms have been proposed to have
98 different functions[25,29] and, while ZAP-L has been reported to have higher antiviral
99 activity against several viruses than ZAP-S[23,24,29,30], the mechanism for this is
100 unknown. Furthermore, of the substantial number of proteins that ZAP has been
101 reported to interact with[3], which of these are required for ZMD, and how they may act
102 in an ordered pathway, has not been defined.
103
104 In this study, we characterised the ZMD pathway to determine how ZAP interacts with
105 multiple enzymes to mediate viral RNA decay. Cells genetically engineered to only
106 express ZAP-L or ZAP-S show that, even though both isoforms contain the RBD, ZAP-
107 L is the predominant isoform that inhibits retroviral replication due to increased binding
108 to viral RNA and KHNYN. TRIM25 promotes ZAP binding to viral RNA, further

109 indicating that the specificity of the ZAP RBD is not sufficient to discriminate ZMD
110 targets. We then show that ZAP and TRIM25 interact with multiple RNA decay
111 enzymes including KHNYN, XRN1, TUT7 and DIS3L2 and order these enzymes into
112 a decay pathway. *In cellulo* identification of KHNYN cleavage sites shows that KHNYN
113 mediates endonucleolytic cleavage of the viral RNA in regions with high levels of ZAP
114 binding. The 5' cleavage fragment is 3' uridylated by the paralogs TUT4/TUT7 and
115 then degraded by DIS3L2. XRN1 degrades the 3' cleavage fragment. Viral infection
116 increases binding of ZAP and TRIM25 to several components of this pathway and
117 promotes regulation of cellular genes by ZAP, TRIM25 and KHNYN. Overall, ZAP acts
118 as a core scaffold for several enzymes to mediate ZMD.
119

120 **RESULTS**

121 **Differential activity between ZAP isoforms correlates with binding viral RNA and**
122 **KHNYN**

123 ZAP restricts a wide range of viruses[4]. For this study, we used HIV-1 as a molecularly
124 tractable model virus to define the ordered events of the ZMD pathway. Wild-type HIV-
125 1 (HIV-WT) is substantially depleted in CpG dinucleotides and the paucity of these is
126 due to strong selective pressure, likely from ZAP[13,22,44,45]. Mutations that introduce a
127 CpG in HIV-1 have a substantial fitness cost and CpG abundance in HIV-1 *env* is
128 linked to disease progression[46-48]. Because HIV-1 is substantially depleted in CpGs,
129 these can be introduced through synonymous genome recoding to create HIV-CpG,
130 which contains a potent ZRE in *env* (Figure 1A)[13,33]. This allows isogenic viruses to be
131 analysed in which ZAP restricts HIV-CpG but not HIV-WT (Figure 1B). This restriction
132 can be quantified, showing that HIV-CpG infectious virus production is ~1% of HIV-WT
133 (Figure 1B). Importantly, both viruses produce similar levels of infectious virus

134 production when ZAP is knocked out. These paired viruses allow components of the
135 ZAP antiviral system to be characterised under conditions in which the *cis*-acting ZRE
136 is either absent or present in the virus, allowing ZMD substrate specificity to be
137 determined in each experiment.
138
139 The relative contribution of the two primary ZAP isoforms (Figure 1C) to ZMD is
140 controversial. While several papers have reported that ZAP-L is more active than ZAP-
141 S, others have proposed that ZAP-S, or shorter fragments of ZAP, are sufficient for
142 antiviral activity[23,24,28,29,34,48,49]. However, many of these experiments were performed
143 by overexpressing ZAP isoforms in wild-type cells or reconstituting ZAP isoforms in
144 CRISPR knockout cells, often by transient transfection, leading to non-physiological
145 protein levels. Furthermore, the mechanism by which ZAP-L may be more active than
146 ZAP-S is unclear. Therefore, we used a set of previously described HEK293T cells
147 that were CRISPR-engineered to only express ZAP-L or ZAP-S by targeting splice
148 sites and poly(A) sites in the endogenous *ZC3HAV1* locus[17]. These cells allow the
149 antiviral activity of ZAP-L or ZAP-S to be characterised without ectopic expression
150 (Figure 1D). ZAP-L knockout eliminated almost all ZAP antiviral activity against the
151 HIV-CpG while ZAP-S knockout had only a minor effect (Figure 1E and S1A-B),
152 demonstrating that in this system, ZAP-L is the essential mediator of viral restriction.
153
154 ZAP-L has previously been reported to interact more strongly with TRIM25 and
155 KHNYN than ZAP-S, though one or more of these proteins was often overexpressed
156 in the experimental conditions[30,38,50]. To determine if ZAP-L and ZAP-S preferentially
157 interact with TRIM25 or KHNYN when all the proteins are expressed from their
158 endogenous locus, we performed co-immunoprecipitation experiments in the wild-

159 type, ZAP-L and ZAP-S knockout cell lines. While ZAP-L or ZAP-S knockout did not
160 affect the interaction of the other ZAP isoform with TRIM25, ZAP-L knockout led to a
161 substantial decrease in ZAP interacting with KHNYN whereas ZAP-S knockout had
162 only a marginal effect (Figure 1F).
163
164 ZAP directly binds KHNYN through its N-terminal RBD[34,35]. Therefore, both ZAP
165 isoforms contain the KHNYN interaction domain even though ZAP-L is more
166 competent to assemble the RNA decay complex than ZAP-S. One possible
167 mechanism for this may be that ZAP-L assembles on the ZRE more efficiently than
168 ZAP-S. To test this hypothesis, we used iCLIP[51] to quantify the binding of each isoform
169 to HIV-WT and HIV-CpG RNA (Figure 1G and S1C-E). There are three major ZAP
170 binding peaks in HIV-CpG, with the first at the 5' end in the viral genome, the second
171 in the 5' region of *env* covering the sequence that has been CpG recoded, and the
172 third in the 3' region of *env*. As expected, ZAP binding is most pronounced in Peak 2
173 covering the engineered ZRE. Interestingly, the CpG clusters underlying Peaks 1 and
174 Peak 3, which are present in both HIV-WT and HIV-CpG, show lower levels of binding
175 in HIV-WT than HIV-CpG RNA (Figure S1C), indicating that high ZAP binding to the
176 engineered region promotes additional ZAP recruitment at non-contiguous regions in
177 the viral RNA containing clustered CpGs. In cells depleted for ZAP-S, high levels of
178 ZAP-L binding to HIV-CpG RNA were observed (Figure 1G and S1E). By contrast, in
179 ZAP-L knockout cells, ZAP-S showed low levels of RNA binding, correlating with its
180 low antiviral activity (Figure 1G and S1D). In sum, ZAP-L is the primary antiviral
181 isoform against HIV-CpG due to increased binding to HIV-CpG RNA and a stronger
182 interaction with KHNYN.
183

184 **Assembly of the ZAP-TRIM25-KHNYN complex is regulated by viral infection**

185 To determine if the presence of a substantial amount of ZAP substrate during HIV-CpG
186 infection affects the interaction between ZAP, TRIM25 and KHNYN, we performed co-
187 immunoprecipitation experiments in mock infected, HIV-WT and HIV-CpG infected
188 HeLa cells. Interestingly, upon HIV-WT or HIV-CpG infection, there was a small
189 increase in the RNase-resistant interaction between ZAP and TRIM25 and, more
190 potently, between TRIM25 and KHNYN (Figure 2A-B), even though the abundance of
191 these proteins was unchanged. Therefore, viral infection appears to promote assembly
192 of the ZAP-TRIM25-KHNYN complex, but this does not require the strong ZRE present
193 in HIV-CpG.

194

195 We then analysed whether TRIM25 or KHNYN modulate ZAP assembly on the HIV-
196 CpG ZRE. In TRIM25 knockout cells, there was a moderate decrease in ZAP bound
197 to HIV-CpG (Figure 2C and S2A-B). By contrast, KHNYN depletion did not
198 substantially affect ZAP binding (Figure 2D). RNA-immunoprecipitation experiments
199 also showed that TRIM25 promotes ZAP binding to HIV-CpG RNA (Figure 2E). Of
200 note, while TRIM25 has been proposed to regulate the specificity of ZMD in addition
201 to its potency[34], we did not observe any substantial change in ZAP antiviral activity
202 against HIV-WT in TRIM25 knockout cells (Figure S2C). Furthermore, the relative
203 position of ZAP binding to HIV-CpG RNA did not change in TRIM25 knockout cells
204 (Figure 2C, Spearman correlation of relative crosslink signal at each nucleotide in
205 CRISPR control and TRIM25 knockout cells = 0.92) and there was no substantial
206 increase in the number of cellular mRNAs with multiple ZAP binding peaks in our iCLIP
207 data (Figure S2D). Therefore, TRIM25 appears to promote ZAP binding to viral RNA
208 but does not change the specificity for the transcripts that ZAP targets.

209

210 As there are three consistent, predominant ZAP-binding peaks on HIV-CpG (Figures
211 1G, 2C-D), we used the iCLIP data to determine the width of each peak and whether
212 ZAP binding spreads beyond the recoded region in CRISPR control and KHNYN
213 knockout cells (Figure S2E-F). Each ZAP binding peak in HIV-CpG contains several
214 clustered CpGs. Peak 1 is ~250 nucleotides (nt) wide in the 5' leader region and *gag*,
215 which contain a highly structured packaging element[52]. Peak 2 encompassing the CpG
216 recoded region in HIV-CpG *env* (Figure 1A, 475 nt long[33]) is ~900 nt long, indicating
217 that ZAP binding extends beyond the recoded region. Peak 3 is ~800 nt long and
218 covers splice acceptor 7 (SA7, nt 7904-7915)[53]. Consistent with ZAP oligomerisation
219 and the hypothesis that multivalent binding of ZAP to multiple CpGs is likely required
220 for its antiviral activity[14,34,54], the iCLIP data implies assembly of ZAP molecules that
221 cover wide, non-contiguous regions on the viral RNA.

222

223 **XRN1 degrades the 3' fragment after KHNYN cleavage**

224 Since KHNYN is an active endoribonuclease that directly binds ZAP, it has been
225 hypothesised that ZAP recruits KHNYN to cleave viral RNA[33-35]. However, KHNYN
226 enzymatic activity has only been tested on *in vitro* substrates using purified
227 KHNYN[34,35,55]. To test this hypothesis directly in cells, we first used iCLIP to map
228 KHNYN binding sites on HIV-CpG. Because we have not found an antibody that
229 efficiently immunoprecipitates KHNYN, we identified KHNYN binding sites on viral
230 RNA using a previously characterised stable cell line expressing an active KHNYN-
231 GFP fusion protein[50]. KHNYN had substantial binding within areas of high CpG
232 abundance, including the CpG recoded region of HIV-CpG (Figure 3A). As KHNYN

233 has no known RNA binding specificity[35,37], this supports the hypothesis that ZAP
234 recruits KHNYN to target RNA.
235
236 We then aimed to identify authentic KHNYN cleavage sites in target viral RNAs *in*
237 *cellulo* using RACE-seq. Importantly, identifying and inactivating the downstream
238 degradation of such cleavage products is likely to substantially increase the chances
239 of detecting the sites of KHNYN cleavage. Furthermore, since many viruses produce
240 substantial amounts of viral RNA, full degradation of these transcripts after KHNYN
241 cleavage would be important to prevent accumulation of RNA fragments that may
242 disrupt cellular function. Therefore, we sought to determine how exonucleolytic RNA
243 decay enzymes modulate ZMD.
244
245 Several proteins that mediate or regulate RNA degradation have been linked to ZAP
246 including CNOT1, PARN, EXOSC5, EXOSC3, EXOSC7, DCP2, XRN1, DDX17 and
247 DDX30[3,31,32,56-59]. To identify whether any of these proteins are required for ZAP
248 antiviral activity on HIV-CpG, we depleted them using RNAi. While siRNA-mediated
249 depletion of ZAP, TRIM25 and KHNYN increased HIV-CpG infectious virus production,
250 none of the other potential ZAP cofactors had an effect (Figure S3A-D), in agreement
251 with a previous study[13], and consistent with the hypothesis that cleavage by KHNYN
252 effectively renders the viral RNA non-infectious and incompetent for translation.
253 However, it is possible that these proteins could exert their effects downstream of
254 KHNYN cleavage by mediating decay of the resulting RNA fragments. Therefore, we
255 designed a probe to potentially identify the 3' cleavage fragment of HIV with the
256 presumption that many KHNYN cleavage sites will be near the major ZAP binding site
257 in the CpG-recoded region (probe 2 (p2) in Figure 3B). Of note, due to the complex

258 HIV-1 alternative splicing program[53], only the unspliced and singly spliced HIV-1
259 transcripts will contain the ZAP-sensitive CpG-recoded region and the fully spliced
260 RNAs are not subject to ZMD[13,33] (Figure 3B). To quantify only ZAP-sensitive
261 transcripts, p2 was designed to span SA7, which is the last sequence that is specific
262 to the singly and unspliced transcripts. Notably, this sequence is within ZAP binding
263 peak 3 (Figure S2E-F). RNAi-mediated depletion of CNOT1, PARN, EXOSC3,
264 EXOSC5, EXOSC7, DDX17 or DHX30 did not affect HIV-CpG RNA abundance using
265 the probe for the 3' fragment (Figure S3E-F), but depletion of the major 5'-3'
266 exonuclease XRN1 led to a significant increase in abundance (Figure 3C and S3D).
267 We then used 5' RACE to identify potential ZAP cleavage sites using a HIV specific
268 primer at the 3' end of the viral genome (Figure 3B). Visualising the RACE products
269 on an agarose gel showed that a novel band appeared for the control siRNA HIV-CpG
270 sample (3' cleavage fragment in Figure 3D). When ZAP was depleted, this band was
271 not present and the abundance of the band increased when XRN1 was knocked down.
272 Therefore, we hypothesised that this band could be a KHNYN cleavage product that
273 is further degraded by XRN1.
274
275 To map the 5' cleavage sites, we used Oxford Nanopore sequencing for the total 5'
276 RACE reactions and identified the 5' end of each read. In XRN1 knockdown cells, HIV-
277 CpG 5' ends clustered in ZAP binding Peaks 2 and 3 (Figure 3E). This indicates that
278 ZAP likely recruits KHNYN to these sites to mediate cleavage and that the ZAP binding
279 in Peak 3 is sufficient for KHNYN cleavage. This also suggests that the increase in 3'
280 cleavage fragment abundance when XRN1 is depleted (Figure 3C) may be
281 underestimated because the probe (Figure 3B) covers the site of endonucleolytic
282 cleavage in ZAP binding Peak 3. However, as described above, this is the most distal

283 3’ sequence that the probe can cover and still be specific for the ZRE-containing RNA.
284 Importantly, the 5’ RACE assay has a 3’ bias in that if KHNYN cuts multiple times in a
285 transcript, the 5’ end closest to the RACE primer will be identified, so the relative
286 frequency of cleavage sites in Peaks 2 and 3 should not be directly compared. The
287 sequence underlying Peak 3 is identical between HIV-WT and HIV-CpG, suggesting
288 that ZAP binding in Peak 2 not only promotes ZAP binding in Peak 3 (Figure S1C and
289 S2A) but that this binding is functional, leading to endonucleolytic cleavage.
290
291 To determine if ZAP and TRIM25 interact with XRN1, we performed co-
292 immunoprecipitation experiments in the absence and presence of HIV-1 infection
293 (Figure 3F-G). Both proteins interact with XRN1 in a RNase-resistant manner and the
294 interaction between TRIM25 and XRN1 is strongly stimulated by either HIV-WT or HIV-
295 CpG infection. Overall, we hypothesise that a ZAP-TRIM25 complex recruits XRN1 to
296 degrade the 3’ KHNYN cleavage product.
297
298 **3’ uridylation by TUT4/7 regulate the abundance of the 5’ cleavage fragment**
299 We then screened the RNA degradation proteins linked to ZAP to determine if they
300 regulate the abundance of the 5’ cleavage fragments using a qRT-PCR assay with a
301 primer-probe set upstream of Peak 2 (probe 1 in Figure S4A). Surprisingly, depletion
302 of three exosome subunits, EXOSC3, EXOSC5 or EXOSC7, had no effect on the
303 abundance of the 5’ cleavage fragment (Figure S4B). Depletion of CNOT1, PARN,
304 XRN1, DDX17 or DHX30 also had no effect, indicating that degradation of the 5’
305 fragment was not regulated by known ZAP-interacting proteins linked to RNA decay.
306 Therefore, we performed 3’ RACE to identify the location and identity of the 3’ ends of
307 the 5’ cleavage fragment (Figure 4A and S4A). A potential 5’ cleavage fragment was

308 observed specifically in cells infected with HIV-CpG and was not present when ZAP
309 was depleted. We cloned this band and sequenced seven clones using Sanger
310 sequencing. Interestingly, reads terminated within ZAP binding Peak 2 and several
311 had non-templated uridines at the 3' end (Figure 4B). This result raised the possibility
312 that the 5' fragment was modified by 3' uridylation.
313
314 Several members of the TENT family of polymerases add non-templated nucleotides
315 at the 3' end of transcripts and TUT1, TUT4 and TUT7 are three proteins that
316 specifically add non-templated uridines[60]. TUT1 is a nuclear protein that had no effect
317 when it was depleted (Figure 4C). In contrast, depletion of two cytoplasmic paralogs
318 that are often functionally redundant, TUT4 and TUT7, promoted an increase in the
319 abundance of the 5' cleavage fragment (Figure 4C and S4C). As expected, TUT4/7
320 depletion had no effect on infectious virus production, indicating that it acts
321 downstream of KHNYN cleavage (Figure 4D), or the abundance of the 3' cleavage
322 fragment (Figure S4D). We then analysed whether ZAP and TRIM25 interacted with
323 TUT7. Both proteins interacted with it in an RNase-resistant manner and this
324 interaction was increased in the presence of HIV-WT or HIV-CpG infection (Figure 4E-
325 F).
326
327 **DIS3L2 degrades the 5' KHNYN cleavage fragment**
328 DIS3L2 is an exosome-independent, 3'-5' exoribonuclease that degrades transcripts
329 with 3' terminal uridylation[61-63]. Depleting its expression increased expression of
330 transcripts using the 5' probe but not the 3' probe (Figure 5A and S5A-B), indicating
331 that it is required to degrade the 5' KHNYN cleavage fragment. DIS3L2 depletion does
332 not affect HIV-CpG infectious virus production (Figure S5C), indicating that it also acts

333 after KHNYN cleavage. ZAP and TRIM25 interact with DIS3L2 in an RNase-resistant
334 manner, and the interaction between TRIM25 and DIS2L2 is increased by HIV
335 infection (Figure 5B-C). We then used 3' RACE-seq to map internal cleavage sites in
336 HIV-CpG after DIS3L2 depletion and found that they are concentrated in ZAP binding
337 Peak 2 (Figure 5D). We also determined proportion of 3' uridylation and length. When
338 DIS3L2 is depleted, ~50% of the reads have at least one non-templated nucleotide
339 and a substantial portion have at least 10 non-templated nucleotides (Figure 5E).
340 These non-templated nucleotides are highly enriched in uridine, indicating a poly-
341 uridine tail. In contrast, when ZAP is depleted, few reads contain non-templated tails
342 and the strong enrichment for uridine is not present. These data indicate that after
343 KHNYN cleavage, 5' cleavage products of the viral RNA become 3'-polyuridylated by
344 TUT4/7 and are then degraded by DIS3L2.
345
346 **HIV-1 infection induces ZMD for cellular transcripts**
347 HIV-1 infection strongly increased the interaction between TRIM25 and KHNYN,
348 XRN1, TUT7 and DIS3L2 (Figures 2B, 3G, 4F, 5C). Gene set enrichment analysis of
349 differentially expressed genes between HIV-WT or HIV-CpG infected cells compared
350 to mock infected cells showed a substantial change in transcripts associated with
351 pathogen signalling including TNFα signalling, inflammatory response, interferon-α
352 and interferon-γ responses (Figure 6A-B). This suggests that the increase in ZMD
353 complex assembly correlates with an overall antiviral response and we posited that
354 HIV-1 infection may stimulate ZMD activity. While the effect of ZMD on viral RNA
355 cannot be analysed in the absence of viral infection, we hypothesised that, if infection
356 induces ZMD, then the number of cellular transcripts regulated by the ZAP-TRIM25-
357 KHNYN pathway would be increased.

358

359 We determined how HIV-WT and HIV-CpG infection affected transcripts regulated by
360 ZMD. Because ZMD leads to RNA degradation, we focused on transcripts upregulated
361 when ZAP, TRIM25 or KHNYN were knocked out. HIV-WT or HIV-CpG infection
362 increased the number of genes expressing transcripts upregulated by ZAP or TRIM25
363 by ~2-3-fold and KHNYN-upregulated transcripts by ~5-fold (Supplemental Figure 6A-
364 C, significantly upregulated transcripts defined by adjusted $p < 0.05$)). Analysis of
365 genes encoding transcripts co-upregulated by ZAP, TRIM25 and KHNYN showed an
366 ~10-fold increase in both HIV-WT and HIV-CpG infected cells compared to mock
367 infected cells (Figure 6C-D). Transcripts encoded by 102 genes were upregulated in
368 ZAP, TRIM25 and KHNYN knockout cells in both HIV-WT and HIV-CpG infected cells.
369 These are enriched for encoding proteins associated with the endoplasmic reticulum
370 (Figure 6E), which correlates with ZAP-L localisation and the previous observation that
371 ZAP-L preferentially regulates transcripts encoding membrane proteins[16,17]. Overall,
372 HIV infection increases TRIM25 interaction with ZAP and KHNYN and this correlates
373 with an increase in the number of genes that are co-regulated by ZAP, TRIM25 and
374 KHNYN, implying that an infectious stimulus modulates the activity of ZMD on host
375 transcripts.

376

377 **DISCUSSION**

378 Overall, these experiments lead to a comprehensive model for ZMD (Figure 7). ZAP-
379 L forms a complex with TRIM25 and KHNYN to direct KHNYN-mediated
380 endonucleolytic cleavage in regions of high ZAP binding. After KHNYN cleavage, ZAP
381 and TRIM25 interact with additional RNA decay machinery to mediate degradation of
382 the resultant fragments. XRN1 degrades the 3' cleavage fragment, likely due to a

383 monophosphate being present on the 5' end after KHNYN cleavage. The 5' cleavage
384 fragment is uridylated by TUT4/7 and then degraded by DIS3L2. This allows the entire
385 viral RNA to be degraded after its recognition by ZAP.
386
387 How ZMD is targeted to specific transcripts remains unclear. A simple model is that
388 ZAP binds CpG dinucleotides through ZnF2 in its RBD and clustered CpGs are
389 sufficient to mediate specificity[13,14]. Supporting this model, ZAP has high affinity for
390 RNAs containing a CpG dinucleotide and CpG recoding can sensitise diverse viruses
391 to ZAP-mediated restriction[11-13,41,42,64]. However, some regions of viral RNA appear to
392 be more sensitive to CpG-mediated ZAP restriction and high levels of CpG abundance
393 sometimes does not sensitise viral RNA to ZAP[42,48,65-68]. In addition, UpA recoding
394 has been reported to sensitise viral RNA to ZAP[41,64]. Our data indicate that, while the
395 RBD is present in both ZAP-L and ZAP-S, only ZAP-L is potent for binding HIV-CpG
396 RNA and targeting it for decay. Furthermore, TRIM25 promotes ZAP binding to CpG
397 clusters in HIV-CpG RNA, consistent with a previous study that showed that TRIM25
398 promotes the interaction of ZAP with a reporter transcript that is regulated by ZAP at
399 the level of translation[39]. Therefore, the simple model in which the ZAP RBD is
400 sufficient to determine substrate specificity by binding clustered CpG dinucleotides
401 likely needs to be revised.
402
403 In addition to the sequence specificity in its RBD, ZAP target recognition could be
404 modulated by protein cofactors or its subcellular localisation. While KHNYN has two
405 KH domains, they show no evidence of RNA binding and form an extended di-KH
406 domain that may instead function as a protein-protein binding domain[35,37]. TRIM25
407 has multiple modes of RNA binding[69-71] but whether this regulates ZAP targeting of

408 viral RNA is unclear. Overall, the relative contribution of TRIM25
409 ubiquitylation/ISGylation activity[38,39,72,73], multimerisation through the N-terminal RING
410 domain[34,74] and RNA binding[69-71] for regulating ZAP antiviral activity remains to be
411 determined, especially considering that RNA binding and E3 ubiquitin ligase activity
412 are linked[69-71]. ZAP-L restriction has been shown to require both the ART-like domain
413 and the CaaX box that mediates S-farnesylation, which are not present in ZAP-
414 S[23,29,30]. While the function of the ART-like domain remains unknown, localising ZAP
415 to the cytoplasmic endomembrane system through S-farnesylation may provide a
416 physical scaffold that promotes ZAP binding or multimerization on viral RNA and as
417 well as interaction with KHNYN. Such compartmentalisation may also be important for
418 restricting the activity of the ZAP and KHNYN to avoid off target effects and reinforce
419 substrate specificity. Therefore, we propose that ZAP substrate specificity is
420 determined by at least three components: the specificity in the ZAP RBD, ZAP
421 cofactors such as TRIM25 and ZAP subcellular localisation determined by S-
422 farnesylation.
423
424 Depletion of the 5' degradation fragment requires TUT4/7 and DIS3L2. An interesting
425 property of DIS3L2 is that it can degrade structured RNA[75]. The HIV-1 genome has
426 several regions of RNA structure[76] and this property of the enzyme may be required
427 to completely degrade the viral RNA. It should be noted that many ZAP-sensitive
428 viruses produce viral RNA that comprise a substantial percentage of total poly(A)
429 RNA[77-79]. Efficient degradation of the RNA may be important to prevent incompletely
430 degraded RNA species from sequestering cellular RNA binding proteins, stimulating
431 the innate immune response or otherwise dysregulating cellular RNA metabolism.
432

433 HIV infection increased ZAP and TRIM25 interaction with KHNYN, XRN1, TUT7 and
434 DIS3L2. Importantly, this is not due to an increase in ZAP-specific RNA substrate
435 because the effect was observed for both HIV-WT and HIV-CpG infection. Increased
436 binding of ZAP and TRIM25 to RNA decay enzymes correlates with increased
437 repression of cellular transcripts by ZAP, TRIM25 and KHNYN. Therefore, it appears
438 that HIV infection activates this antiviral pathway and this may both restrict viral
439 replication and alter cellular gene expression.
440

441 **METHODS**

442 **Cell lines, viruses and infection**

443 HEK293T CRISPR cells, HeLa, HeLa CRISPR cells and TZM-bl cells were maintained
444 in high-glucose DMEM supplemented with GlutaMAX (Thermo Fisher Scientific), 10%
445 fetal bovine serum, 100 U/mL penicillin and 100 µg/mL streptomycin and grown with
446 5% CO2 at 37°C.
447

448 HIV-WT and HIV-CpG stocks were produced by co-transfecting pHIV-WT or pHIV-CpG
449 with pVSV-G into HEK293T CRISPR ZAP cells, harvested at 48h post-transfection
450 and titered on TZM-bl cells as described below.
451

452 Unless otherwise stated, cells were infected at MOI=3. The medium was replaced 18
453 hours (h) post inoculation. Cell lysates and supernatants were collected at 48 hours
454 post-infection. Cell-free supernatants were produced by centrifugation at 300 ×*g* for
455 five minutes.
456

457 **siRNA transfection and infection**

458 siRNA were purchased from Horizon Discovery and transfected using Lipofectamine
459 RNAiMAX (Invitrogen). siRNA were reverse-transfected with 1 uL lipofectamine at 25
460 nM final concentration for 90,000 HeLa cells in a 24-well plate. 24 hours later, the
461 medium was replaced, and cells were transfected again with siRNA at 25 nM final
462 concentration with 1uL lipofectamine. 6-8 hours later, cells were infected with HIV at
463 MOI=3. The medium was replaced 18 hours post inoculation. Supernatant and cell
464 lysates were collected 48 hours post-infection as described above.
465
466 **TZM-bl infectivity assay**
467 The TZM-bl cell line was used to quantify the amount of infectious virus[80-82]. 1E+4 cells
468 per well were seeded in 96-well plates and infected by incubation with virus stocks or
469 supernatants of infected cells. At 40-48 hours post-infection, the cells were lysed and
470 infectivity was measured by β-galactosidase expression using the Galacto-Star assay
471 system as per the manufacturer instructions. β-galactosidase activity was quantified
472 as relative light units per second using a PerkinElmer luminometer.
473
474 **Immunoblotting**
475 Cell lysates for immunoblotting were prepared in Laemmli buffer. Virions were pelleted
476 from culture supernatant by centrifugation at 20,000 x *g* for 2 hours at 4°C through a
477 20% sucrose cushion in phosphate-buffered saline (PBS). Viral pellets were
478 resuspended in 2x Laemmli buffer. Samples were boiled for 10 min at 95°C then
479 separated on 8-16% gradient precast gels then transferred to nitrocellulose or PVDF
480 membranes. Membranes were briefly rinsed in PBS then blocked in blocking buffer
481 (PBS, 5% non-fat milk, 0.1% Tween 20) for 30 minutes at RT. Membranes were
482 incubated with primary antibodies diluted in blocking buffer at 4°C overnight, then

483 washed 3 times for 5 min in PBS-T (PBS, 0.1% Tween 20), then incubated with
484 secondary antibodies diluted in blocking buffer for 1h at RT, then washed for 3 times
485 for 5 min in PBS-T. Membranes were imaged using an Odyssey FC (LI-COR)
486 measuring secondary antibody fluorescence, or using Amersham ECL Prime Western
487 Blotting Detection reagent (Cytiva) for HRP-linked antibodies with an Amersham
488 ImageQuant 800 (Cytiva). Antibodies used in this study are reported in the STAR
489 methods table.
490
491 **Coimmunoprecipitation**
492 1.5 mg Dynabeads protein A (Invitrogen) per sample were incubated in lysis buffer
493 (0.5% Nonidet P 40 substitute, 10mM HEPES pH=7.5, 150mM KCl, 3mM MgCL2,
494 protease inhibitor) supplemented with 5% BSA for 30 min at RT and under gentle
495 agitation. 4 ug of antibody per sample was bound to the beads for 30 min at RT and
496 under gentle agitation. The unbound antibody was then washed off and the beads
497 were resuspended in 50 uL lysis buffer per sample. Cells cultivated in a 10-cm dish
498 were lysed on ice in 1mL lysis buffer. In case of infected cells (MOI=3), cells were lysed
499 48 hours post-infection. Lysates were gently mixed on a rotator for 30 min at 4°C, then
500 centrifuged at 20,000 ×*g* for 10 min to pellet the nuclear fraction. 75 µL of the post-
501 nuclear supernatant was saved as the input lysate and mixed with 1 volume of
502 Laemmli buffer. 900 uL of the remaining post-nuclear supernatant was incubated with
503 the prepared beads under gentle agitation over night at 4°C. The samples were then
504 washed four times in wash buffer (0.05% Nonidet P 40 substitute, 10mM HEPES
505 pH=7.5, 150mM KCl, 3mM MgCL2, protease inhibitor) before being resuspended in
506 90 uL Laemmli buffer. For experiments including an RNase I treatment, the samples
507 were washed twice in wash buffer after the overnight incubation with the prepared

508 beads, then resuspended in 100 uL wash buffer. The samples were incubated with
509 300 U RNase I for 30 min at 37°C and washed three times in wash buffer before being
510 resuspended in 90 uL Laemmli buffer. Input lysates and samples were analysed by
511 immunoblotting as described above.
512
513 **qRT-PCR**
514 Total RNA was extracted using the RNeasy Mini Kit (Qiagen). cDNA was synthesised
515 using the High-Capacity cDNA RT kit (Applied Biosystems) following the
516 manufacturer's instructions. cDNA was diluted 20 times in RNase-free water and
517 quantified by qPCR using Applied Biosystems TaqMan Universal PCR Master Mix
518 (Applied Biosystems) in a QuantStudio 5 qPCR System (Applied Biosystems).
519 Relative cDNA quantities were determined using the $\Delta\Delta^{Ct}$ method. Primer-probes sets
520 used in this study are reported in the STAR methods table.
521
522 **RNA Immunoprecipitation**
523 HeLa CRISPR control/TRIM25/KHNYN cells were plated in 6-well plates and
524 transfected using TransIT-LT1 transfection reagent with 500 ng of HIV-WT or HIV-CpG
525 proviral DNA, and medium was replaced 4-6 hours later. 48 hours post-transfection
526 cells were washed with PBS prior to 'on-dish' irradiation with 400mJ/cm2 using a UV
527 Stratalinker 2400 (Stratagene). Cells were then pelleted and lysed with RIPA buffer
528 (50mM Tris-HCl pH 7.4; 150mM NaCl; 0.1% SDS; 0.5% sodium deoxycholate, 1%
529 Nonidet P 40 substitute and protease inhibitors), and then sonicated. Cleared lysates
530 were immunoprecipitated overnight at 4°C with anti-ZAP antibody (Abcam) and
531 Dynabeads protein G (Invitrogen). Following 3 washes with RIPA buffer, beads were
532 resuspended in 100 µl of RIPA and boiled for 10min to decouple protein/RNA

533 complexes from the beads. Finally, input and pulldown samples were incubated with
534 proteinase K (Thermo Fisher) for 1 hr at 37°C, and then boiled for 10 minutes to
535 inactivate the enzyme. Samples were stored at -20°C for downstream RNA extraction
536 and qRT-PCR analysis.
537
538 RNA from the input and pulldown samples was extracted using QIAzol (QIAGEN).
539 cDNA synthesis was performed using the High-Capacity cDNA RT kit as described
540 above. Of the final reaction, 5 µl were subjected to qPCR using the primer/probe sets
541 for human GAPDH and HIV-1 genomic RNA reported in the STAR methods table.
542 Serial dilutions of HIV-WT proviral DNA were used in standard curves as calibrator for
543 quantification of HIV-1 RNA copies in the pull-down fractions.
544
545 **ZAP and KHNYN iCLIP**
546 Cells were infected at MOI=3 and media was replaced 18 h post-inoculation with
547 DMEM supplemented with 2% fetal bovine serum. At 48hpi, cells were washed with
548 ice-cold PBS, irradiated once with 150 mJ/cm2 in a Stratalinker 2400 at 254 nm,
549 harvested by scraping in PBS, pelleted by centrifugation then resuspended in lysis
550 buffer (50 mM Tris-HCl, pH 7.4; 100 mM NaCl; 1% Igepal CA-630; 0.1% SDS; 0.5%
551 sodium deoxycholate; protease inhibitor) and stored at -80°C until use.
552
553 Individual cross-linking and immunoprecipitation (iCLIP) was performed according to
554 the previously published protocol[83], with the following details. For ZAP iCLIP, 4 µg of
555 anti-ZAP antibody (GeneTex, GTX120134) was used for immunoprecipitation per
556 sample. For the KHNYN KO, TRIM25 KO and ZAP-L KO samples, the membrane was
557 cut above 110 kDa. In the ZAP-S KO experiments, the membrane was cut from 130

558 kDa upwards. For KHNYN-GFP iCLIP, 4 µg anti-GFP antibody (Roche, ref.
559 11814460001) was used for immunoprecipitation per sample. The membrane was cut
560 from 130 kDa upwards. RNA-seq libraries were sequenced by Novogene as paired-
561 end 150 bp reads on Illumina Novaseq 6000. Each iCLIP experiment was performed
562 with three biological replicates.
563
564 **iCLIP data processing**
565 Reads were mapped to the human GRCh38.p13 genome build with GENCODE v38
566 annotation. The HIV-1_{WT} or HIV-1_{CpG}, genome and annotation were merged with the
567 human genome build and annotation files. The reads from biological replicates for
568 each sample set were pooled, and all subsequent analyses were performed on the
569 combined datasets. Pair-end reads were pre-processed using a custom script written
570 in Bash. Briefly, the unique molecular identifiers (UMIs) were extracted using *UMI-*
571 *tools extract*[84] (https://github.com/CGATOxford/UMI-tools,*--extract-method=regex*)
572 and the paired-end reads merged with *BBTools' BBMerge*[85]
573 (https://sourceforge.net/projects/bbmap/, *mininsert=0, minoverlap=5*). The processed
574 reads were analysed according to the standardised iCLIP workflow with the *nf-*
575 *core/clipseq v1.0.0* pipeline[86,87] (https://doi.org/10.5281/zenodo.4723017), which was
576 run with *Nextflow v22.10.0*[88] and the following parameters: *--deduplicate true, --*
577 *smrna_org human, --peakcaller paraclu*[89]*.* In addition, the crosslink BED files
578 generated during the iCLIP analysis were used to call peaks on the HIV-1 genome
579 with *Clippy v1.5.0* tool (https://github.com/ulelab/clippy). The following parameters
580 were used: *--window_size 100, --min_prom_adjust 1, --min_height_adjust 1.5, --*
581 *min_gene_counts 20, --min_peak_counts 5.*
582

583 To account for the variability in library size and RNA abundance across experiments,
584 the crosslink signal was normalised in two steps. First, crosslink counts were scaled
585 to counts per million (CPM) using the *clipplotr v1.0.0* command-line tool[90]
586 (https://github.com/ulelab/clipplotr) to adjust for differences in library size. Second, to
587 control for variations in HIV-1 RNA abundance between samples, CPM-normalised
588 crosslink counts were further adjusted using mean HIV-1 RNA abundance values,
589 calculated from the DESeq2 normalised RNA-seq data across all three biological
590 replicates. The standard error of the mean, shown as the error band on the iCLIP plots,
591 was calculated from the standard deviation of three biological replicates. The data was
592 smoothed using a rolling mean of 100.
593
594 **RNA-seq**
595 RNA was extracted using the Direct-zol RNA Miniprep Plus Kit (Zymo Pure) from cells
596 grown and infected in the same conditions as the iCLIP experiments. Poly(A)+ RNA
597 library preparation and sequencing were performed by GENEWIZ. Sequencing reads
598 were processed using the *nf-core/rnaseq v3.17.0* pipeline[86]
599 (https://doi.org/10.5281/zenodo.13986791) run on *Nextflow v24.04.2*[88] with default
600 parameters. Differentially expressed genes were identified with DESeq2 using the *nf-*
601 *core/differentialabundance v1.2.0* pipeline (https://doi.org/10.5281/zenodo.7846648).
602 To account for transcript length bias, the analysis utilized the raw gene count matrix
603 (salmon.merged.gene_counts.tsv) and transcript length matrix
604 (salmon.merged.gene_lengths.tsv) generated by the *nf-core/rnaseq* pipeline. The
605 pipeline was run with *Nextflow v24.04.2*[88] using default settings. Differentially
606 expressed genes were defined as adjusted $p < 0.05$ and a basemean >500.
607

608 Gene set enrichment analysis (GSEA) was performed using the R package
609 *clusterProfiler v4.12.6*[91] and human gene sets from the 50 Hallmark pathways
610 obtained using the package *msigdbr v7.5.1* (https://CRAN.R-
611 project.org/package=msigdbr). Gene expression relationships between different
612 sample sets were analysed and visualised using the R packages *VennDiagram*
613 *v1.7.3*[92] and *eulerr v7.0.2* (https://CRAN.R-project.org/package=eulerr). The
614 packages *ggplot2 v3.5.1* (https://ggplot2.tidyverse.org) and *pheatmap v1.0.12*
615 (https://github.com/raivokolde/pheatmap) were used to create heat maps of gene
616 expression.
617
618 **5' and 3' RACE-seq**
619 Total RNA was extracted using the Direct-zol RNA Miniprep Plus Kit. The 5' RACE was
620 performed from 1 ug total RNA using the SMARTer RACE 5'/3' Kit according to the
621 manufacturer's instructions. In order to make cleaved RNA compatible with the 3'
622 RACE cDNA synthesis step, the total RNA was poly(A) tailed using *E. coli* Poly(A)
623 polymerase (NEB) and then purified with the Monarch Spin RNA Cleanup Kit (NEB).
624 The 3' RACE was performed from 500 ng purified poly(A) tailed RNA using the
625 SMARTer RACE 5'/3' Kit according to the manufacturer's instructions. The HIV specific
626 primers used for 5' and 3' RACE are reported in the STAR methods table. RACE
627 sequencing libraries were prepared using the Native Barcoding Kit 24 V14 (Oxford
628 Nanopore Technologies). Sequencing of 5' RACE product libraries was performed on
629 R10.4.1 flowcells (FLO-MIN114, Oxford Nanopore, UK). 3' RACE sequencing libraries
630 were sequenced on PromethION with R10.4.1 flowcells.
631

632 For the 5' RACE-seq, basecalling of *FAST5* files for the first replicate was completed
633 using *guppy_basecaller* (v6.4.6). Dorado (v0.5.1) was used to basecall *POD5* files
634 generated from 5' RACE-seq replicates 2 and 3 as well as the 3' RACE-seq replicates.
635 Reads were basecalled and demultiplexed using the Super High Accuracy (SUP)
636 model (v4.3.0). Additional command-line arguments were implemented to ensure that
637 only reads with identical nanopore barcodes at both ends were considered, as well as
638 excluding any reads containing barcodes in the middle. Bash scripts submitted to
639 KCL's CREATE HPC for basecalling are available on Github (5' RACE-seq:
640 https://github.com/Tycour/HIV1-ZAP_RACEseq; 3' RACE-seq:
641 https://github.com/GregaGD/ZAP_HIV-1_3-prime_RACE-seq), along with the bash
642 scripts and custom Python and R scripts used in the downstream analysis.
643
644 Following basecalling, the FASTQ files were filtered with *NanoFilt v2.8.0*[93] to eliminate
645 low-quality reads and retain reads with an average quality score > Q20. The quality
646 metrics of the filtered reads were evaluated using *NanoPlot v1.42.0*[93]. A custom
647 Python script was implemented to identify and retain reads containing primer
648 sequences at both ends that were used in the RACE reaction. For the 5' RACE
649 reaction, the 5' primer was
650 CTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGTACATGGG and the 3'
651 primer was CCAGCAGCAGATGGGGTGGGAGCAGTAT. For the 3' RACE reaction,
652 the 5' primer was TGGTGGGCGGGGATCAAGCAGGAATTTGGC and the 3' primer
653 was GTACTCTGCGTTGATACCACTGCTTGCCCTATAGTGAG. Reads without
654 internal primer sequences were filtered out to ensure specificity in the downstream
655 analysis.
656

657 Based on the infection conditions, the processed reads were aligned to either the HIV-
658 1_{WT} or the HIV-1_{CpG} reference genome using *Minimap2*[94,95] *v2.17-r941*, with the splice-
659 aware argument (*--splice*). The alignment files in SAM format were converted to a
660 sorted and indexed BAM format using *samtools (v1.10 or* v1.19). Each read was
661 oriented to the leading or lagging strand using a custom Python script to facilitate
662 downstream processing.
663
664 *Bedtools* (v2.31.1) converted the *BAM* file containing deduplicated reads to *BED*
665 format. For the 5' RACE-seq, an *awk* (v20200816) command was employed to extract
666 the 5' start positions for reads. These positions were then analysed using a custom
667 script that maps and plots the distribution of read starts along the reference genome.
668 To examine the cleavage sites and detect non-templated nucleotides for the 3' RACE-
669 seq, the alignment files in BAM format were analysed using a custom R script. Using
670 the *GenomicAlignments* (Lawrence et al., 2013) and *Biostrings*
671 (https://bioconductor.org/packages/Biostrings) packages, the final mapped position of
672 the read was determined by parsing the CIGAR string. Specifically, the last mapped
673 segment was extracted from the CIGAR string and the corresponding 3′ end nucleotide
674 position was recorded. This was used in later steps to determine if any nucleotides
675 beyond the mapped part of the read were non-templated.
676
677 Similarly, the *GenomicAlignments*[96] and *Biostrings*
678 (https://bioconductor.org/packages/Biostrings) packages were used to identify
679 stretches of poly(A) nucleotides added during the 3' RACE protocol. The 3′ end soft-
680 clipped segment of the read was extracted based on the CIGAR string and analysed
681 to detect poly(A) segments of at least 15 consecutive adenosines, allowing up to two

682 mismatches. The position of the poly(A) start in the read sequence was recorded and
683 only reads with recognised poly(A) segments were selected for downstream analysis.
684 The non-templated nucleotides were defined as the sequence between the final
685 mapped position and the start of the poly(A) sequence in each read. The length
686 distribution and nucleotide composition of these regions were analysed, and the length
687 distribution was visualised with bar charts, while the nucleotide composition per
688 sequence length was visualised with stacked bar charts. The package *ggplot2 v3.5.1*
689 (https://ggplot2.tidyverse.org) was used to generate all charts.
690

691 **Statistics**

692 Unless otherwise stated, data are represented as mean ± standard deviation.
693 Statistical significance for infectious virus production data and qRT-PCR data was
694 determined by one-way ANOVA from log-transformed data using GraphPad Prism10
695 (GraphPad Software Inc.). Significance was ascribed to p values < 0.05 compared to
696 control conditions.
697

698 **Data and code**

699 The RNA-seq data reported in this paper are available at ArrayExpress (E-MTAB-
700 15071, E-MTAB-15072, E-MTAB-15077, E-MTAB-15065, E-MTAB-15056, E-MTAB-
701 15061, E-MTAB-15060, E-MTAB-15095, E-MTAB-15064). Custom scripts for the 5'
702 and 3' RACE analysis are available at https://github.com/Tycour/HIV1-ZAP_RACEseq
703 and https://github.com/GregaGD/ZAP_HIV-1_3-prime_RACE-seq.
704

705 **AUTHOR CONTRIBUTIONS**

706 Conceptualization: C.M.S, S.J.D.N, J.U; Software: A.M.C.; Formal analysis: C.B.,
707 G.G.D., M.J.L, R.P.G, T.C., H.W., P.H., H.M., A.M.C., Investigation: C.B., G.G.D.,
708 M.J.L., R.P.G, H.W.; Writing - Original Draft: C.M.S, C.B., G.G.D., Writing - Review &
709 Editing: all authors; Visualization: C.B., G.G.D, M.J.L, R.P.G., T.C.; Supervision: J.U.,
710 S.J.D.N, C.M.S.; Funding acquisition: M.P., J.U., S.J.D.N., C.M.S.
711

712 **DECLARATION OF INTERESTS**

713 The authors declare no competing interests.
714

715 **ACKNOWLEDGEMENTS**

716 We thank members of the Ule, Neil and Swanson laboratories for helpful discussions.
717 The following reagents were obtained through the NIH AIDS Research and Reference
718 Reagent Program, Division of AIDS, NIAID, NIH: TZM-bl from Dr. John C. Kappes, Dr.
719 Xiaoyun Wu and Tranzyme Inc; HIV-1 p24 Hybridoma (183-H12-5C) from Dr. Bruce
720 Chesebro. The Antiserum to HIV-1 gp120 #20 (ARP421) was obtained from the NIBSC
721 Centre for AIDS Reagents. We thank Xavier Roca for the ZAP-L and ZAP-S knockout
722 cells. We gratefully acknowledge the support of King's Genomics, Social Genetic &
723 Development Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience,
724 King's College London for the Oxford Nanopore Technologies PromethION
725 sequencing and the HPC CREATE at King’s College London. These studies were
726 funded by Medical Research Council grants MR/S000844/1 and MR/W018519/1 to
727 SJDN and CMS, Wellcome Early Career Award (225081/Z/22/Z) to AMC, Sir Henry
728 Dale Fellowship from the Wellcome Trust and the Royal Society (218537/Z/19/Z) to
729 HEM, Wellcome Trust Senior Research Fellowship (WT098049AIA) to SJDN, the
730 European Research Council (ERC) under the European Union’s Horizon 2020

research and innovation programme (grant agreement No 835300) to JU, and Slovenian Research and Innovation Agency ARIS grant P3-00083 to MP and GGD. The Slovenian Research Agency financially supported GGD through the Young Researcher Training Program for four years (2021–2025; contract number 802-7/2021-1).

REFERENCES

1. Chan, C.P., and Jin, D.Y. (2022). Cytoplasmic RNA sensors and their interplay with RNA-binding partners in innate antiviral response: theme and variations. RNA *28*, 449-477. 10.1261/rna.079016.121.
2. Yang, E., and Li, M.M.H. (2020). All About the RNA: Interferon-Stimulated Genes That Interfere With Viral RNA Processes. Front Immunol *11*, 605024. 10.3389/fimmu.2020.605024.
3. Ficarelli, M., Neil, S.J.D., and Swanson, C.M. (2021). Targeted Restriction of Viral Gene Expression and Replication by the ZAP Antiviral System. Annu Rev Virol *8*, 265-283. 10.1146/annurev-virology-091919-104213.
4. Shao, R., Visser, I., Fros, J.J., and Yin, X. (2024). Versatility of the Zinc-Finger Antiviral Protein (ZAP) As a Modulator of Viral Infections. Int J Biol Sci *20*, 4585-4600. 10.7150/ijbs.98029.
5. Gao, G., Guo, X., and Goff, S.P. (2002). Inhibition of retroviral RNA production by ZAP, a CCCH-type zinc finger protein. Science *297*, 1703-1706. 10.1126/science.1074276.
6. Bick, M.J., Carroll, J.W., Gao, G., Goff, S.P., Rice, C.M., and MacDonald, M.R. (2003). Expression of the zinc-finger antiviral protein inhibits alphavirus replication. J Virol *77*, 11555-11562. 10.1128/jvi.77.21.11555-11562.2003.

7. Zhu, Y., Wang, X., Goff, S.P., and Gao, G. (2012). Translational repression precedes and is required for ZAP-mediated mRNA decay. EMBO J *31*, 4236-4246. 10.1038/emboj.2012.271.

8. Zimmer, M.M., Kibe, A., Rand, U., Pekarek, L., Ye, L., Buck, S., Smyth, R.P., Cicin-Sain, L., and Caliskan, N. (2021). The short isoform of the host antiviral protein ZAP acts as an inhibitor of SARS-CoV-2 programmed ribosomal frameshifting. Nature communications *12*, 7193. 10.1038/s41467-021-27431-0.

9. Guo, X., Carroll, J.W., Macdonald, M.R., Goff, S.P., and Gao, G. (2004). The zinc finger antiviral protein directly binds to specific viral mRNAs through the CCCH zinc finger motifs. J Virol *78*, 12781-12787. 10.1128/JVI.78.23.12781-12787.2004.

10. Chen, S., Xu, Y., Zhang, K., Wang, X., Sun, J., Gao, G., and Liu, Y. (2012). Structure of N-terminal domain of ZAP indicates how a zinc-finger protein recognizes complex RNA. Nat Struct Mol Biol *19*, 430-435. 10.1038/nsmb.2243.

11. Meagher, J.L., Takata, M., Goncalves-Carneiro, D., Keane, S.C., Rebendenne, A., Ong, H., Orr, V.K., MacDonald, M.R., Stuckey, J.A., Bieniasz, P.D., and Smith, J.L. (2019). Structure of the zinc-finger antiviral protein in complex with RNA reveals a mechanism for selective targeting of CG-rich viral sequences. Proc Natl Acad Sci U S A *116*, 24303-24309. 10.1073/pnas.1913232116.

12. Luo, X., Wang, X., Gao, Y., Zhu, J., Liu, S., Gao, G., and Gao, P. (2020). Molecular Mechanism of RNA Recognition by Zinc-Finger Antiviral Protein. Cell reports *30*, 46-52 e44. 10.1016/j.celrep.2019.11.116.

781 13. Takata, M.A., Goncalves-Carneiro, D., Zang, T.M., Soll, S.J., York, A., Blanco-
782 Melo, D., and Bieniasz, P.D. (2017). CG dinucleotide suppression enables
783 antiviral defence targeting non-self RNA. Nature *550*, 124-127.
784 10.1038/nature24039.
785 14. Goncalves-Carneiro, D., Mastrocola, E., Lei, X., DaSilva, J., Chan, Y.F., and
786 Bieniasz, P.D. (2022). Rational attenuation of RNA viruses with zinc finger
787 antiviral protein. Nat Microbiol *7*, 1558-1567. 10.1038/s41564-022-01223-8.
788 15. Karlin, S., Doerfler, W., and Cardon, L.R. (1994). Why is CpG suppressed in
789 the genomes of virtually all small eukaryotic viruses but not in those of large
790 eukaryotic viruses? J Virol *68*, 2889-2897. 10.1128/JVI.68.5.2889-2897.1994.
791 16. Todorova, T., Bock, F.J., and Chang, P. (2014). PARP13 regulates cellular
792 mRNA post-transcriptionally and functions as a pro-apoptotic factor by
793 destabilizing TRAILR4 transcript. Nature communications *5*, 5362.
794 10.1038/ncomms6362.
795 17. Ly, P.T., Xu, S., Wirawan, M., Luo, D., and Roca, X. (2022). ZAP isoforms
796 regulate unfolded protein response and epithelial- mesenchymal transition.
797 Proc Natl Acad Sci U S A *119*, e2121453119. 10.1073/pnas.2121453119.
798 18. Goncalves-Carneiro, D., Mastrocola, E., Lei, X., and Bieniasz, P.D. (2025).
799 Modulation of host gene expression by the zinc finger antiviral protein. Proc
800 Natl Acad Sci U S A *122*, e2420819122. 10.1073/pnas.2420819122.
801 19. Busa, V.F., Ando, Y., Aigner, S., Yee, B.A., Yeo, G.W., and Leung, A.K.L.
802 (2024). Transcriptome regulation by PARP13 in basal and antiviral states in
803 human cells. iScience *27*, 109251. 10.1016/j.isci.2024.109251.
804 20. Cooper, D.N., and Gerber-Huber, S. (1985). DNA methylation and CpG
805 suppression. Cell Differ *17*, 199-205. 10.1016/0045-6039(85)90488-9.

806 21. Shaw, A.E., Rihn, S.J., Mollentze, N., Wickenhagen, A., Stewart, D.G., Orton,
807 R.J., Kuchi, S., Bakshi, S., Collados, M.R., Turnbull, M.L., et al. (2021). The
808 antiviral state has shaped the CpG composition of the vertebrate interferome
809 to avoid self-targeting. PLoS Biol *19*, e3001352.
810 10.1371/journal.pbio.3001352.
811 22. Nchioua, R., Kmiec, D., Krchlikova, V., Mattes, S., Noettger, S., Bibollet-
812 Ruche, F., Russell, R.M., Sparrer, K.M.J., Charpentier, T., Tardy, F., et al.
813 (2025). Host ZAP activity correlates with the levels of CpG suppression in
814 primate lentiviruses. Proc Natl Acad Sci U S A *122*, e2419489122.
815 10.1073/pnas.2419489122.
816 23. Kerns, J.A., Emerman, M., and Malik, H.S. (2008). Positive selection and
817 increased antiviral activity associated with the PARP-containing isoform of
818 human zinc-finger antiviral protein. PLoS genetics *4*, e21.
819 10.1371/journal.pgen.0040021.
820 24. Li, M.M.H., Aguilar, E.G., Michailidis, E., Pabon, J., Park, P., Wu, X., de Jong,
821 Y.P., Schneider, W.M., Molina, H., Rice, C.M., and MacDonald, M.R. (2019).
822 Characterization of Novel Splice Variants of Zinc Finger Antiviral Protein
823 (ZAP). J Virol *93*, e00715-00719. 10.1128/JVI.00715-19.
824 25. Hayakawa, S., Shiratori, S., Yamato, H., Kameyama, T., Kitatsuji, C., Kashigi,
825 F., Goto, S., Kameoka, S., Fujikura, D., Yamada, T., et al. (2011). ZAPS is a
826 potent stimulator of signaling mediated by the RNA helicase RIG-I during
827 antiviral responses. Nat Immunol *12*, 37-44. 10.1038/ni.1963.
828 26. Xue, G., Braczyk, K., Goncalves-Carneiro, D., Dawidziak, D.M., Sanchez, K.,
829 Ong, H., Wan, Y., Zadrozny, K.K., Ganser-Pornillos, B.K., Bieniasz, P.D., and

830 Pornillos, O. (2022). Poly(ADP-ribose) potentiates ZAP antiviral activity. PLoS
831 Pathog *18*, e1009202. 10.1371/journal.ppat.1009202.
832 27. Kuttiyatveetil, J.R.A., Soufari, H., Dasovich, M., Uribe, I.R., Mirhasan, M.,
833 Cheng, S.J., Leung, A.K.L., and Pascal, J.M. (2022). Crystal structures and
834 functional analysis of the ZnF5-WWE1-WWE2 region of PARP13/ZAP define
835 a distinctive mode of engaging poly(ADP-ribose). Cell reports *41*, 111529.
836 10.1016/j.celrep.2022.111529.
837 28. Charron, G., Li, M.M., MacDonald, M.R., and Hang, H.C. (2013). Prenylome
838 profiling reveals S-farnesylation is crucial for membrane targeting and antiviral
839 activity of ZAP long-isoform. Proc Natl Acad Sci U S A *110*, 11085-11090.
840 10.1073/pnas.1302564110.
841 29. Schwerk, J., Soveg, F.W., Ryan, A.P., Thomas, K.R., Hatfield, L.D., Ozarkar,
842 S., Forero, A., Kell, A.M., Roby, J.A., So, L., et al. (2019). RNA-binding protein
843 isoforms ZAP-S and ZAP-L have distinct antiviral and immune resolution
844 functions. Nat Immunol *20*, 1610-1620. 10.1038/s41590-019-0527-6.
845 30. Kmiec, D., Lista, M.J., Ficarelli, M., Swanson, C.M., and Neil, S.J.D. (2021).
846 S-farnesylation is essential for antiviral activity of the long ZAP isoform
847 against RNA viruses with diverse replication strategies. PLoS Pathog *17*,
848 e1009726. 10.1371/journal.ppat.1009726.
849 31. Guo, X., Ma, J., Sun, J., and Gao, G. (2007). The zinc-finger antiviral protein
850 recruits the RNA processing exosome to degrade the target mRNA. Proc Natl
851 Acad Sci U S A *104*, 151-156. 10.1073/pnas.0607063104.
852 32. Zhu, Y., Chen, G., Lv, F., Wang, X., Ji, X., Xu, Y., Sun, J., Wu, L., Zheng, Y.T.,
853 and Gao, G. (2011). Zinc-finger antiviral protein inhibits HIV-1 infection by

854 selectively targeting multiply spliced viral mRNAs for degradation. Proc Natl
855 Acad Sci U S A *108*, 15834-15839. 10.1073/pnas.1101676108.
856 33. Ficarelli, M., Wilson, H., Pedro Galao, R., Mazzon, M., Antzin-Anduetza, I.,
857 Marsh, M., Neil, S.J., and Swanson, C.M. (2019). KHNYN is essential for the
858 zinc finger antiviral protein (ZAP) to restrict HIV-1 containing clustered CpG
859 dinucleotides. Elife *8*, e46767. 10.7554/eLife.46767.
860 34. Bohn, J.A., Meagher, J.L., Takata, M.A., Goncalves-Carneiro, D., Yeoh, Z.C.,
861 Ohi, M.D., Smith, J.L., and Bieniasz, P.D. (2024). Functional anatomy of zinc
862 finger antiviral protein complexes. Nature communications *15*, 10834.
863 10.1038/s41467-024-55192-z.
864 35. Yeoh, Z.C., Meagher, J.L., Kang, C.Y., Bieniasz, P.D., Smith, J.L., and Ohi,
865 M.D. (2024). A minimal complex of KHNYN and zinc-finger antiviral protein
866 binds and degrades single-stranded RNA. Proc Natl Acad Sci U S A *121*,
867 e2415048121. 10.1073/pnas.2415048121.
868 36. OhAinle, M., Helms, L., Vermeire, J., Roesch, F., Humes, D., Basom, R.,
869 Delrow, J.J., Overbaugh, J., and Emerman, M. (2018). A virus-packageable
870 CRISPR screen identifies host factors mediating interferon inhibition of HIV.
871 Elife *7*, e39823. 10.7554/eLife.39823.
872 37. Youle, R.L., Lista, M.J., Bouton, C., Kunzelmann, S., Wilson, H., Cottee, M.A.,
873 Purkiss, A.G., Morris, E.R., Neil, S.J.D., Taylor, I.A., and Swanson, C.M.
874 (2025). Structural and functional characterization of the extended-diKH
875 domain from the antiviral endoribonuclease KHNYN. J Biol Chem, 108336.
876 10.1016/j.jbc.2025.108336.
877 38. Li, M.M., Lau, Z., Cheung, P., Aguilar, E.G., Schneider, W.M., Bozzacco, L.,
878 Molina, H., Buehler, E., Takaoka, A., Rice, C.M., et al. (2017). TRIM25

879 Enhances the Antiviral Action of Zinc-Finger Antiviral Protein (ZAP). PLoS
880 Pathog *13*, e1006145. 10.1371/journal.ppat.1006145.
881 39. Zheng, X., Wang, X., Tu, F., Wang, Q., Fan, Z., and Gao, G. (2017). TRIM25
882 Is Required for the Antiviral Activity of Zinc Finger Antiviral Protein. J Virol *91*,
883 e00088-00017. 10.1128/JVI.00088-17.
884 40. Kim, M., Pyo, Y., Hyun, S.I., Jeong, M., Choi, Y., and Kim, V.N. (2025).
885 Exogenous RNA surveillance by proton-sensing TRIM25. Science *388*,
886 eads4539. 10.1126/science.ads4539.
887 41. Fros, J.J., Visser, I., Tang, B., Yan, K., Nakayama, E., Visser, T.M., Koenraadt,
888 C.J.M., van Oers, M.M., Pijlman, G.P., Suhrbier, A., and Simmonds, P. (2021).
889 The dinucleotide composition of the Zika virus genome is shaped by
890 conflicting evolutionary pressures in mammalian hosts and mosquito vectors.
891 PLoS Biol *19*, e3001201. 10.1371/journal.pbio.3001201.
892 42. Sharp, C.P., Thompson, B.H., Nash, T.J., Diebold, O., Pinto, R.M., Thorley, L.,
893 Lin, Y.T., Sives, S., Wise, H., Clohisey Hendry, S., et al. (2023). CpG
894 dinucleotide enrichment in the influenza A virus genome as a live attenuated
895 vaccine development strategy. PLoS Pathog *19*, e1011357.
896 10.1371/journal.ppat.1011357.
897 43. Lin, Y., Zhang, H., Liang, J., Li, K., Zhu, W., Fu, L., Wang, F., Zheng, X., Shi,
898 H., Wu, S., et al. (2014). Identification and characterization of alphavirus M1
899 as a selective oncolytic virus targeting ZAP-defective human cancers. Proc
900 Natl Acad Sci U S A *111*, E4504-4512. 10.1073/pnas.1408759111.
901 44. Kypr, J., Mrazek, J., and Reich, J. (1989). Nucleotide composition bias and
902 CpG dinucleotide content in the genomes of HIV and HTLV 1/2. Biochim
903 Biophys Acta *1009*, 280-282. 10.1016/0167-4781(89)90114-0.

904 45. Shpaer, E.G., and Mullins, J.I. (1990). Selection against CpG dinucleotides in
905 lentiviral genes: a possible role of methylation in regulation of viral expression.
906 Nucleic Acids Res *18*, 5793-5797. 10.1093/nar/18.19.5793.
907 46. Wasson, M.K., Borkakoti, J., Kumar, A., Biswas, B., and Vivekanandan, P.
908 (2017). The CpG dinucleotide content of the HIV-1 envelope gene may predict
909 disease progression. Sci Rep *7*, 8162. 10.1038/s41598-017-08716-1.
910 47. Theys, K., Feder, A.F., Gelbart, M., Hartl, M., Stern, A., and Pennings, P.S.
911 (2018). Within-patient mutation frequencies reveal fitness costs of CpG
912 dinucleotides and drastic amino acid changes in HIV. PLoS genetics *14*,
913 e1007420. 10.1371/journal.pgen.1007420.
914 48. Kmiec, D., Nchioua, R., Sherrill-Mix, S., Sturzel, C.M., Heusinger, E., Braun,
915 E., Gondim, M.V.P., Hotter, D., Sparrer, K.M.J., Hahn, B.H., et al. (2020). CpG
916 Frequency in the 5' Third of the env Gene Determines Sensitivity of Primary
917 HIV-1 Strains to the Zinc-Finger Antiviral Protein. mBio *11*, e02903-02919.
918 10.1128/mBio.02903-19.
919 49. Goncalves-Carneiro, D., Takata, M.A., Ong, H., Shilton, A., and Bieniasz, P.D.
920 (2021). Origin and evolution of the zinc finger antiviral protein. PLoS Pathog
921 *17*, e1009545. 10.1371/journal.ppat.1009545.
922 50. Lista, M.J., Ficarelli, M., Wilson, H., Kmiec, D., Youle, R.L., Wanford, J.,
923 Winstone, H., Odendall, C., Taylor, I.A., Neil, S.J.D., and Swanson, C.M.
924 (2023). A Nuclear Export Signal in KHNYN Required for Its Antiviral Activity
925 Evolved as ZAP Emerged in Tetrapods. J Virol *97*, e0087222.
926 10.1128/jvi.00872-22.
927 51. Konig, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J.,
928 Luscombe, N.M., and Ule, J. (2010). iCLIP reveals the function of hnRNP

929 particles in splicing at individual nucleotide resolution. Nat Struct Mol Biol *17*,
930 909-915. 10.1038/nsmb.1838.
931 52. Keane, S.C., and Summers, M.F. (2016). NMR Studies of the Structure and
932 Function of the HIV-1 5'-Leader. Viruses *8*, 338. 10.3390/v8120338.
933 53. Sertznig, H., Hillebrand, F., Erkelenz, S., Schaal, H., and Widera, M. (2018).
934 Behind the scenes of HIV-1 replication: Alternative splicing as the dependency
935 factor on the quiet. Virology *516*, 176-188. 10.1016/j.virol.2018.01.011.
936 54. Law, L.M., Albin, O.R., Carroll, J.W., Jones, C.T., Rice, C.M., and Macdonald,
937 M.R. (2010). Identification of a dominant negative inhibitor of human zinc
938 finger antiviral protein reveals a functional endogenous pool and critical
939 homotypic interactions. J Virol *84*, 4504-4512. 10.1128/JVI.02018-09.
940 55. Hong, S., and Choe, J. (2024). Crystallization and biochemical studies of the
941 NYN domain of human KHNYN. Acta Crystallogr F Struct Biol Commun *80*,
942 67-72. 10.1107/S2053230X24000943.
943 56. Youn, J.Y., Dunham, W.H., Hong, S.J., Knight, J.D.R., Bashkurov, M., Chen,
944 G.I., Bagci, H., Rathod, B., MacLeod, G., Eng, S.W.M., et al. (2018). High-
945 Density Proximity Mapping Reveals the Subcellular Organization of mRNA-
946 Associated Granules and Bodies. Mol Cell *69*, 517-532 e511.
947 10.1016/j.molcel.2017.12.020.
948 57. Chen, G., Guo, X., Lv, F., Xu, Y., and Gao, G. (2008). p72 DEAD box RNA
949 helicase is required for optimal function of the zinc-finger antiviral protein.
950 Proc Natl Acad Sci U S A *105*, 4352-4357. 10.1073/pnas.0712276105.
951 58. Ye, P., Liu, S., Zhu, Y., Chen, G., and Gao, G. (2010). DEXH-Box protein
952 DHX30 is required for optimal function of the zinc-finger antiviral protein.
953 Protein Cell *1*, 956-964. 10.1007/s13238-010-0117-8.

954 59. Goodier, J.L., Pereira, G.C., Cheung, L.E., Rose, R.J., and Kazazian, H.H., Jr.
955 (2015). The Broad-Spectrum Antiviral Protein ZAP Restricts Human
956 Retrotransposition. PLoS genetics *11*, e1005252.
957 10.1371/journal.pgen.1005252.
958 60. Yu, S., and Kim, V.N. (2020). A tale of non-canonical tails: gene regulation by
959 post-transcriptional RNA tailing. Nat Rev Mol Cell Biol *21*, 542-556.
960 10.1038/s41580-020-0246-8.
961 61. Chang, H.M., Triboulet, R., Thornton, J.E., and Gregory, R.I. (2013). A role for
962 the Perlman syndrome exonuclease Dis3l2 in the Lin28-let-7 pathway. Nature
963 *497*, 244-248. 10.1038/nature12119.
964 62. Lubas, M., Damgaard, C.K., Tomecki, R., Cysewski, D., Jensen, T.H., and
965 Dziembowski, A. (2013). Exonuclease hDIS3L2 specifies an exosome-
966 independent 3'-5' degradation pathway of human cytoplasmic mRNA. EMBO
967 J *32*, 1855-1868. 10.1038/emboj.2013.135.
968 63. Malecki, M., Viegas, S.C., Carneiro, T., Golik, P., Dressaire, C., Ferreira, M.G.,
969 and Arraiano, C.M. (2013). The exoribonuclease Dis3L2 defines a novel
970 eukaryotic RNA degradation pathway. EMBO J *32*, 1842-1854.
971 10.1038/emboj.2013.63.
972 64. Odon, V., Fros, J.J., Goonawardane, N., Dietrich, I., Ibrahim, A.,
973 Alshaikhahmed, K., Nguyen, D., and Simmonds, P. (2019). The role of ZAP
974 and OAS3/RNAseL pathways in the attenuation of an RNA virus with elevated
975 frequencies of CpG and UpA dinucleotides. Nucleic Acids Res *47*, 8061-8083.
976 10.1093/nar/gkz581.
977 65. Ficarelli, M., Antzin-Anduetza, I., Hugh-White, R., Firth, A.E., Sertkaya, H.,
978 Wilson, H., Neil, S.J.D., Schulz, R., and Swanson, C.M. (2020). CpG

979 Dinucleotides Inhibit HIV-1 Replication through Zinc Finger Antiviral Protein
980 (ZAP)-Dependent and -Independent Mechanisms. J Virol *94*, e01337-01319.
981 10.1128/JVI.01337-19.
982 66. Groenke, N., Trimpert, J., Merz, S., Conradie, A.M., Wyler, E., Zhang, H.,
983 Hazapis, O.G., Rausch, S., Landthaler, M., Osterrieder, N., and Kunec, D.
984 (2020). Mechanism of Virus Attenuation by Codon Pair Deoptimization. Cell
985 reports *31*, 107586. 10.1016/j.celrep.2020.107586.
986 67. Nguyen, L.P., Aldana, K.S., Yang, E., Yao, Z., and Li, M.M.H. (2023).
987 Alphavirus Evasion of Zinc Finger Antiviral Protein (ZAP) Correlates with CpG
988 Suppression in a Specific Viral nsP2 Gene Sequence. Viruses *15*.
989 10.3390/v15040830.
990 68. Sertkaya, H., Hidalgo, L., Ficarelli, M., Kmiec, D., Signell, A.W., Ali, S., Parker,
991 H., Wilson, H., Neil, S.J.D., Malim, M.H., et al. (2021). Minimal impact of ZAP
992 on lentiviral vector production and transduction efficiency. Mol Ther Methods
993 Clin Dev *23*, 147-157. 10.1016/j.omtm.2021.08.008.
994 69. Choudhury, N.R., Heikel, G., Trubitsyna, M., Kubik, P., Nowak, J.S., Webb, S.,
995 Granneman, S., Spanos, C., Rappsilber, J., Castello, A., and Michlewski, G.
996 (2017). RNA-binding activity of TRIM25 is mediated by its PRY/SPRY domain
997 and is required for ubiquitination. BMC Biol *15*, 105. 10.1186/s12915-017-
998 0444-9.
999 70. Sanchez, J.G., Sparrer, K.M.J., Chiang, C., Reis, R.A., Chiang, J.J., Zurenski,
1000 M.A., Wan, Y., Gack, M.U., and Pornillos, O. (2018). TRIM25 Binds RNA to
1001 Modulate Cellular Anti-viral Defense. J Mol Biol *430*, 5280-5293.
1002 10.1016/j.jmb.2018.10.003.

1003 71. Alvarez, L., Haubrich, K., Iselin, L., Gillioz, L., Ruscica, V., Lapouge, K.,
1004 Augsten, S., Huppertz, I., Choudhury, N.R., Simon, B., et al. (2024). The
1005 molecular dissection of TRIM25's RNA-binding mechanism provides key
1006 insights into its antiviral activity. Nature communications *15*, 8485.
1007 10.1038/s41467-024-52918-x.
1008 72. Galao, R.P., Wilson, H., Schierhorn, K.L., Debeljak, F., Bodmer, B.S., Goldhill,
1009 D., Hoenen, T., Wilson, S.J., Swanson, C.M., and Neil, S.J.D. (2022). TRIM25
1010 and ZAP target the Ebola virus ribonucleoprotein complex to mediate
1011 interferon-induced restriction. PLoS Pathog *18*, e1010530.
1012 10.1371/journal.ppat.1010530.
1013 73. Yang, E., Huang, S., Jami-Alahmadi, Y., McInerney, G.M., Wohlschlegel, J.A.,
1014 and Li, M.M.H. (2022). Elucidation of TRIM25 ubiquitination targets involved in
1015 diverse cellular and antiviral processes. PLoS Pathog *18*, e1010743.
1016 10.1371/journal.ppat.1010743.
1017 74. Sanchez, J.G., Chiang, J.J., Sparrer, K.M.J., Alam, S.L., Chi, M., Roganowicz,
1018 M.D., Sankaran, B., Gack, M.U., and Pornillos, O. (2016). Mechanism of
1019 TRIM25 Catalytic Activation in the Antiviral RIG-I Pathway. Cell reports *16*,
1020 1315-1325. 10.1016/j.celrep.2016.06.070.
1021 75. Meze, K., Axhemi, A., Thomas, D.R., Doymaz, A., and Joshua-Tor, L. (2023).
1022 A shape-shifting nuclease unravels structured RNA. Nat Struct Mol Biol *30*,
1023 339-347. 10.1038/s41594-023-00923-x.
1024 76. Watts, J.M., Dang, K.K., Gorelick, R.J., Leonard, C.W., Bess, J.W., Jr.,
1025 Swanstrom, R., Burch, C.L., and Weeks, K.M. (2009). Architecture and
1026 secondary structure of an entire HIV-1 RNA genome. Nature *460*, 711-716.
1027 10.1038/nature08237.

1028 77. Garcia-Moreno, M., Noerenberg, M., Ni, S., Jarvelin, A.I., Gonzalez-Almela,
1029 E., Lenz, C.E., Bach-Pages, M., Cox, V., Avolio, R., Davis, T., et al. (2019).
1030 System-wide Profiling of RNA-Binding Proteins Uncovers Key Regulators of
1031 Virus Infection. Mol Cell *74*, 196-211 e111. 10.1016/j.molcel.2019.01.017.
1032 78. Rozman, B., Nachshon, A., Levi Samia, R., Lavi, M., Schwartz, M., and Stern-
1033 Ginossar, N. (2022). Temporal dynamics of HCMV gene expression in lytic
1034 and latent infections. Cell reports *39*, 110653. 10.1016/j.celrep.2022.110653.
1035 79. Burgess, H.M., Grande, R., Riccio, S., Dinesh, I., Winkler, G.S., Depledge,
1036 D.P., and Mohr, I. (2023). CCR4-NOT differentially controls host versus virus
1037 poly(a)-tail length and regulates HCMV infection. EMBO Rep *24*, e56327.
1038 10.15252/embr.202256327.
1039 80. Derdeyn, C.A., Decker, J.M., Sfakianos, J.N., Wu, X., O'Brien, W.A., Ratner,
1040 L., Kappes, J.C., Shaw, G.M., and Hunter, E. (2000). Sensitivity of human
1041 immunodeficiency virus type 1 to the fusion inhibitor T-20 is modulated by
1042 coreceptor specificity defined by the V3 loop of gp120. J Virol *74*, 8358-8367.
1043 10.1128/jvi.74.18.8358-8367.2000.
1044 81. Wei, X., Decker, J.M., Liu, H., Zhang, Z., Arani, R.B., Kilby, J.M., Saag, M.S.,
1045 Wu, X., Shaw, G.M., and Kappes, J.C. (2002). Emergence of resistant human
1046 immunodeficiency virus type 1 in patients receiving fusion inhibitor (T-20)
1047 monotherapy. Antimicrob Agents Chemother *46*, 1896-1905.
1048 10.1128/aac.46.6.1896-1905.2002.
1049 82. Platt, E.J., Wehrly, K., Kuhmann, S.E., Chesebro, B., and Kabat, D. (1998).
1050 Effects of CCR5 and CD4 cell surface concentrations on infections by
1051 macrophagetropic isolates of human immunodeficiency virus type 1. J Virol
1052 *72*, 2855-2864. 10.1128/JVI.72.4.2855-2864.1998.

83. Lee, F.C.Y., Chakrabarti, A.M., Hänel, H., Monzón-Casanova, E., Hallegger, M., Militti, C., Capraro, F., Sadée, C., Toolan-Kerr, P., Wilkins, O., et al. (2021). An improved iCLIP protocol. bioRxiv, 2021.2008.2027.457890. 10.1101/2021.08.27.457890.

84. Smith, T., Heger, A., and Sudbery, I. (2017). UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res *27*, 491-499. 10.1101/gr.209601.116.

85. Bushnell, B., Rood, J., and Singer, E. (2017). BBMerge - Accurate paired shotgun read merging via overlap. PLoS One *12*, e0185056. 10.1371/journal.pone.0185056.

86. Ewels, P.A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M.U., Di Tommaso, P., and Nahnsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines. Nat Biotechnol *38*, 276-278. 10.1038/s41587-020-0439-x.

87. West, C., Capitanchik, C., Cheshire, C., Luscombe, N.M., Chakrabarti, A., and Ule, J. (2023). nf-core/clipseq - a robust Nextflow pipeline for comprehensive CLIP data analysis. Wellcome Open Res *8*, 286. 10.12688/wellcomeopenres.19453.1.

88. Di Tommaso, P., Chatzou, M., Floden, E.W., Barja, P.P., Palumbo, E., and Notredame, C. (2017). Nextflow enables reproducible computational workflows. Nat Biotechnol *35*, 316-319. 10.1038/nbt.3820.

89. Frith, M.C., Valen, E., Krogh, A., Hayashizaki, Y., Carninci, P., and Sandelin, A. (2008). A code for transcription initiation in mammalian genomes. Genome Res *18*, 1-12. 10.1101/gr.6831208.

1077 90. Chakrabarti, A.M., Capitanchik, C., Ule, J., and Luscombe, N.M. (2023).
1078 clipplotr-a comparative visualization and analysis tool for CLIP data. RNA *29*,
1079 715-723. 10.1261/rna.079326.122.
1080 91. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R
1081 package for comparing biological themes among gene clusters. Omics : a
1082 journal of integrative biology *16*, 284-287. 10.1089/omi.2011.0118.
1083 92. Chen, H., and Boutros, P.C. (2011). VennDiagram: a package for the
1084 generation of highly-customizable Venn and Euler diagrams in R. BMC
1085 bioinformatics *12*, 35. 10.1186/1471-2105-12-35.
1086 93. De Coster, W., D'Hert, S., Schultz, D.T., Cruts, M., and Van Broeckhoven, C.
1087 (2018). NanoPack: visualizing and processing long-read sequencing data.
1088 Bioinformatics *34*, 2666-2669. 10.1093/bioinformatics/bty149.
1089 94. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences.
1090 Bioinformatics *34*, 3094-3100. 10.1093/bioinformatics/bty191.
1091 95. Li, H. (2021). New strategies to improve minimap2 alignment accuracy.
1092 Bioinformatics *37*, 4572-4574. 10.1093/bioinformatics/btab705.
1093 96. Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman,
1094 R., Morgan, M.T., and Carey, V.J. (2013). Software for computing and
1095 annotating genomic ranges. PLoS computational biology *9*, e1003118.
1096 10.1371/journal.pcbi.1003118.
1097

1098 **FIGURE LEGENDS**

1099 **Figure 1: ZAP isoforms differ in their activity to restrict HIV-CpG and interact**
1100 **with KHNYN and viral RNA.**

1101 **(A)** Genomic organisation and localisation of CpG dinucleotides of HIV-WT and HIV-
1102 CpG.
1103 **(B)** Relative infectious HIV production (left) in HEK293T CRISPR control and ZAP
1104 cells, and HIV-CpG production expressed as % infectious HIV-WT production (right).
1105 Error bars represent standard deviation. *p <0.05 calculated by one-way ANOVA on
1106 log-transformed data, compared to HIV-CpG production in HEK293T CRISPR control
1107 cells.
1108 **(C)** Schematic representation of ZAP's main isoforms, ZAP-L and ZAP-S. ZnF: zinc
1109 finger. ART-like: ADP-ribosyltransferase-like domain. CaaX: CaaX box that mediates
1110 S-farnesylation.
1111 **(D)** Immunoblotting of uninfected isoform-specific HEK293T CRISPR ZAP cells.
1112 **(E)** Infectious HIV-CpG production, expressed as % infectious HIV-WT production, in
1113 isoform-specific HEK293T CRISPR ZAP cells. Error bars represent standard
1114 deviation. *p <0.05 calculated by one-way ANOVA on log-transformed data, compared
1115 to HIV-CpG production in HEK293T cells.
1116 **(F)** Immunoblotting of ZAP coimmunoprecipitations and corresponding input lysates
1117 obtained from isoform-specific HEK293T CRISPR ZAP cells at steady state level.
1118 **(G)** Normalised ZAP iCLIP signal mapped to HIV-CpG gRNA in isoform-specific
1119 HEK293T CRISPR ZAP cells. The average signal of three independent experiments
1120 and the SEM (shading) are plotted. Localisation of CpG dinucleotides on HIV-CpG
1121 gRNA are indicated by vertical bars. Pink bars represent the 36 added CpG
1122 dinucleotides in HIV-CpG.
1123
1124 **Figure 2: TRIM25 modulates ZAP binding to HIV RNA.**

1125 **(A)** Immunoblotting of ZAP coimmunoprecipitations and corresponding input lysates
1126 obtained from infected HeLa cells. Samples were treated with RNase I after pulldown.
1127 **(B)** Immunoblotting of TRIM25 coimmunoprecipitations and corresponding input
1128 lysates obtained from infected HeLa cells. Samples were treated with RNase I after
1129 pulldown.
1130 **(C)** ZAP iCLIP signal mapped to HIV-CpG gRNA in HeLa CRISPR TRIM25 and control.
1131 The average signal of three independent experiments and the SEM (shading) are
1132 plotted.
1133 **(D)** ZAP iCLIP signal mapped to HIV-CpG gRNA in HeLa CRISPR KHNYN and control.
1134 The average signal of three independent experiments and the SEM (shading) are
1135 plotted.
1136 **(E)** Relative levels of ZAP-associated HIV gRNA measured by qRT-PCR in HeLa
1137 CRISPR control (left), CRISPR TRIM25 (middle) and CRISPR KHNYN (right) cells
1138 transfected with proviral DNA. Error bars represent standard deviation. *p <0.05
1139 calculated by unpaired two-tailed t test, compared to ZAP-associated HIV-WT gRNA.
1140
1141 **Figure 3: Degradation of HIV-CpG 3' cleavage fragment by XRN1.**
1142 **(A)** KHNYN iCLIP signal mapped to HIV-CpG gRNA in HeLa CRISPR KHNYN + GFP
1143 and HeLa CRISPR KHNYN + CRISPR-resistant KHNYN-GFP. The average signal of
1144 three independent experiments and the SEM (shading) are plotted.
1145 **(B)** Genomic organisation of HIV-WT and HIV-CpG and localisation of CpG
1146 dinucleotides. The qRT-PCR probe used to quantify the 3' cleavage fragment (probe
1147 2 (p2)), and the HIV specific primer used for the 5' RACE (5'R) are shown.
1148 **(C)** Level of HIV-CpG 3' cleavage fragment measured by qRT-PCR in cells transfected
1149 with siRNA targeting XRN1. Results are expressed as % 3' cleavage fragment in HIV-

1150 WT infected cells. The dotted line represents the level of 3' cleavage fragment in cells
1151 treated with a non-targeting siRNA control. Error bars represent standard deviation. *p
1152 <0.05 calculated by one-way ANOVA on log-transformed data, compared to the level
1153 of 3' cleavage fragment in cells treated with a non-targeting siRNA control.
1154 **(D)** Agarose gel electrophoresis of 5' RACE reactions. Sizes in kilobases are indicated
1155 on the left.
1156 **(E)** 5' RACE DNA was sequenced by long-read sequencing and the 5' end of each
1157 molecule was identified and mapped onto HIV gRNA. The number of 5' end at every
1158 position was determined and compared between experimental conditions. iCLIP
1159 signals for ZAP (from Figure S2B) and KHNYN (from panel A) are shown for reference.
1160 **(F)** Immunoblotting of ZAP coimmunoprecipitations and corresponding input lysates
1161 obtained from infected HeLa cells. Samples were treated with RNase I after pulldown.
1162 **(G)** Immunoblotting of TRIM25 coimmunoprecipitations and corresponding input
1163 lysates obtained from infected HeLa cells. Samples were treated with RNase I after
1164 pulldown. Note that the experiment shown here was performed alongside the
1165 experiment shown in Figure 2C, thus the input blots for TRIM25, PABPC1, GAPDH,
1166 and gp160/gp120 are identical between the two figures.
1167
1168 **Figure 4: HIV-CpG 5' cleavage fragment abundance is regulated by TUT4/TUT7.**
1169 **(A)** Agarose gel electrophoresis of 3' RACE reactions. Sizes in kilobases are indicated
1170 on the left.
1171 **(B)** Non-templated nucleotides are found in the 3' end of HIV-CpG 5' cleavage
1172 fragments amplified by 3' RACE and determined by Sanger sequencing. The 3' end
1173 positions shown here all fall within the CpG-enriched region of HIV-CpG.

1174 **(C)** Level of HIV-CpG 5' cleavage fragment measured by qRT-PCR in cells transfected
1175 with siRNA targeting TUTs. Results are expressed as % 5' cleavage fragment in HIV-
1176 WT infected cells. The dotted line represents the level of 5' cleavage fragment in cells
1177 treated with a non-targeting siRNA control. Error bars represent standard deviation. *p
1178 <0.05 calculated by one-way ANOVA on log-transformed data, compared to the level
1179 of 5' cleavage fragment in cells treated with a non-targeting siRNA control.
1180 **(D)** Infectious HIV-CpG production, expressed as % infectious HIV-WT production in
1181 HeLa cells transfected with siRNA targeting TUTs. The dotted line represents HIV-CpG
1182 production in cells transfected with a non-targeting control siRNA. Error bars represent
1183 standard deviation. *p <0.05 calculated by one-way ANOVA on log-transformed data,
1184 compared to HIV-CpG production in cells treated with a non-targeting siRNA control.
1185 **(E)** Immunoblotting of ZAP coimmunoprecipitations and corresponding input lysates
1186 obtained from infected HeLa cells. Samples were treated with RNase I after pulldown.
1187 **(F)** Immunoblotting of TRIM25 coimmunoprecipitations and corresponding input
1188 lysates obtained from infected HeLa cells. Samples were treated with RNase I after
1189 pulldown.
1190
1191 **Figure 5: HIV-CpG 5' cleavage fragment is degraded by DIS3L2.**
1192 **(A)** Level of HIV-CpG 5' cleavage fragment measured by qRT-PCR in cells transfected
1193 with siRNA targeting DIS3L2. Results are expressed as % 5' cleavage fragment in
1194 HIV-WT infected cells. The dotted line represents the level of 5' cleavage fragment in
1195 cells treated with a non-targeting siRNA control. Error bars represent standard
1196 deviation. *p <0.05 calculated by one-way ANOVA on log-transformed data, compared
1197 to the level of 5' cleavage fragment in cells treated with a non-targeting siRNA control.

1198 Note that this experiment was performed alongside the experiment shown in Figure
1199 5C and therefore shared the same control and ZAP values.
1200 **(B)** Immunoblotting of ZAP coimmunoprecipitations and corresponding input lysates
1201 obtained from infected HeLa cells. Samples were treated with RNase I after pulldown.
1202 **(C)** Immunoblotting of TRIM25 coimmunoprecipitations and corresponding input
1203 lysates obtained from infected HeLa cells. Samples were treated with RNase I after
1204 pulldown.
1205 **(D)** 3' RACE DNA was sequenced by long-read sequencing and the 3' end of each
1206 molecule was identified and mapped onto HIV gRNA. The number of 3' end at every
1207 position was determined and compared between experimental conditions. iCLIP
1208 signals for ZAP (from Figure S2B) and KHNYN (from Figure 3A) are shown for
1209 information.
1210 **(E)** The sequence and length of non-templated nucleotides was extracted for every
1211 sequenced 3' end in HIV-CpG infected cells transfected with siRNA control or siRNA
1212 targeting ZAP or DIS3L2.
1213
1214 **Figure 6: HIV infection increases the number of cellular genes that are**
1215 **coregulated by ZAP, TRIM25, and KHNYN.**
1216 **(A-B)** Hallmark gene set enrichment plot of differentially expressed genes in HIV-WT
1217 **(A)** or HIV-CpG **(B)** infected HeLa CRISPR control cells. Black dots are non-
1218 significant, blue and pink dots are significant. Pink dots show functional groups
1219 highlighted in the text.
1220 **(C)** Heat map indicating differentially expressed genes that are co-upregulated in ZAP,
1221 TRIM25 and KHNYN CRISPR cells compared to CRISPR control cells in mock, HIV-
1222 WT or HIV-CpG infected cells.

1223 **(D)** Venn-diagram showing differentially expressed genes that are co-upregulated in
1224 ZAP, TRIM25 and KHNYN CRISPR cells compared to CRISPR control cells in mock,
1225 HIV-WT or HIV-CpG infected cells.
1226 **(E)** Gene ontology analysis of the 102 genes that are upregulated in the absence of
1227 ZAP, TRIM25 and KHNYN upon HIV-WT and HIV-CpG infection. Black dots are non-
1228 significant, blue and pink dots are significant. Pink dots show functional groups
1229 highlighted in the text.
1230
1231 **Figure 7: Model of the ZAP-mediated RNA decay**
1232 ZAP triggers the assembly of ZAP-TRIM25-KHNYN complexes around CpG
1233 dinucleotides. This leads to RNA cleavage by KHNYN. The 3' cleavage fragment is
1234 then degraded by the 5'-3' exoribonuclease XRN1, while the 5' cleavage fragment is
1235 uridylated by TUT4/7 and then degraded by the 3'-5' exoribonuclease DIS3L2. Created
1236 in BioRender. Bouton, C. (2025) https://BioRender.com/f8zm1lm
1237
1238 **SUPPLEMENTAL FIGURE LEGENDS**
1239 **Figure S1: ZAP isoforms differ in their activity and their interaction with viral**
1240 **RNA.**
1241 **(A)** Relative infectious virus production in isoform-specific HEK293T CRISPR ZAP
1242 cells. The dotted line represents HIV-CpG production in HEK293T cells. Error bars
1243 represent standard deviation.
1244 **(B)** Immunoblotting of cell lysates and virions from HIV-WT- or HIV-CpG-infected
1245 isoform-specific HEK293T CRISPR ZAP cells.

1246 **(C-E)** Normalised ZAP iCLIP signal mapped to HIV gRNA in HEK293T cells **(C)**,
1247 HEK293T ZAP-L KO **(D)** and HEK293T ZAP-S KO **(E)**. The average of three
1248 independent experiments and the SEM are plotted.
1249
1250 **Figure S2: TRIM25 modulates ZAP binding to HIV RNA but does not alter its**
1251 **target specificity.**
1252 **(A-B)** Normalised ZAP iCLIP signal mapped to HIV gRNA in HeLa CRISPR control **(A)**
1253 or CRISPR TRIM25 cells **(B)**.
1254 **(C)** HIV-WT or HIV-CpG infectious virus production from CRISPR control, ZAP,
1255 TRIM25 or KHNYN knockout cells.
1256 **(D)** Venn-diagrams of genes encoding transcripts with >5 ZAP peaks in HeLa CRISPR
1257 control and TRIM25 knockout cells infected with HIV-WT (top) or HIV-CpG (bottom).
1258 **(E-F)** ZAP binding peaks on HIV-CpG RNA in HeLa CRISPR control **(E)** and KHNYN
1259 knockout cells **(F)**. Peaks were called by Clippy and their coordinates are indicated.
1260
1261 **Figure S3: Depletion of CNOT1, PARN, EXOSC3, EXOSC5, EXOSC7, DDX17,**
1262 **DHX30 or DCP2 does not affect HIV-CpG infectious virus production.**
1263 **(A)** Infectious HIV-CpG production, expressed as % infectious HIV-WT production in
1264 HeLa cells transfected with siRNA targeting ZAP, TRIM25 or KHNYN. The dotted line
1265 represents HIV-CpG production in cells transfected with a non-targeting control siRNA.
1266 Error bars represent standard deviation. *p <0.05 calculated by one-way ANOVA on
1267 log-transformed data, compared to HIV-CpG production in cells treated with a non-
1268 targeting siRNA control.
1269 **(B-C)** Infectious HIV-CpG production, expressed as % infectious HIV-WT production
1270 in HeLa cells transfected with siRNA targeting potential ZAP cofactors. The dotted line

1271 represents HIV-CpG production in cells transfected with a non-targeting control siRNA.
1272 Error bars represent standard deviation. *p <0.05 calculated by one-way ANOVA on
1273 log-transformed data, compared to HIV-CpG production in cells treated with a non-
1274 targeting siRNA control.
1275 **(D)** Immunoblotting of HeLa cells transfected with siRNA targeting the genes in (B)
1276 and (C), then infected with HIV-WT.
1277 **(E-F)** Level of HIV-CpG 3' cleavage fragment measured by qRT-PCR in cells
1278 transfected with siRNA targeting potential ZAP cofactors. Results are expressed as %
1279 3' cleavage fragment in HIV-WT infected cells. The dotted line represents the level of
1280 3' cleavage fragment in cells treated with a non-targeting siRNA control. Error bars
1281 represent standard deviation. *p <0.05 calculated by one-way ANOVA on log-
1282 transformed data, compared to the level of 3' cleavage fragment in cells treated with
1283 a non-targeting siRNA control.
1284
1285 **Figure S4: Depletion of CNOT1, PARN, EXOSC3, EXOSC5, EXOSC7, XRN1,**
1286 **DDX17, or DHX30 does not affect the abundance of the 5' HIV-CpG RNA cleavage**
1287 **fragment.**
1288 **(A)** Genomic organisation of HIV-WT and HIV-CpG and localisation of CpG
1289 dinucleotides. The qRT-PCR primer/probe set used to quantify the 5' cleavage
1290 fragment (probe 1 (p1)) and the HIV-specific primer used for the 3' RACE (3'R) are
1291 shown.
1292 **(B)** Level of HIV-CpG 5' cleavage fragment measured by qRT-PCR in cells transfected
1293 with siRNA targeting potential ZAP cofactors. Results are expressed as % 5' cleavage
1294 fragment in HIV-WT infected cells. The dotted line represents the level of 5' cleavage
1295 fragment in cells treated with a non-targeting siRNA control. Error bars represent

1296 standard deviation. *p <0.05 calculated by one-way ANOVA on log-transformed data,
1297 compared to the level of 3' cleavage fragment in cells treated with a non-targeting
1298 siRNA control.
1299 **(C)** Immunoblotting of HeLa cells transfected with siRNA targeting TUT4 and TUT7
1300 then infected with HIV-WT.
1301 **(D)** Level of HIV-CpG 3' cleavage fragment measured by qRT-PCR in cells transfected
1302 with siRNA targeting TUTs. Results are expressed as % 3' cleavage fragment in HIV-
1303 WT infected cells. The dotted line represents the level of 3' cleavage fragment in cells
1304 treated with a non-targeting siRNA control. Error bars represent standard deviation. *p
1305 <0.05 calculated by one-way ANOVA on log-transformed data, compared to the level
1306 of 3' cleavage fragment in cells treated with a non-targeting siRNA control.
1307
1308 **Figure S5: Depletion of DIS3L2 does not affect HIV-CpG infectious virus**
1309 **production or abundance of the 3' cleavage fragment.**
1310 **(A)** Immunoblotting of HeLa cells transfected with siRNA targeting DIS3L2 then
1311 infected with HIV-WT.
1312 **(B)** Level of HIV-CpG 3' cleavage fragment measured by qRT-PCR in cells transfected
1313 with siRNA targeting DIS3L2. Results are expressed as % 3' cleavage fragment in
1314 HIV-WT infected cells. The dotted line represents the level of 3' cleavage fragment in
1315 cells treated with a non-targeting siRNA control. Error bars represent standard
1316 deviation. *p <0.05 calculated by one-way ANOVA on log-transformed data, compared
1317 to the level of 5' cleavage fragment in cells treated with a non-targeting siRNA control.
1318 This experiment was performed alongside the experiment shown in figure S5D and
1319 therefore shared the same control and ZAP values.

1320 **(C)** Infectious HIV-CpG production, expressed as % infectious HIV-WT production in
1321 HeLa cells transfected with siRNA targeting DIS3L2. The dotted line represents HIV-
1322 CpG production in cells transfected with a non-targeting control siRNA. Error bars
1323 represent standard deviation. *p <0.05 calculated by one-way ANOVA on log-
1324 transformed data, compared to HIV-CpG production in cells treated with a non-
1325 targeting siRNA control. This experiment was performed alongside the experiment
1326 shown in figure 5D and therefore shared the same control and ZAP values.
1327
1328 **Figure S6: HIV-1 infection increases the number of ZAP, TRIM25 or KHNYN-**
1329 **regulated genes.**
1330 **(A)** Heat map indicating differentially expressed genes that are upregulated in ZAP,
1331 TRIM25 or KHNYN CRISPR cells compared to CRISPR control cells in mock, HIV-WT
1332 or HIV-CpG infected cells.
1333 **(B)** Number of upregulated genes in ZAP, KHNYN or TRIM25 knockout cells compared
1334 to CRISPR control cells in mock, HIV-WT or HIV-CpG infected cells.
1335 **(C)** Venn-diagram showing differentially expressed genes that are upregulated in ZAP,
1336 TRIM25 or KHNYN CRISPR cells compared to CRISPR control cells in mock, HIV-WT
1337 or HIV-CpG infected cells.

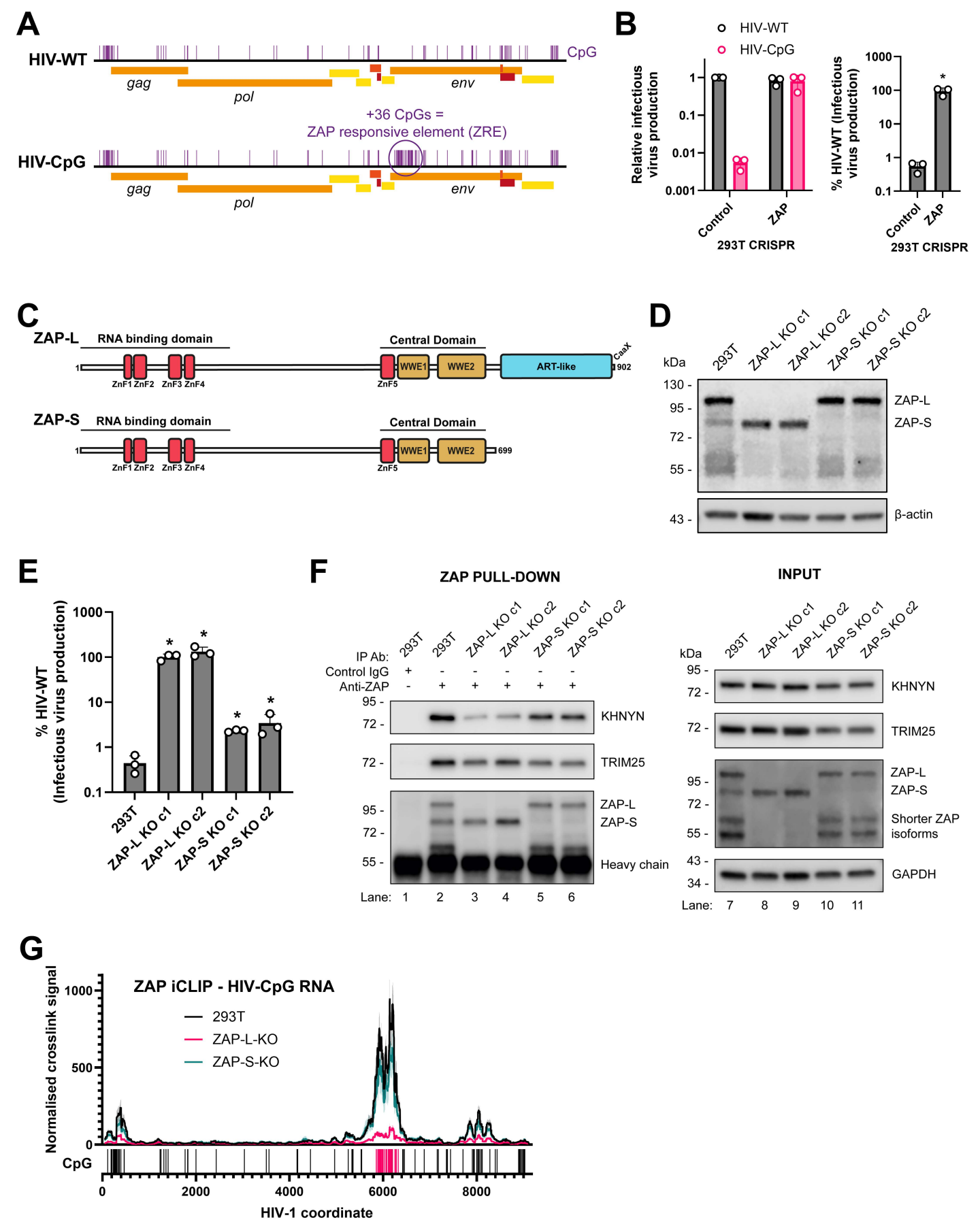

A
HIV-WT
CpG
gag
pol
env
+36 CpGs =
ZAP responsive element (ZRE)
HIV-CpG
gag
pol
env
B
HIV-WT
HIV-CpG
Relative infectious virus production
1
0.1
0.01
0.001
Control
ZAP
293T CRISPR
% HIV-WT (Infectious virus production)
1000
100
10
1
0.1
*
Control
ZAP
293T CRISPR
C
ZAP-L
RNA binding domain
Central Domain
1
ZnF1 ZnF2
ZnF3 ZnF4
ZnF5
WWE1
WWE2
ART-like
CaaX
902
ZAP-S
RNA binding domain
Central Domain
1
ZnF1 ZnF2
ZnF3 ZnF4
ZnF5
WWE1
WWE2
699
D
kDa
293T
ZAP-L KO c1
ZAP-L KO c2
ZAP-S KO c1
ZAP-S KO c2
130 -
95 -
72 -
55 -
ZAP-L
ZAP-S
43 -
β-actin
E
% HIV-WT
(Infectious virus production)
1000
100
10
1
0.1
*
*
*
*
293T
ZAP-L KO c1
ZAP-L KO c2
ZAP-S KO c1
ZAP-S KO c2
F
ZAP PULL-DOWN
IP Ab:
Control IgG
Anti-ZAP
293T
293T
ZAP-L KO c1
ZAP-L KO c2
ZAP-S KO c1
ZAP-S KO c2
95 -
72 -
KHNYN
72 -
TRIM25
95 -
72 -
55 -
ZAP-L
ZAP-S
Heavy chain
Lane: 1 2 3 4 5 6
INPUT
kDa
293T
ZAP-L KO c1
ZAP-L KO c2
ZAP-S KO c1
ZAP-S KO c2
95 -
72 -
KHNYN
72 -
TRIM25
95 -
72 -
55 -
ZAP-L
ZAP-S
Shorter ZAP isoforms
43 -
34 -
GAPDH
Lane: 7 8 9 10 11
G
ZAP iCLIP - HIV-CpG RNA
293T
ZAP-L-KO
ZAP-S-KO
Normalised crosslink signal
1000
500
0
CpG
0
2000
4000
6000
8000
HIV-1 coordinate

Figure 1

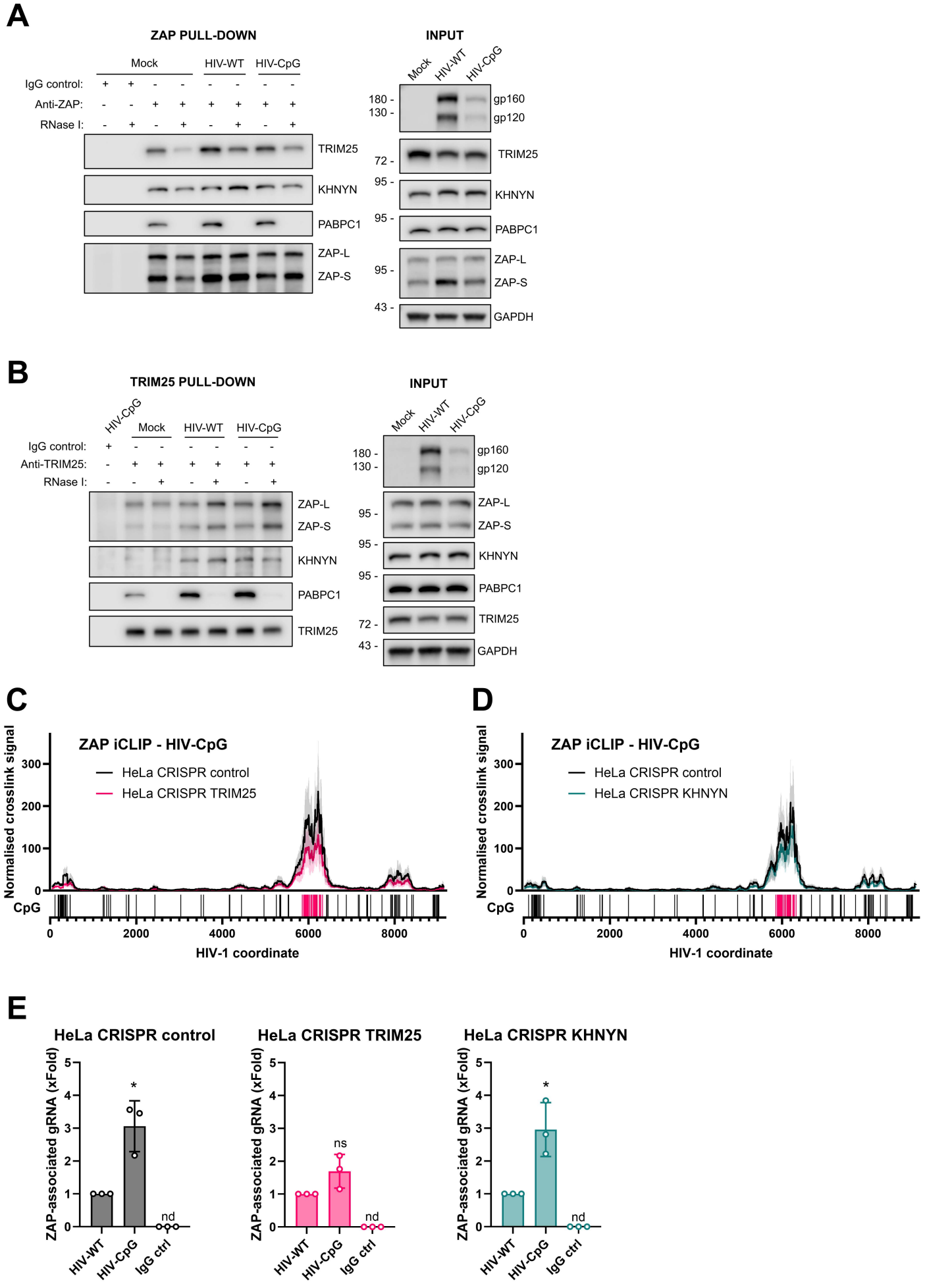

A
ZAP PULL-DOWN
Mock
HIV-WT
HIV-CpG
IgG control:
Anti-ZAP:
RNase I:
TRIM25
KHNYN
PABPC1
ZAP-L
ZAP-S
INPUT
Mock
HIV-WT
HIV-CpG
180 -
130 -
gp160
gp120
72 -
95 -
43 -
GAPDH
B
TRIM25 PULL-DOWN
HIV-CpG
Mock
HIV-WT
HIV-CpG
IgG control:
Anti-TRIM25:
RNase I:
ZAP-L
ZAP-S
KHNYN
PABPC1
TRIM25
INPUT
Mock
HIV-WT
HIV-CpG
180 -
130 -
gp160
gp120
95 -
72 -
43 -
GAPDH
C
ZAP iCLIP - HIV-CpG
HeLa CRISPR control
HeLa CRISPR TRIM25
Normalised crosslink signal
CpG
HIV-1 coordinate
D
ZAP iCLIP - HIV-CpG
HeLa CRISPR control
HeLa CRISPR KHNYN
Normalised crosslink signal
CpG
HIV-1 coordinate
E
HeLa CRISPR control
HeLa CRISPR TRIM25
HeLa CRISPR KHNYN
ZAP-associated gRNA (xFold)
*
ns
nd
HIV-WT
HIV-CpG
IgG ctrl

Figure 2

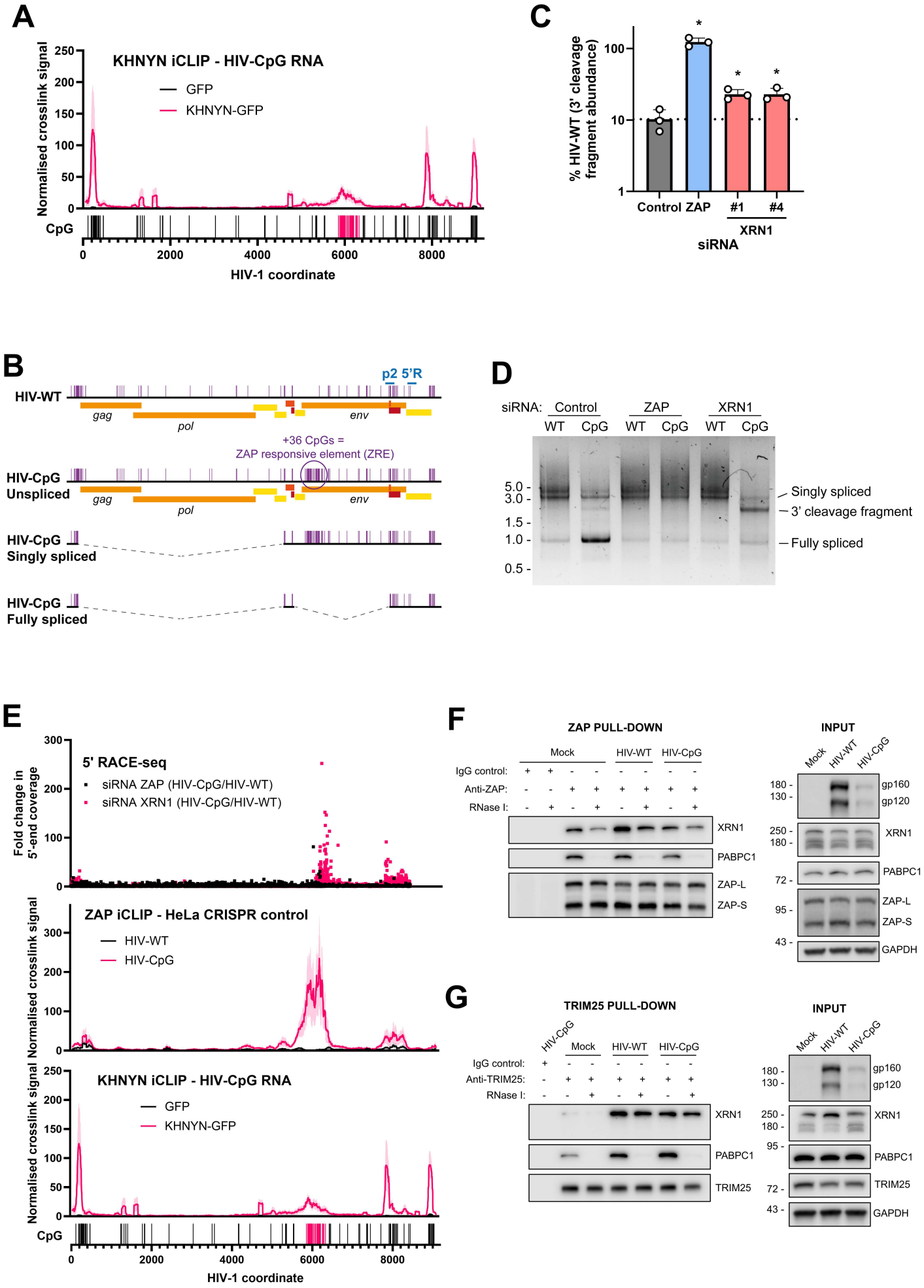
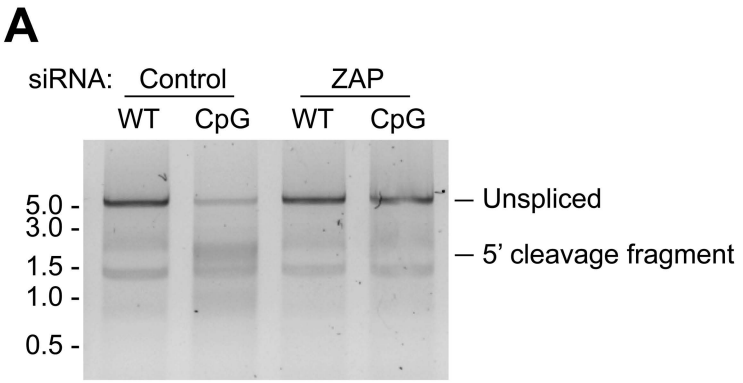
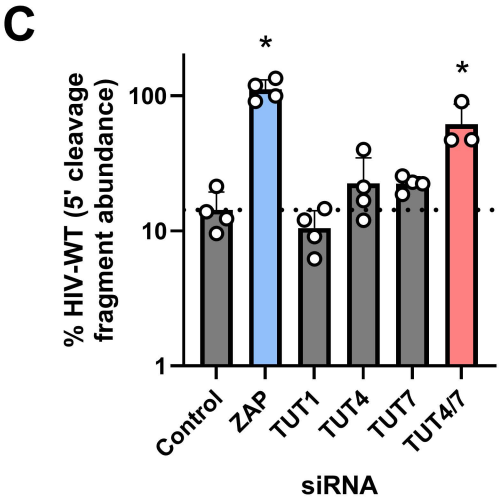

A
KHNYN iCLIP - HIV-CpG RNA
GFP
KHNYN-GFP
Normalised crosslink signal
CpG
HIV-1 coordinate
B
HIV-WT
p2 5'R
gag
pol
env
+36 CpGs =
ZAP responsive element (ZRE)
HIV-CpG
Unspliced
HIV-CpG
Singly spliced
HIV-CpG
Fully spliced
C
% HIV-WT (3' cleavage fragment abundance)
Control
ZAP
#1
#4
XRN1
siRNA
D
siRNA:
Control
ZAP
XRN1
WT
CpG
Singly spliced
3' cleavage fragment
Fully spliced
E
5' RACE-seq
siRNA ZAP (HIV-CpG/HIV-WT)
siRNA XRN1 (HIV-CpG/HIV-WT)
Fold change in 5'-end coverage
ZAP iCLIP - HeLa CRISPR control
HIV-WT
HIV-CpG
KHNYN iCLIP - HIV-CpG RNA
GFP
KHNYN-GFP
F
ZAP PULL-DOWN
INPUT
Mock
HIV-WT
HIV-CpG
IgG control:
Anti-ZAP:
RNase I:
XRN1
PABPC1
ZAP-L
ZAP-S
gp160
gp120
GAPDH
G
TRIM25 PULL-DOWN
INPUT
Anti-TRIM25:
TRIM25

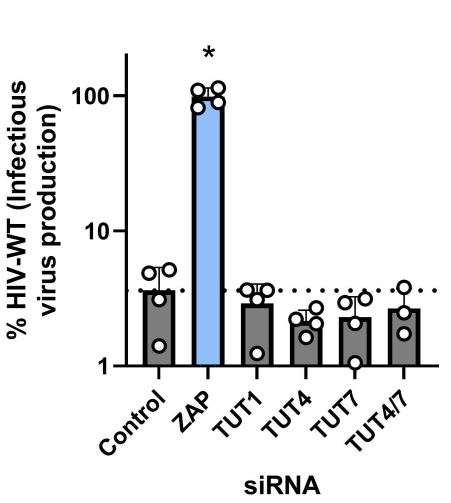
Figure 3

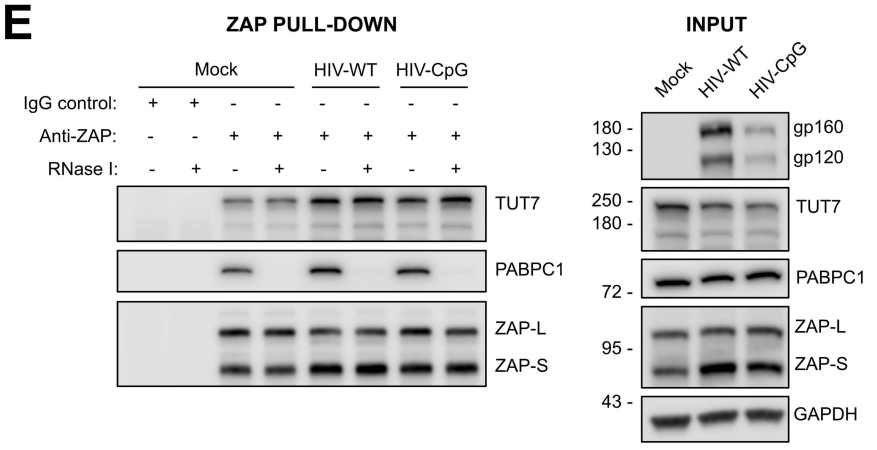


B

3' end position	3' non-templated nucleotides
5686	C
5746	none
5948	none
6196	UUUUU
6284	UUUU
6299	UUUU
6322	U







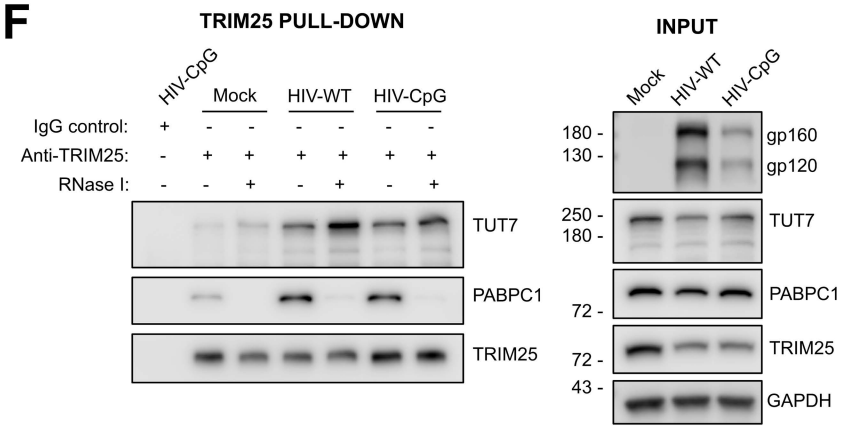


Figure 4

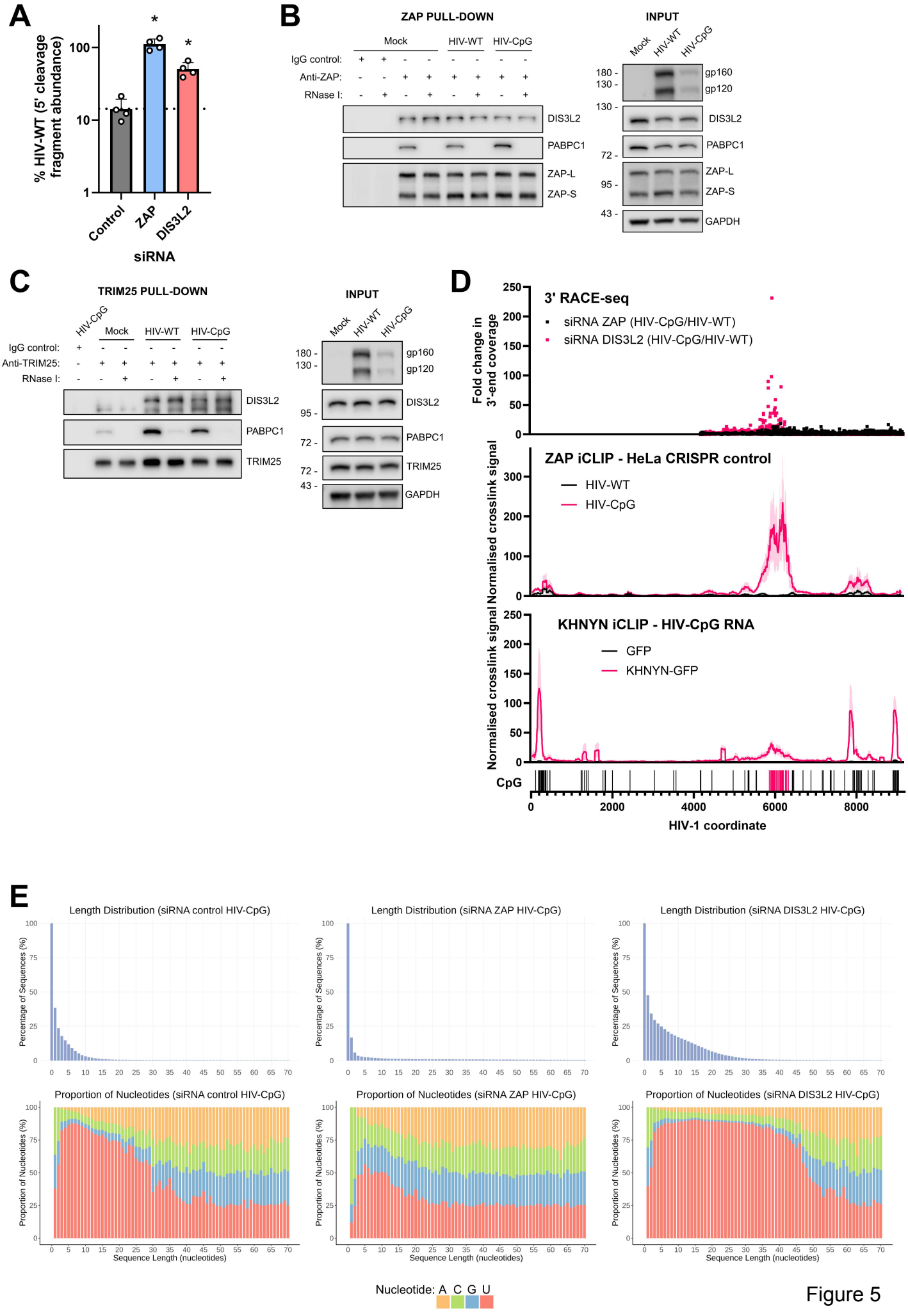

A
% HIV-WT (5' cleavage fragment abundance)
100
10
1
*
*
Control
ZAP
DIS3L2
siRNA
B
ZAP PULL-DOWN
Mock
HIV-WT
HIV-CpG
IgG control:
Anti-ZAP:
RNase I:
DIS3L2
PABPC1
ZAP-L
ZAP-S
INPUT
Mock
HIV-WT
HIV-CpG
180 -
130 -
gp160
gp120
130 -
DIS3L2
PABPC1
72 -
ZAP-L
95 -
ZAP-S
43 -
GAPDH
C
TRIM25 PULL-DOWN
HIV-CpG
Mock
HIV-WT
HIV-CpG
IgG control:
Anti-TRIM25:
RNase I:
DIS3L2
PABPC1
TRIM25
INPUT
Mock
HIV-WT
HIV-CpG
180 -
130 -
gp160
gp120
95 -
DIS3L2
72 -
PABPC1
72 -
TRIM25
43 -
GAPDH
D
3' RACE-seq
siRNA ZAP (HIV-CpG/HIV-WT)
siRNA DIS3L2 (HIV-CpG/HIV-WT)
Fold change in 3'-end coverage
ZAP iCLIP - HeLa CRISPR control
HIV-WT
HIV-CpG
Normalised crosslink signal
KHNYN iCLIP - HIV-CpG RNA
GFP
KHNYN-GFP
Normalised crosslink signal
CpG
0
2000
4000
6000
8000
HIV-1 coordinate
E
Length Distribution (siRNA control HIV-CpG)
Length Distribution (siRNA ZAP HIV-CpG)
Length Distribution (siRNA DIS3L2 HIV-CpG)
Percentage of Sequences (%)
Proportion of Nucleotides (siRNA control HIV-CpG)
Proportion of Nucleotides (siRNA ZAP HIV-CpG)
Proportion of Nucleotides (siRNA DIS3L2 HIV-CpG)
Proportion of Nucleotides (%)
Sequence Length (nucleotides)
Nucleotide: A C G U

Figure 5

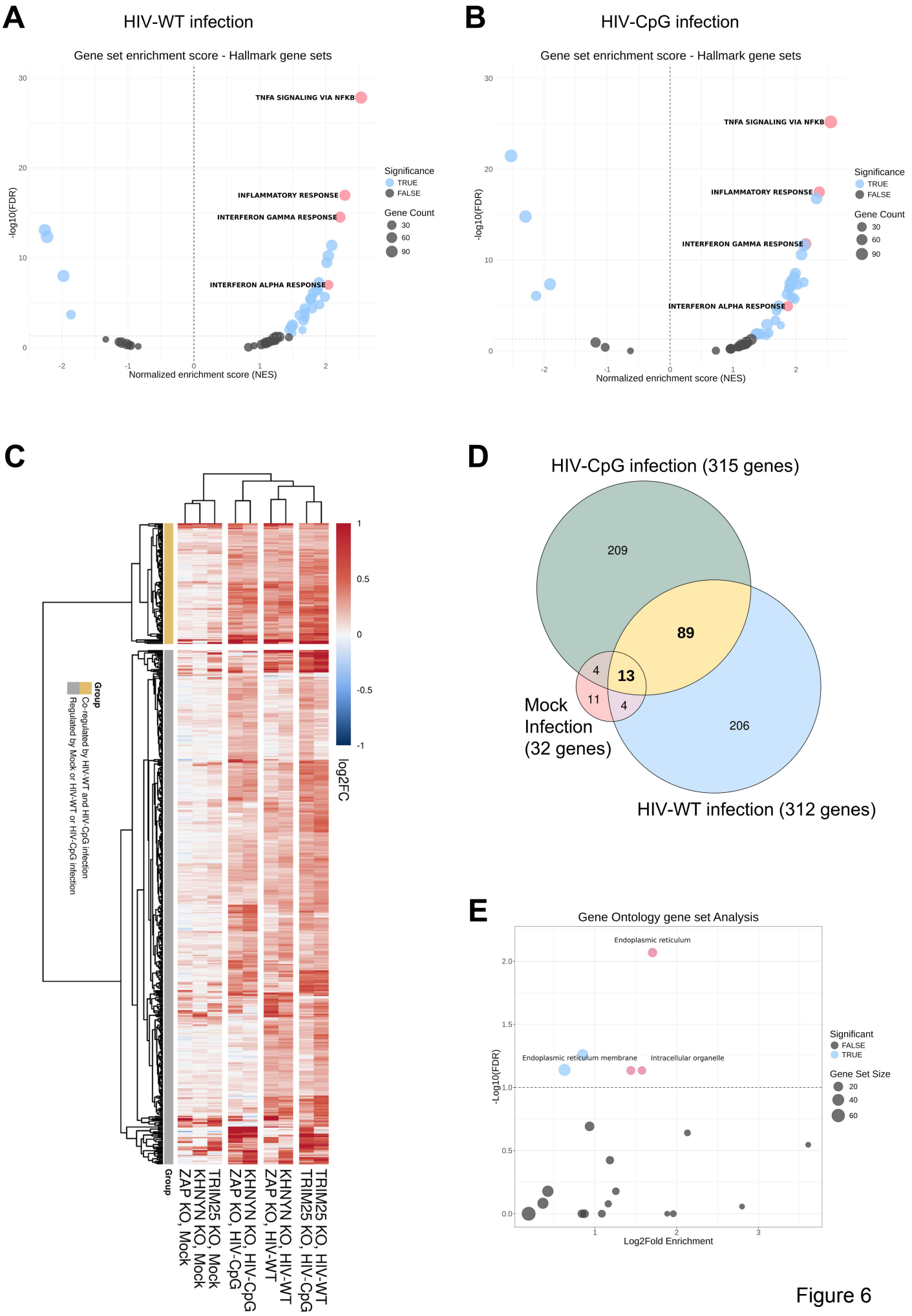

A
HIV-WT infection
Gene set enrichment score - Hallmark gene sets
TNFA SIGNALING VIA NFKB
INFLAMMATORY RESPONSE
INTERFERON GAMMA RESPONSE
INTERFERON ALPHA RESPONSE
-log10(FDR)
Normalized enrichment score (NES)
Significance
TRUE
FALSE
Gene Count
30
60
90
B
HIV-CpG infection
Gene set enrichment score - Hallmark gene sets
TNFA SIGNALING VIA NFKB
INFLAMMATORY RESPONSE
INTERFERON GAMMA RESPONSE
INTERFERON ALPHA RESPONSE
-log10(FDR)
Normalized enrichment score (NES)
Significance
TRUE
FALSE
Gene Count
30
60
90
C
Group
Co-regulated by HIV-WT and HIV-CpG infection
Regulated by Mock or HIV-WT or HIV-CpG infection
log2FC
ZAP KO, Mock
KHNYN KO, Mock
TRIM25 KO, Mock
ZAP KO, HIV-CpG
KHNYN KO, HIV-CpG
ZAP KO, HIV-WT
KHNYN KO, HIV-WT
TRIM25 KO, HIV-CpG
TRIM25 KO, HIV-WT
D
HIV-CpG infection (315 genes)
209
89
4
13
11
4
206
Mock Infection (32 genes)
HIV-WT infection (312 genes)
E
Gene Ontology gene set Analysis
Endoplasmic reticulum
Endoplasmic reticulum membrane
Intracellular organelle
-Log10(FDR)
Log2Fold Enrichment
Significant
FALSE
TRUE
Gene Set Size
20
40
60

Figure 6

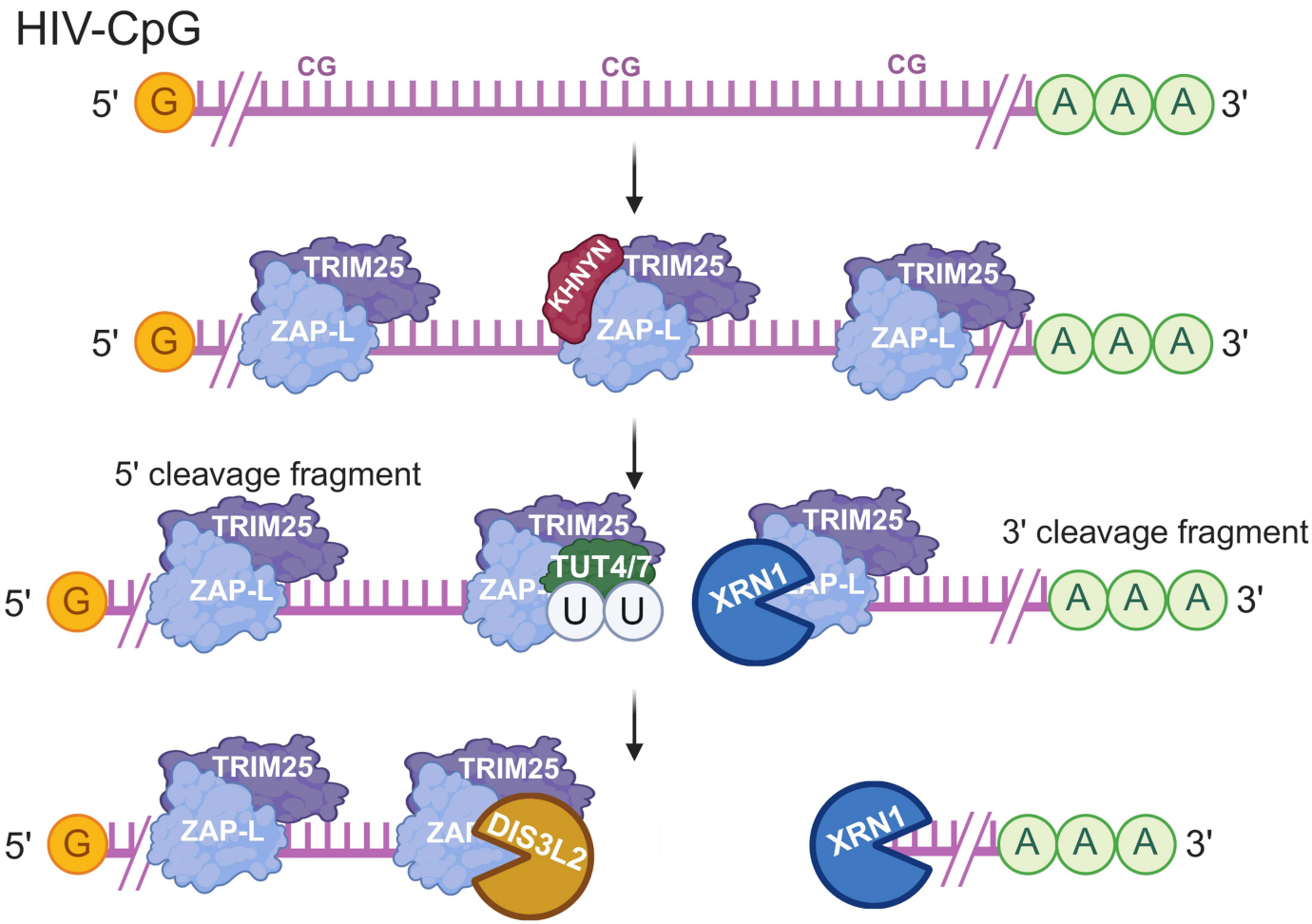

HIV-CpG
5'
G
CG
CG
CG
A A A
3'
TRIM25
ZAP-L
KHNYN
5' cleavage fragment
TUT4/7
U U
XRN1
3' cleavage fragment
DIS3L2

Figure 7