CpG enrichment as a live attenuated vaccine development strategy for influenza A virus

Colin P Sharp[1], Beth H Thompson[1], Tessa J Nash[1], Luke Thorley[1], Sara Clohisey Hendry[1], Yao-Tang Lin[1], Samantha Sives[1], Helen Wise[2], Ola Diebold[1], Lonneke Vervelde[1], Finn Grey[1], Peter Simmonds[3], Paul Digard[1], Eleanor R Gaunt[1*]

*Corresponding author

[1] The Roslin Institute, The University of Edinburgh, Easter Bush Campus, Midlothian, EH25 9RG

[2] Royal Infirmary of Edinburgh, NHS Lothian, Edinburgh, EH16 4SA

[3] Nuffield Department of Medicine, University of Oxford, South Parks Road, Oxford, OX1 3SY

ABSTRACT

Synonymous recoding of RNA virus genomes is a promising approach for generating attenuated viruses to use as vaccines. However, such recoding typically hinders virus growth. A potential solution to this is CpG dinucleotide enrichment. CpGs are recognised by cellular zinc-finger antiviral protein (ZAP), leading to viral RNA degradation and/or type I interferon stimulation and consequently reduced viral yields. Therefore, in principle, removing the ZAP CpG sensor from a virus propagation system will reverse attenuation of a CpG-enriched virus. Using influenza A viruses (IAVs) engineered for increased CpG content in different genome regions, we show that ZAP mediated virus attenuation is dependent on the viral segment to which CpGs are added. In ZAP-expressing human cells, CpG introduction invariably impaired transcript and protein production, but ZAP-sensitivity did not convey type I interferon activation. Importantly for vaccine development, CpG-enriched viruses were genetically stable during serial passage. Furthermore, the ZAP-sensitive virus was fully replication competent in MDCK cells and in embryonated hens' eggs, both of which are used to propagate live attenuated influenza vaccines. Thus, ZAP-sensitive CpG enriched viruses that are defective in human systems can yield high titre in vaccine propagation systems, providing a realistic, economically viable platform for live attenuated vaccine development.

INTRODUCTION

Single stranded RNA (ssRNA) viruses typically under-represent CpG dinucleotides in their genomes (1-3). Over the past decade, many groups have investigated the impact of synonymously increasing the CpG content in the genomes of RNA viruses. Without exception, these studies have determined that genomic CpG

30 enrichment causes viral replication deficiencies (reviewed in (4)). To explain this, it was hypothesised that a
31 cellular antiviral sensor may detect CpG motifs in viral RNAs (5), and in 2017 this idea was confirmed through
32 identification of the cellular protein ZAP (zinc-finger antiviral protein) as a selective binder of CpG motifs in viral
33 RNA transcripts (6). ZAP bound CpG motifs in transcripts made by defective HIV-1 viruses enriched for CpG
34 and knockdown of ZAP restored virus fitness (6).

35 Two major isoforms of ZAP – ZAPS (short) and ZAPL (long) are produced by alternative splicing (7), with a
36 common N-terminal region that contains four zinc-finger motifs with RNA-binding function (8-10). While CpG-
37 binding activity was originally found to be specific to the short isoform (6), CpG-specific antiviral activity has
38 since also been demonstrated for ZAPL (11). Therefore, a consensus mechanism for CpG sensing by ZAP
39 remains unresolved. ZAPS is an interferon-stimulated gene (ISG) (10,12,13), whereas ZAPL is constitutively
40 expressed (7). For antiviral function, a critical interaction between ZAPS and its binding partner, the RNA-
41 binding E3 ubiquitin ligase-like tripartite motif containing protein 25 (TRIM25) must take place (14,15). The
42 current understanding is that ZAPS binds CpG motifs on viral RNA, and upon subsequent binding by TRIM25
43 and catalytic activation, a downstream cascade leads to the induction of type I interferon (IFN) (12). ZAPS may
44 also directly bind RNA leading to exonuclease-mediated degradation (16), and so the antiviral activity of ZAPS
45 appears multi-faceted (reviewed in (4)). A newly discovered ZAPS binding partner, KHNYN, has so far only
46 been found to be required for ZAP-mediated antiviral activity during lentivirus infections (17,18). While ZAPS
47 has so far been more extensively studied, interactions of ZAPL with TRIM25 and KHNYN have been shown
48 to be more efficient than those of ZAPS (11). CpG-specific binding by ZAP is likely an evolutionarily conserved
49 trait across tetrapod species that is not retained in avian species (19); nevertheless, chickens express ZAPL
50 but not the shorter isoform (20).

51 Due to the replication defects so reliably imparted by CpG introduction into viral genomes, CpG enrichment
52 has widely been proposed as a potential strategy for the development of live attenuated vaccines. However, if
53 viral growth is impaired, then the large-scale virus propagation required for vaccine manufacturing (21) is not
54 viable. The identification of ZAPS as a CpG sensor (6) delivered the possibility that CpG-enriched viruses
55 could be grown in ZAPS knockout (KO) systems, thus restoring viral titre in an artificial system, with the
56 replication defect maintained in a vaccine recipient (4).

57 Live attenuated vaccines for IAV are used in children in the USA and Europe (22,23), with up to 14 million
58 doses delivered annually (24). Influenza A virus (IAV) represents an ideal virus to target using vaccines
59 engineered to incorporate additional CpGs, as this approach has the potential to overcome some of the

60 significant challenges arising during live attenuated IAV vaccine manufacture. IAV live vaccines are
61 propagated in either embryonated hens' eggs or MDCK cells. For egg based vaccines, unpredictable supply
62 and ethical challenges mean that there is a consensus about the need to move towards cell culture generated
63 vaccines such as those propagated in MDCK cells (25,26). However, MDCK based vaccines are more
64 expensive and large scale manufacture will require set up of new production pipelines (27,28); therefore, this
65 transition is not straightforward and egg-based vaccines will be propagated for the foreseeable future.

66 The IAV genome is (relatively) straightforward to recode (i.e., synonymously alter the genetic sequence of)
67 due to the facile reverse genetics system (29,30), in which the 8 segments of negative sense RNA genome
68 are encoded on plasmids that upon co-transfection into permissive cells yield infectious virus. The IAV genome
69 itself is thoroughly characterised, with well-defined packaging signals (31-33) and gene annotation (including
70 accessory proteins, reviewed in (34)), providing comparative clarity about which genome regions may be
71 suitable for recoding and those that must remain unaltered. Furthermore, prior to the discovery of ZAPS as a
72 CpG sensor (6), we found that synonymous addition of CpGs to IAV genome segment 5 (encoding
73 nucleoprotein (NP)) significantly attenuated the virus but that antibody and T cell responses to infection were
74 equivalent in mice, thereby demonstrating that attenuation mediated by CpG enrichment need not impair
75 development of adaptive immunity (35). Additionally, both IAV vaccine propagation systems - eggs or MDCK
76 cells - can be subjected to genome editing (36,37) to improve virus yield.

77 To further develop the strategy of CpG enrichment as a route to a live attenuated vaccine for IAV, we tested
78 the effects of adding CpG to another viral segment, and assessed the impact of this on various aspects of
79 virus replication. Surprisingly, while both segment 1 (encoding the polymerase protein PB2) and segment 5
80 CpG-high viruses were attenuated in human cell systems, only the segment 1 CpG-high virus could have
81 wildtype fitness restored by ZAP KO. The ZAP-sensitive virus was genetically stable in ZAP-positive cells, and
82 unexpectedly showed no growth defect in the embryonated hens' egg or MDCK cell systems used for vaccine
83 virus propagation. Therefore, we have provided a pathway for the development of CpG-enriched live
84 attenuated IAV viruses that can be propagated to high yield using standard approaches.

85

86 **RESULTS**

87 **Adding CpG dinucleotides to the IAV genome can cause a ZAPS-dependent or independent defect in**
88 **replication.**

89 We have previously shown that synonymous addition of CpGs to segment 5 of the IAV genome inhibited virus
90 replication (35). This work was undertaken prior to the discovery that ZAP is a cellular restriction factor that
91 recognises CpG-enriched transcripts (6). We therefore investigated whether adding CpGs to a different IAV
92 segment led to a replication deficiency and if CpG-enriched IAVs in general could be restored to wildtype (WT)
93 virus fitness in ZAP -/- cells. To do this, CpGs were synonymously added into segment 1 of the PR8 strain of
94 IAV genome (+126 CpGs) to provide a comparator to our existing segment 5 CpG-high (+86 CpGs) virus (35).
95 As before, compensatory mutations to maintain individual base frequencies across the recoded regions of
96 segment 1 were also incorporated, while a control ('CDLR') virus that was synonymously recoded in the same
97 region by rearrangement of codons was also engineered, so that any attenuation imparted by disruption of e.g.
98 RNA structures or alternative open reading frames would be detected. The panel of viral mutants is
99 summarised (**Table 1**).

100 When CpG motifs were added into segment 1 of the virus (encoding the polymerase component PB2), the
101 generated mutant ('seg1-CpGH') displayed a significant ~1.5-$\log_{10}$ defect in production of infectious virus
102 progeny in A549 cells, whereas no defect was apparent in the control-recoded seg1-CDLR virus (**Fig. 1A**). In
103 ZAP -/- cells that were confirmed negative for ZAP expression by western blotting (**Fig. S1A**) (38), the
104 replication defect of seg1-CpGH was lost (**Fig. 1A**), suggesting that the additional CpGs rendered this virus
105 susceptible to ZAP activity in WT cells. When the same infection experiments were performed using the seg5-
106 CpGH virus (35), a similar replication defect of ~1.5 $\log_{10}$ was observed (**Fig. 1B**) (as reported previously (35)).
107 However, the defect imparted by CpG enrichment was not altered in ZAP -/- cells, indicating that the seg5-
108 CpGH virus had a replication defect not mediated by ZAP. Notably, the seg5-CpGH virus generated
109 considerably smaller plaques in MDCK cells, whereas the plaque size reduction phenotype was minimal for
110 seg1-CpGH (**Fig. S2**).

111 ZAP binding partner TRIM25 is required for ZAP antiviral activity (14,15) and so a further comparison of the
112 replication of WT and seg1- and seg5-CpGH mutants was undertaken in TRIM25-/- cells (validated in **Fig.**
113 **S1B**) (39). In WT HEK293 cells, both the seg1-CpGH and seg5-CpGH viruses were again significantly
114 impaired in their replication (**Fig. 1C, D**). As with ZAP KO, TRIM25 KO also resulted in rescued replication of
115 seg1-CpGH (**Fig. 1C**) but not seg5-CpGH viruses (**Fig. 1D**).

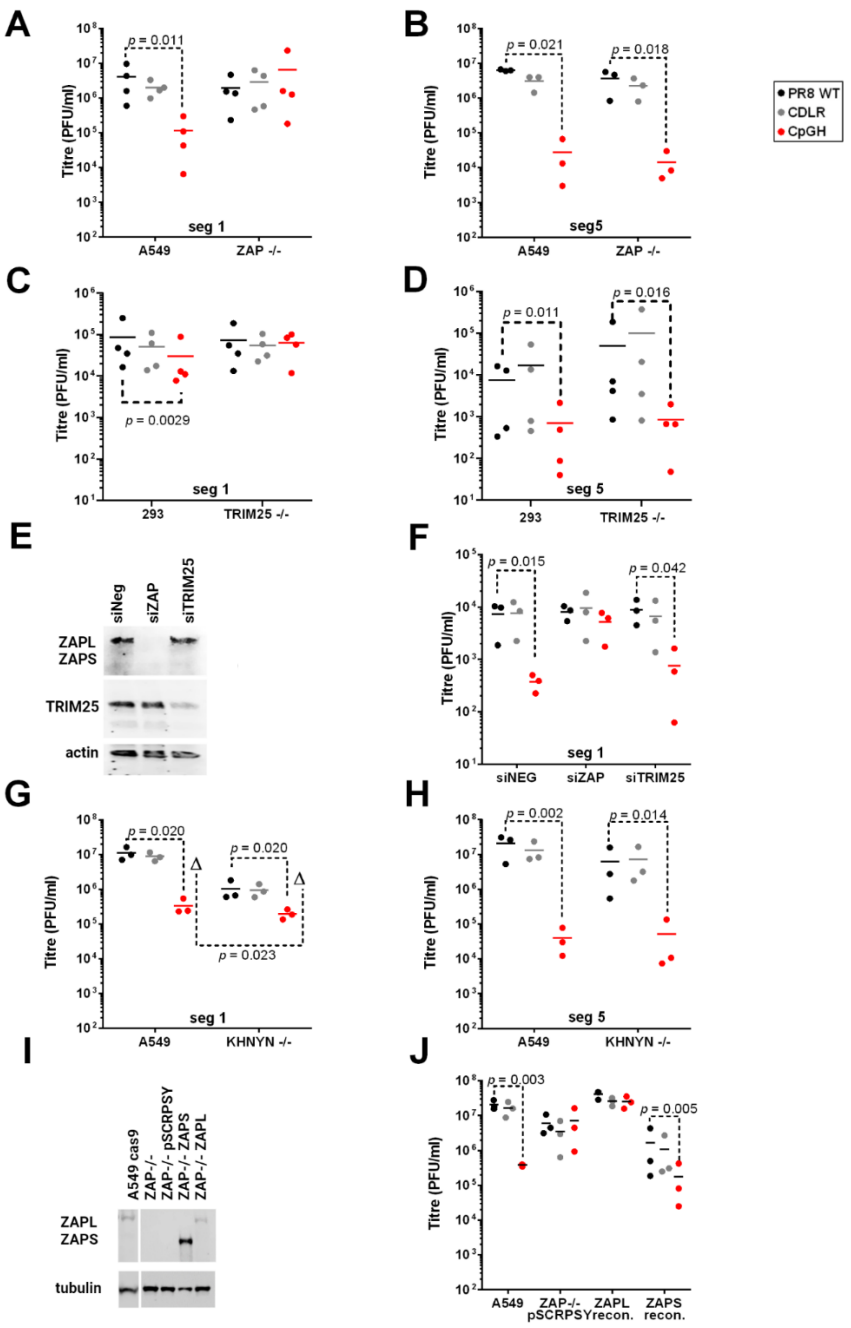


Figure 1. ***Influenza A viruses with CpG enrichment in segment 1 or segment 5 have replication defects mediated by ZAP-dependent and ZAP-independent mechanisms.*** WT PR8, synonymously permuted control virus (CDLR) and CpG enriched (CpGH) viruses were used to infect various cell types at low MOI for 48 hours. Synonymous recoding was performed either in segment 1 (seg1 virus panel) or segment 5 (seg5 virus panel). **A.** Seg1 virus panel was grown in A549 and paired ZAP -/- cells. **B.** Seg5 virus panel was grown in A549 and paired ZAP -/- cells. **C.** Seg1 virus panel was grown in 293 and paired TRIM25 -/- cells. **D.** Seg 5 virus panel was grown in 293 and paired TRIM25 -/- cells. **E.** siRNA knockdown of ZAP or TRIM25 in NHDF primary cells was confirmed by western blot. **F.** Impact of ZAP or TRIM25 knockdown on replication of the seg1 virus panel in NHDFs. **G.** Seg1 virus panel was grown in A549 and paired KHNYN -/- cells. **H.** Seg5 virus panel was grown in A549 and paired KHNYN -/- cells. **I.** ZAPL/ ZAPS expression was reconstituted in A549 cells using lentivirus expression constructs, and isoform-specific expression was confirmed by western blotting. **J.** A549 ZAP-/- cells were reconstituted for isoform-specific ZAPL or ZAPS expression, and titres of seg1 virus panel were determined. For validation of KO phenotypes, please refer to **Fig. S1**. Plaque phenotypes of the seg1 and seg5 virus panels are presented in **Fig. S2**. Confocal visualisation of ZAP reconstitution is presented in **Fig. S3**.

129 All experiments so far had been performed in immortalised cells and so whether the observed phenotypes
130 could be recapitulated in primary cells was investigated. siRNA knockdown of ZAP and TRIM25 was applied
131 in NHDF cells (primary human foreskin fibroblasts) (as previously described (40)) and successful knockdown
132 was confirmed by western blotting (**Fig. 1E**). While ZAP knockdown recapitulated rescue of seg1-CpGH
133 fitness, TRIM25 knockdown did not (**Fig. 1F**). As TRIM25 knockdown was incomplete (Fig. 1E), this may
134 indicate that sub-physiological levels of TRIM25 are needed for its role during CpG recognition. Replication of
135 the seg5-CpGH virus was too defective in NHDF cells to perform the same set of experiments for the seg5
136 virus panel (data not shown).

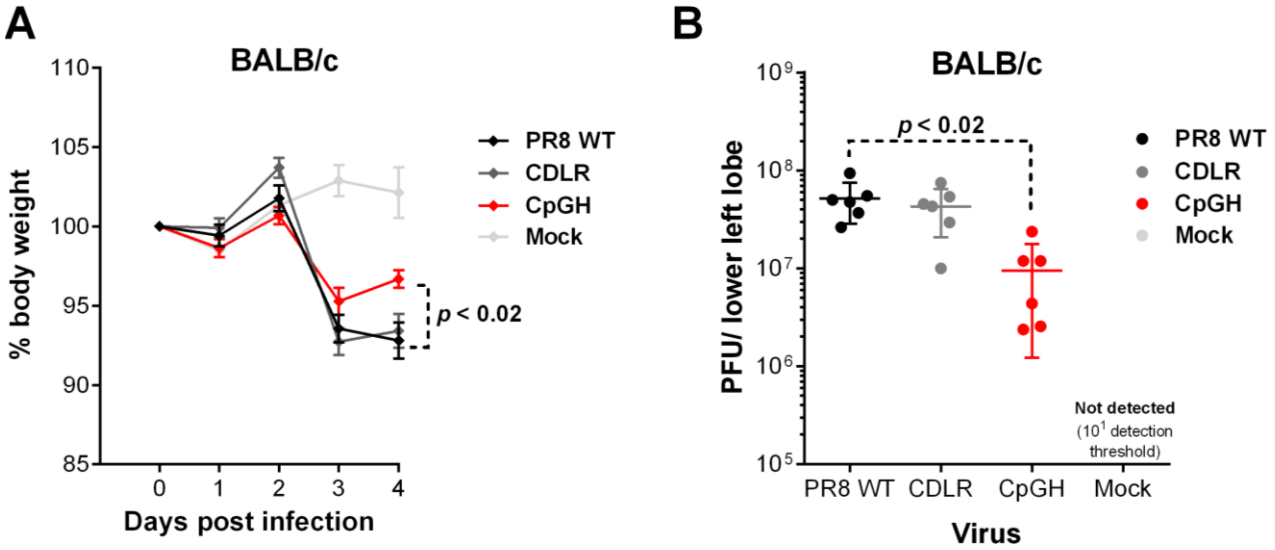
137 Next, we tested the impact of KO of KHNYN, another ZAP-binding partner (17), on replication of the seg1 and
138 seg5 virus panels; KHNYN KO in A549 cells was confirmed by western blot (**Fig. S1C**). In WT A549 cells, a
139 significant defect in the replication of seg1-CpGH and seg5-CpGH was seen again (**Fig. 1G**). In KHNYN-/-
140 cells, seg1-CpGH virus titres remained significantly lower than those of WT PR8. However, the fold difference
141 in titre between WT PR8 and seg1-CpGH was significantly less in the absence of KHNYN, indicating a partial
142 rescue of seg1-CpGH replication in the absence of KHNYN. Consistent with other gene KOs, that of KHNYN
143 did not affect the replication defect of the seg5-CpGH virus (p = ns for change in fold difference) (**Fig. 1H**).

144 To determine whether ZAP sensitivity of the seg1-CpGH virus was mediated by a specific ZAP isoform,
145 expression of ZAPL or ZAPS was reconstituted in A549 ZAP -/- cells using lentivirus vectors expressing
146 individual ZAP isoforms. Isoform-specific expression was confirmed by western blotting (**Fig. 1I**). Even levels
147 of cellular expression for ZAPS were confirmed by confocal microscopy, while ZAPL expression was more
148 variable (**Fig. S3**). This panel of cells was then infected with the seg1 virus panel, and the replication defect
149 reported above (Fig. 1A) was recapitulated (**Fig. 1J**). In ZAP-/- cells expressing an empty pSCRPSY lentivirus
150 vector there was no difference in replication between WT, seg1-CDLR and seg1-CpGH viruses in the absence
151 of ZAP, indicating that the lentivirus vector did not impair seg1-CpGH virus replication. When ZAPL expression
152 was reconstituted, a modest increase in the replication of all viruses in the panel was observed, and the titres
153 of seg1-CpGH remained similar to those of WT PR8. However, when ZAPS was overexpressed, all viruses in
154 the panel replicated less well than in WT cells, but the titres of seg1-CpGH relative to WT PR8 virus were
155 significantly lower, indicating that ZAPS is the CpG sensor for the seg1-CpGH virus.

156 CpG introduction attenuates IAV replication, but not necessarily though ZAP. When CpG introduction confers
157 ZAP sensitivity, repression requires TRIM25 and ZAPS in particular but is less dependent on KHNYN.

158 **CpG enrichment in segment 1 causes viral attenuation in mice.**

159 Previously, we showed that adding CpGs into segment 5 of PR8 caused a fitness loss in mice (35). However,
160 since this attenuation is independent of ZAP (Fig. 1B), it was possible that a ZAP-sensitive CpG-enriched virus
161 might show a different phenotype in mice. Therefore, mice were infected with the seg1 virus panel under the
162 same conditions as previously reported (200 PFU virus were inoculated into 6-week old female BALB/c mice).
163 All virus-infected animals lost weight after day 2 post infection but mice infected with seg1-CpGH lost
164 significantly less weight than mice infected with WT PR8 (**Fig. 2A**). In addition, seg1-CpGH viral loads in the
165 lung were significantly lower (**Fig. 2B**). Therefore, ZAP-dependent CpG-driven attenuation *in vitro* also leads
166 to attenuation in an *in vivo* system.



167

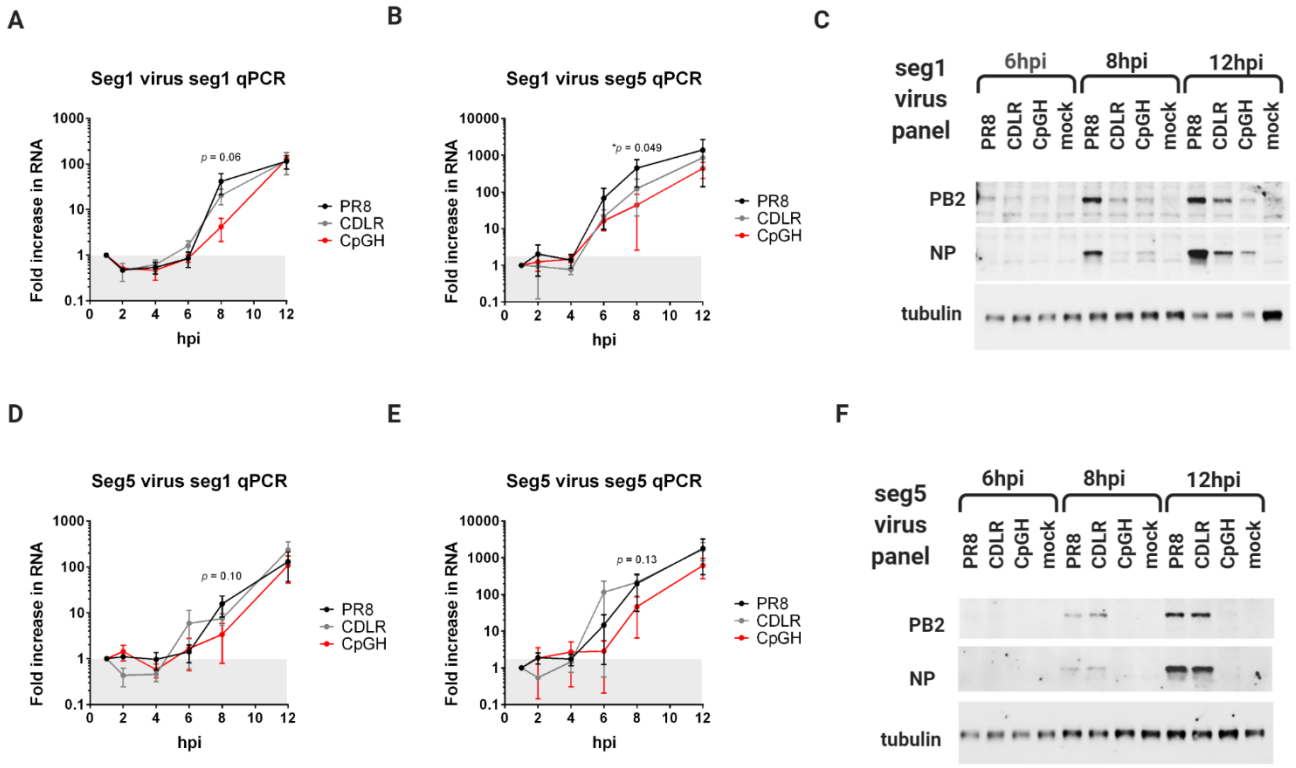
168 **Figure 2. Seg1-CpGH virus is attenuated in mice.** Groups of six, 6 week-old female BALB/c mice were infected with 200PFU of either
169 WT PR8 virus, seg1-CDLR permuted control, seg1-CpGH virus or mock infected (n = 2) for 5 days; weight change measured daily (**A**).
170 On day 5, lower left lung lobes were harvested and the presence of infectious virus was quantified (**B**).

171 **CpG enrichment may affect transcriptional or translational efficiency for IAV.**

172 Defective viral replication was imparted by addition of CpGs into either segment 1 or segment 5 of IAV, but
173 only the segment 1 mutant had fitness restored by ZAP depletion. We hypothesised that adding CpGs may
174 impact different stages of the virus replication cycle in different ways, depending on where in the viral genome
175 the CpGs were added. The key stages of virus replication we considered most likely to be affected by genomic
176 CpG enrichment were transcription and translation, post-transcriptional modifications, type I IFN induction and
177 virus assembly. We first examined transcription and translation.

178 We first assessed whether viral transcript and/or protein production were affected by CpG enrichment during
179 virus replication. Time-course experiments were performed at high MOI with samples collected at 1, 2, 4, 6, 8
180 and 12hpi for quantification of total viral RNAs and proteins. Lower gene and protein expression levels (for

181 both PB2 and NP) were seen for the seg1-CpGH virus compared with WT PR8 and paired CDLR control virus
182 (**Fig. 3 A-C**). For the seg5-CpGH virus (**Fig. 3 D-F**), CpGH transcript levels were also lower than WT or CDLR
183 counterparts, but these data were borderline statistically significant. However, by western blotting, protein
184 levels were obviously affected.



185

186 **Figure 3. CpG enrichment in the IAV genome results in reduced transcriptional levels during virus replication.** Transcription and
187 translation were investigated in the context of infection using high MOI timecourses. Seg1 virus panel infections were analysed for segment
188 1 (**A**) or segment 5 (**B**) RNA production, and for protein production from the genes corresponding to these segments (PB2 and NP
189 respectively) (**C**). Seg5 virus panel infections were analysed for segment 1 (**D**) or segment 5 (**E**) RNA production, and for protein production
190 from the genes corresponding to these segments (PB2 and NP respectively) (**F**). Throughout, *n* = 3; for blots, representative data are
191 shown.

192 The reduced transcript levels for both CpGH mutants might be explained by inefficiencies of the viral
193 polymerase. Therefore, polymerase functionality was assessed using vRNP reconstitution assays using
194 luciferase under an IAV promoter, thereby coupling IAV transcription and translation in a cell-based assay in
195 the absence of infection (41). Four plasmids corresponding to the viral RNP proteins (PB2, PB1, PA and NP,
196 encoded on genome segments 1, 2, 3 and 5 respectively) that complex with vRNA to make vRNPs were
197 transfected into HEK293T cells. As PB2 is encoded on segment 1, and NP on segment 5, the impact of CpG
198 enrichment on segment 1 and segment 5 was directly compared by titration of the corresponding plasmid in
199 this assay. For the seg1-CpGH construct, luciferase activity was significantly lower than for WT PR8 and CDLR
200 controls at non-saturating plasmid concentrations (**Fig. 4A**). This reduced luciferase production was confirmed

201 by western blotting (**Fig. 4B**). Although only segment 1 was recoded, production of both PB2 and NP were
202 also reduced, most likely indicating that vRNPs were formed at a lower efficiency as a consequence of reduced
203 transcription/translation of the seg1-CpGH construct. For seg5-CpGH, luciferase signal was slightly but not
204 significantly lower than the WT PR8 and CDLR controls (**Fig. 4C**); correspondingly, luciferase protein levels
205 titrated similarly (**Fig. 4D**). NP protein levels were lower for the seg5-CpGH construct, but PB2 levels were not.
206 Thus, luciferase reporter signals indicated that seg1-CpGH vRNPs had reduced polymerase activity, whereas
207 the seg5-CpGH polymerase activity was equivalent to WT; however, there was an apparent defect in protein
208 production for this construct.

209 Possibly, the impaired polymerase and translational defects so far identified for seg1-CpGH and seg5-CpGH
210 resulted from factors independent of cellular processes (e.g., ZAP sensing or proteasome activity). To test this,
211 transcription and translational efficiencies of the seg1 and seg5 constructs were assayed in cell-free conditions
212 (i.e. free of cellular proteins such as ZAP and its cofactors). Equimolar amounts of WT PR8, seg1-CDLR, seg1-
213 CpGH, seg5-CDLR and seg5-CpGH amplicons were used as input for *in vitro* transcription assays, and the
214 amount of RNA produced was measured up to 90 minutes later. In the absence of cellular factors (and using
215 a bacteriophage RNA polymerase rather than the IAV one), no differences in transcript production were seen
216 across either the seg1 panel or the seg5 panels (**Fig. 4E, F**). Next, translational efficiencies were tested by
217 titrating equimolar amounts of RNA in *in vitro* translation assays. No notable differences in protein production
218 were apparent for either panel of viral mRNAs (**Fig. 4G, H**). Thus, synonymous recoding imparted no inherent
219 defect on transcription or translation in the absence of nucleic acid sensing pathways.

220 Overall, CpG enrichment caused no intrinsic defect to transcription or translation as evidenced in cell-free
221 systems. However, evidence for diminished transcript or protein levels were apparent when viral transcripts
222 were expressed in cell based systems. CpG enrichment caused reduced transcript (seg1-CpGH) and protein
223 (seg5-CpGH) levels, correlating with ZAP-dependent and independent attenuation respectively.

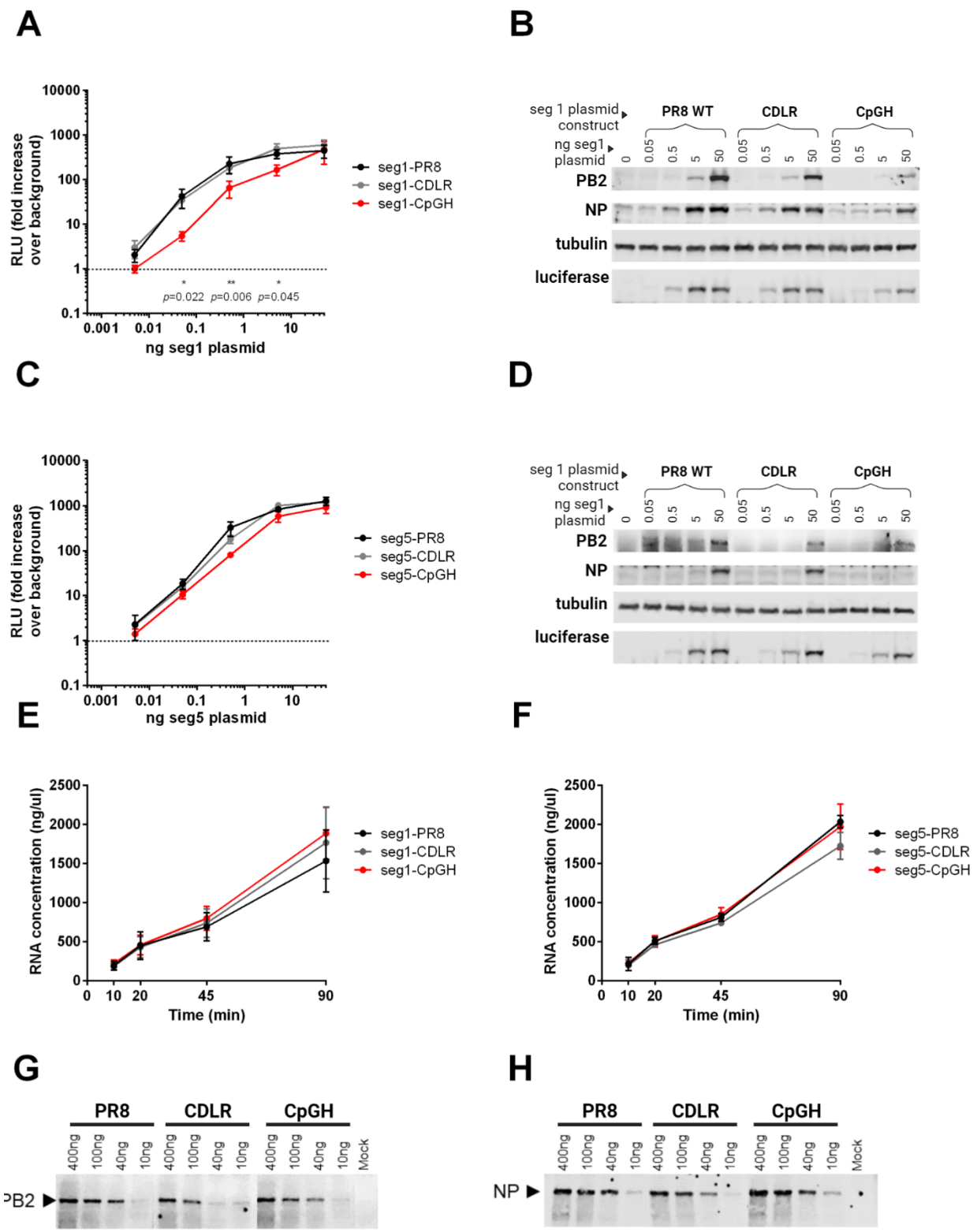


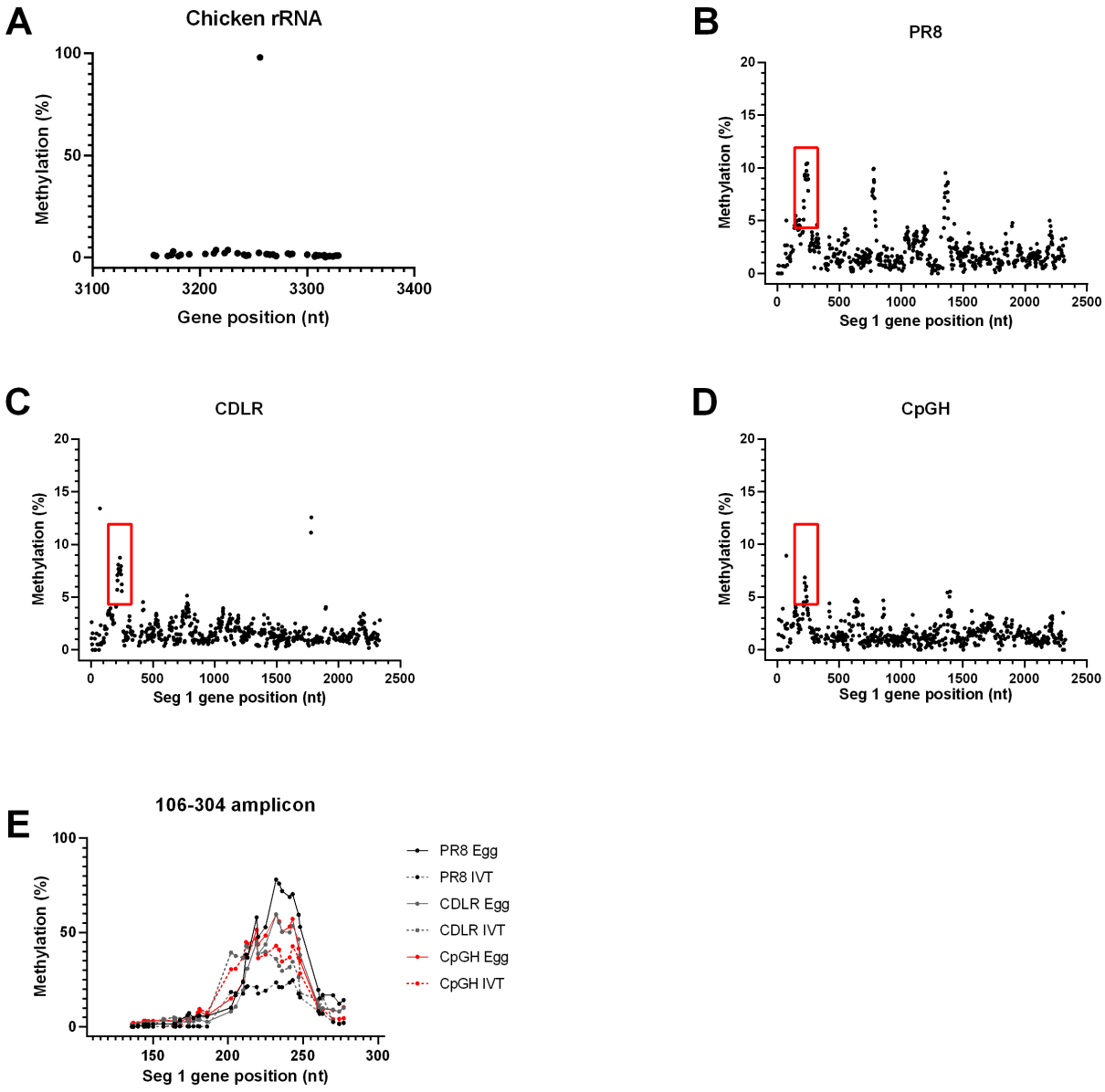
Figure 4. CpG modification may impact viral transcript production in cells but it does not impact the efficiency of transcription and translation in cell-free systems. Minireplicon assays that reconstitute the viral polymerase were performed using a luciferase reporter for the seg1 panel (**A**, luciferase reporter signal; **B**, protein signal by western blot) and the seg5 panel (**C**, luciferase reporter signal; **D**, protein signal by western blot). **E and F.** Equal amounts of segment 1 **(E)** or segment 5 **(F)** WT (PR8), permuted control (CDLR) or CpG-enriched (CpGH) amplicons under T7 promoters were used in *in vitro* transcription assays to assess whether synonymous recoding impacted transcriptional efficiency in a cell free system. The efficiency of translation of RNAs was then assessed by titrating the synthesised RNAs in *in vitro* translation assays for the segment 1 panel **(G)** and the segment 5 panel **(H)**.

232 **CpG enrichment does not impact methylation status of IAV vRNAs.**

233 Methylation of cytosines in CpG dinucleotides of DNA regulates transcription (42), and host factors imparting
234 cytosine methylation in RNA have been described (43,44). We therefore hypothesised that cytosines of CpG
235 motifs in RNA may be preferentially methylated. DNA-encoded cytosines in CpG conformations are prone to
236 spontaneously deaminate, creating a cytosine to thymine nucleotide substitution (45), and if the same were
237 true for CpGs in RNA, CpG enrichment of viral RNA genomes could result in hypermutation. The seg1-CpGH
238 virus was chosen for this analysis as it had the highest number of CpG motifs introduced. Bisulphite conversion
239 reactions, in which cytosines in purified RNA are converted to uracil unless protected from conversion by a
240 methyl group, were performed on viral genomic RNA purified from virions. Converted RNAs were then
241 sequenced and compared to non-converted sequences to identify potential methylation sites. RNA extracted
242 from virus stocks grown in eggs was used, and a strongly methylated site at position 3256 on chicken 28S
243 ribosomal RNA (GenBank accession # XR_003078040.1), analogous to position 4447 on human 28S
244 ribosomal RNA served as a positive control. Methylation frequencies of 97.0 - 98.8% were detected at this site
245 across all samples, indicating that the procedure worked efficiently (**Fig. 5A**).

246 Deep sequencing reads of bisulphite converted RNA mapped to the *in silico* converted reference sequences
247 for the seg1 virus panel yielded no individual sites with methylation signals of greater than 15% (**Fig. 5B-D**).
248 However, three obvious peaks in methylation signal in WT PR8 were identified, each occurring over regions
249 of ~50 bases, that were evident in up to approximately 10% of sequences (**Fig. 4B**). The first of these clusters
250 was identified in a region which was not mutated in seg1-CDLR or seg1-CpGH and the clustered signal was
251 maintained across all three runs, indicating a biologically reproducible phenomenon. The second two clusters
252 were detected in regions that were modified in seg1-CDLR and seg1-CpGH, and these were not retained in
253 the mutant viruses (**Fig. 5C&D**), indicating that mutation at these sites removed the methylation signals. Unlike
254 the single base methylation seen in the chicken rRNA, the sites of apparent viral RNA methylation were
255 broader, centred around ~50 nucleotide regions with diminishing signal in cytosines around the peak. We
256 considered the possibility that these clustered methylation signals were false positives (generated for example
257 by RNA secondary structures which did not allow RNA unfolding, thus protecting cytosines from base
258 conversion). To determine whether this was the case, *in vitro* transcribed RNAs corresponding to seg1-WT
259 PR8, seg1-CDLR and seg1-CpGH were generated to establish whether the methylation signal was retained
260 in RNAs synthesised in a cell-free system where post-transcriptional modifications cannot occur. These RNAs,
261 along with paired vRNAs, were then subject to bisulphite conversion, PCR amplification and sequencing across
262 the selected region. Methylation signals were retained in amplicons generated from the *in vitro* transcribed

263 RNAs (**Fig. 5E**), suggesting that the methylation clusters were artefactual. Notably, the conversion rate was
264 higher from the targeted sequencing, most likely because PCR amplification had biased representation of the
265 non-converted signal. Through whole genome sequencing of bisulphite converted RNA, we found no evidence
266 of high efficiency methylation at any site in the PR8 genome (**Fig. S4**), and CpG enrichment did not introduce
267 methylation sites.



268

269 **Figure 5. CpG enrichment of viral RNA does not alter the methylation profile.** Segment 1 modified viruses were processed through
270 bisulphite conversion reactions and then deep sequenced to determine the frequency of methylation in unmodified versus CpG-enriched
271 transcripts. **A.** As virus reconstitution was performed in embryonated hens' eggs, chicken ribosomal RNA was sequenced from viral stocks
272 as a positive control due to a known methylation site in this RNA. **B.** 5-methylcytosine signal intensity across PR8 segment 1 vRNA. **C.** 5-
273 methylcytosine signal intensity across CDLR segment 1 vRNA. **D.** 5-methylcytosine signal intensity across CDLR segment 1 vRNA. **E.** 5-
274 methylcytosine signal intensity across PCR amplicons of either vRNA (PR8 egg/ CDLR egg/ CpGH egg) or *in vitro* transcribed (IVT) RNA,
275 across segment 1 positions 106-304. This region was selected due to the spike in signal observed across segment 1 vRNAs for the
276 segment 1 virus panel (B-D, red boxes). For PR8 whole genome methylation plots please refer to **Fig. S4**.

277 **ZAP-mediated inhibition of CpG enriched viruses is independent of the type I IFN response.**

278 In the data presented in Fig. 3, a defect in transcript abundance for seg1-CpGH was identified, attributable to
279 ZAPS activity (Fig. 1). As well as recruiting exosome machinery for transcript degradation, ZAP is a known
280 ISG, and so the potential for CpG-enriched IAVs to induce IFN (46,47) was assayed. Supernatants from
281 infected A549 cells were used in a HEK-Blue assay, which gives a colorimetric signal proportional to the levels
282 of IFN-α/β present.

283 As expected in the presence of an intact NS1 (48), WT PR8 did not induce a detectable IFN response (**Fig.**
284 **6A**), whereas a control virus mutated in segment 8 to impair the ability of NS1 to control the IFN response
285 (R38K41A; (49)) strongly induced type I IFN signalling. The seg1-CDLR control and seg1-CpGH viruses also
286 did not induce substantial amounts of type I IFN. This suggests that the regulation of seg1-CpGH by ZAP is
287 independent of any IFN-stimulating activity by ZAPS (12).

288 In contrast, a significant induction of IFN was observed for the seg5-CpGH but not the seg5-CDLR virus,
289 providing a further insight into a cause of the inherent defectiveness of the seg5-CpGH virus. While the
290 defective seg5-CpGH virus was not rescued by ZAP KO (Fig. 1B), it is possible that this virus could have both
291 ZAP-independent and underlying ZAP-dependent defects that are masked by another layer of attenuation
292 introduced by recoding. To determine whether ZAP was required for triggering the seg5-CpGH-induced IFN
293 signalling, the same panel of viruses were grown in ZAP -/- cells, and the assay was repeated (**Fig. 6B**). The
294 NS1 mutant virus still induced high levels of type I IFN, but the difference between the levels of IFN induced
295 by WT and seg5-CpGH viruses was no longer significant (p = 0.145), although seg5-CpGH associated IFN
296 levels were still slightly higher than those of the WT or CDLR viruses. Therefore, the induction of IFN by seg5-
297 CpGH may be partially mediated by ZAP.

298 It was possible that IAV NS1 suppressed ZAPS activity sufficiently to block ZAP-mediated IFN induction by
299 seg1-CpGH. To test this, we introduced mutations into NS1 that have been described to specifically inhibit its
300 anti-ZAP or anti-TRIM25 activity ((50,51); **Table 2**). Introduction of these NS1 mutations into the seg1-CpGH
301 virus should alleviate NS1-dependent suppression of the ZAP/TRIM25 pathway, and if NS1 was inhibiting
302 ZAP's ability to sense CpGs, the defect in replication of the seg1-CpGH virus would be exacerbated by these
303 mutations. However, this was not the case; in low MOI infections in A549 cells (**Fig. 6C**), the seg1 viruses with
304 mutations in NS1 at positions 95/99 (reportedly blocking anti-TRIM25 activity (50)) all had similar titres to those
305 for viruses without any NS1 mutations. For viruses with mutations in NS1 at positions 96/97 (reportedly
306 impairing anti-ZAP activity), titres were down by 2-3 $\log_{10}$, indicating a general replication defect consistent

307 with other reports indicating these mutations have a deleterious effect on NS1 stability (50,52). For viruses
308 with mutations in NS1 at positions 107-110 (targeting anti-ZAP functionality), titres were similar to those for
309 viruses without mutations in NS1. The replication defect of seg1-CpGH virus relative to WT PR8 was consistent
310 regardless of the NS1 mutations introduced, indicating a lack of specificity of these NS1 sites for ZAP/ TRIM25
311 regulation. For seg1-CpGH with no modifications to NS1, there was an increase in ZAPS protein expression
312 compared with cells infected with WT virus, suggesting some induction of innate immune signalling (**Fig. 6D**).
313 Seg1-CpGH viral protein accumulation was also lower than for WT virus, reflecting the defective replication of
314 this virus. In cells infected with the NS1 mutant viruses, ZAPS expression was also higher and viral protein
315 expression (especially NP) was lower. However, ZAPS expression levels were consistent whether segment 1
316 was WT PR8, CDLR control or CpG-high, indicating that the NS1 mutations were generally failing to control
317 ISG expression rather than providing evidence for specific CpG-mediated ZAPS upregulation. Additionally,
318 TRIM25 levels were consistent throughout, supporting the conclusion that the mutations in NS1 do not
319 specifically target this pathway.

320 IFN induction by the seg1/seg8 mutant virus panel was then examined (**Fig. 6E**). For viruses with mutations
321 introduced into NS1 amino acid residues 95 and 99, no significant type I IFN activity was detected in any of
322 the seg1 virus panel. Conversely, when residues 96 and 97 were mutated, all seg1 viruses induced IFN to
323 similar levels. NS1 mutations at positions 107-110 delivered detectable IFN signal for the seg1-CpGH virus,
324 but also for the seg1-CDLR virus (though, curiously, not for the seg1-PR8). The same virus panel was also
325 grown in ZAP -/- cells, giving similar results (**Fig. 6F**). Therefore, while seg1-CpGH is ZAP sensitive (Fig. 1),
326 we could not link ZAP anti-CpG activity with the type I IFN pathway in this system.

327 **CpG enrichment causes little or no impairment to virus assembly.**

328 While mutations introduced into the PR8 genome were designed to avoid classically described packaging
329 signals present at the termini of viral genome segments (31,53), recent work has highlighted the potential
330 requirement for inter-segment RNA-RNA interactions occurring in the middle of segments during virion
331 assembly (33). Therefore, the possibility that the mutations made in segment 1 or segment 5 may have
332 disrupted viral genome packaging was investigated.

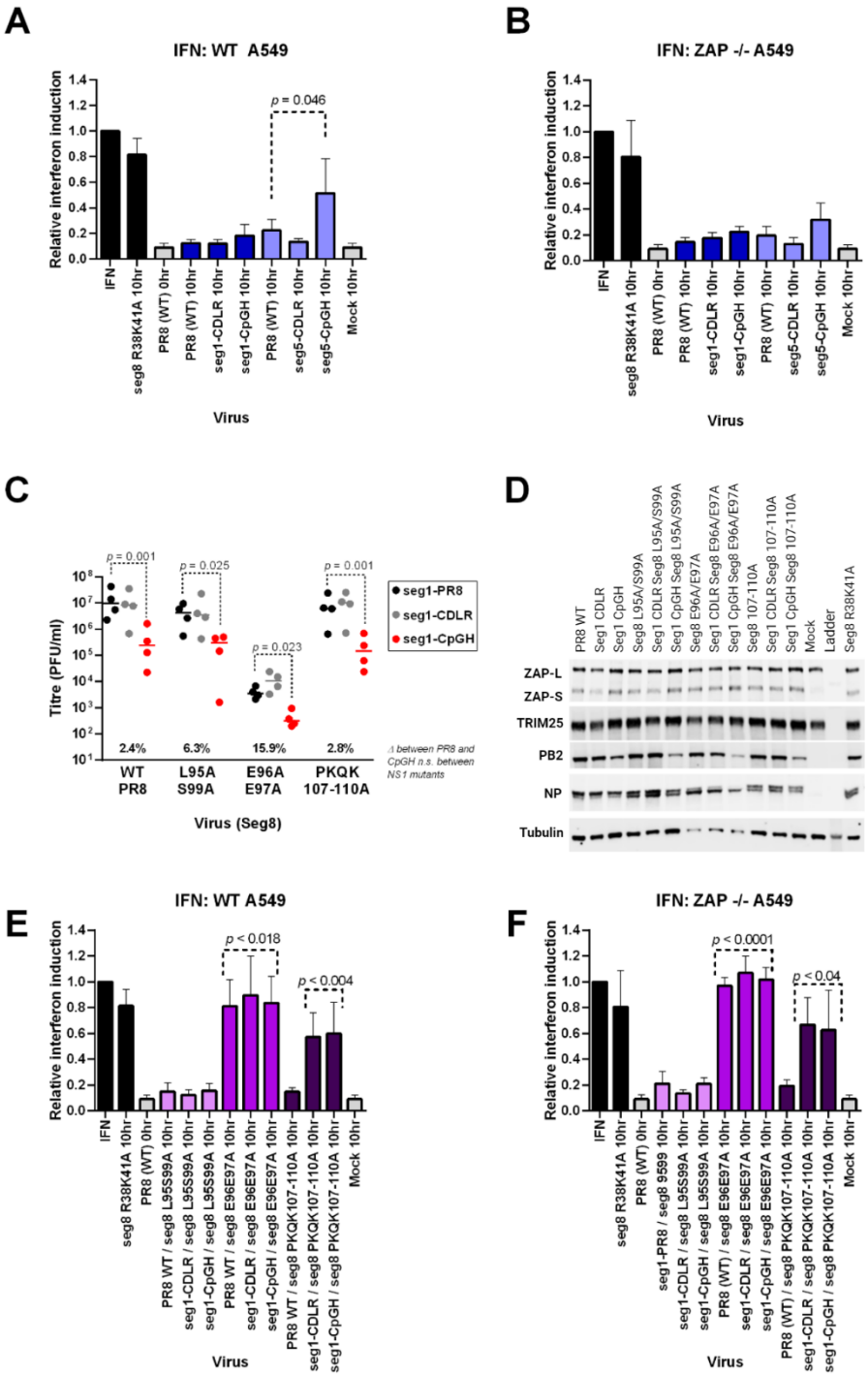


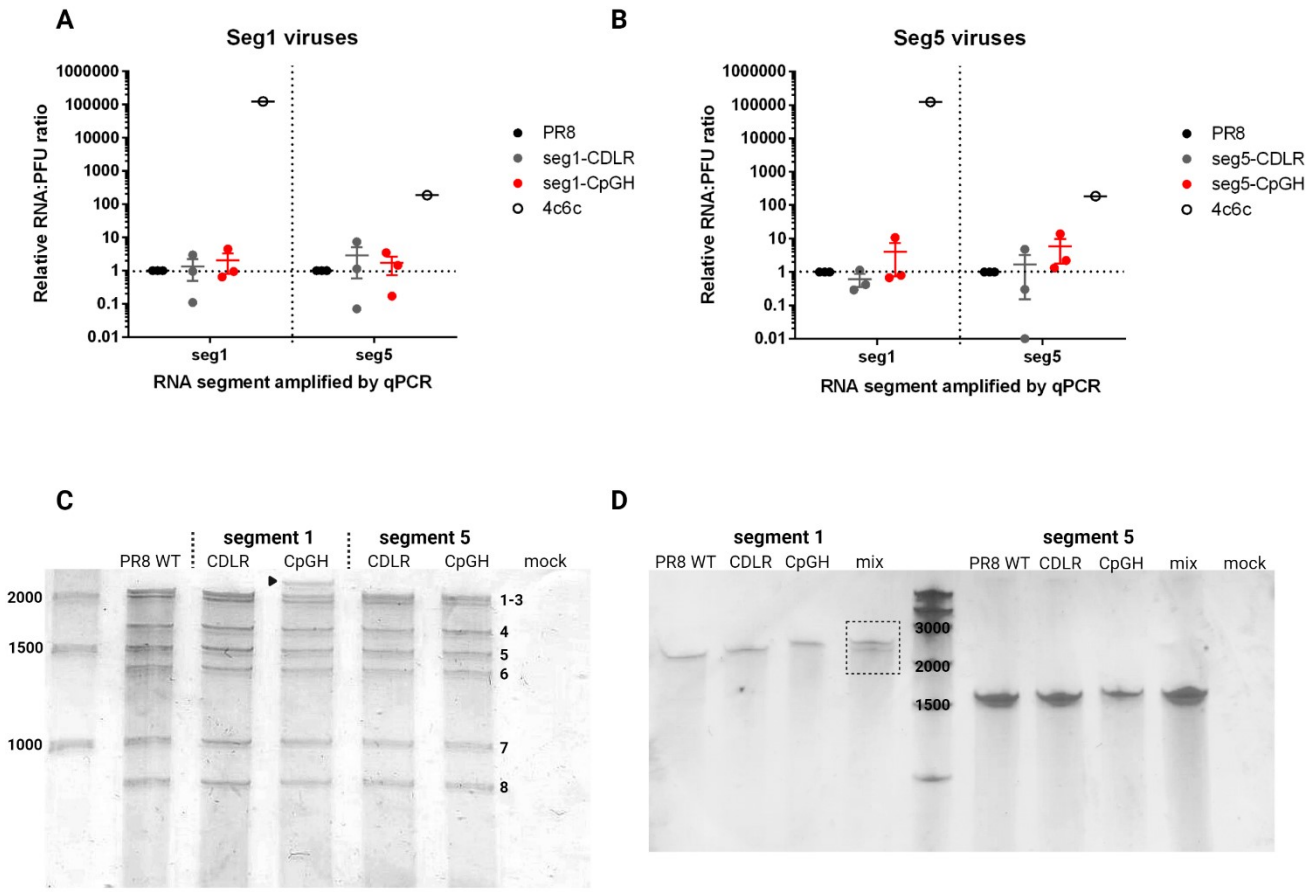
Figure 6. Interaction of CpG-enriched viruses with the IFN pathway. A. Seg1 and seg5 virus panels were used to infect A549 cells at high MOI for 24 hours, and then IFN induction was quantified by HEK-Blue assay after 10 hours. IFN was used as a standard, and an IAV mutated in segment 8 to inhibit IFN induction (R38K41A) was used as a positive control. **B.** As in A, except that infections were performed in ZAP -/- A549 cells. **C.** IAV WT PR8 virus, or segment 1 modified control CDLR or CpGH viruses were further mutated in segment 8 to abrogate the ZAP/TRIM25 targeting activity of NS1 encoded on this segment. Mutations introduced into segment 8 were L95S99A, reportedly defective in TRIM25 targeting (50), E96/97A, reportedly defective in ZAP targeting (51), and PKQK107-110A, reportedly defective in ZAP targeting (51). Seg1/8 mutant viruses were used to infect A549 cells at MOI 0.01, for 48h, and viral titres were quantified. **D.** Western blots examining ZAP and TRIM25 expression in the same infections. **E.** The seg1/8 mutant virus panel was used to infect A549 cells at high MOI for 10 hours, and then IFN induction was quantified by HEK-Blue assay (A640 read-out). **F.** As in D, except that infections were performed in ZAP -/- A549 cells.

344 Firstly, genome copy:PFU ratios for virus stocks were determined for the seg1 and seg5 virus panels. For both
345 panels, RNA copy numbers for both segment 1 and segment 5 were analysed. A previously characterised
346 packaging mutant, '4c6c' (31,54) was included as a control. Relative to WT PR8, no significant impact on
347 RNA:PFU ratios was observed for either the seg1-CDLR or seg1-CpGH viruses (**Fig. 7A**), or the seg5-CDLR
348 and seg5-CpGH viruses (**Fig. 7B**).

349 Secondly, electrophoretic separation of RNA extracted from purified virus particles was performed, allowing
350 visual determination of whether the amount of RNA successfully packaged for each segment was altered by
351 synonymous recoding. For each virus, equal titres of virus (in PFU) were pelleted through sucrose, then RNA
352 was extracted from viral pellets and separated by urea-PAGE. As expected, the PR8 WT virus yielded 7 bands
353 corresponding with 8 segments of RNA (segments 1 and 2 co-migrate as they are the same size; **Fig. 7C**).
354 The CDLR control band densities were visually similar to those of the WT virus. For seg1-CpGH, while it did
355 not appear that there were any differences in band densities (in keeping with the genome copy:PFU ratios),
356 an apparent retardation in the electrophoretic mobility of seg1-CpGH was observed. No differences in band
357 density were immediately apparent for seg5-CDLR or seg5-CpGH compared with seg5-WT PR8, but upon
358 close inspection, there was an observable retardation in electrophoretic mobility of seg5-CpGH compared with
359 seg5-WT, but to a much lesser degree than observed for the seg1-CpGH RNA.

360 The altered electrophoretic mobility of segments with added CpGs was unexpected as these constructs have
361 been compositionally corrected so that all bases are equally represented, and therefore the masses should be
362 identical. RNAs were run on denaturing gels, and so RNA structure should also not influence mobility.
363 Presence of the most likely post-transcriptional modification (cytosine methylation) had also been ruled out
364 (Fig. 5) (m6a methylation was considered an unlikely explanation as the frequency of the consensus DRACH
365 motif required for this modification (55) across the recoded region was 32 for PR8 WT, 31 for seg1-CDLR and
366 33 for seg1-CpGH). To determine whether an alternative post-transcriptional modification might explain the
367 shift in electrophoretic mobility of CpG-enriched RNAs, the seg1- and seg5-PR8, -CDLR and -CpGH transcripts
368 were synthesised by *in vitro* transcription, so that no cellular-derived post-transcriptional modifications could
369 occur. These RNAs were also subjected to electrophoretic separation, with equimolar CDLR and CpGH
370 constructs mixed together to confirm whether their migration was different or not (**Fig. 7D**). In the mix of seg1-
371 CDLR and seg1-CpGH, two discrete species were clearly visible, indicating that these sequences migrated
372 differently for reasons that cannot be explained by base composition or RNA modification. In the seg5-CDLR
373 and seg5-CpGH mix, two products could not be resolved but a slight difference in migration was apparent (as
374 a 'fatter' band), implying that altered electrophoretic mobility due to CpG enrichment is a reproducible trait.

375 Most likely, CpG introduction has altered secondary structures in the CpG-enriched RNA transcripts that
376 survived denaturation and retarded electrophoretic migration. Overall, however, we saw no evidence that
377 altering CpG frequencies deleteriously affected vRNA packaging.



378

379 **Figure 7. CpG enrichment does not affect packaging but may alter electrophoretic density.** Virus stocks of both seg1 and seg5
380 panel of viruses were titred by plaque assay, and then the relative copy numbers of segment 1 and segment 5 were determined by qPCR.
381 RNA:PFU ratios were then calculated for the seg1 virus panel **(A)** and the seg5 virus panel **(B)**. A '4c6c' mutant virus (54) was used as a
382 packaging mutant control. To visualise packaged viral RNAs, Viral stocks were purified by ultracentrifugation and subjected to urea-PAGE,
383 thereby separating the 8 segments of the viral genome **(C)**. A shift in the electrophoretic mobility of segment 1 of the seg1-CpGH virus
384 was observed (larger black arrow), and in order to determine whether this was due to RNA modification, RNAs generated through *in vitro*
385 transcription (i.e. a cell-free system, free from RNA modifying enzymes) were run under the same conditions **(D)**. For 'mix' samples,
386 equimolar combinations of CDLR and CpGH constructs were combined and run to unambiguously demonstrate whether a mobility shift
387 was apparent in CpGH constructs.

388 **Large scale recoding offers a genetically stable method of virus attenuation.**

389 Synonymous recoding of viral genomes as a method of attenuation for vaccine development is reliant on the
390 modifications being genetically stable. To investigate the stability of CpG enrichment in the IAV genome, the
391 seg1 and seg5 virus panels were serially passaged ten times in A549 cells. Passages were performed four
392 times, with input virus for two experiments being derived from egg-grown stocks and two from MDCK-grown

393 stocks. For the seg1 virus panel, the ~1 $\log_{10}$ reduction in titre for seg1-CpGH compared with WT or seg1-
394 CDLR virus was maintained over the course of the passage experiments, for both egg-derived (**Fig. 8A**) and
395 MDCK-derived (**Fig. 8B**) virus stocks. Seg5-CpGH similarly maintained its replication defect for both egg-
396 derived (**Fig. 8C**) and MDCK-derived (**Fig. 8D**) virus stocks.

397 After the tenth passage, the resulting viruses were subjected to full-genome deep amplicon sequencing to
398 analyse nucleotide changes occurring in their genomes. Sequencing coverage across the twenty viruses (4
399 replicates each of WT PR8, seg1-CDLR, seg1-CpGH, seg5-CDLR and seg5-CpGH) ranged from 127X-
400 139,000X (**Fig. S5**).

401 To establish background signal from PCR and sequencing error, amplifications and sequencing were
402 performed on the PR8 plasmids used for virus rescue. In this control sample (**Fig. S6U**), the highest mismatch
403 versus database reference was a single position at 2.3% (data not shown), and so this was used as the
404 threshold above which mutations were considered real. The sum totals of mutations over this threshold arising
405 in all egg-derived (**Fig. 8E**) and MDCK-derived (**Fig. 8F**) virus stocks after ten passages were calculated to
406 include proportional representation for substitutions that had not reached fixation (for example, if a mutation
407 was present in 10% of sequences, this was counted as 0.1 mutations). Consistent numbers of mutations were
408 observed across both seg1 and seg5 virus panels despite the lower viral titres for the two CpGH viruses, and
409 so more mutations were detected per infectious virion for CpG enriched viruses. However, the overall number
410 of nucleotide changes identified in recoded viruses did not exceed those for WT PR8, indicating a genetic
411 stability equivalent to WT virus. The numbers of mutations arising were broken down by segment (**Fig. 8G**);
412 the highest density of mutations occurred in segments 4 and 6 (encoding the viral surface proteins) and
413 segment 8 (encoding NS1 and NEP), as expected (56).

414 Nucleotide changes occurring at recoded sites after ten passages were analysed for seg1-CpGH (**Fig. 8H**)
415 and the seg5-CpGH (**Fig. 8I**). In the recoded segment 1, no nucleotide changes at altered sites became fixed
416 in any of the four replicates. For the seg5-CpGH, one change in both replicates of the egg-derived virus went
417 to fixation (A312T, represented in 98.8 and 99.0% of sequences) by passage ten; another was identified in
418 64.3% of the reads from one egg-derived replicate (A1146G). One mutation from an MDCK-derived virus was
419 present in 36.6% of the sequence reads of one replicate (A315G). All of these mutations were synonymous,
420 and none occurred at CpG sites; the third described mutation actually introduced a CpG site. Full genome
421 maps of sequencing changes for all viruses are presented in Supplementary **Fig. S6**. Overall, the data
422 indicated that CpG enrichment is a genetically stable recoding strategy for IAV.

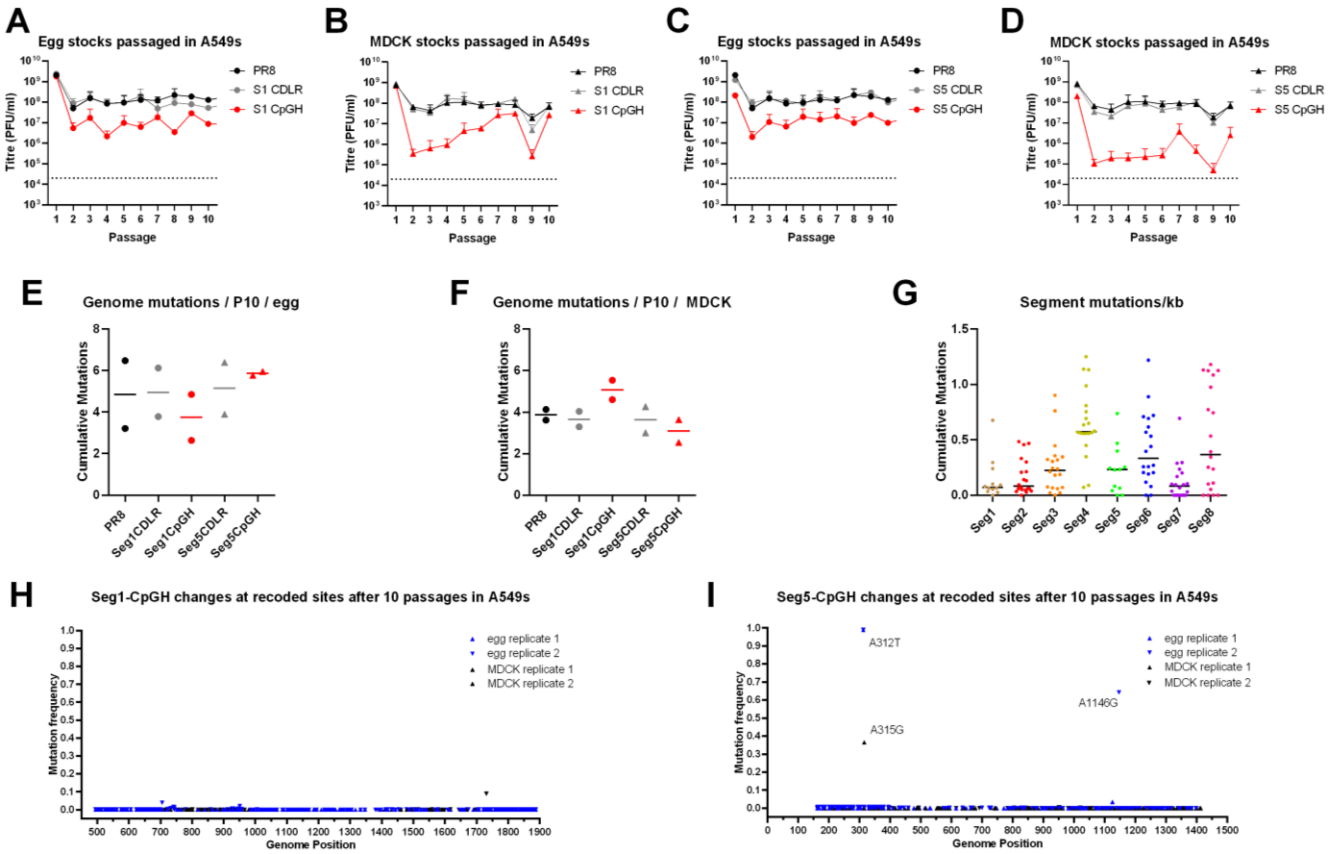


Figure 8. Serial passage of CpG modified viruses does not enable deselection of CpG motifs. Segment 1 (**A**) and segment 5 (**B**) egg derived virus stocks were passaged at an MOI of 0.01 ten times in A549 cells, and titred after each passage (*n* = 2). Segment 1 (**C**) and segment 5 (**D**) MDCK derived virus stocks were passaged at an MOI of 0.01 ten times in A549 cells, and titred after each passage (*n* = 2). After ten passages, viruses were whole genome sequenced to test for reversion or epistatic compensatory mutations. No mutations became fixed in the recoded region of segment 1 after 10 passages in any experiment. (**E**). Total cumulative genome mutations after ten passages were calculated, with proportional representation of mutations that had not become fixed (e.g. if 10% of sequences contained a mutation at a given position, this scored as 0.1 mutations) calculated for egg virus stock passages (**F**) and MDCK virus stock passages. (**G**). Cumulative mutations were calculated for each segment for each of the five viruses in the panel across four replicates at passage 10. (**H**). Mutation frequency at recoded positions across segment 1. The sequences of the four serially passaged replicates are overlaid. (**I**). As in (H) for segment 5. Three nucleotide changes that were at or heading towards fixation are indicated, but each were present in only one of four replicates. No mutations at CpG sites were observed. For sequence coverage, please refer to **Fig. S5** and for full genome mutational profiles to **Fig. S6**.

432 **ZAP-dependent CpG sensitivity is overcome in vaccine propagation systems**

433 Live attenuated IAV vaccines are traditionally propagated in eggs, but there is an increasing impetus to move
434 towards cell culture based propagation systems for ethical and reliability reasons (28,57). The impact of CpG
435 enrichment on viral titres was therefore examined in both systems. MDCKs are the preferred cell line for IAV
436 vaccine virus propagation (28), and so viral growth was examined in these cells. Although seg1-CpGH yielded
437 slightly smaller plaques in these cells compared with PR8 WT (Fig. S2), titres from low MOI infections were
438 equivalent across the seg1 virus panel; conversely, the seg5-CpGH virus was attenuated (**Fig. 9A**).

439 Viral rescues were performed in embryonated hens' eggs, and it was noted that the titres obtained for seg1-
440 CpGH virus (48 hours post infection) were reproducibly equivalent to those for WT PR8, whereas those of the
441 seg5-CpGH virus were lower (**Fig. 9B**). Possibly, the replication defect in seg1-CpGH was not detected in this
442 system as titres had become saturated. To test this, eggs were inoculated at embryonic day 6 (ED6) and day
443 10 (ED10) with 100 PFU virus and harvested at 24 hours. While virus replicated to higher levels in eggs at
444 ED10, no differences in titres between WT PR8 and seg1-CpGH viruses were seen at either developmental
445 time point (**Fig. 9C**). Next, embryonic survival was observed over four days between ED10-14 when eggs were
446 inoculated with ten-fold serially diluted virus doses, but again no differences were apparent between the WT
447 PR8, seg1-CDLR and seg1-CpGH viruses, with all embryos dying at similar rates (**Fig. S7**). Nonlinear
448 regression curves were fit for lethality across the dose range used (**Fig. 9D**); r2 values of 0.92-0.99 were
449 obtained. LD50 values were within a five-fold range (PR8 = 0.29PFU, CDLR = 1.27PFU and CpGH =
450 0.55PFU), occupying a range within the limits of a single dilution factor for this experiment and therefore
451 confirming a lack of difference.

452 To determine whether the lack of fitness penalty for the seg1-CpGH virus was recapitulated in other chicken
453 cells, chicken lung epithelial cells (CLEC213s) (58) were infected with the seg1 and seg5 virus panels, and the
454 same phenotype was observed (**Fig. 9E**). As IAV is tropic for the gastrointestinal tract in birds (59), replication
455 of the seg1 virus panel was also tested in chicken enteroids (60). Again, seg1-CpGH replicated to similar titres
456 to the WT virus (**Fig. 9F**) and although viral protein production was variable, there was no evidence of a defect
457 for the CpGH virus (**Fig. 9G**). In this system, there was also no evidence of differential infection levels either
458 by confocal imaging (**Fig. S8**) or by qPCR array to analyse expression of 88 innate immune genes in chickens
459 (**Fig. S9**).

460 Thus, while seg1-CpGH virus had a ZAP-mediated defect in human derived cell lines, no defect in replication
461 was evident in MDCK cells or in embryonated hens' eggs used for vaccine virus propagation.

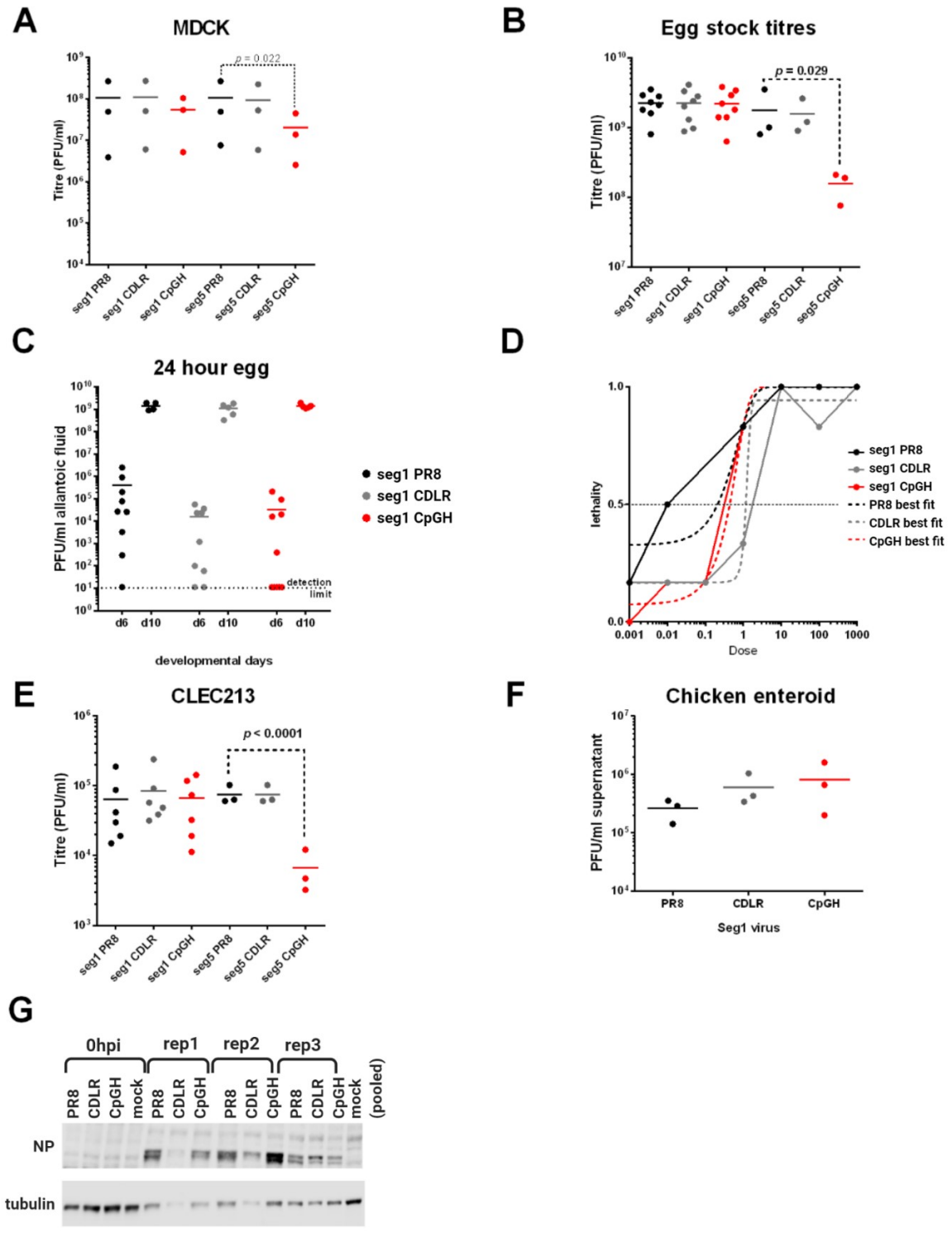


Figure 9. A virus enriched for CpG with ZAP-mediated attenuation in mammalian systems replicates equivalently to WT in chicken systems. A. Seg1 and seg5 viruses rescued in embryonated hens' eggs were titred. **B.** Embroyonated hens' eggs were inoculated with 100PFU virus at developmental day 5 or day 9, and viral titre in allantoic fluid was assessed 24 hours later (day 6 or day 10). **C.** Lethality of the seg1 virus panel in eggs was fit to nonlinear regression curves for LD50 calculations. **D.** The seg1 and seg5 virus panels were grown in chicken epithelial cell line CLEC213 and viral titres were quantified by plaque assay. **E.** The seg1 virus panel was used to infect chicken enteroids at an MOI of 10 for 24 hours, and the production of infectious virus, and viral protein production **(F)** was assayed. For confocal imaging of infected enteroids, please refer to **Fig. S8**. For innate gene expression analyses from infected enteroids, please refer to **Fig. S9**.

471 **DISCUSSION**

472 Synonymous recoding of viral genomes has been proposed as a potential strategy for live attenuated vaccine
473 development. However, such recoding inevitably impairs virus replication and so reduces yield. Here, we have
474 demonstrated that CpG enrichment is a viable vaccine development strategy for IAV as we have made a
475 replication-defective CpG-enriched virus whose propagation is unimpaired in systems used for vaccine virus
476 synthesis (MDCK cells and embryonated hens' eggs). Furthermore (to our knowledge for the first time), we
477 have demonstrated that a ZAP-sensitive CpG-enriched virus is attenuated in an animal model.

478 We tested CpG enrichment in two discrete regions of the IAV genome – segment 1 and segment 5. The
479 mechanisms of attenuation were different for the two constructs, as seg1-CpGH conferred ZAP sensitivity,
480 whereas this was not the case for seg5-CpGH. CpG enrichment impaired transcription and translation for both
481 constructs, likely via both ZAP-dependent and ZAP-independent mechanisms. Nevertheless, for the ZAP-
482 sensitive CpG-enriched virus, we found no evidence that CpG enrichment alters other stages of virus
483 replication, including methylation, packaging, or triggering of the IFN response. Furthermore, CpG-enrichment
484 is a genetically stable recoding strategy in IAV. Altogether, the data presented herein support this approach
485 for live attenuated vaccine development.

486 The lack of ZAP sensitivity demonstrated by the seg5-CpGH virus is currently unexplained. In plaque assays,
487 seg5-CpGH gave considerably smaller plaques (35), whereas this was not the case for seg1-CpGH. Seg5-
488 CpGH had an apparent translational defect in cell-based systems under the IAV promoter that was not seen
489 in an *in vitro* system under a bacterial promoter, and was also capable of inducing an IFN response, not seen
490 with seg1-CpGH (Fig. 6). We favour two, non-exclusive hypotheses to explain these phenotypes. Firstly, we
491 may have introduced an unfavourable sequence that acts as a pathogen-associated molecular pattern for
492 another cellular sensor; alternatively, we may have disrupted an unknown functional element in the virus. It is
493 possible that the offending sequence elements may not have been introduced/ disrupted in the seg5-CDLR
494 control; while recoding has been performed over the same ~1kb region, individual mutations are by necessity
495 not matched at individual base positions. Serial passage of the seg5-CpGH virus identified three mutations at
496 recoded sites that were either at fixation or trending towards selection. These mutations did not result in any
497 increase in virus replication but require characterisation in future studies. This scenario illustrates the
498 importance of understanding the mechanism of attenuation when considering the development of synthetically
499 recoded live attenuated vaccines; if we do not know the attenuation mechanism, it is possible that a heavily

500 recoded virus is actually only a few mutations away from reversion to WT virulence. For this reason, multiple
501 methods of viral attenuation integrated into a single platform should be considered.

502 During these investigations, we found no evidence of m5C-methylation in the IAV genome. Cytosine
503 methylation has been reported in the transcriptomes of RNA viruses including HIV-1 (61,62), and is observed
504 in a range of cellular RNA species (44). We only examined negative sense vRNA for evidence of methylation;
505 possibly, positive sense RNAs may undergo cytosine methylation more readily than their negative sense
506 counterpart.

507 It has previously been reported that the IAV NS1 protein can specifically inhibit ZAPS or its binding partner
508 TRIM25 (50,51) and that this can be ablated by mutation. However, we were unable to exacerbate CpG-
509 mediated attenuation by mutations in NS1 proposed to disrupt ZAP/TRIM25 targeting. Possibly, our assays
510 were not sensitive enough to detect accentuation of a defect. If ZAP/TRIM25 act through triggering of type I
511 IFN signalling, other aspects of NS1 function not affected by the mutations we added may still inhibit this
512 pathway. Nonetheless, the NS1 96/97 pair of mutations have previously been reported to lack specificity, likely
513 due to NS1 misfolding (50,52).

514 Here, we show a fundamental difference between human and chicken systems with respect to viral CpG
515 sensing. While a homolog of ZAPL has been identified in chickens (20), no ZAPS homolog has been identified.
516 That chicken ZAP may not have the same CpG binding specificity as human ZAP has previously been reported
517 using RNA-protein binding assays (19). However, this is the first demonstration that a ZAP-sensitive CpG-
518 enriched virus can be propagated equivalently to a WT virus in chicken systems. Possibly, chicken ZAP has
519 lost CpG-specificity (19). Species-specific type I IFN signalling cascades may also contribute; for example,
520 chickens do not have RIG-I (63) and if CpG-specific responses signal via this pathway, chickens may have
521 lost the ability to sense aberrant CpGs. Biochemical studies investigating RNA protein interactions of ZAP and
522 its cofactors in different species are needed in order to understand this further.

523 CpG enrichment is increasingly being proposed as a strategy for live attenuated vaccine development. In IAV,
524 CpG enrichment does not significantly impair viral growth in embryonated hens' eggs or cultured MDCK cells
525 – the main vectors used for propagation of live attenuated IAV vaccines. This means that, if genomic CpG
526 frequency is increased in order to attenuate a virus in a ZAP-sensitive manner, vaccine virus titres can be
527 produced that are equivalent to WT virus, whilst still being attenuated in humans. More widely, any live
528 attenuated vaccines generated using a reverse genetics system and propagated in mammalian cells can be

529 CpG-enriched and grown in ZAP-/- cells; therefore, large scale recoding is extrapolatable to other virus
530 systems.

531

532 **MATERIALS AND METHODS**

533 **Cells.** Madin-Darby canine kidney (MDCK), human adenocarcinoma A549, human embryonic kidney
534 (HEK)293T (all Sigma, Glasgow, UK), Normal Human Dermal Fibroblasts (NHDF; Thermo Fisher) and
535 HEK293 cells and were grown in Dulbecco's Modified Eagle Medium (DMEM) (Sigma) supplemented with
536 10% v/v foetal calf serum (FCS) (Thermo Fisher Scientific, Waltham, MA) and 1% v/v penicillin/ streptomycin
537 (Thermo Fisher Scientific). CLEC213 cells were grown in 1:1 Ham's F21 nutrient broth/ DMEM supplemented
538 with 5% v/v FCS and 1% v/v penicillin/ streptomycin. Cells were maintained by twice-weekly passage. A549
539 ZAP -/- cells were a kind gift from the laboratory of Professor Sam Wilson (38), HEK293 TRIM25 -/- cells were
540 a kind gift from the laboratory of Professor Gracjan Michlewski (39), and A549 KHNYN -/- cells were a kind gift
541 from the laboratory of Dr Chad Swanson (17). CLEC213 cells were a kind gift from Dr Sascha Trapp and
542 generated by the laboratory of Dr Pascale Quéré (58). Cells were checked for mycoplasma contamination
543 regularly using a MycoAlert mycoplasma detection kit (Lonza, UK).

544 **Plasmids and plasmid mutagenesis.** IAV reverse genetic plasmids corresponding with each of the 8
545 segments of IAV genome were a derivative of the UK National Institute of Biological Standards & Control
546 A/Puerto Rico/8/1934 (PR8) vaccine strain (64). For synonymous introduction of additional CpGs into segment
547 1 of the viral genome with compensatory mutations to maintain base frequencies ('seg1-CpGH') or to generate
548 a permuted control construct synonymously recoded over the same genome region but with nucleotide,
549 dinucleotide and codon frequencies maintained ('seg1-CDLR'), constructs were designed using SSE software
550 Simmonics v1.2 'mutate sequences' function (65). The region of segment 1 targeted for mutation was selected
551 to avoid previously characterised genome packaging signals and splice donor and acceptor sites for PB2-S1
552 (66). Full segment sequences (**Table S1**) are summarised in comparison with our previously described
553 segment 5 virus panel (**Table 1**) (35). For recoding of 1.42 and 1.25kb regions of segment 1 and segment 5
554 respectively, synthetic plasmid constructs were ordered from GeneArt (Thermo Fisher) flanked by BsmBI sites
555 for cloning into pDUAL plasmids used for virus rescue, as previously described (64).

556 Mutagenesis of segment 8 was also performed to disrupt interactions of NS1 with CpG sensing pathways or
557 various components of innate immune signalling; these are summarised (**Table 2**). Mutations were made on

558 pDUAL plasmids encoding the PR8 gene (64) using a QuikChange II site directed mutagenesis kit (Agilent)
559 according to manufacturer's instructions. Primers used are tabulated (**Table S2A**).

560 **Table 1. Summary of the genetic properties of CpG modified influenza A viruses that were recoded in segment 1**
561 **or segment 5.** CAI, codon adaptation index. CAI scores have been calculated relative to the human genome (A/Puerto
562 Rico/8/1934) in https://www.biologicscorp.com/tools/CAICalculator. For full construct sequences, plese refer to Table S1.

Virus backbone	Mutation	No. CpGs	Δ CpG	No. UpAs	Δ UpA	CAI
A/Puerto Rico/8/1934	Wildtype (seg1)	51	-	108	-	0.72
A/Puerto Rico/8/1934	Seg1-CDLR	52	+1	108	-	0.72
A/Puerto Rico/8/1934	Seg1-CpGH	177	+126 (347%)	103	-5	0.64
A/Puerto Rico/8/1934	Wildtype (seg5)	43	-	52	-	0.74
A/Puerto Rico/8/1934	Seg5-CDLR	43	-	52	-	0.75
A/Puerto Rico/8/1934	Seg5-CpGH	129	+86 (300%)	54	+2	0.62

563

564 **Table 2. Previously characterised mutations in NS1 that reverse ZAP/TRIM25-mediated IAV attenuation**

NS1 target	Abrogating mutation	Mechanism disrupted	Reference
TRIM25	L95/S99A	RIG-I ubiquitination/ vRNP binding	(50)
ZAP	E96/E97A	Can no longer prevent TRIM25-mediated ubiquitination of ZAP	(51)
ZAP	PKQK107-110A	Can no longer prevent TRIM25-mediated ubiquitination of ZAP	(51)
Interferon	R38K41A	NS1 can no longer bind RNA	(49)

565

566 **Virus rescues.** Viruses were rescued as previously described (67,68). HEK293T cells in 6 well plates were
567 transfected in Opti-MEM reduced serum medium (Thermo Fisher) with 250ng each of the 8 plasmids
568 corresponding with each segment of viral genome (or 7 plasmids for mock) and 4 µl Lipofectamine 2000
569 transfection reagent (Thermo Fisher) per well. pDUAL plasmids were used for all segments, which are bi-
570 directional; vRNA is under a polI promoter, and in the opposite orientation, positive sense RNA is under a polII
571 promoter. After 24 hours, medium was replaced with serum free medium containing 1 µg/ml tosyl phenylalanyl
572 chloromethyl ketone (TPCK)-treated trypsin (Sigma) and 0.14% BSA (w/v) (Sigma) (viral growth medium). At
573 48-72 hours post-transfection, supernatants were collected and passaged, either (1) onto MDCK cells in viral
574 growth medium, with supernatants collected 48 hours later; or (2) 100 µl was inoculated into the allantoic fluid

575 of an embyronated hen's egg at days 11-12 development. A packaging mutant, '4c6c' (multiple nucleotide
576 changes in segment 4 and segment 6, encoding the haemagglutinin and neuraminidase viral surface proteins
577 respectively) for use as a control for RNA:PFU ratio assays was propagated in MDCK cells rather than eggs
578 to avoid reversion (31,54), as was an NS1 (segment 8) R38K41A mutant (defective in regulating interferon)
579 (49).

580 **Egg infections.** Dekalb White fertilised eggs were purchased from Henry Stewart (Norfolk, UK) and incubated
581 at 37.5ºC, 50% relative humidity until infection at day 11-12 (unless stated). Eggs were candled to confirm
582 viability. Under sterile conditions, an egg punch was used to create two small holes in the shell – one in the air
583 cavity and one in the locality of the allantoic fluid. 100 µl virus was injected into the allantoic fluid and both
584 holes were sealed with sticky tape. For preparation of virus stocks, infected eggs were left for 48 hours prior
585 to allantoic fluid harvest, which was spun down for 5 minutes at 800g to pellet cellular debris, aliquoted and
586 used as neat virus stock; presence of virus was confirmed by HA assay (69). For comparison of viral titres
587 between days 6 and 10, eggs were inoculated with 100 PFU of virus and were left for 24 hours prior to
588 harvesting of allantoic fluid for virus titration. For kill curves, eggs were inoculated at day 10 with 0.001 – 1000
589 PFU virus and incubated until day 14 unless embryonic death was observed before that time. At experimental
590 endpoints, eggs were chilled overnight and embryonic death was confirmed by membrane disruption.

591 **Virus titrations.** Viral titres were quantified by plaque assays; 800 µl of ten-fold serial dilutions of viral stocks/
592 experiment supernatants/ allantoic fluid were incubated on sub-confluent MDCKs in 6-well plates for an hour,
593 and then 2 ml overlay comprising 1:1 viral growth medium:2.4% cellulose (Sigma) was added. Plates were
594 incubated without disturbance for 48 hours, after which time 1 ml 10% formalin was added to each well and
595 cells were fixed for at least 20 minutes. Overlays were removed and cells were stained with 0.1% toluidine
596 blue (Sigma).

597 **Virus infections.** Cells were infected at either a high multiplicity of infection (MOI) of 3 (single cycle infections)
598 or low MOI of 0.01-0.1 (cell line dependent, multi-cycle infections) for 24 or 48 hours respectively unless
599 otherwise stated. Viruses were added to cells and incubated for 1 hour at 37°C, and then inoculum was
600 removed, cells were washed with serum free medium, and then either serum free medium (single cycle
601 infection) or viral growth medium (multi-cycle infection) was added. Infected cells were incubated at 37°C for
602 the duration of the experiment.

603 **Generation of lentivirus vectors expressing ZAPL and ZAPS.** In order to reconstitute isoform-specific ZAP
604 expression in ZAP -/- A549 cells, lentiviruses expressing ZAPL or ZAPS isoforms under a polII promoter were

605 produced using a pSCRPSY (accession number KT368137.1) vector-based system (a kind gift from the
606 laboratory of Prof Sam Wilson). Both the A549 Cas9+ ZAP -/- cells and pSCRPSY carry puromycin resistance,
607 and so a blasticidin S-resistance gene was cloned into the NheI site of pSCRPSY. pSCRPSY ZAPL and
608 pSCRPSY ZAPS constructs were produced by cloning a cDNA of the open reading frame of ZAPL (accession
609 number NM_020119.4) or ZAPS (accession number NM_024625.4) into the dual SfiI sites of pSCRPSY. For
610 lentivirus production, the growth media on HEK293Ts at 80% confluency in a T175 flask was replaced with
611 25ml of Opti-MEM Reduced Serum Medium (Thermo Fisher). 6 µg PMD2.G (Addgene, 12259), 18µg PsPAX2
612 (Addgene, 12260) and 12 µg of pSCRPSY (pSCRPSY, pSCRPSY-ZAPL or pSCRPSY-ZAPS) were added to
613 450µl of Opti-MEM. 72µl of Lipofectamine 2000 (Thermo Fisher, 11668019) was added to 450µl of Opti-MEM.
614 Diluted plasmids and Lipofectamine were mixed and incubated for 30 minutes at room temperature. The
615 transfection mixture was added to the flask and cells were then incubated at 37°C 5% CO_2 for 18h. Opti-MEM
616 was then replaced with 25ml of 10% FBS D-MEM. After a further 48h, supernatant containing P0 lentivirus
617 was collected and centrifuged to remove cell debris. Due to low titre of ZAPL-expressing lentivirus, 75 ml of
618 cell culture supernatant containing lentivirus was concentrated by ultracentrifugation (50,000 *g*, 2 hours). The
619 viral pellet was then resuspended in 1 ml of 10% FCS D-MEM. Lentivirus was aliquoted and stored at -80°C.

620 **Generation of A549 cells expressing specific ZAP isoforms.** A549 ZAP-/- cells at 70% confluency in a 6
621 well plate were transduced with ZAPL- or ZAPS-expressing pSCRPSY lentiviruses. 1ml of lentivirus
622 (pSCRPSY empty vector control, concentrated ZAPL, or ZAPS) or 10% FBS D-MEM as control, supplemented
623 with polybrene (Merck, TR-1003-G) at 8 µg/ml, was added to each well of a 6 well plate. Spinfection was
624 performed (500 *g*, 1 hour). 48 hours after transduction, the inoculum was removed and cells were passaged
625 in 10% FBS D-MEM supplemented with 50 µg/ml blasticidin to select for transduced cells. Expression of the
626 appropriate ZAP isoforms was confirmed by western blotting and confocal imaging.

627 **Western blotting.** Samples for western blotting were collected in 2X homemade Laemmli buffer (70), boiled
628 for 10 minutes, cooled and loaded onto 10% pre-cast gels (Bio-Rad) for separation by SDS-PAGE. Proteins
629 were transferred onto nitrocellulose membranes using wet transfer (100V for 90 minutes on ice) and
630 membranes were blocked using 5% milk (Morrisons, Edinburgh, UK) for at least 30 minutes. Membranes were
631 washed three times with TBS-T (Tris-buffered-saline/0.1% Tween 20 (Sigma)) and incubated overnight at 4°C
632 with primary antibodies in 2% BSA/TBS-T (0.1%). Commercial antibodies were used for β-tubulin (clone YL1/2,
633 Bio-Rad, Watford, UK; 1:5000), β-actin (Abcam ab8227; 1:2000), ZAP (Invitrogen PA5-31650; 1:1000),
634 TRIM25 (Abcam ab167154; 1:1000) and KHNYN (Insight Bio sc-514168; 1:1000). In-house rabbit polyclonal
635 antibodies raised against NP and PB2 are previously described (71) and were both diluted 1:1000. Membranes

636 were then washed 3 times and incubated with species-specific secondary antibodies conjugated to Alexafluor-
637 680 or -800 (Thermo Fisher) diluted 1:5000 in 2% BSA/TBS-T (0.1%) for 90 minutes. Membranes were washed
638 3 times and visualised using a LICOR Odyssey Fc imaging system.

639 **Confocal imaging.** Immunofluorescence was used to confirm ZAP expression in lentivirus-transduced A549
640 cells, and presence of viral NP in chicken enteroids. For imaging of A549 cells, A549 pSCRPSY-transduced
641 cells at 95% confluency on coverslips in 24 well plates were fixed using 500 µl of 3.7% neutral buffered formalin
642 for 15 minutes and then washed with PBS. Cells were permeabilised using 500 µl of 0.02% Triton X100 for 5
643 minutes and blocked for 30 minutes in 2% FCS. Anti-ZC3HAV1 Antibody (16820-1-AP) used 1:500 was diluted
644 in 2% FCS/ PBS and incubated on cells overnight at 4°C. 200 µl of 1:1000 AlexaFluor-488 anti-rabbit
645 secondary antibody (Thermo Fisher, A32731) was added for 1 hour. Cell nuclei were stained using 1:5000
646 Hoechst in PBS for 10 minutes. Between addition of each reagent, cells were washed 3 times using 500 µl of
647 PBS. All images were acquired with Zeiss LSM 710 confocal microscope fitted with 40X/1.4 oil-immersion
648 objective lens. Controls for background signal were mock infected cells and infected cells stained using
649 secondary antibodies only.

650 For organoid imaging, 50 organoids in a 24 well plate were fixed in 4% neutral buffered formalin for 45 minutes
651 and washed with PBS. Organoids were then permeabilised and blocked with 5% v/v goat serum (Sigma) in
652 permeabilisation solution (0.5% v/v bovine serum albumin and 0.1% v/v Saponin (Merck, Glasgow, UK) in
653 PBS) for 30 minutes. Antibodies were diluted in permeabilisation solution and incubated with organoids (mouse
654 monoclonal NP antibody AA5H clone (Abcam, Cambridge, UK ab20343) used 1:500 overnight at 4°C,
655 phalloidin-647 (Thermo Fisher A22287) used 1:50 and incubated for 1 hour at room temperature). 200ul of
656 1:1000 AlexaFluor-488 anti-mouse secondary antibody (Thermo Fisher A32723) was added for 1 hour. Cell
657 nuclei were stained using 1:5000 Hoechst in PBS for 10 minutes. Between addition of each reagent, cells
658 were washed 3x using 500 µl of PBS. All images were acquired with Zeiss LSM 880 confocal microscope fitted
659 with 40X/1.4 oil-immersion objective lens. Secondary-only and isotype (mouse IgG2b) controls were used.

660 **Mouse infections.** Mouse experiments were approved by the Roslin Institute Animal Welfare and Ethical
661 Review Board under Home Office project license PF795AF60. Five-week-old female BALB/c mice, maintained
662 in groups of six, were purchased from Harlan (Oxon, UK) and housed for one week prior to infections. Mice
663 were inoculated intranasally with 200 PFU virus (either PR8, seg1-CDLR, seg1-CpGH or mock) under brief
664 isofluorane anaesthesia. Mice were weighed daily until day 5, when mice were euthanised by CO_2
665 asphyxiation. Lower left lung lobes were harvested in 2 ml Eppendorf (Stevenage, UK) Safe-Lock micro test

666 tubes containing 500 µl DMEM and a stainless-steel carbide bead (Qiagen), and snap frozen on dry ice until
667 virus titration. For titration, lungs were thawed on wet ice and fragmented using a Qiagen TissueLyser II at 200
668 oscillations/minute for 2 minutes or until tissue was visibly homogenised. Suspended tissue was then serially
669 diluted and titrated by plaque assay.

670 **Minigenome assays.** Minigenome assays were performed as previously described (41). Culture medium on
671 70-90% confluent HEK293T cells in 24 well plates was replaced with Opti-MEM reduced serum medium. Cells
672 were then transfected with 50 ng/well (unless stated) of plasmids encoding IAV segments 1, 2, 3, and 5
673 (sufficient for viral ribonucleoprotein (vRNP) formation) with positive orientation under a PolII promoter (the
674 same pDUAL plasmid constructs used for virus rescue were used here), along with 20 ng/ well of a reporter
675 construct expressing an IAV-like vRNA with a luciferase coding sequence in place of an IAV open reading
676 frame. At 48 hours post-transfection, cells were harvested in 100 µl 5X reporter lysis buffer (Promega,
677 Chilworth, UK) and refrigerated to allow cell debris to settle to the bottom of the tube. 60 µl supernatant was
678 added to 25 µl 600 µM beetle luciferin in a white walled 96 well plate, and luminescence was measured on a
679 GloMax luminometer (Promega). Each data point was represented by four technical repeats, and each assay
680 was undertaken for at least four biological repeats.

681 ***In vitro* transcription assays.** PCR amplicons were generated for the production of positive or negative
682 sense, full-length segment 1 or segment 5 RNA by the addition of a T7 promoter sequence to the forward or
683 reverse primer, respectively (**Table S2B**). PCRs were performed using Q5 DNA polymerase kit (NEB) with the
684 appropriate primer pair and 25 ng of plasmid as template (Seg1-PR8, Seg1-CDLR, Seg1-CpGH, Seg5-PR8,
685 Seg5-CDLR or Seg5-CpGH). Cycling conditions were 95°C for 5 minutes, then 30 cycles of 95°C for 30
686 seconds, 50°C for 30 seconds then 72°C for 2 minutes, with a final extension of 72°C for 5 minutes. Amplicons
687 were purified by agarose gel extraction using the MinElute Gel Extraction kit (Qiagen) and quantified using the
688 Qubit dsDNA BR Assay kit (Thermo Fisher). RNA transcripts were generated with 40 ng of these amplicons
689 as template using the MEGAscript T7 Transcription kit (Thermo Fisher) in half reaction volumes according to
690 manufacturer's instructions for the indicated times. Following incubation, transcript concentrations were
691 determined using the Qubit RNA HS Assay kit (Thermo Fisher).

692 ***In vitro* translation assays.** *In vitro* generated RNA transcripts above were used templates for *in vitro*
693 translation assays using a Rabbit Reticulocyte Lysate kit (Promega) at the indicated concentrations. Reactions
694 were performed at 1/5th volume with the inclusion of Transcend Biotin-Lysyl-tRNA (Promega) to allow antibody
695 independent detection of all translated products. Samples were assayed by western blot (as above) with the

696 exception that membranes were blocked with 5% BSA/TBS for 1 hour, incubated with IRDye 800CW
697 Streptavidin (LICOR) for 1 hour, washed 3 times and visualised using LICOR Odyssey Fc imaging system.

698 **Reverse transcription (RT)-quantitative (q)PCR**. RT-qPCR was performed as previously described (35).
699 RNA was extracted from cells using Qiagen QiAmp Viral RNA mini kits. Nucleic acid extracts were treated
700 using RQ1 RNAse-free DNAse (Promega) according to manufacturer's instructions. qPCR was performed in
701 technical triplicates using BioLine (London, UK) SensiFAST One Step RT-qPCR Lo-Rox kit (BIO-82020), with
702 primers targeting a conserved region of segment 1 or segment 5 of IAV PR8 strain (**Table S2C**). Cycling
703 conditions were 45°C for 10 minutes, then 40 cycles of 95°C for 10 seconds then 60°C for 30 seconds. Melt
704 curves were performed between 50°C-99°C with one degree increments at second intervals.

705 **Bisulphite conversion.** RNA was extracted from ~$5x10^8$ IAV virions (egg virus stock) using a Qiagen RNEasy
706 mini extraction kit, followed by Promega RQ1 DNAse treatment. Bisulphite conversion reactions were
707 performed on extracted egg stock RNAs or *in vitro* generated genomic RNA (gRNA; negative sense) transcripts
708 using the Zymo (Freiburg, Germany) RNA bisulphite conversion kit (EZ RNA Methylation Kit, RS001). In the
709 reaction, cytosines are converted to uracils unless protected by 5-methylcytosine (5-mc) methylation. Following
710 conversion, RNA libraries for sequencing were prepared using the Takara (Saint-Germain-en-Laye, France)
711 low input RNAseq kit and sequenced on a MiSeq Micro v2 150PE flowcell at Edinburgh Genomics (Edinburgh
712 UK). Sequencing data handling and analyses were performed using the Galaxy platform (72). Adapter
713 sequences (AGATCGGAAGAGCACACGTCTGAACTCCAGTCA and
714 AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT) were trimmed from reads using cutadapt and output
715 sequences were joined using fastq-join. Sequences were aligned to an *in silico* converted PR8 segment 1
716 (where all cytosines in the negative sense were replaced with thymines) using Bowtie2 (73) and resultant BAM
717 datasets were converted to tabular pileup format to allow for calculation of the proportion of retained cytosines
718 (indicating presence of 5-mc).

719 Specific regions selected for further analysis were amplified by RT-PCR from the converted RNA. RNAs were
720 reverse transcribed using SuperScript III (Invitrogen) with random hexamer primers according to
721 manufacturer's instructions. PCRs were performed using specific primers (**Table S2D**) and Q5 High-Fidelity
722 Polymerase (NEB, Hitchin, UK) according to manufacturer's instructions. Cycling conditions were 98°C for 30
723 seconds, then 45 cycles of 98°C for 10 seconds, 58°C for 21 seconds then 72°C for 30 seconds with a final
724 extension of 72°C for 2 minutes. Amplicons were sent for sequencing using the Amplicon-EZ service (Genewiz,

725 Leipzig, Germany). Sequences were analysed as above except that adapter trimming was performed using
726 the primer sequences.

727 **Urea-PAGE RNA gels.** IAV virions equivalent to 10^{10} PFU from egg stocks were semi-purified by centrifugation
728 through a buffered 25% sucrose cushion (w/v) (100 mM NaCl, 10 mM Tris-HCl [pH 7], 1mM EDTA) using a
729 SW32Ti rotor in a Beckman Coulter Optima Max-E ultracentrifuge at 280,000 *g* for 90 minutes at 4°C. RNAs
730 were extracted from virions using an RNEasy mini extraction kit (Qiagen) following resuspension of the pellets
731 directly into 350 µl of buffer RLT. Extracted RNA was separated by PAGE using homemade 5% urea
732 polyacrylamide gels in 1X Tris-borate-EDTA (TBE) buffer (89mM Tris-borate, 2mM EDTA [pH 8.3]) run for 6
733 hours at 120 V. Viral gene segments were visualised by silver staining using the Silver Stain Plus Kit (Bio-Rad)
734 according to manufacturer's instructions and imaged using a Samsung Xpress C480FW scanner.

735 **HEK Blue assay.** To assay type I IFN production, A549 cells in 24 well plates were infected at an MOI of 10
736 as described above but in the absence of trypsin. At 10 hours post infection, supernatants were harvested
737 and viruses were inactivated by exposure to 120 mJ/cm^2 of UV for 10 minutes in a UVP CL-1000 UV
738 crosslinker. 20 µl of inactivated sample or type I IFN standard (Abcam, 500, 50 or 5 pg/ µl) was added to HEK-
739 Blue™ IFNα/β cells (InvivoGen hkb-ifnab) at 4x10^4 cells/ well in 96 well format. HEK-Blue cells have a gene
740 for secreted alkaline phosphatase incorporated under an ISG54 promoter (that is robustly induced in the
741 presence of type I IFN). Cells were incubated with supernatant from A549 infections for 24 hours. HEK-Blue
742 cell supernatants were then collected and added to QuantiBlue reagent (InvivoGen) according to
743 manufacturer's instructions. After 15-30 minutes, when a colour change was visible in the positive control
744 (titrated IFN), absorbance was read at 620 nm, as a correlate of IFN concentration.

745 **Serial passage.** Viruses were serially passaged ten times at an MOI of 0.01 in A549 cells. Viruses were titred
746 by plaque assay after each passage. Biological duplicates were performed with viruses rescued from both
747 eggs and from MDCK cells. At the end of ten passages, viruses were sequenced. RNA was extracted from
748 culture supernatants using Viral RNA mini kits (Qiagen) according to manufacturer's instructions. Extracted
749 RNAs were reverse transcribed using SuperScript III (Invitrogen) with the IAV vRNA specific Uni12 primer
750 (AGCAAAAGCAGG) (74) according to manufacturer's instructions. PCR primers were designed to amplify
751 approximately 400 base fragments with approximately 60 base overlaps across the genome (**Table S2E**).
752 Additional primers were designed specific to the segment 1 and segment 5 recoded regions as appropriate.
753 Four sets of primer pools each containing 10 non-overlapping sets of primer pairs were generated (1A-3E odd,
754 1B-3F even, 3G-8B odd and 4A-8C even) with each primer at a final concentration of 5 µM. PCRs were

755 performed using specific primers and Q5 High-Fidelity Polymerase (NEB) according to manufacturer's
756 instructions with primers at a concentration of 500 nM each in the final reactions. Cycling conditions were 98°C
757 for 30 seconds, then 45 cycles of 98°C for 20 seconds, 55°C for 20 seconds then 72°C for 30 seconds with a
758 final extension of 72°C for 2 minutes. The 4 multiplexed amplicon sets for each sample were purified using a
759 PureLink PCR Purification kit (Invitrogen), quantified using a Qubit dsDNA BR assay kit (Invitrogen) and pooled
760 at equimolar ratios. Amplicon pools were sent for sequencing using the Amplicon-EZ service (Genewiz).
761 Sequencing data handling and analyses were performed using the Galaxy platform (72). Primer sequences
762 were trimmed from reads using cutadapt and output sequences were joined using fastq-join. Sequences were
763 aligned to the PR8 reference genome or genome sequences with recoded segments 1 or 5, as appropriate,
764 using Bowtie2 (73). Variants and coverage levels in the resultant BAM datasets were analysed using iVar (75)
765 variants with tabular outputs.

766 **Chicken enteroid infections.** Chicken enteroids were prepared as described previously (60). Briefly, small
767 intestinal tissue was removed from Hy-Line Brown embryos supplied by the National Avian Research Facility,
768 Edinburgh, UK (Home Office Project Licenses PE263A4FA) and cut into 5 mm sections. Tissue was digested
769 and villi were isolated and cultured. At day 2 of culture, approximately 200 enteroids were infected with 10^5
770 virus particles/ enteroid for 1 hour, washed 3 times in PBS and then incubated for 48 hours in the presence of
771 2 µg/mL TPCK trypsin for virus quantification by plaque assay, confocal imaging, or arrayed gene expression
772 analyses.

773 **Fluidigm analysis of innate immune response gene expression in infected chicken enteroids.** At 48 hpi,
774 enteroids were suspended in RNALater and RNA was isolated using Tru-Seq total RNA Sample Preparation
775 v2 kit (Illumina, Great Abingdon, UK) according to manufacturer's instructions. cDNA was synthesized from
776 100 ng of RNA using High Capacity Reverse Transcription kit (Life Technologies, UK) according to
777 manufacturer's instructions with a random hexamer primer and oligo(dT). cDNA was preamplified as previously
778 described (76) and analysed through quantitative PCR (qPCR) with the microfluidic 96.96 Dynamic array
779 (Fluidigm, San Francisco, CA) performed on the BioMark HD instrument (BioMark, Boise, ID). Raw Cq values
780 were processed with GenEx6 and GenEx Enterprise (MultiD Analyses AB), with correction for primer efficiency
781 and reference gene normalisation. Reference genes used were TATA box binding protein (TBP), Tubulin alpha
782 chain (TUBA8B), beta-actin (ACTB), beta-glucuronidase (GUSB), glyceraldehyde-3-phosphate
783 dehydrogenase (GAPDH) and ribosomal 28S (r28S).

784

785 **ACKNOWLEDGEMENTS**

786 We are grateful to the labs of Prof Sam Wilson (MRC-University of Glasgow Centre for Virus Research, UK)
787 for provision of A549 ZAP-/- cell lines and pSCRPSY lentivirus vector, Prof Gracjan Michlewski (University of
788 Edinburgh, UK/ International Institute of Molecular and Cell Biology, Poland) for the provision of TRIM25-/-
789 HEK293 cells, Dr Chad Swanson (King's College London, UK) for A549 KHNYN-/- cells and Dr Sascha Trapp
790 (French National Institute for Agriculture, Food and Environment, France) for the provision of CLEC213 cells.
791 We are indebted to the staff in the Roslin Biological Research Facility for technical support for mouse infection
792 experiments and mouse line maintenance, to the staff of the Central Services Unit, Bioimaging facilities, and
793 to the National Avian Research Facility for supply of eggs in the Roslin Institute, University of Edinburgh. FG,
794 BMD, LV, PD and EG are supported by a BBSRC Institute Strategic Programme grant (BB/P013740/1). BT
795 was supported by a Microbiology Society Harry Smith Vacation studentship (GA002550) and is supported by
796 Wellcome Trust ISSF3 funding. FG and PD are supported by BBSRC grant no BB/S00114X/1. PS is supported
797 by Wellcome Investigator Award (WT103767MA). LV and PD are also supported by the European Union's
798 Horizon 2020 research and innovation programme under grant agreement no. 727922 (DELTA-FLU) and SS
799 under grant agreement no. 731014 (VetBioNet). TN was supported by an iCase studentship from the BBSRC
800 (BB/MO14819). EG is also supported by a Wellcome Trust/ Royal Society Sir Henry Dale Fellowship
801 (211222_Z_18_Z).

802 We are grateful to Christina Vrettou (The Roslin Institute, UK) for technical assistance, to Dr James Glover
803 (The Roslin Institute, UK) for critical feedback on the manuscript and to Dr Edward Hutchinson (MRC-
804 University of Glasgow Centre for Virus Research, UK) for helpful discussions.

805

806 **COMPETING INTERESTS**

807 None declared.

808 **SUPPLEMENTARY TABLES AND FIGURES**

809 **Table S1. Construct sequences.**

810 >A/human/PuertoRico/8_seg1_wildtype

811 AGCGAAAGCAGGTCAATTATATTCAATATGGAAAGAATAAAAGAACTAAGAAATCTAATGTCGCAGTCTCGCACCCGCGAGATACTC
812 ACAAAAACCACCGTGGACCATATGGCCATAATCAAGAAGTACACATCAGGAAGACAGGAGAAGAACCCAGCACTTAGGATGAAATG
813 GATGATGGCAATGAAATATCCAATTACAGCAGACAAGAGGATAACGGAAATGATTCCTGAGAGAAATGAGCAAGGACAAACTTTATG
814 GAGTAAAATGAATGATGCCGGATCAGACCGAGTGATGGTATCACCACTGGCTGTGACATGGTGGAATAGGAATGGACCAATAACAA
815 ATACAGTTCATTATCCAAAAATCTACAAAACTTATTTTGAAAGAGTCGAAAGGCTAAAGCATGGAACCTTTGGCCCTGTCCATTTTAG
816 AAACCAAGTCAAAATACGTCGGAGAGTTGACATAAATCCTGGTCATGCAGATCTCAGTGCCAAGGAGGCACAGGATGTAATCATGG
817 AAGTTGTTTTCCCTAACGAAGTGGGAGCCAGGATACTAACATCGGAATCGCAACTAACGATAACCAAAGAGAAGAAAGAAGAACTC
818 CAGGATTGCAAAATTTCTCCTTTGATGGTTGCATACATGTTGGAGAGAGAACTGGTCCGCAAAACGAGATTCCTCCCAGTGGCTGGT
819 GGAACAAGCAGTGTGTACATTGAAGTGTTGCATTTGACTCAAGGAACATGCTGGGAACAGATGTATACTCCAGGAGGGGAAGTGAG
820 GAATGATGATGTTGATCAAAGCTTGATTATTGCTGCTAGGAACATAGTGAGAAGAGCTGCAGTATCAGCAGATCCACTAGCATCTTT
821 ATTGGAGATGTGCCACAGCACACAGATTGGTGGAATTAGGATGGTAGACATCCTTAGGCAGAACCCAACAGAAGAGCAAGCCGTG
822 GATATATGCAAGGCTGCAATGGGACTGAGAATTAGCTCATCCTTCAGTTTTGGTGGATTCACATTTAAGAGAACAAGCGGATCATCA
823 GTCAAGAGAGAGGAAGAGGTGCTTACGGGCAATCTTCAAACATTGAAGATAAGAGTGCATGAGGGATATGAAGAGTTCACAATGGT
824 TGGGAGAAGAGCAACAGCCATACTCAGAAAAGCAACCAGGAGATTGATTCAGCTGATAGTGAGTGGGAGAGACGAACAGTCGATT
825 GCCGAAGCAATAATTGTGGCCATGGTATTTTCACAAGAGGATTGTATGATAAAAGCAGTCAGAGGTGATCTGAATTTCGTCAATAGG
826 GCGAATCAGCGATTGAATCCTATGCATCAACTTTTAAGACATTTTCAGAAGGATGCGAAAGTGCTTTTTCAAAATTGGGGAGTTGAAC
827 CTATCGACAATGTGATGGGAATGATTGGGATATTGCCCGACATGACTCCAAGCATCGAGATGTCAATGAGAGGAGTGAGAATCAGC
828 AAAATGGGTGTAGATGAGTACTCCAGCACGGAGAGGGTAGTGGTGAGCATTGACCGTTTTTTGAGAATCCGGGACCAACGAGGAA
829 ATGTACTACTGTCTCCCGAGGAGGTCAGTGAAACACAGGGAACAGAGAAACTGACAATAACTTACTCATCGTCAATGATGTGGGAG
830 ATTAATGGTCCTGAATCAGTGTTGGTCAATACCTATCAATGGATCATCAGAAACTGGGAAACTGTTAAAATTCAGTGGTCCCAGAAC
831 CCTACAATGCTATACAATAAAATGGAATTTGAACCATTTCAGTCTTTAGTACCTAAGGCCATTAGAGGCCAATACAGTGGGTTTGTAA
832 GAACTCTGTTCCAACAAATGAGGGATGTGCTTGGGACATTTGATACCGCACAGATAATAAAACTTCTTCCCTTCGCAGCCGCTCCAC
833 CAAAGCAAAGTAGAATGCAGTTCTCCTCATTTACTGTGAATGTGAGGGGATCAGGAATGAGAATACTTGTAAGGGGCAATTCTCCTG
834 TATTCAACTATAACAAGGCCACGAAGAGACTCACAGTTCTCGGAAAGGATGCTGGCACTTTAACTGAAGACCCAGATGAAGGCACA
835 GCTGGAGTGGAGTCCGCTGTTCTGAGGGGATTCCTCATTCTGGGCAAAGAGAACAAGAGATATGGGCCAGCACTAAGCATCAATG
836 AACTGAGCAACCTTGCGAAAGGAGAGAAGGCTAATGTGCTAATTGGGCAAGGAGACGTGGTGTTGGTAATGAAACGGAAACGGGA
837 CTCTAGCATACTTACTGACAGCCAGACAGCGACCAAAAGAATTCGGATGGCCATCAATTAGTGTCGAATAGTTTAAAAACGACCTTG
838 TTTCTACT

839 > A/human/PuertoRico/8_seg1_CDLR

840 AGCGAAAGCAGGTCAATTATATTCAATATGGAAAGAATAAAAGAACTAAGAAATCTAATGTCGCAGTCTCGCACCCGCGAGATACTC
841 ACAAAAACCACCGTGGACCATATGGCCATAATCAAGAAGTACACATCAGGAAGACAGGAGAAGAACCCAGCACTTAGGATGAAATG
842 GATGATGGCAATGAAATATCCAATTACAGCAGACAAGAGGATAACGGAAATGATTCCTGAGAGAAATGAGCAAGGACAAACTTTATG
843 GAGTAAAATGAATGATGCCGGATCAGACCGAGTGATGGTATCACCACTGGCTGTGACATGGTGGAATAGGAATGGACCAATAACAA
844 ATACAGTTCATTATCCAAAAATCTACAAAACTTATTTTGAAAGAGTCGAAAGGCTAAAGCATGGAACCTTTGGCCCTGTCCATTTTAG

845 AAACCAAGTCAAAATACGTCGGAGAGTTGACATAAATCCTGGTCATGCAGATCTCAGTGCCAAGGAGGCTCAGGATGTAATCATGG
846 AAGTGGTGTTCCCAAACGAGGTTGGAGCAAGAATCCTCACTTCTGAGTCGCAACTAACAATAACTAAAGAGAAAAAAGAGGAGCTA
847 CAAGATTGCAAAATATCACCTTTGATGGTTGCATACATGTTGGAGAGAGAACTGGTCCGAAAGACGAGATTCCTACCCGTGGCCGG
848 TGGGACCAGCAGTGTATACATAGAAGTGTTGCATTTGACTCAAGGGACGTGCTGGGAACAGATGTATACACCAGGGGGGGAAGTG
849 AGAAATGATGATGTTGACCAGAGCTTGATAATTGCCGCGAGGAACATTGTGAGAAGAGCCGCCGTATCGGCAGATCCTCTAGCATC
850 CTTATTGGAAATGTGTCACAGCACACAGATTGGTGGAATAAGAATGGTTGACATCCTTAGACAGAATCCCACAGAAGAGCAGGCTGT
851 GGATATTTGCAAGGCTGCCATGGGGCTGAGAATCAGTTCATCATTTAGTTTTGGTGGATTCACTTTCAAGAGAACAAGCGGGTCATC
852 AGTCAAGAGAGAGGAAGAAGTGCTAACGGGGAATCTCCAAACATTGAAAATAAGAGTTCATGAGGGCTATGAAGAGTTCACGATGG
853 TTGGGAGGAGAGCAACCGCCATACTTAGAAAAGCAACGAGGAGATTGATACAACTGATAGTGAGTGGCAGAGACGAGCAATCTATT
854 GCAGAAGCAATAATTGTGGCTATGGTGTTTTCCCAGGAAGATTGTATGATTAAAGCAGTCAGAGGAGACCTGAACTTTGTTAATAGG
855 GCAAATCAACGATTGAACCCTATGCATCAACTTTTAAGACATTTTCAAAAGGATGCAAAAGTGCTTTTTCAAAATTGGGGAGTTGAAC
856 CAATCGACAATGTGATGGGCATGATAGGAATTTTGCCAGACATGACTCCAAGCATCGAGATGTCAATGAGAGGTGTCAGAATCAGC
857 AAAATGGGTGTAGATGAATACTCGAGCACTGAGAGGGTAGTGGTCAGCATTGACCGATTTTTGAGAATTCGGGATCAACGAGGGAA
858 TGTTCTTCTGTCACCGGAAGAGGTGAGTGAAACTCAGGGAACTGAGAAACTGACAATAACATATTCCTCATCAATGATGTGGGAGAT
859 TAATGGCCCAGAATCAGTATTGGTGAATACATATCAATGGATCATCAGAAACTGGGAGACTGTTAAAATTCAGTGGTCTCAGAATCC
860 CACTATGCTATACAATAAAATGGAGTTCGAGCCCTTCCAATCTTTAGTGCCTAAGGCCATTAGAGGCCAGTACAGTGGATTCGTAAG
861 AACACTGTTCCAACAGATGAGAGATGTACTTGGGACATTTGATACAGCACAGATTATAAAACTGCTTCCATTTGCAGCCGCTCCACC
862 AAAGCAAAGTAGAATGCAGTTCTCCTCATTTACTGTGAATGTGAGGGGATCAGGAATGAGAATACTTGTAAGGGGCAATTCTCCTGT
863 ATTCAACTATAACAAGGCCACGAAGAGACTCACAGTTCTCGGAAAGGATGCTGGCACTTTAACTGAAGACCCAGATGAAGGCACAG
864 CTGGAGTGGAGTCCGCTGTTCTGAGGGGATTCCTCATTCTGGGCAAAGAGAACAAGAGATATGGGCCAGCACTAAGCATCAATGAA
865 CTGAGCAACCTTGCGAAAGGAGAGAAGGCTAATGTGCTAATTGGGCAAGGAGACGTGGTGTTGGTAATGAAACGGAAACGGGACT
866 CTAGCATACTTACTGACAGCCAGACAGCGACCAAAAGAATTCGGATGGCCATCAATTAGTGTCGAATAGTTTAAAAACGACCTTGTT
867 TCTACT

868 > A/human/PuertoRico/8_seg1_CpG-high

869 AGCGAAAGCAGGTCAATTATATTCAATATGGAAAGAATAAAAGAACTAAGAAATCTAATGTCGCAGTCTCGCACCCGCGAGATACTC
870 ACAAAAACCACCGTGGACCATATGGCCATAATCAAGAAGTACACATCAGGAAGACAGGAGAAGAACCCAGCACTTAGGATGAAATG
871 GATGATGGCAATGAAATATCCAATTACAGCAGACAAGAGGATAACGGAAATGATTCCTGAGAGAAATGAGCAAGGACAAACTTTATG
872 GAGTAAAATGAATGATGCCGGATCAGACCGAGTGATGGTATCACCACTGGCTGTGACATGGTGGAATAGGAATGGACCAATAACAA
873 ATACAGTTCATTATCCAAAAATCTACAAAACTTATTTTGAAAGAGTCGAAAGGCTAAAGCATGGAACCTTTGGCCCTGTCCATTTTAG
874 AAACCAAGTCAAAATACGTCGGAGAGTTGACATAAATCCTGGTCATGCAGATCTCAGTGCAAAAGAAGCTCAAGATGTGATCATGGA
875 AGTAGTTTTTCCGAATGAAGTCGGCGCGCGAATACTAACGTCGGAATCGCAACTGACGATAACGAAAGAGAAAAAAGAAGAACTTC
876 AGGATTGCAAAATTTCGCCATTAATGGTCGCGTATATGTTGGAACGCGAACTCGTTCGAAAAACGCGATTTCTTCCGGTTGCCGGAG
877 GAACGAGCAGCGTGTACATAGAAGTATTACATTTGACGCAAGGAACGTGTTGGGAACAAATGTATACGCCGGGCGGCGAAGTTCGA
878 AACGATGACGTAGATCAAAGTTTGATTATAGCGGCTCGAAATATCGTGCGACGCGCGGCGGTTTCCGCCGATCCGCTCGCGTCATT
879 ATTGGAAATGTGTCACAGCACGCAGATAGGAGGAATCCGAATGGTCGACATTCTACGGCAAAATCCGACGGAAGAACAAGCAGTCG
880 ACATTTGCAAGGCCGCGATGGGACTGCGAATCAGTTCGTCTTTTAGTTTTGGCGGATTCACGTTCAAACGTACGAGTGGATCGTCA
881 GTGAAGCGCGAAGAAGAAGTTCTGACGGGAAATCTTCAAACATTGAAAATCCGAGTGCATGAAGGATACGAAGAATTCACTATGGT
882 CGGTCGACGCGCGACAGCGATACTACGAAAAGCGACGCGTCGATTAATTCAGCTGATAGTTAGCGGACGGGACGAACAGTCGATC
883 GCCGAAGCGATCATCGTAGCGATGGTGTTTTCGCAAGAAGATTGCATGATAAAAGCCGTGCGAGGCGATCTAAATTTCGTAAATCG

884 AGCGAATCAACGATTGAATCCGATGCATCAACTTTTGCGACATTTTCAGAAAGACGCGAAAGTGCTTTTTCAAAATTGGGGAGTCGA
885 ACCGATCGACAACGTAATGGGAATGATAGGTATTTTACCCGACATGACGCCGAGCATCGAAATGTCGATGCGCGGCGTTCGAATCA
886 GCAAAATGGGCGTCGACGAATATTCAAGCACAGAGCGAGTCGTTGTAAGCATTGATCGTTTTTTACGAATACGCGATCAACGAGGT
887 AACGTTCTTCTTTCGCCGGAAGAAGTAAGTGAAACGCAAGGTACGGAGAAACTTACGATCACGTATTCGTCGTCAATGATGTGGGA
888 GATCAACGGTCCAGAATCCGTCTTGGTAAACACGTATCAATGGATCATCCGAAATTGGGAAACAGTGAAAATTCAGTGGTCGCAAAA
889 TCCGACGATGCTTTACAACAAAATGGAATTCGAACCGTTTCAATCATTAGTTCCAAAAGCGATTCGCGGTCAATATAGTGGATTTGTT
890 CGAACGCTGTTTCAACAAATGCGTGACGTACTCGGAACGTTCGACACCGCGCAAATTATTAAACTTCTACCGTTCGCAGCCGCTCC
891 ACCAAAGCAAAGTAGAATGCAGTTCTCCTCATTTACTGTGAATGTGAGGGGATCAGGAATGAGAATACTTGTAAGGGGCAATTCTCC
892 TGTATTCAACTATAACAAGGCCACGAAGAGACTCACAGTTCTCGGAAAGGATGCTGGCACTTTAACTGAAGACCCAGATGAAGGCA
893 CAGCTGGAGTGGAGTCCGCTGTTCTGAGGGGATTCCTCATTCTGGGCAAAGAGAACAAGAGATATGGGCCAGCACTAAGCATCAAT
894 GAACTGAGCAACCTTGCGAAAGGAGAGAAGGCTAATGTGCTAATTGGGCAAGGAGACGTGGTGTTGGTAATGAAACGGAAACGGG
895 ACTCTAGCATACTTACTGACAGCCAGACAGCGACCAAAAGAATTCGGATGGCCATCAATTAGTGTCGAATAGTTTAAAAACGACCTT
896 GTTTCTACT

897 >A/human/PuertoRico/8_seg5_wildtype

898 AGCAAAAGCAGGGTAGATAATCACTCACTGAGTGACATCAAAATCATGGCGTCTCAAGGCACCAAACGATCTTACGAACAGATGGA
899 GACTGATGGAGAACGCCAGAATGCCACTGAAATCAGAGCATCCGTCGGAAAAATGATTGGTGGAATTGGACGATTCTACATCCAAA
900 TGTGCACCGAACTCAAACTCAGTGATTATGAGGGACGGTTGATCCAAAACAGCTTAACAATAGAGAGAATGGTGCTCTCTGCTTTTG
901 ACGAAAGGAGAAATAAATACCTTGAAGAACATCCCAGTGCGGGGAAAGATCCTAAGAAAACTGGAGGACCTATATACAGGAGAGTA
902 AACGGAAAGTGGATGAGAGAACTCATCCTTTATGACAAAGAAGAAATAAGGCGAATCTGGCGCCAAGCTAATAATGGTGACGATGC
903 AACGGCTGGTCTGACTCACATGATGATCTGGCATTCCAATTTGAATGATGCAACTTATCAGAGGACAAGAGCTCTTGTTCGCACCGG
904 AATGGATCCCAGGATGTGCTCTCTGATGCAAGGTTCAACTCTCCCTAGGAGGTCTGGAGCCGCAGGTGCTGCAGTCAAAGGAGTT
905 GGAACAATGGTGATGGAATTGGTCAGAATGATCAAACGTGGGATCAATGATCGGAACTTCTGGAGGGGTGAGAATGGACGAAAAAC
906 AAGAATTGCTTATGAAAGAATGTGCAACATTCTCAAAGGGAAATTTCAAACTGCTGCACAAAAAGCAATGATGGATCAAGTGAGAGA
907 GAGCCGGAACCCAGGGAATGCTGAGTTCGAAGATCTCACTTTTCTAGCACGGTCTGCACTCATATTGAGAGGGTCGGTTGCTCACA
908 AGTCCTGCCTGCCTGCCTGTGTGTATGGACCTGCCGTAGCCAGTGGGTACGACTTTGAAAGGGAGGGATACTCTCTAGTCGGAATA
909 GACCCTTTCAGACTGCTTCAAAACAGCCAAGTGTACAGCCTAATCAGACCAAATGAGAATCCAGCACACAAGAGTCAACTGGTGTG
910 GATGGCATGCCATTCTGCCGCATTTGAAGATCTAAGAGTATTAAGCTTCATCAAAGGGACGAAGGTGCTCCCAAGAGGGAAGCTTT
911 CCACTAGAGGAGTTCAAATTGCTTCCAATGAAAATATGGAGACTATGGAATCAAGTACACTTGAACTGAGAAGCAGGTACTGGGCCA
912 TAAGGACCAGAAGTGGAGGAAACACCAATCAACAGAGGGCATCTGCGGGCCAAATCAGCATACAACCTACGTTCTCAGTACAGAGA
913 AATCTCCCTTTTGACAGAACAACCATTATGGCAGCATTCAATGGGAATACAGAGGGGAGAACATCTGACATGAGGACCGAAATCATA
914 AGGATGATGGAAAGTGCAAGACCAGAAGATGTGTCTTTCCAGGGGCGGGGAGTCTTCGAGCTCTCGGACGAAAAGGCAGCGAGC
915 CCGATCGTGCCTTCCTTTGACATGAGTAATGAAGGATCTTATTTCTTCGGAGACAATGCAGAGGAGTACGACAATTAAAGAAAAATA
916 CCCTTGTTTCTACT

917 >A/human/PuertoRico/8_seg5_CDLR

918 AGCAAAAGCAGGGTAGATAATCACTCACTGAGTGACATCAAAATCATGGCGTCCCAAGGCACCAAACGATCTTACGAACAGATGGA
919 GACTGATGGAGAACGCCAGAATGCCACTGAAATCAGAGCATCCGTCGGAAAAATGATTGGTGGAATTGGACGGTTCTACATTCAAA
920 TGTGCACCGAACTCAAACTCAGTGACTATGAAGGACGGTTGATTCAAAACAGCTTAACCATAGAGAGAATGGTCCTCTCTGCTTTTG
921 ACGAAAGAAGAAATAAATACCTTGAAGAACATCCCAGTGCTGGGAAGGATCCTAAGAAAACTGGAGGCCCAATCTACAGGAGGGTA

922 AACGGAAAGTGGATGAGAGAACTCATTCTGTATGATAAAGAAGAGATAAGGCGGATCTGGAGACAAGCCAATAATGGTGACGATGC
923 AACGGCTGGCCTGACTCACATGATGATATGGCATTCTAATTTGAATGATGCAACATATCAGAGAACTAGAGCTCTGGTGAGGACCG
924 GAATGGATCCCAGAATGTGCTCTCTGATGCAAGGCTCAACTCTTCCTAGAAGATCAGGTGCAGCAGGTGCTGCCGTCAAAGGAGTT
925 GGGACAATGGTGATGGAATTGGTTAGAATGATCAAACGAGGGATCAATGATCGGAACTTTTGGAGAGGTGAAAATGGACGGAAAAC
926 AAGAATTGCATATGAGAGGATGTGCAACATCCTCAAAGGAAAATTTCAAACTGCTGCTCAAAAAGCTATGATGGATCAAGTGCGAGA
927 AAGCAGGAATCCAGGGAACGCTGAATTCGAAGATCTCACCTTCCTAGCACGGTCTGCACTCATATTGAGAGGGTCTGTTGCACACA
928 AGTCCTGTCTCCCTGCGTGTGTCTATGGTCCGGCGGTTGCCAGTGGGTATGATTTTGAGAGAGAGGGGTACTCACTAGTCGGGATA
929 GACCCTTTCAGACTGCTTCAGAACAGCCAAGTGTACAGCCTAATCAGACCTAATGAAAATCCTGCTCACAAAAGCCAACTAGTGTGG
930 ATGGCTTGCCATTCTGCAGCTTTTGAGGATCTAAGGGTATTAAGTTTCATAAAAGGGACGAAAGTACTGCCTAGAGGAAAGCTCTCA
931 ACAAGAGGAGTCCAAATTGCTTCCAATGAAAACATGGAGACCATGGAGTCAAGTACACTTGAACTGAGGAGCAGATACTGGGCAAT
932 AAGGACCCGAAGTGGAGGGAATACCAACCAACAGAGGGCTTCTGCCGGACAAATCAGCATCCAACCCACATTTTCAGTGCAGCGA
933 AATCTACCTTTTGACAGAACAACAATCATGGCAGCATTCAATGGCAACACGGAGGGAAGAACATCCGACATGAGGACAGAAATCAT
934 CAGGATGATGGAAAGTGCAAGACCCGAAGATGTGTCTTTCCAGGGGCGGGGAGTCTTCGAGCTCTCGGACGAAAAGGCAGCGAG
935 CCCGATCGTGCCTTCCTTTGACATGAGTAATGAAGGATCTTATTTCTTCGGAGACAATGCAGAGGAGTACGACAATTAAAGAAAAAT
936 ACCCTTGTTTCTACT

937 >A/human/PuertoRico/8_seg5_CpG-high

938 AGCAAAAGCAGGGTAGATAATCACTCACTGAGTGACATCAAAATCATGGCGTCCCAAGGCACCAAACGATCTTACGAACAGATGGA
939 GACTGATGGAGAACGCCAGAATGCCACTGAAATCAGAGCATCCGTCGGAAAAATGATTGGTGGAATTGGACGATTTTACATACAAAT
940 GTGCACGGAACTAAAACTCAGTGATTACGAAGGACGTTTGATTCAAAACAGTTTAACGATTGAGCGAATGGTTCTTTCGGCATTTGA
941 CGAACGAAGAAACAAATATCTTGAAGAACATCCGAGCGCGGGTAAAGATCCAAAAAAAACTGGAGGACCCATTTATCGACGAGTTA
942 ATGGAAAATGGATGCGCGAACTAATACTTTACGACAAAGAAGAAATACGACGAATCTGGCGACAAGCGAACAATGGTGACGACGCG
943 ACTGCTGGTCTCACGCACATGATGATATGGCATTCGAATTTGAACGATGCGACGTATCAGCGAACAAGAGCGCTTGTTCGAACGGG
944 AATGGATCCTCGAATGTGCTCGCTGATGCAAGGATCGACGCTTCCAAGACGATCGGGTGCTGCGGGAGCCGCGGTTAAAGGAGTA
945 GGAACAATGGTAATGGAATTGGTCCGAATGATAAAACGCGGGATAAATGATCGGAACTTTTGGCGAGGTGAAAACGGACGAAAAAC
946 ACGAATTGCGTATGAAAGAATGTGCAACATTCTAAAAGGAAAATTTCAAACCGCGGCGCAAAAAGCAATGATGGATCAAGTTAGAGA
947 GAGCCGTAACCCGGGAAATGCGGAATTTGAAGATCTTACGTTTCTTGCGCGTTCGGCGCTCATTTTGAGAGGTTCGGTAGCTCACA
948 AATCATGTCTGCCGGCATGCGTGTATGGACCAGCGGTAGCGAGCGGTTATGATTTCGAAAGAGAAGGATATTCACTCGTAGGAATA
949 GATCCTTTTCGACTTCTTCAAAACAGCCAAGTATACAGCCTCATTCGACCGAATGAAAATCCGGCACATAAAAGTCAACTAGTCTGG
950 ATGGCGTGTCATTCAGCGGCGTTTGAAGATCTTCGTGTTTTGAGCTTCATAAAAGGAACGAAAGTCCTTCCACGCGGAAAACTTTCG
951 ACGCGAGGCGTTCAAATAGCTTCAAACGAAAACATGGAAACGATGGAATCGAGTACGCTAGAACTTCGAAGTCGTTACTGGGCGAT
952 CAGAACGCGAAGTGGCGGAAATACGAATCAACAAAGAGCGTCGGCCGGACAAATTAGTATTCAACCAACGTTTTCGGTGCAGCGAA
953 ATCTCCCGTTCGACAGAACGACAATCATGGCGGCATTCAACGGAAACACTGAAGGACGAACATCGGACATGCGAACAGAAATTATC
954 CGAATGATGGAAAGTGCGCGACCAGAAGACGTTTCTTTCCAGGGGCGGGGAGTCTTCGAGCTCTCGGACGAAAAGGCAGCGAGC
955 CCGATCGTGCCTTCCTTTGACATGAGTAATGAAGGATCTTATTTCTTCGGAGACAATGCAGAGGAGTACGACAATTAAAGAAAAATA
956 CCCTTGTTTCTACT

957

958 **Table S2. Primers used for mutagenesis and sequencing.**

959

960 **Table S2A.** Primers used for site directed mutagenesis of IAV PR8 segment 8.

Target	F primer	R primer
Seg8 L95S99A	TGACATGACTGCTGAGGAAATGGCAAGG	CCTTGCCATTTCCTCAGCAGTCATGTCA
Seg8 E9697A	ATGACTCTTGCGGCAATGTCAAGGG	CCCTTGACATTGCCGCAAGAGTCAT
Seg8 P107A	CATGCTCATAGCCAAGCAGAAAG	CTTTCTGCTTGGCTATGAGCATG
Seg8 P107K108A	GCTCATAGCCGCGCAGAAAGTGG	CCACTTTCTGCGCGGCTATGAGC
Seg8 P107K108Q109A	CATAGCCGCGGCGAAAGTGGCAG	CTGCCACTTTCGCCGCGGCTATG
Seg8 P107K108Q109K110A	AGCCGCGGCGGCAGTGGCAGGC	GCCTGCCACTGCCGCCGCGGCT

961

962

963 **Table S2B.** Primers for generation of *in vitro* transcription amplicon templates

Target	Forward primer	Reverse primer
T7 Seg1 +sense	TAATACGACTCACTATAGGGAGCGAAAGCAGGTCAATTATATTC	AGTAGAAACAAGGTCGTTTTTAAACTA
T7 Seg1 -sense	AGCGAAAGCAGGTCAATTATATTC	TAATACGACTCACTATAGGGAGTAGAAACAAGGTCGTTTTTAAACTA
T7 Seg5 +sense	TAATACGACTCACTATAGGGAGCAAAAGCAGGGTAGATAATCA	AGTAGAAACAAGGGTATTTTTCTTTAATTGT
T7 Seg5 -sense	AGCAAAAGCAGGGTAGATAATCA	TAATACGACTCACTATAGGGAGTAGAAACAAGGGTATTTTTCTTTAATTGT

964

965 **Table S2C.** Primers for amplification of influenza A virus A/Puerto Rico/8/1934 (PR8) viral transcripts by qPCR.

Target	Forward primer	Reverse primer
PR8 segment 1	GGCTAATGTGCTAATTGGGCAAG	GGCTGTCAGTAAGTATGCTAGAG
PR8 segment 5	ATCATGGCGTCTCAAGGCACCAA	GCTCTGATTTCAGTGGCATTCTGG

966

967 **Table S2D.** Primers for amplification of specific bisulphite converted RNA regions.

Target	Forward primer	Reverse primer
PR8 106-304	CATATAACCATAATCAAAAAATACACATC	GTGATATTATTATTTGGTTTGATTTGG
Chicken rRNA 3163-3307	GTGATTTTTGTTTAGTGTTTTGAATG	AAATCTCATTCATCCATTCATACAC

968

969 **Table S2E.** Primers for overlapping full genome sequencing.

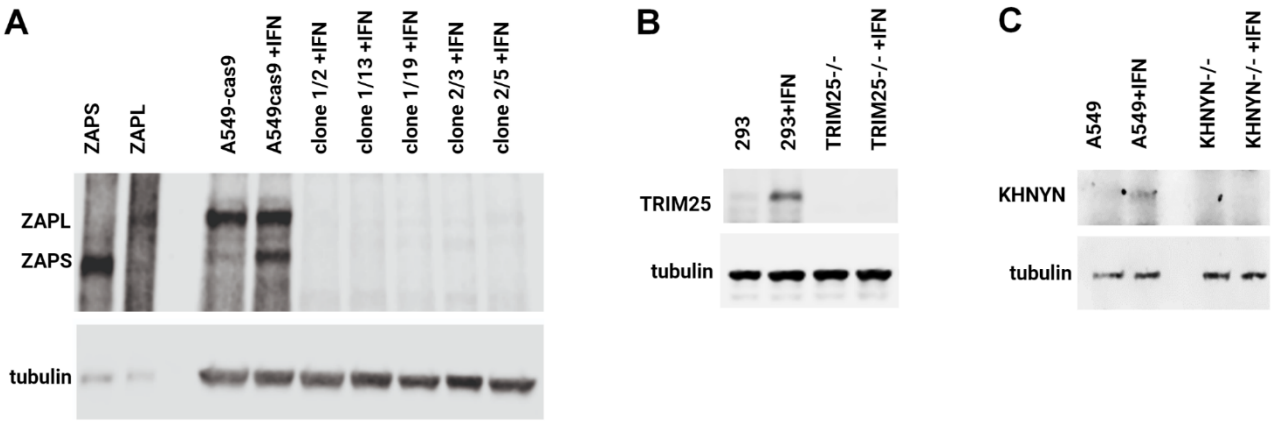
Target	Forward primer	Reverse primer
Seg1 amplicon A	AGCGAAAGCAGGTCAATTAT	CCATGCTTTAGCCTTTCGAC
Seg1 amplicon B	ACATGGTGGAATAGGAATGG	TGCAACACTTCAATGTACAC
Seg1 amplicon C	CAAAACGAGATTCCTCCCAG	TGAAGATTGCCCGTAAGCAC
Seg1 amplicon D	AAGGCTGCAATGGGACTGAG	AGCACTTTCGCATCCTTCTG
Seg1 amplicon E	GCGATTGAATCCTATGCATC	TTGTAGGGTTCTGGGACCAC
Seg1 amplicon F	ATGGATCATCAGAAACTGGG	GGACTCCACTCCAGCTGTGC
Seg1 amplicon G	GCCGCTCCACCAAAGCAAAG	TTCGACACTAATTGATGGCC
Seg2 amplicon A	GAAAGCAGGCAAACCATTTG	TCATAGGTCTGTCGGCCTTG
Seg2 amplicon B	AACGATGGAGGTTGTTCAGC	CCCTGGGGTTGCAATTGCTC
Seg2 amplicon C	CAATGACCAAAGATGCTGAG	CTTGCTCTCAAACATATACCC
Seg2 amplicon D	AATCAGCCCGAATGGTTCAG	GTCGACTCCGGCTTGAATCC
Seg2 amplicon E	TGGTCTTCAATCCTCTGACG	GGGTTTGCTCCCACAGTTTC
Seg2 amplicon F	ATGCCATAGAGGTGACACAC	AATTGCAGCACCTTTGGTAC
Seg2 amplicon G	CCCATTTGTCAGCCATAAAG	GAAGGACAAGCTAAATTCAC
Seg3 amplicon A	AGCGAAAGCAGGTACTGATC	ACTTCTCTCCTTGTTACTCC
Seg3 amplicon B	GGCTGAGAAACCAAAGTTTC	GAATCCATCCACATAGGCTC
Seg3 amplicon C	TTGCCGACCAAAGTCTCCCG	TCATTCTCAATGTCCTGCAG
Seg3 amplicon D	GTTAAACCACACGAAAAGGG	ATGTATTCTGTGGCTCTGCAG
Seg3 amplicon E	ACATTGCAAGCATGAGAAGG	GAGGTTCCATTTGTTCTCAC
Seg3 amplicon F	AAGAAGTGCCATAGGCCAGG	CTCAATTGCTTCATATAGCC
Seg3 amplicon G	ATGATTGAAGCTGAGTCCTC	TTTTGGACAGTATGGATAGC
Seg4 amplicon A	CAACCAAAATGAAGGCAAAC	GGCCATGAGCTTTCTTTGGG
Seg4 amplicon B	GAGCTCAGTGTCATCATTCG	AATATTATTGTGTCTCCGGG
Seg4 amplicon C	TAAGAGATCAAGCTGGGAGG	AGCCTGATCCCTGTTCATTC
Seg4 amplicon D	ACTGGAATGATAGATGGATG	CATTTCTTACACTTTCCATGC
Seg4 amplicon E	GGATTTCTGGACATTTGGAC	ATTCTGCACTGCAAAGATCC
Seg5 amplicon A	CAAAAGCAGGGTAGATAATC	TCACCATTATTAGCTTGGCG

Seg5 amplicon B	GAGAGAACTCATCCTTTATG	TTCCGGCTCTCTCTCACTTG
Seg5 amplicon C	GAAATTTCAAACTGCTGCAC	TTGGAAGCAATTTGAACTCC
Seg5 amplicon D	CCAAGAGGGAAGCTTTCCAC	GTACTCCTCTGCATTGTCTC
Seg6 amplicon A	TGGATCAATCTGTCTGGTAG	ACAGGGCAGCTCATTAAGGCC
Seg6 amplicon B	TTCAAGTGGGACTGTTAAGG	GGTATCAGGGTAACAGGAAC
Seg6 amplicon C	GAGTTGAATGCACCTAATTC	CCGCTATACCCTGACCAATC
Seg6 amplicon D	GTGTATGTTGATGGAGCAAAC	GAACAGACTACTTGTCAATG
Seg7 amplicon A	GATGAGTCTTCTAACCGAGG	TGCCACTTCAGTGGTCACAG
Seg7 amplicon B	CCAAATAACATGGACAAAGC	GCTAGGATGAGTTCCAATGG
Seg7 amplicon C	GCTATGGAGCAAATGGCTGG	TTTTTACTCCAGCTCTATGC
Seg8 amplicon A	CATAATGGATCCAAACACTG	TCTCCAGCCGGTCAAAAATC
Seg8 amplicon B	AGAAGAATCCGATGAGGCAC	TTACTGCTTCTCCAAGCGAATC
Seg8 amplicon C	GACTCTAATATTGCTAAGGG	AAGCTGAAACGAGAAAGTTC
Seg1 CDLR amplicon B	ACATGGTGGAATAGGAATGG	TGCAACACTTCTATGTATAC
Seg1 CDLR amplicon C	AAAGACGAGATTCCTACCCG	TGGAGATTCCCCGTTAGCAC
Seg1 CDLR amplicon D	AAGGCTGCCATGGGGCTGAG	AGCACTTTTGCATCCTTTTG
Seg1 CDLR amplicon E	ACGATTGAACCCTATGCATC	TAGTGGGATTCTGAGACCAC
Seg1 CDLR amplicon F	ATGGATCATCAGAAACTGGG	GGACTCCACTCCAGCTGTGC
Seg1 CpGH amplicon B	ACATGGTGGAATAGGAATGG	TAATACTTCTATGTACACGC
Seg1 CpGH amplicon C	AAAAACGCGATTTCTTCCGG	TGAAGATTTCCCGTCAGAAC
Seg1 CpGH amplicon D	AAGGCCGCGATGGGACTGCG	AGCACTTTCGCGTCTTTCTG
Seg1 CpGH amplicon E	ACGATTGAATCCGATGCATC	TCGTCGGATTTTGCGACCAC
Seg1 CpGH amplicon F	ATGGATCATCCGAAATTGGG	GGACTCCACTCCAGCTGTGC
Seg5 CDLR amplicon A	CAAAAGCAGGGTAGATAATC	TCGTCACCATTATTGGCTTG
Seg5 CDLR amplicon B	GAGAGAACTCATTCTGTATG	TTCCTGCTTTCTCGCACTTG
Seg5 CDLR amplicon C	AAAATTTCAAACTGCTGCTC	TTGGAAGCAATTTGGACTCC
Seg5 CDLR amplicon D	CCTAGAGGAAAGCTCTCAAC	GTACTCCTCTGCATTGTCTC
Seg5 CpGH amplicon A	CAAAAGCAGGGTAGATAATC	TCGTCACCATTGTTCGCTTG
Seg5 CpGH amplicon B	GCGCGAACTAATACTTTACG	TTACGGCTCTCTCTAACTTG
Seg5 CpGH amplicon C	AAAATTTCAAACCGCGGCGC	TTTGAAGCTATTTGAACGCC

Seg5 CpGH amplicon D	CATAAAAGGAACGAAAGTCC	GTACTCCTCTGCATTGTCTC

970

971



972

973 **Fig. S1. Validation of ZAP pathway inhibition assays. A.** ZAP -/- A549 cells were validated by western blotting. **B.** Western blot to

974 confirm the absence of ZAP in CRISPR-cas9 modified ZAP -/- 293T cells. **C.** KHNYN -/- A549 cells were validated by western blotting.

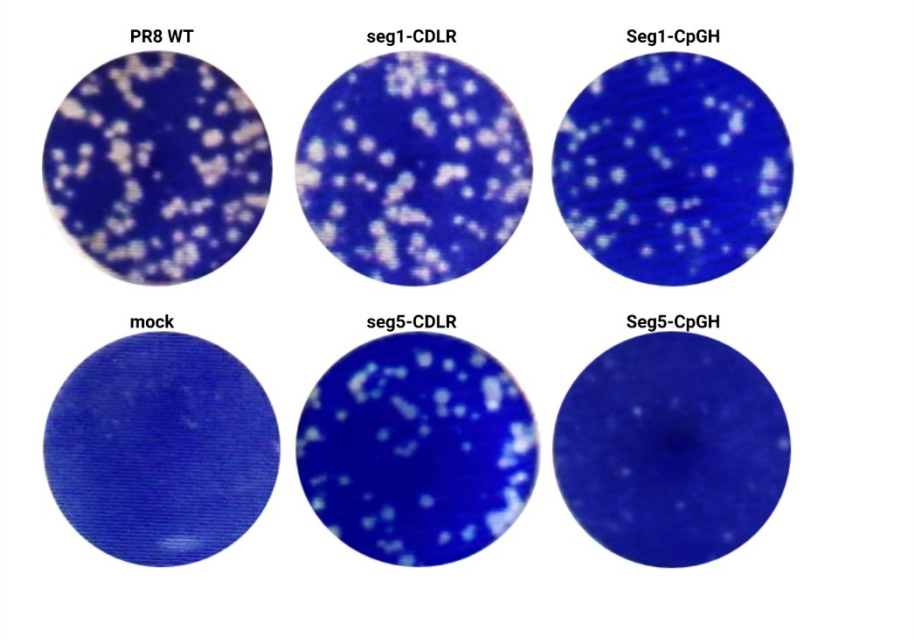
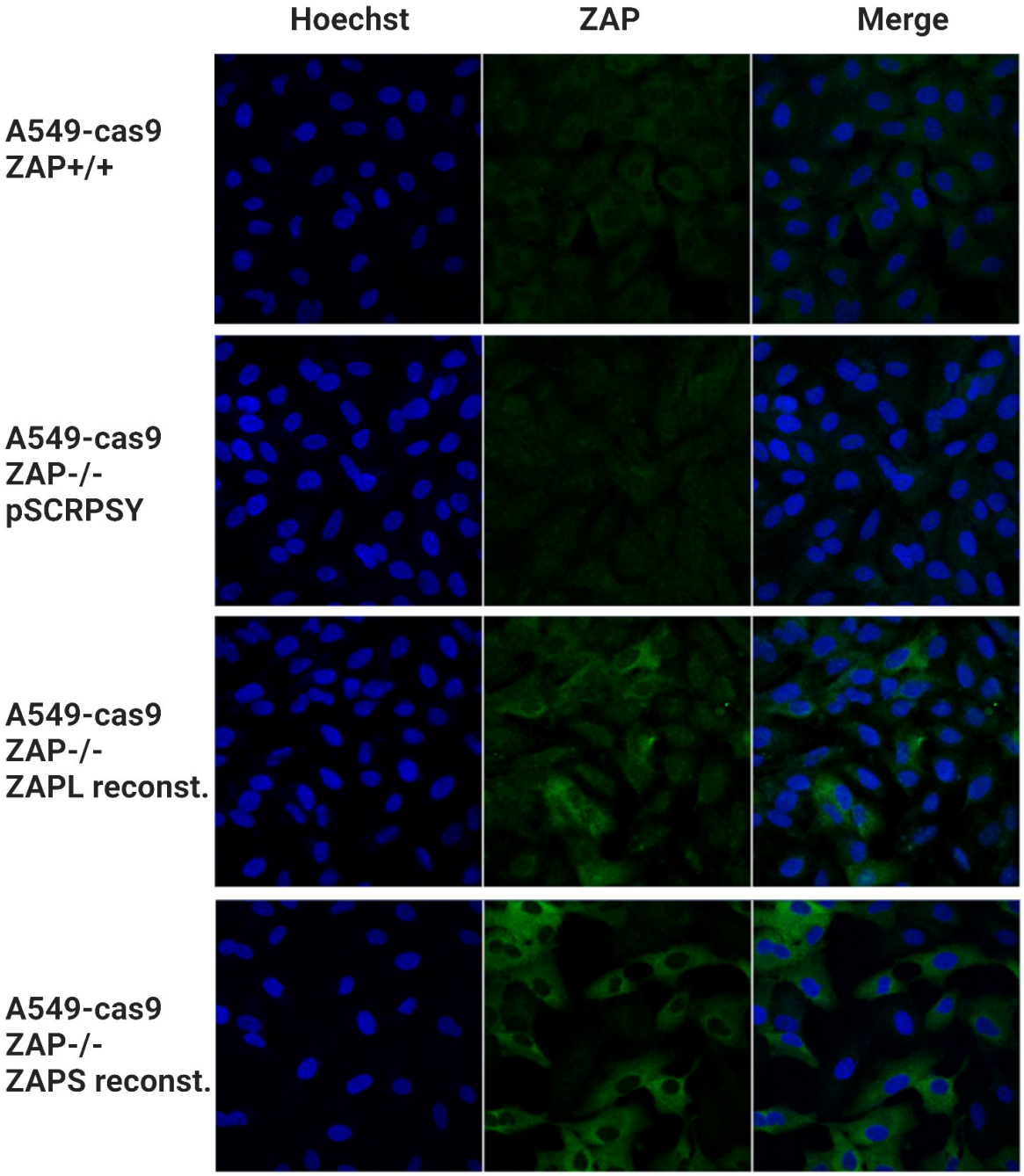
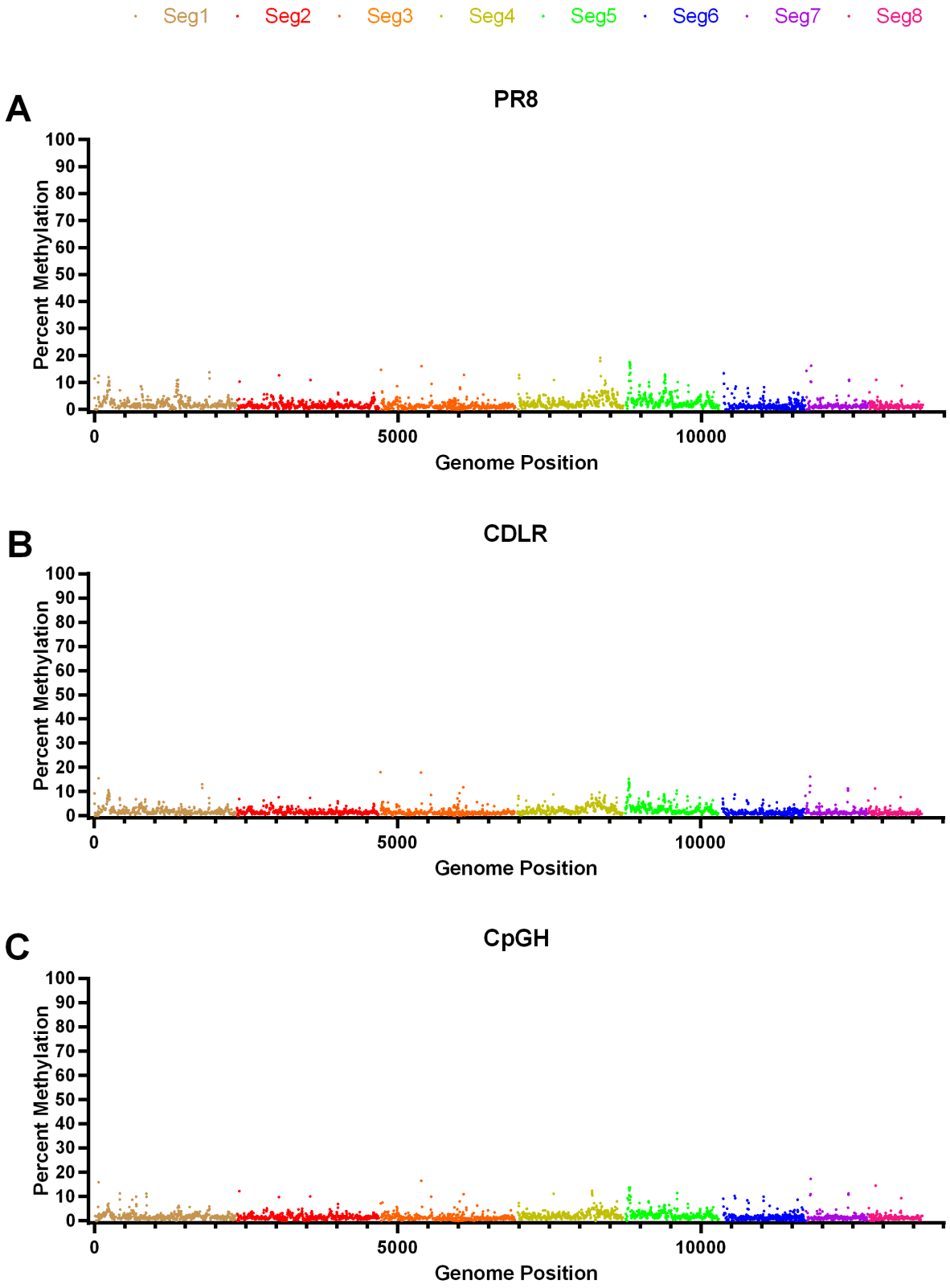


Fig. S2. Plaque phenotype of synonymously recoded viruses. Viruses were rescued in embryonated hens' eggs and titred by plaque assay in MDCK cells. Representative images taken from plaque assays performed at the same time (NB dilutions shown vary due to the variation in titres across the virus panel).



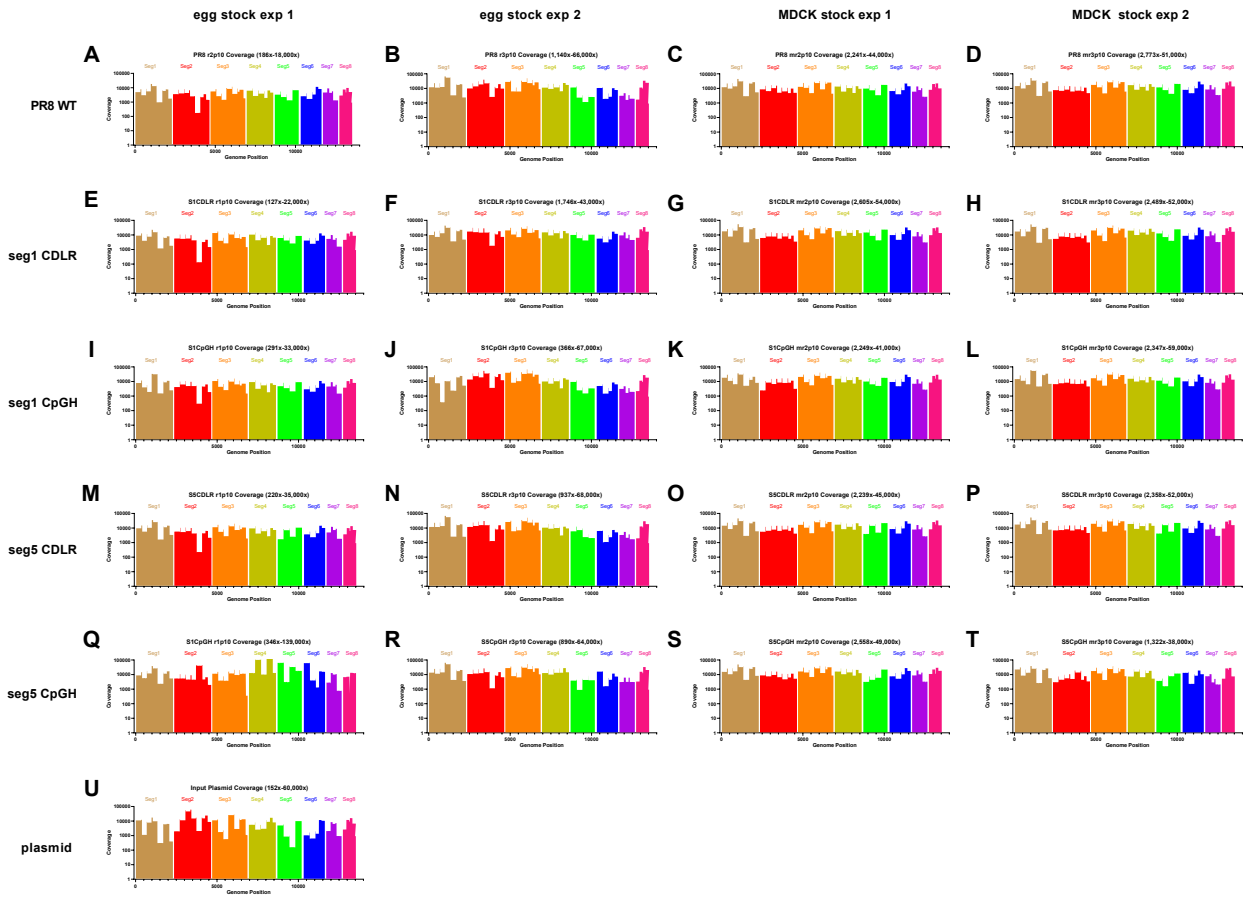
979

980 **Fig. S3. Validation of ZAP expression reconstitution in ZAP-/- A549 cells delivered by lentivirus vector.**



981

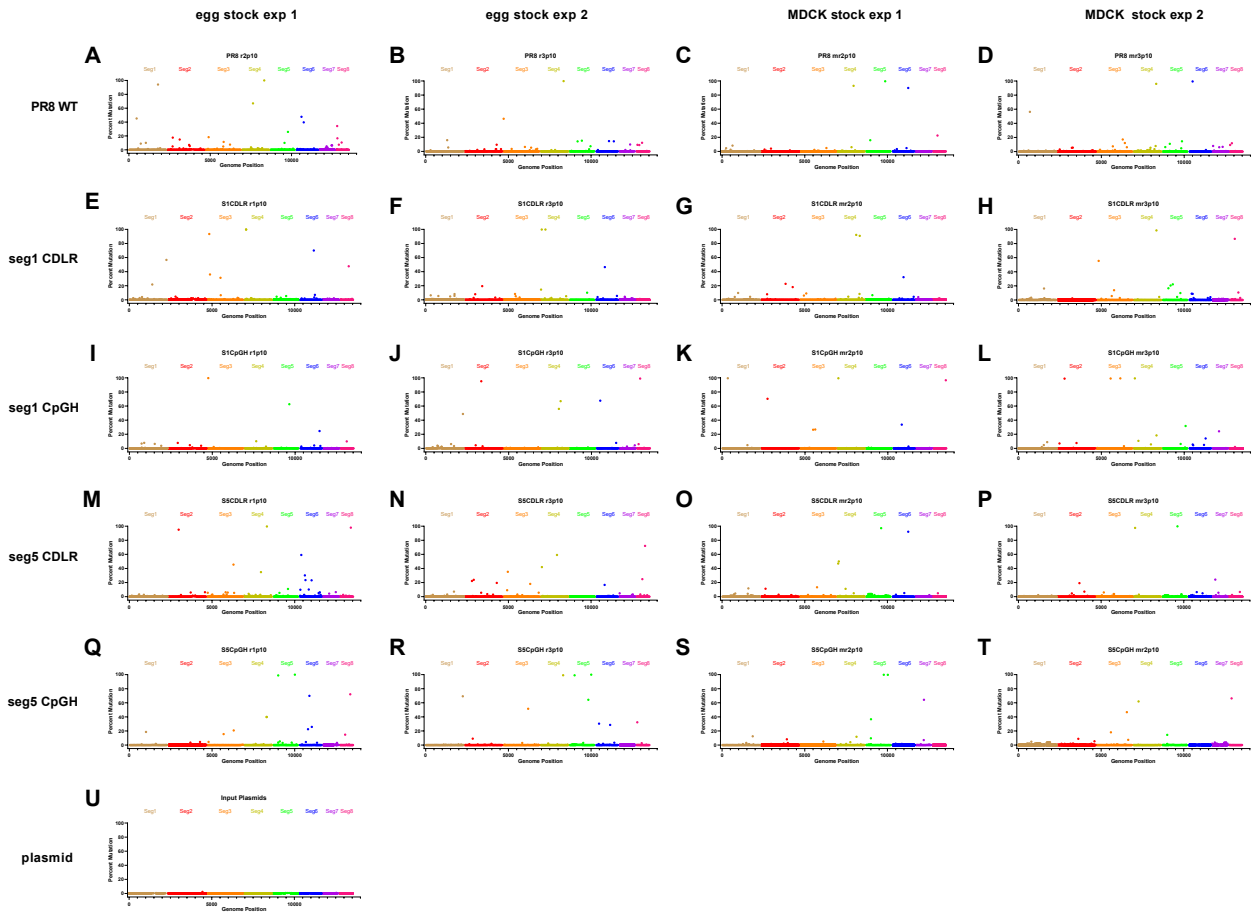
982 **Fig. S4. Bisulphite conversion does not identify any conserved m5c methylation sites in the IAV genome, and CpG enrichment**
983 **does not introduce any such methylation events.** Methylation signal identified across the genome of wildtype PR8 **(A)**, seg1-CDLR
984 **(B)** and seg1-CpGH **(C)**. While the segment 1 sequence was different in A-C, the genome sequence across segments 2-8 was identical,
985 and so for these segments, the data are representative of 3 biological repeats.



986

987 **Fig. S5. Sequence coverage depth for deep sequenced serially passaged viruses. A-D**, PR8 wildtype virus; **E-H**, seg1-CDLR; **I-L**,

988 seg1-CpGH; **M-P**, seg5-CDLR; **Q-T**, seg5-CpGH, **U,** input plasmid.



989

990 **Fig. S6. Full genome mutation traces for all serially passaged viruses. A-D**, PR8 wildtype virus; **E-H**, seg1-CDLR; **I-L**, seg1-CpGH;

991 **M-P**, seg5-CDLR; **Q-T**, seg5-CpGH, **U,** input plasmid.

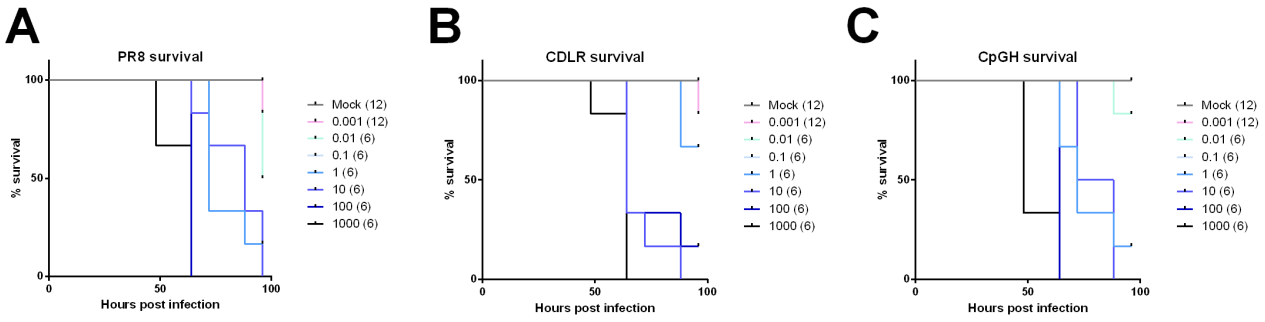


Fig S7. Survival of chicken embryos infected at day 10 post fertilisation. Embryonated hens' eggs at developmental day 10 were inoculated with ten-fold serial dilutions of virus stock (A, PR8; B, seg1-CDLR; C, seg1-CpGH) and survival was monitored by candling for 96 hours. The number of eggs inoculated at each dose is indicated in brackets.

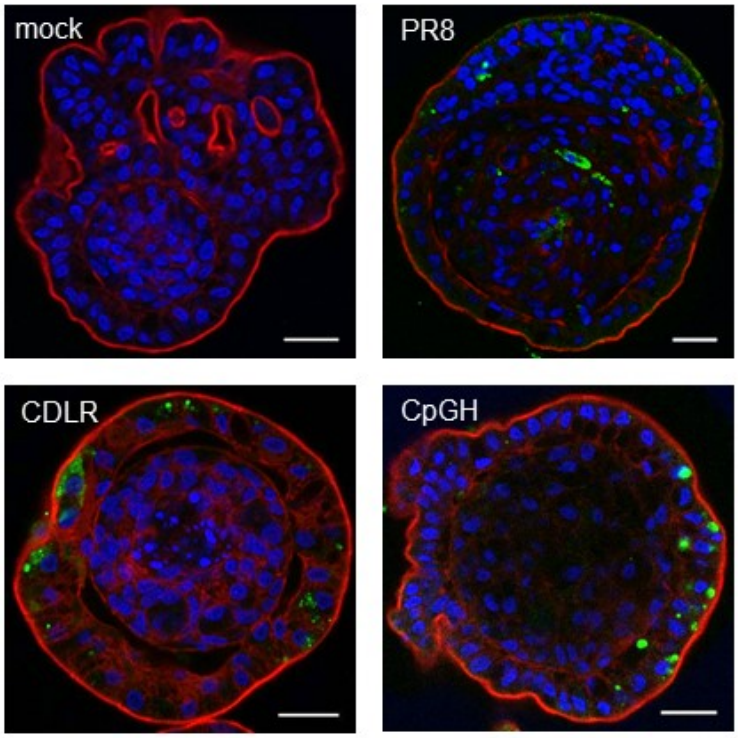
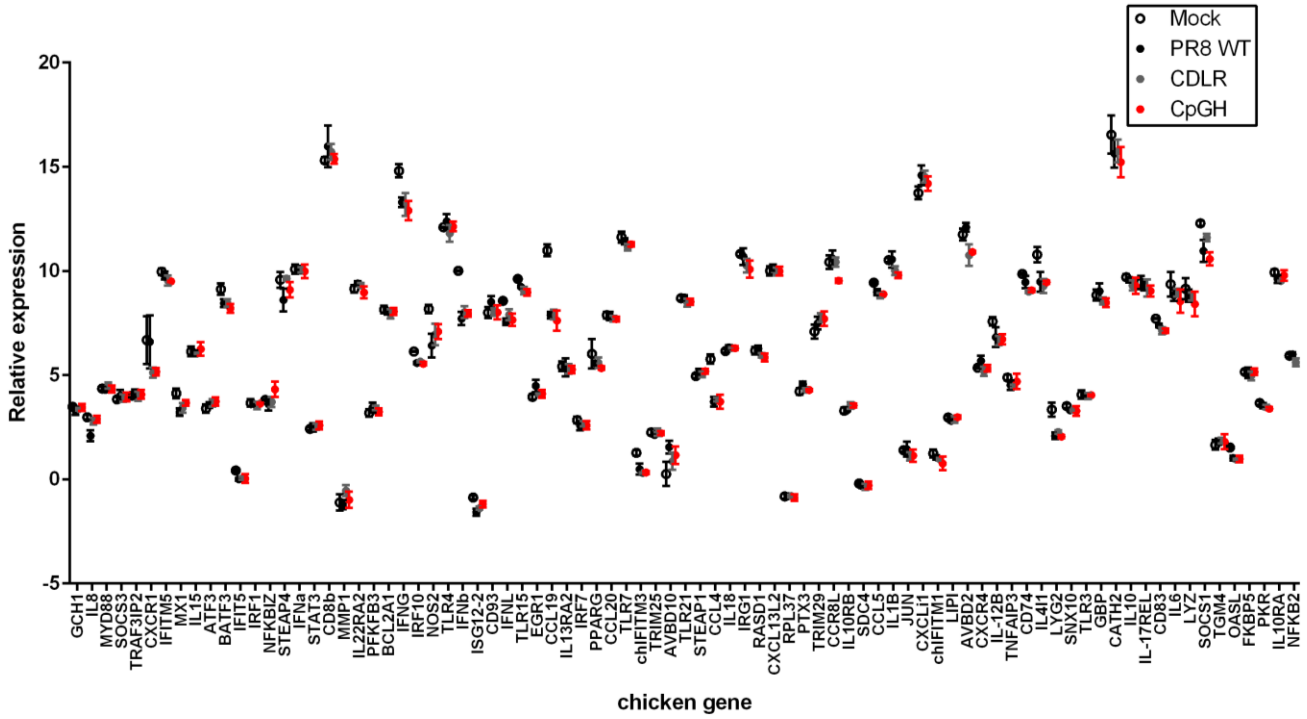


Fig. S8. The seg1 virus panel was used to infect chicken enteroids at 10^5 PFU/enteroid for 48 hours**,** and viral protein production was assessed visually using confocal microscopy. Representative z-axis projections of chicken enteroid whole-mount stained to detect cell nuclei (DAPI, blue), F-actin-expressing brush border (red) and virus nucleoprotein (green); Scale bar: 20 µm.



1000

1001 **Fig. S9. Innate immune gene expression levels of seg1 virus panel in chicken organoids.**

1002 **REFERENCES**

1003 1. Simmonds, P., Xia, W., Baillie, J.K. and McKinnon, K. (2013) Modelling mutational and selection
1004 pressures on dinucleotides in eukaryotic phyla –selection against CpG and UpA in cytoplasmically
1005 expressed RNA and in RNA viruses. *BMC Genomics*, **14**.

1006 2. Rima, B.K. and McFerran, N.V. (1997) Dinucleotide and stop codon frequencies in single-stranded
1007 RNA viruses. *Journal of General Virology*, **78**, 2859-2870.

1008 3. Karlin, S., Doerfler, W. and Cardon, L.R. (1994) Why is CpG suppressed in the genomes of virtually
1009 all small eukaryotic viruses but not in those of large eukaryotic viruses? *Journal of Virology*, **68**,
1010 2889-2897.

1011 4. Gaunt, E.R. and Digard, P. (2021) Compositional biases in RNA viruses: Causes, consequences and
1012 applications. *WIREs RNA*, e1679.

1013 5. Atkinson, N., Witteveldt, J., Evans, D. and Simmonds, P. (2014) The influence of CpG and UpA
1014 dinucleotide frequencies on RNA virus replication and characterization of the innate cellular
1015 pathways underlying virus attenuation and enhanced replication. *Nucleic Acids Research*, **42**, 4527-
1016 4545.

1017 6. Takata, M.A., Gonçalves-Carneiro, D., Zang, T.M., Soll, S.J., York, A., Blanco-Melo, D. and Bieniasz,
1018 P.D. (2017) CG dinucleotide suppression enables antiviral defence targeting non-self RNA. *Nature*,
1019 **550**, 124-127.

1020 7. Li, M.M.H., Aguilar, E.G., Michailidis, E., Pabon, J., Park, P., Wu, X., de Jong, Y.P., Schneider, W.M.,
1021 Molina, H., Rice, C.M. *et al.* (2019) Characterization of Novel Splice Variants of Zinc Finger Antiviral
1022 Protein (ZAP). *Journal of Virology*, **93**, e00715-00719.

1023 8. Chen, S., Xu, Y., Zhang, K., Wang, X., Sun, J., Gao, G. and Liu, Y. (2012) Structure of N-terminal
1024 domain of ZAP indicates how a zinc-finger protein recognizes complex RNA. *Nature structural &*
1025 *molecular biology*, **19**, 430-435.

1026 9. Todorova, T., Bock, F.J. and Chang, P. (2014) PARP13 regulates cellular mRNA post-transcriptionally
1027 and functions as a pro-apoptotic factor by destabilizing TRAILR4 transcript. *Nat Commun*, **5**, 5362.

1028 10. Gao, G., Guo, X. and Goff, S.P. (2002) Inhibition of Retroviral RNA production by ZAP, a CCCH-type
1029 zinc finger protein. **297**, 1703-1706.
1030 11. Kmiec, D., Lista, M.J., Ficarelli, M., Swanson, C.M. and Neil, S.J.D. (2021) S-farnesylation is essential
1031 for antiviral activity of the long ZAP isoform against RNA viruses with diverse replication strategies.
1032 *PLoS Pathog*, **17**, e1009726.
1033 12. Hayakawa, S., Shiratori, S., Yamato, H., Kameyama, T., Kitatsuji, C., Kashigi, F., Goto, S., Kameoka, S.,
1034 Fujikura, D., Yamada, T. *et al.* (2011) ZAPS is a potent stimulator of signaling mediated by the RNA
1035 helicase RIG-I during antiviral responses. *Nature immunology*, **12**, 37-44.
1036 13. Guo, X., Carroll, J.-W.N., MacDonald, M.R., Goff, S.P. and Gao, G. (2004) The Zinc Finger Antiviral
1037 Protein Directly Binds to Specific Viral mRNAs through the CCCH Zinc Finger Motifs. *Journal of*
1038 *Virology*, **78**, 12781.
1039 14. Li, M.M.H., Lau, Z., Cheung, P., Aguilar, E.G., Schneider, W.M., Bozzacco, L., Molina, H., Buehler, E.,
1040 Takaoka, A., Rice, C.M. *et al.* (2017) TRIM25 enhances the antiviral action of zinc-finger antiviral
1041 protein (ZAP). *PLoS Pathogens*, **13**, e1006145.
1042 15. Zheng, X., Wang, X., Tu, F., Wang, Q., Fan, Z. and Gao, G. (2017) TRIM25 is required for the antiviral
1043 activity of zinc finger antiviral protein. *Journal of Virology*, **91**, e00088-00017.
1044 16. Zhu, Y., Chen, G., Lv, F., Wang, X., Ji, X., Xu, Y., Sun, J., Wu, L., Zheng, Y.T. and Gao, G. (2011) Zinc-
1045 finger antiviral protein inhibits HIV-1 infection by selectively targeting multiply spliced viral mRNAs
1046 for degradation. *Proc Natl Acad Sci U S A*, **108**, 15834-15839.
1047 17. Ficarelli, M., Wilson, H., Pedro Galão, R., Mazzon, M., Antzin-Anduetza, I., Marsh, M., Neil, S.J.D.
1048 and Swanson, C.M. (2019) KHNYN is essential for the zinc finger antiviral protein (ZAP) to restrict
1049 HIV-1 containing clustered CpG dinucleotides. *Elife*, **8**, e46767.
1050 18. Ficarelli, M., Antzin-Anduetza, I., Hugh-White, R., Firth, A.E., Sertkaya, H., Wilson, H., Neil, S.J.D.,
1051 Schulz, R. and Swanson, C.M. (2020) CpG dinucleotides inhibit HIV-1 replication through zinc finger
1052 antiviral protein (ZAP)-dependent and -independent mechanisms. *Journal of Virology*, **94**, e01337-
1053 01319.

1054 19. Gonçalves-Carneiro, D., Takata, M.A., Ong, H., Shilton, A. and Bieniasz, P.D. (2021) Origin and
1055 evolution of the zinc finger antiviral protein. *PLoS Pathogens*, **17**, e1009545.
1056 20. Goossens, K.E., Karpala, A.J., Ward, A. and Bean, A.G. (2014) Characterisation of chicken ZAP.
1057 *Developmental and comparative immunology*, **46**, 373-381.
1058 21. Gomez, P.L. and Robinson, J.M. (2018) Vaccine Manufacturing. *Plotkin's Vaccines*, 51-60.e51.
1059 22. Grohskopf, L.A., Alyanak, E., Ferdinands, J.M., Broder, K.R., Blanton, L.H., Talbot, H.K. and Fry, A.M.
1060 (2021) Prevention and Control of Seasonal Influenza with Vaccines: Recommendations of the
1061 Advisory Committee on Immunization Practices, United States, 2021-22 Influenza Season.
1062 *Morbidity and mortality weekly report. Recommendations and reports*, **70**, 1-28.
1063 23. Singanayagam, A., Zambon, M., Lalvani, A. and Barclay, W. (2018) Urgent challenges in
1064 implementing live attenuated influenza vaccine. *The Lancet Infectious Diseases*, **18**, e25-e32.
1065 24. Gould, P.S., Easton, A.J. and Dimmock, N.J. (2017) Live Attenuated Influenza Vaccine contains
1066 Substantial and Unexpected Amounts of Defective Viral Genomic RNA. *Viruses*, **9**.
1067 25. Manini, I., Trombetta, C.M., Lazzeri, G., Pozzi, T., Rossi, S. and Montomoli, E. (2017) Egg-
1068 Independent Influenza Vaccines and Vaccine Candidates. **5**, 18.
1069 26. Milián, E. and Kamen, A.A. (2015) Current and Emerging Cell Culture Manufacturing Technologies
1070 for Influenza Vaccines. *BioMed Research International*, **2015**, 504831.
1071 27. Chen, J.-R., Liu, Y.-M., Tseng, Y.-C. and Ma, C. (2020) Better influenza vaccines: an industry
1072 perspective. *Journal of Biomedical Science*, **27**.
1073 28. Hegde, N.R. (2015) Cell culture-based influenza vaccines: A necessary and indispensable investment
1074 for the future. *Hum Vaccin Immunother*, **11**, 1223-1234.
1075 29. Fodor, E., Devenish, L., Engelhardt, O.G., Palese, P., Brownlee, G.G. and García-Sastre, A. (1999)
1076 Rescue of Influenza A Virus from recombinant DNA. *Journal of Virology*, **73**, 9679-9682.
1077 30. Hoffmann, E., Neumann, G., Kawaoka, Y., Hobom, G. and Webster, R.G. (2000) A DNA transfection
1078 system for generation of influenza A virus from eight plasmids. *Proceedings of the National*
1079 *Academy of Sciences*, **97**, 6108.

1080 31. Gog, J.R., Afonso, E.D.S., Dalton, R.M., Leclercq, I., Tiley, L., Elton, D., von Kirchbach, J.C., Naffakh,
1081 N., Escriou, N. and Digard, P. (2007) Codon conservation in the influenza A virus genome defines
1082 RNA packaging signals. *Nucleic Acids Research*, **35**, 1897-1907.
1083 32. Wise, H., Barbezange, C., Jagger, B.W., Dalton, R.M., Gog, J.R., Curran, M.D., Taubenberger, J.K.,
1084 Anderson, E.C. and Digard, P. (2011) Overlapping signals for translational regulation and packaging
1085 of influenza A virus segment 2. *Nucleic acids research*, **39**, 7775-7790.
1086 33. Dadonaite, B., Gilbertson, B., Knight, M.L., Trifkovic, S., Rockman, S., Laederach, A., Brown, L.E.,
1087 Fodor, E. and Bauer, D.L.V. (2019) The structure of the influenza A virus genome. *Nature*
1088 *Microbiology*, **4**, 1781-1789.
1089 34. Pinto, R.M., Lycett, S., Gaunt, E. and Digard, P. (2020) Accessory Gene Products of Influenza A Virus.
1090 *Cold Spring Harb Perspect Med*, **1**, a038380.
1091 35. Gaunt, E., Wise, H.M., Zhang, H., Lee, L.N., Atkinson, N.J., Nicol, M.Q., Highton, A.J., Klenerman, P.,
1092 Beard, P.M., Dutia, B.M. *et al.* (2016) Elevation of CpG frequencies in influenza A genome
1093 attenuates pathogenicity but enhances host response to infection. *Elife*, **5**, e12735-e12735.
1094 36. Ran, F.A., Hsu, Patrick D., Lin, C.-Y., Gootenberg, Jonathan S., Konermann, S., Trevino, A.E., Scott,
1095 David A., Inoue, A., Matoba, S., Zhang, Y. *et al.* (2013) Double Nicking by RNA-Guided CRISPR Cas9
1096 for Enhanced Genome Editing Specificity. *Cell*, **154**, 1380-1389.
1097 37. Idoko-Akoh, A., Taylor, L., Sang, H.M. and McGrew, M.J. (2018) High fidelity CRISPR/Cas9 increases
1098 precise monoallelic and biallelic editing events in primordial germ cells. *Scientific Reports*, **8**.
1099 38. Shaw, A.E., Rihn, S.J., Mollentze, N., Wickenhagen, A., Stewart, D.G., Orton, R.J., Kuchi, S., Bakshi, S.,
1100 Collados, M.R., Turnbull, M.L. *et al.* (2021) The antiviral state has shaped the CpG composition of
1101 the vertebrate interferome to avoid self-targeting. *PLoS biology*, **19**, e3001352.
1102 39. Choudhury, N.R., Heikel, G., Trus, I., Dos Santos Pinto, R.M., Trubitsyna, M., Gaunt, E., Digard, P.
1103 and Michlewski, G. (2021) TRIM25 inhibits Influenza A infection by destabilising its mRNA and is
1104 redundant for the RIG-I pathway. *bioRxiv*, 2021.2009.2013.460052.
1105 40. Lin, Y.-T., Chiweshe, S., McCormick, D., Raper, A., Wickenhagen, A., DeFillipis, V., Gaunt, E.,
1106 Simmonds, P., Wilson, S.J. and Grey, F. (2020) Human cytomegalovirus evades ZAP detection by

1107 suppressing CpG dinucleotides in the major immediate early genes. *PLoS Pathogens*,
1108 2020.2001.2007.897132.
1109 41. Wise, H.M., Foeglein, A., Sun, J., Dalton, R.M., Patel, S., Howard, W., Anderson, E.C., Barclay, W.S.
1110 and Digard, P. (2009) A complicated message: Identification of a novel PB1-related protein
1111 translated from influenza A virus segment 2 mRNA. *J Virol*, **83**, 8021-8031.
1112 42. Moore, L.D., Le, T. and Fan, G. (2013) DNA Methylation and Its Basic Function.
1113 *Neuropsychopharmacology*, **38**, 23-38.
1114 43. Goll, M.G., Kirpekar, F., Maggert, K.A., Yoder, J.A., Hsieh, C.L., Zhang, X., Golic, K.G., Jacobsen, S.E.
1115 and Bestor, T.H. (2006) Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2.
1116 *Science*, **311**, 395-398.
1117 44. Brzezicha, B., Schmidt, M., Makałowska, I., Jarmołowski, A., Pieńkowska, J. and Szweykowska-
1118 Kulińska, Z. (2006) Identification of human tRNA:m5C methyltransferase catalysing intron-
1119 dependent m5C formation in the first position of the anticodon of the pre-tRNA Leu (CAA). *Nucleic*
1120 *Acids Research*, **34**, 6034-6043.
1121 45. Cooper, D.N. and Krawczak, M. (1989) Cytosine methylation and the fate of CpG dinucleotides in
1122 vertebrate genomes. *Human Genetics*, **83**, 181-188.
1123 46. Khabar, K.S.A. and Young, H.A. (2007) Post-transcriptional control of the interferon system.
1124 *Biochimie*, **89**, 761-769.
1125 47. Atasheva, S., Frolova Elena, I. and Frolov, I. (2014) Interferon-Stimulated Poly(ADP-Ribose)
1126 Polymerases Are Potent Inhibitors of Cellular Translation and Virus Replication. *Journal of Virology*,
1127 **88**, 2116-2130.
1128 48. García-Sastre, A., Egorov, A., Matassov, D., Brandt, S., Levy, D.E., Durbin, J.E., Palese, P. and Muster,
1129 T. (1998) Influenza A Virus Lacking the NS1 Gene Replicates in Interferon-Deficient Systems.
1130 *Virology*, **252**, 324-330.
1131 49. Talon, J., Horvath Curt, M., Polley, R., Basler Christopher, F., Muster, T., Palese, P. and García-
1132 Sastre, A. (2000) Activation of Interferon Regulatory Factor 3 Is Inhibited by the Influenza A Virus
1133 NS1 Protein. *Journal of Virology*, **74**, 7989-7996.

1134 50. Koliopoulos, M.G., Lethier, M., van der Veen, A.G., Haubrich, K., Hennig, J., Kowalinski, E., Stevens,
1135 R.V., Martin, S.R., Reis e Sousa, C., Cusack, S. *et al.* (2018) Molecular mechanism of influenza A NS1-
1136 mediated TRIM25 recognition and inhibition. *Nature communications*, **9**, 1820-1820.
1137 51. Tang, Q., Wang, X. and Gao, G. (2017) The Short Form of the Zinc Finger Antiviral Protein Inhibits
1138 Influenza A Virus Protein Expression and Is Antagonized by the Virus-Encoded NS1. **91**, e01909-
1139 01916.
1140 52. Pereira, C.F., Wise, H.M., Kurian, D., Pinto, R.M., Amorim, M.J., Gill, A.C. and Digard, P. (2018)
1141 Effects of mutations in the effector domain of influenza A virus NS1 protein. *BMC Research Notes*,
1142 **11**, 673.
1143 53. Li, X., Gu, M., Zheng, Q., Gao, R. and Liu, X. (2021) Packaging signal of influenza A virus. *Virology*
1144 *Journal*, **18**, 36.
1145 54. Wise, H., Gaunt, E., Ping, J., Holzer, B., Jasim, S., Lycett, S.J., Murphy, L., Livesey, A., Brown, R.,
1146 Smith, N. *et al.* (2019) An alternative AUG codon that produces an N-terminally extended form of
1147 the influenza A virus NP is a virulence factor for a swine-derived virus. *bioRxiv*, 738427.
1148 55. Csepany, T., Lin, A., Baldick, C.J. and Beemon, K. (1990) Sequence specificity of mRNA N6-adenosine
1149 methyltransferase. *Journal of Biological Chemistry*, **265**, 20117-20122.
1150 56. Saitou, N. and Nei, M. (1986) Polymorphism and evolution of influenza A virus genes. *Molecular*
1151 *Biology and Evolution*, **3**, 57-74.
1152 57. Pérez Rubio, A. and Eiros, J.M. (2018) Cell culture-derived flu vaccine: Present and future. *Hum*
1153 *Vaccin Immunother*, **14**, 1874-1882.
1154 58. Esnault, E., Bonsergent, C., Larcher, T., Bed'hom, B., Vautherot, J.-F., Delaleu, B., Guigand, L.,
1155 Soubieux, D., Marc, D. and Quéré, P. (2011) A novel chicken lung epithelial cell line:
1156 Characterization and response to low pathogenicity avian influenza virus. *Virus Research*, **159**, 32-
1157 42.
1158 59. Hinshaw, V.S., Webster, R.G., Naeve, C.W. and Murphy, B.R. (1983) Altered tissue tropism of
1159 human-avian reassortant influenza viruses. *Virology*, **128**, 260-263.

1160 60. Nash, T.J., Morris, K.M., Mabbott, N.A. and Vervelde, L. (2021) Inside-out chicken enteroids with
1161 leukocyte component as a model to study host–pathogen interactions. *Communications Biology*, **4**,
1162 377.
1163 61. Courtney, D.G., Tsai, K., Bogerd, H.P., Kennedy, E.M., Law, B.A., Emery, A., Swanstrom, R., Holley,
1164 C.L. and Cullen, B.R. (2019) Epitranscriptomic Addition of m5C to HIV-1 Transcripts Regulates Viral
1165 Gene Expression. *Cell Host & Microbe*, **26**, 217-227.e216.
1166 62. Dev, R.R., Ganji, R., Singh, S.P., Mahalingam, S., Banerjee, S. and Khosla, S. (2017) Cytosine
1167 methylation by DNMT2 facilitates stability and survival of HIV-1 RNA in the host cell during
1168 infection. *Biochemical Journal*, **474**, 2009-2026.
1169 63. Magor, K.E., Miranzo Navarro, D., Barber, M.R.W., Petkau, K., Fleming-Canepa, X., Blyth, G.A.D. and
1170 Blaine, A.H. (2013) Defense genes missing from the flight division. *Developmental and comparative*
1171 *immunology*, **41**, 377-388.
1172 64. de Wit, E., Spronken, M.I., Bestebroer, T.M., Rimmelzwaan, G.F., Osterhaus, A.D. and Fouchier, R.A.
1173 (2004) Efficient generation and growth of influenza virus A/PR/8/34 from eight cDNA fragments.
1174 *Virus Res*, **103**, 155-161.
1175 65. Simmonds, P. (2012) SSE: a nucleotide and amino acid sequence analysis platform. *BMC Research*
1176 *Notes*, **5**, 50.
1177 66. Yamayoshi, S., Watanabe, M., Goto, H. and Kawaoka, Y. (2016) Identification of a Novel Viral
1178 Protein Expressed from the PB2 Segment of Influenza A Virus. *J Virol*, **90**, 444-456.
1179 67. Hutchinson E. C., Curran Martin, D., Read Eliot, K., Gog Julia, R. and Digard, P. (2008) Mutational
1180 Analysis of cis-Acting RNA Signals in Segment 7 of Influenza A Virus. *Journal of Virology*, **82**, 11869-
1181 11879.
1182 68. Wise Helen, M., Foeglein, A., Sun, J., Dalton Rosa, M., Patel, S., Howard, W., Anderson Emma, C.,
1183 Barclay Wendy, S. and Digard, P. (2009) A Complicated Message: Identification of a Novel PB1-
1184 Related Protein Translated from Influenza A Virus Segment 2 mRNA. *Journal of Virology*, **83**, 8021-
1185 8031.

1186 69. Hirst , G.K. (1942) The quantitative determination of influenza virus and antibodies by means of red
1187 cell agglutination. *Journal of Experimental Medicine*, **75**, 49-64.
1188 70. Cold Spring Harbor. (2006) Laemmli sample buffer (2X). pdb.rec10424.
1189 71. Carrasco, M., Amorim, M.J. and Digard, P. (2004) Lipid raft-dependent targeting of the influenza A
1190 virus nucleoprotein to the apical plasma membrane. *Traffic (Copenhagen, Denmark)*, **5**, 979-992.
1191 72. Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., Chilton, J., Clements, D.,
1192 Coraor, N., Grüning, B.A. *et al.* (2018) The Galaxy platform for accessible, reproducible and
1193 collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, **46**, W537-W544.
1194 73. Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient
1195 alignment of short DNA sequences to the human genome. *Genome Biology*, **10**, R25.
1196 74. Hoffmann, E., Stech, J., Guan, Y., Webster, R.G. and Perez, D.R. (2001) Universal primer set for the
1197 full-length amplification of all influenza A viruses. *Arch Virol*, **146**, 2275-2289.
1198 75. Grubaugh, N.D., Gangavarapu, K., Quick, J., Matteson, N.L., De Jesus, J.G., Main, B.J., Tan, A.L., Paul,
1199 L.M., Brackney, D.E., Grewal, S. *et al.* (2019) An amplicon-based sequencing framework for
1200 accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biology*, **20**, 8.
1201 76. Borowska, D., Kuo, R., Bailey, R.A., Watson, K.A., Kaiser, P., Vervelde, L. and Stevens, M.P. (2019)
1202 Highly multiplexed quantitative PCR-based platform for evaluation of chicken immune responses.
1203 *PLoS ONE*, **14**, e0225658.

1204