Mini viral RNAs act as innate immune agonists during influenza virus infection

Aartjan J.W. te Velthuis[1,2,§,*], Joshua C. Long[1,§], David L.V. Bauer[1,§], Rebecca L.Y. Fan[3], Hui-Ling Yen[3], Jane Sharps[1], Jurre Y. Siegers[4], Marian J. Killip[1,5], Hollie French[2], Maria José Oliva-Martín[1], Richard E. Randall[5], Emmie de Wit[6], Debby van Riel[4], Leo L.M. Poon[3], Ervin Fodor[1,*]

[1] Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, United Kingdom.

[2] Division of Virology, Department of Pathology, University of Cambridge, Addenbrooke's Hospital, Hills Road, Cambridge CB2 0QQ, United Kingdom.

[3] School of Public Health, University of Hong Kong, 21 Sassoon Road, Pokfulam, Hong Kong SAR, China.

[4] Department of Viroscience, Erasmus Medical Centre, 3015 CN, Rotterdam, The Netherlands.

[5] Biomedical Sciences Research Complex, University of St Andrews, North Haugh, St Andrews, Fife KY16 9ST, United Kingdom.

[6] Laboratory of Virology, Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, Montana, USA

§ These authors contributed equally to this work.

* Corresponding authors.

1

Influenza A virus infection usually causes a mild to moderately severe respiratory disease in humans. However, infection with the 1918 H1N1 pandemic or highly pathogenic avian influenza viruses (HPAIV) of the H5N1 subtype, can lead to viral pneumonia, systemic disease and death. The molecular processes that determine the outcome of influenza virus infection are multifactorial and involve a complex interplay between host, viral, and bacterial factors[1]. However, it is generally accepted that a strong innate immune dysregulation known as 'cytokine storm' contributes to the pathology of pandemic and avian influenza virus infections[2-4]. The RNA sensor Retinoic acid-inducible gene I (RIG-I) plays an important role in sensing viral infection and initiating a signalling cascade that leads to interferon (IFN) expression[5]. Here we show that short aberrant RNAs (mini viral RNAs; mvRNAs), produced by the viral RNA polymerase during the replication of the viral RNA genome, bind and activate the intracellular pathogen sensor RIG-I, and lead to the expression of interferon-β. We find that erroneous polymerase activity, dysregulation of viral RNA replication, or the presence of avian-specific amino acids underlie mvRNA generation and cytokine expression in mammalian cells and propose an intramolecular copy-choice mechanism for mvRNA generation. By deep-sequencing RNA samples from lungs of ferrets infected with influenza viruses we show that mvRNAs are generated during infection of animal models. We propose that mvRNAs act as main agonists of RIG-I during influenza virus infection and the ability of influenza virus strains to generate mvRNAs should be considered when assessing their virulence potential.

The negative sense viral RNA (vRNA) genome segments of influenza A viruses, as well as the complementary RNA (cRNA) replicative intermediates, contain 5′ triphosphates and partially complementary 5′ and 3′ termini that serve as the viral promoter for replication and transcription of the viral RNA genome[6]. RIG-I has been shown to bind and be activated by the dsRNA structure formed by the termini of influenza virus RNAs[7,8]. However, it remains unclear how RIG-I gains access to this dsRNA structure. Both vRNA and cRNA are assembled into ribonucleoprotein complexes (vRNP and cRNP, respectively) in which the viral RNA polymerase, a heterotrimeric complex of the viral proteins PB1, PB2 and PA, associates with the partially complementary termini, while the rest of the RNA is bound by oligomeric nucleoprotein (NP)[6] (Fig. 1a). The tight binding of the 5′ and 3′ termini of vRNA and

60   cRNA by the RNA polymerase[9] is likely to preclude an interaction with RIG-I.

61   Moreover, it has been demonstrated that IFN expression is triggered only in a fraction

62   of influenza virus infected cells[10,11], suggesting that influenza viruses efficiently hide

63   their genome segments during infection by replicating them in the context of RNPs[11].

64   This led to the proposal that an aberrant RNA replication product might be binding to

65   RIG-I and triggering IFN expression[12]. Indeed, the influenza virus polymerase is known

66   to generate defective interfering (DI) RNAs, which are ≥178 nt long subgenomic RNAs

67   generated during high multiplicity infections[13], and small viral RNAs (svRNAs), which

68   are 22-27 nt long and correspond to the 5′ end of vRNA segments. However, svRNAs

69   have been shown not to be involved in the induction of antiviral cellular defences[14] and

70   DI RNAs assemble into RNP structures (Fig. 1a), as demonstrated for a 248 nt long DI

71   RNA[15], potentially precluding their interaction with RIG-I. Therefore, it remains

72   unclear what kind of viral RNA species is recognised by RIG-I (Fig. 1A) and why

73   different influenza virus strains trigger dramatically different levels of IFN

74   expression[2,3,16].

75   Engineered viral RNAs shorter than 149 nt but containing both the 5′ and 3′

76   termini of vRNAs can be transcribed and replicated in cells by the viral polymerase in

77   the absence of NP[17], suggesting that they do not form canonical RNP structures. We call

78   these short viral RNAs mvRNAs (Fig. 1a). To investigate which class of viral RNA is

79   responsible for triggering IFN expression, we expressed a full-length segment 4

80   (hemagglutinin encoding) vRNA (1775 nt long) or its truncated versions, a 245 nt long

81   DI RNA and 77 nt long mvRNA, in HEK 293T cells together with viral polymerase,

82   NP and a luciferase reporter to measure the activation of the IFN-β promoter (Fig. 1b).

83   We found that the expression of mvRNAs induced significantly higher IFN expression

84   than full-length vRNA or DI RNA, comparable to the levels induced by transfection of

85   2 μg of poly(I:C), a known activator of IFN expression[18]. Similar results were obtained

86   with segment 5 and 6 (NP and neuraminidase encoding, respectively) vRNAs and their

87   truncated DI RNA and mvRNA versions (Fig. 1b). To determine the optimal mvRNA

88   length that triggers activation of the IFN-β promoter, we expressed 47 to 246 nt long

89   vRNAs derived from segment 5 together with viral polymerase and NP and measured

90   the activity of the IFN-β promoter. We found that the replication of 56 to 125 nt long

91   mvRNAs resulted in significantly higher IFN-β promoter activity than the replication

92   of RNAs shorter than 56 nt or longer than 125 nt (Fig. 1c, Fig. S1a and b).

93   To address whether these engineered short mvRNAs triggered IFN expression

94   via RIG-I, we co-expressed viral RNAs with polymerase and NP in HEK 293T RIG-I

95   knockout (RIG-I -/-) or control (RIG-I +/+) cells engineered to express luciferase in

96   response to the activation of the IFN-β promoter. We found that 56 to 125-nt long

97   mvRNAs induced only background levels of luciferase in RIG-I -/- cells, even though

98   the transfection of a plasmid expressing RIG-I or 2 μg of poly(I:C) resulted in

99   significant activation of the IFN-β promoter (Fig. 1d, Fig. S1c). By contrast, significant

100   levels of luciferase activity were detected in RIG-I +/+ cells (Fig. 1d). mvRNAs of 56

101   to 125 nt induced the strongest activation of the IFN-β promoter, in agreement with the

102   data above (Fig. 1c and d, Fig. S1c and d). To address whether mvRNAs trigger the

103   activation of IFN-β expression through binding to RIG-I, we immunoprecipitated myc-

104   RIG-I from cells expressing RNAs of 47 to 583 nt and analysed the amounts of co-

105   immunoprecipiated RNAs. We observed that mvRNAs of 56 to 125 nt were specifically

106   enriched in RIG-I immunoprecipitates (Fig. 1e, Fig. S1e). No mvRNAs were detected

107   in mouse myc-EGF immunoprecipitates used as a negative control (Fig. 1e). To test if

108   mvRNAs also activate RIG-I, we incubated purified myc-RIG-I (Fig. S1f) with an *in*

109   *vitro* transcribed and gel purified 76 nt mvRNA and measured $^{32}P_i$ release in an ATPase

110   assay. We found that a triphosphorylated 76 nt mvRNA induced higher levels of

111   ATPase activity than a dephosphorylated 76 nt mvRNA, while no ATPase activity was

112   observed when we incubated a RIG-I mutant with the triphosphorylated 76 nt mvRNA

113   (Fig. S1g). Overall, these results demonstrate that mvRNAs longer than 47 and shorter

114   than 125 nt are bound by RIG-I, which results in RIG-I activation and the induction of

115   IFN-β expression. This is in agreement with findings that reconstitution of full-length

116   influenza virus vRNPs leads to only low levels of IFN expression unless the cells are

117   pre-treated with IFN[19] and the hypothesis that aberrant replication products trigger the

118   IFN induction cascade[12].

119   We next asked whether mvRNAs are made during influenza virus infection. We

120   infected HEK 293T cells with influenza A/WSN/33 (H1N1) (abbreviated as WSN) and

121   analysed viral RNAs by RT-PCR of segment 1 (encoding PB2), RT-PCR of all

122   segments using universal primers (Fig. S2a), or deep-sequencing of the total small RNA

123   fraction (RNAs 17 to 200 nt in length) (Fig. S2b). We found only very low levels of

124   mvRNAs and, consistently, observed no significant IFN expression (Fig. 2a and b). We

125   hypothesised that mvRNAs are only generated as a consequence of dysregulated viral

RNA replication. To test this, we overexpressed viral RNA polymerase from plasmids prior to infection to generate an imbalance between polymerase and NP levels, which is known to induce innate immune signalling[20]. As shown in Fig. 2a and b, under this condition we found significantly higher levels of mvRNAs and IFN expression, while simultaneous overexpression of NP and polymerase reduced mvRNA and IFN production (Fig. 2a). We verified the identity of mvRNAs using gel isolation and Sanger sequencing (Fig. S2c) as well as deep sequencing (Fig. 2b). We found that the majority of mvRNAs were derived from the PB1-, HA-, NP- and NA-encoding vRNA segments (Fig. 2c) and that mvRNAs had a size distribution with a peak around 55 to 64 nt (Fig. 2d, Fig. S2d). In addition to mvRNAs, we also identified complementary mini viral RNAs (mcRNAs).

Analysis of mvRNA sequences suggests that mvRNAs are generated via an intramolecular copy-choice mechanism that tolerates 3' mismatches or skipped bases (Fig. 2e and f). As shown in Fig. 2g, the generation of mvRNAs can be explained by the separation of the template and nascent product RNAs by backtracking[21], followed by template translocation until base pairing between template and nascent product RNA is re-established (Fig. 2f and g). This process may be induced by an imbalance between viral polymerase and NP levels (Fig. 2a, Fig. S3a). In light of these results, it is important to note that the existence of mvRNAs has likely been overlooked till now, because i) RNA isolation protocols vary in their capacity to recover small RNAs, ii) RT-PCR products from mvRNAs form a diffuse fast-migrating band on standard agarose gels that may be mistaken for primer-dimers, iii) conventional RNA deep sequencing protocols discard short library fragments, and iv) standard ligation-based deep sequencing protocols for small RNAs require a 5′ monophosphate and do not detect viral transcripts with a 5′-triphosphate group.

In humans, infection with the 1918 H1N1 pandemic virus or H5N1 HPAIV lead to strong innate immune activation[2,3,16]. To address whether mvRNAs could contribute to this phenomenon, we investigated the replication of a 246 nt RNA by the polymerase of these viruses. We found that the polymerases of the highly virulent A/Brevig Mission/1/18 (H1N1) (abbreviated as BM18) pandemic virus and the A/duck/Fujian/01/02 (H5N1) (abbreviated as FJ02) HPAIV generated higher levels of mvRNAs than the polymerases of WSN and A/Northern Territory/60/68 (H3N2) (abbreviated as NT60) viruses, even in the presence of high NP concentrations (Fig. 3a). No mvRNAs were observed in a control with an inactive WSN polymerase that

160   had two point mutations in the polymerase active site (PB1a). Using gel isolation and

161   Sanger sequencing we confirmed that the mvRNAs produced by the BM18 polymerase

162   were similar to the WSN mvRNAs (Fig. S3b). Isolation of total RNA from cells

163   expressing polymerase of the BM18 or FJ02 virus and its subsequent transfection into

164   HEK 293T cells resulted in significantly higher IFN-β promoter activity compared to

165   when RNA from cells expressing WSN, NT60, or active site mutant WSN PB1a

166   polymerase was transfected (Fig. 3a).

167   The identification of mismatches and base-skipping during the generation of

168   mvRNAs (see Fig. 2e) suggests that mvRNA production might be dependent on

169   polymerase fidelity. To investigate this further, we introduced a V43I mutation, which

170   has been shown to confer high-fidelity on an H5N1 influenza virus polymerase[22], into

171   the PB1 subunit of the BM18 polymerase (BM18[hf]). We found that mvRNA levels were

172   significantly reduced in the presence of BM18[hf], with a corresponding reduction in IFN-

173   β promoter activity (Fig. 3a). Together, the observations in Fig. 2a, 3a and Fig. S3a

174   suggest that dysregulation of viral RNA replication, e.g. by limiting NP availability,

175   and replication by highly pathogenic avian influenza virus polymerases in mammalian

176   cells generates mvRNAs by employing an error-prone copy-choice mechanism, such

177   as proposed for recombination in positive-strand RNA viruses[23].

178   Having observed a significant difference in mvRNA production by the WSN

179   and BM18 polymerases, we next asked whether a particular subunit is the determinant

180   of mvRNA production. Therefore, we replaced individual polymerase subunits of the

181   BM18 polymerase with subunits of the WSN polymerase in the 246 nt RNA replication

182   assay. We found that particularly replacement of the BM18 PB2 subunit with the WSN

183   PB2 subunit eliminated the generation of mvRNAs (Fig. S4a). Interestingly, the BM18

184   influenza PB2 subunit has been linked to the enhancement of both the kinetics and the

185   magnitude of the host response to viral infection, leading to the induction of strong

186   inflammatory responses with increased cellular infiltration in the lungs of infected

187   mice[24]. To identify PB2 amino acid residues involved in the formation of mvRNAs, we

188   aligned the BM18, WSN, NT60 and FJ02 PB2 sequences and found four amino acid

189   positions that distinguish the BM18 and FJ02 polymerases from the WSN and NT60

190   polymerases: 9 (D→N), 64 (M→T), 81 (T→M), and 661 (A→T) (Fig. S4a). Each of

191   these amino acids has been implicated in avian to mammalian host adaptation[25] and,

192   interestingly, all three N-terminal PB2 adaptive amino acids map to the template exit

6

193   channel of the RNA polymerase heterotrimer (Fig. 3b) [6]. We generated single mutations

194   N9D, T64M, M81T, and double mutations N9D+T64M and N9D+M81T in the PB2

195   subunit of the WSN polymerase and found that mutants N9D and M81T, and the double

196   mutants N9D+T64M and N9D+M81T, significantly increased mvRNA formation (Fig.

197   3c) and IFN-β promoter activity (Fig. 3c and d). However, the levels of mvRNAs

198   generated by these mutants did not reach the levels generated by the BM18 polymerase

199   indicating that further amino acids are involved in determining the ability of a

200   polymerase to produce mvRNAs. In line with our observations, WSN viruses that

201   contain a PB2 N9D substitution or other PB2 mutations near the template exit channel

202   have been reported to induce higher IFN-β expression than wild-type WSN[26,27].

203   To address whether mvRNAs form during infection of mammalian cells, we

204   infected A549 cells with WSN, the highly pathogenic avian strain A/Vietnam/1203/04

205   (H5N1) (abbreviated as VN04), and the VN04 virus with the PB1 V43I high-fidelity

206   mutation (abbreviated as VN04[hf]). PAGE and deep sequencing analysis using RT-PCR

207   universal primers (Fig. S2a) showed that infections with VN04 resulted in high levels

208   of mvRNAs, while WSN infections produced only very low levels (Fig. 4a). Infections

209   with VN04[hf] resulted in significantly reduced mvRNA levels compared to the wild-type

210   VN04 virus. These results demonstrate that mvRNAs are formed during influenza virus

211   infection of lung epithelial cells and that polymerase fidelity is an important

212   determinant of mvRNA formation (Fig. 4a).

213   To investigate whether there is a link between mvRNA production and virus-

214   induced innate immune responses we performed RNAseq of cells infected with VN04

215   and VN04[hf] viruses, which differ only at a single amino acid residue, and examined

216   which genes were differentially expressed in response to mvRNA levels. Despite

217   significantly different mvRNA levels produced by VN04 and VN04[hf], viral mRNA

218   levels were similar (Data S1), in agreement with previous findings that the V43I

219   mutation has only a marginal effect on virus replication[22]. Gene Ontology (GO)

220   analysis (Fig. 4b) showed that basic cellular functions (e.g. translation and

221   transcription) were significantly compromised in VN04 infection relative to VN04[hf],

222   consistent with a greater level of cell death, which is known to exacerbate

223   inflammation[28]. In addition, we observed that genes associated with innate immune

224   responses showed a significant increase in expression in response to higher mvRNA

225   levels (Data S1, Fig. 4b). Overall, these observations are indicative of a link between

erroneous polymerase activity, mvRNA synthesis, and innate immune activation and the induction of cell death. Furthermore, as VN04 exhibited a 10-fold higher lethality compared to V04[hf] in mice[22], our data also suggest a link between mvRNA levels and virulence.

To address whether mvRNAs are also produced in infection of animal models, we investigated mvRNA generation in ferrets, which are regarded as the gold-standard model for influenza virus infection studies. Specifically, we analysed RNA samples from ferret lungs one and three days after infection with highly pathogenic avian A/Indonesia/5/2005 (H5N1) (abbreviated as IN05), 2009 swine-origin pandemic A/Netherlands/602/2009 (H1N1) (abbreviated as NL09) or the BM18 pandemic virus[29,30]. mvRNAs were present in all infected lung samples one day after infection, with mvRNA levels particularly high in the BM18 infected ferret lungs (Fig. 4c, Fig. S5). RNAseq and differential gene expression analysis, followed by GO analysis on the ferret lung samples taken one and three days post infection, showed an up-regulation of apoptosis and innate immune responses as function of mvRNA level, independently of viral titre or the day post infection (Fig. 4d).

In summary, we identify mvRNAs, a novel class of influenza virus RNAs, that act as the main agonists of the pathogen recognition receptor RIG-I during influenza virus infection (Fig. 4e). mvRNAs are produced as a result of aberrant replication of the viral RNA genome by the viral RNA polymerase. Polymerase fidelity and host-specific amino acids are determinants of the ability of the viral polymerase to produce mvRNAs, which are distinct from DI RNAs and full length viral RNA segments in that they can be efficiently replicated in the absence of NP and do not form canonical RNPs[17]. These features of mvRNAs are likely to be critical for their preferential recognition by RIG-I over DI RNAs and full length RNA segments. We further demonstrate that mvRNA production is linked to increased cytokine expression and cell death. Since high levels of mvRNAs lead to more innate immune activation and cell death, our observations strongly suggest that mvRNAs are a contributing factor to influenza virus virulence. We speculate that production of high levels of mvRNAs by the polymerases of the 1918 pandemic and highly pathogenic avian influenza viruses and the resulting increased innate immune activation contribute to the cytokine storm phenomenon underlying the high virulence of these viral strains. The effects of mvRNAs are likely modulated by viral factors, such as the immunomodulatory NS1 and PB1-F2 proteins[12]

8

259   (Fig. 4e). Further studies are required to assess mvRNA levels generated by various

260   influenza virus strains, including seasonal strains, and their effect on virulence.

261

262

263   **Methods**

264   **Ethics and biosafety**

265   All work with highly pathogenic H5N1 viruses in A549 cells was conducted in the

266   Biosafety Level-3 laboratory at the LKS Faculty of Medicine, The University of Hong

267   Kong, under guidelines and ethics approved by the Committee on the Use of Live

268   Animals in Teaching and Research (CULATR). Ferret experiments with IN05 and

269   NL09 were described previously[29] and conducted in the Biosafety Level-3 laboratory

270   of the Erasmus Medical Centre in compliance with European guidelines (EU directive

271   on animal testing 86/609/EEC) and Dutch legislation (Experiments on Animals Act,

272   1997), after approval by the independent animal experimentation ethical review

273   committee of the Netherlands Vaccine Institute (permit number 200900201). Ferret

274   experiments with BM18 were described previously[30] and approved by Institutional

275   Animal Care and Use Committee of Rocky Mountain Laboratories, National Institutes

276   of Health, and conducted in an Association for Assessment and Accreditation of

277   Laboratory Animal Care international-accredited facility according to the guidelines

278   and basic principles in the United States Public Health Service Policy on Humane Care

279   and Use of Laboratory Animals, and the Guide for the Care and Use of Laboratory

280   Animals. Sample inactivation and shipment was performed according to standard

281   operating procedures for the removal of specimens from high containment and

282   approved by the Institutional Biosafety Committee.

283

284   **Plasmids**

285   Plasmids expressing the three polymerase subunits and NP of influenza A/WSN/33

286   (H1N1) [31], A/Northern Territory/60/68 (H3N2)[32], A/duck/Fujian/01/02 (H5N1)[32] (all

287   pcDNA3-based), and A/Brevig Mission/1/18 (H1N1) [33] (pCAGGS-based) have been

288   described. A PB2 E627K mutation was introduced into the A/duck/Fujian/01/02

289   (H5N1) PB2 subunit to enable the FJ02 polymerase to efficiently replicate vRNA in

290   mammalian cells. Plasmids expressing mutant PB1a (D445A/D446A) [34], and mutant

291 PB2 (N9D) [26], of influenza A/WSN/33 (H1N1) virus have been described previously.

292 Full-length or internally truncated vRNAs were expressed from plasmids under the

293 control of cellular RNA polymerase I promoter[35]. Luciferase reporter plasmid under the

294 control of the IFN-β promoter (pIFΔ(-116)lucter), the β-galactosidase reporter plasmid

295 (pJatLacZ) under the control of a constitutive promoter (β-gal), pcDNA-Myc-RIG-I

296 expressing myc-tagged RIG-I, and pcDNA-myc-proEGF have been described

297 previously[36,37]. To construct plasmids expressing mutant PB1, PB2 proteins and myc-

298 RIG-I (myc-RIG-I mut; which contains the mutations K851A, K858A and K861A), the

299 plasmids expressing wild-type proteins were subjected to site-directed mutagenesis

300 using the primers listed in Table S1.

301

302 **Cells and antibodies**

303 Human embryonic kidney HEK 293T cells were originally sourced from the ATCC,

304 stored in the Dunn School cell bank at the University of Oxford, and mycoplasma tested,

305 but not authenticated prior to our experiments. A549 cells were originally sourced from

306 the ATCC and cultured at the University of Hong Kong. Cells were cultured in DMEM

307 (Sigma-Aldrich) and 10% FCS. Western blots were performed using NP antibody

308 GTX125989 (GeneTex), Myc antibody GTX115046 (GeneTex), RIG-I antibody

309 GTX85488 (GeneTex), and PB2 antibody GTX125926 (GeneTex). Wild-type and

310 RIG-I knockout HEK 293T cells expressing luciferase in response to the activation of

311 the IFN-β promoter were described previously[38].

312

313 **Statistical testing**

314 In all figures, error bars indicate standard deviation with sample sizes as indicated in

315 figures or figure legends. Evaluation of the statistical significance between group means

316 was performed across all experiments according to the following criteria: (i) in the case

317 where a comparison of a single variable was made between only two groups, an

318 unpaired t-test was used; (ii) in the case of comparisons between three of more groups

319 of measurements derived from a single independent variable (e.g. IFN-β induction as a

320 function of RNA length), one-way ANOVA was used and P-values were corrected for

321 multiple comparisons using either Dunnett's test (when a single group was taken as a

322 reference/control to which all other groups were compared) or the Bonferroni method

323 (when specific pairs of groups were compared to one another); (iii) in the case of

comparisons between three of more groups of measurements derived from two independent variables (e.g. IFN-β induction as a function of RNA length and RIG-I expression), two-way ANOVA was used and P-values were corrected for multiple testing using the Bonferroni method; (iv) in the case of comparisons between three of more groups of log-distributed data (e.g. viral titres), measured values were first $\log_{10}$ transformed and then compared using one-way ANOVA, with P-values corrected for multiple comparisons by controlling the false discovery rate (FDR) to be <0.05 using the two-stage step-up method of Benjamini, Kreiger, and Yekutieli. For the evaluation of the statistical significance of the relationship between two measured values (e.g. fold increase in IFN-β induction *vs.* mvRNA level), linear regression analysis was used, with the P-value indicating the probability of the null hypothesis (no linear relationship), and the goodness of fit reported as r^2. Statistical testing related to differential gene expression analysis is detailed below, and was performed in R; all other statistical tests were performed using GraphPad Prism.

RNP reconstitution assays and quantitative RNA analysis

RNP reconstitution assays were carried out in 24-well plates out as described previously[34,39]. Briefly, 0.25 μg of the plasmids pcDNA3-NP, pcDNA3-PB2, pcDNA3-PB1, pcDNA3-PA, and a pPOLI plasmid encoding full-length or truncated vRNA templates (for list of vRNA templates used see Table S2) were transfected into HEK 293T cells using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Twenty-four hours post transfection, RNA was extracted using TRI Reagent (Sigma-Aldrich) and dissolved in RNase free water. For quantitative primer extensions, reverse transcription was carried out using SuperScript III reverse transcriptase (Thermo Fisher Scientific) with ^{32}P-labelled oligonucleotides complementary to vRNA-derived RNA species and ribosomal 5S rRNA (for primers see Table S3). cDNA synthesis was stopped with 10 μl loading dye (90% formamide, 10 mM EDTA, xylene cyanole, bromophenol blue) and ^{32}P-labelled cDNAs generated with primer NP- were resolved by 12% denaturing PAGE (19:1 acrylamide/bis-acrylamide, 1x TBE buffer, 7 M urea). ^{32}P-labelled cDNAs generated with primer NP-2 were resolved by 20% denaturing PAGE. The radiolabelled signals were imaged using phosphorimaging on a FLA-5000 scanner (Fuji), and analysed using AIDA (RayTek) and Prism 7 (GraphPad). In all experiments, the apparent RNA levels were

background corrected using the PB1 active site mutant (PB1a) signal and normalised to the 5S rRNA control. Statistical analysis of data from at least three independent experiments was carried out using ANOVA.

RNP reconstitution assays and qualitative RNA analysis

For segment-specific qualitative RNA analysis by RT-PCR, RNA was treated with DNase (Promega) for 10 min according to the manufacturer's instructions and reverse transcribed using SuperScript III and the PB2 primers listed in Table S3. cDNA was amplified using Q5 polymerase (NEB) and the primers listed in Table S3. PCR products were analysed on 1.5% agarose gels in 0.5x Tris-acetate-EDTA (TAE) buffer. For qualitative RT-PCR using universal primers, DNase treated RNA was reverse transcribed using the Lv3aa and Lv3ga primers listed in Table S3 and Superscript III at 37 ºC for 30 min. Second strands synthesis was performed with primer Lv5 and Q5 polymerase (NEB) at 47 ºC for 10 min, followed by a further extension at 72 ºC for 3 min. The primer excess in the reactions was removed by incubating the second strand reaction with 1 U of exonuclease VII (NEB) at 37 degrees Celsius for 1 h. Following inactivation of the exonuclease at 95 ºC for 10 min, the DNA was amplified using Q5 polymerase, and primers P5 and i7 for 25 cycles. PCR products were analysed by 6% PAGE.

Luciferase-based interferon expression assays

For luciferase assays, RNP reconstitutions were performed in wild-type HEK 293T cells or HEK 293T cells engineered to express luciferase from the IFN-β promoter. RNP reconstitutions were performed in a 24-well format by transfecting 0.25 μg of the plasmids pcDNA3-NP, pcDNA3-PB2, pcDNA3-PB1, pcDNA3-PA, a pPOLI plasmid encoding full-length or truncated vRNA templates using lipofectamine2000 (Invitrogen). For RNP reconstitutions in wild-type HEK 293T cells, 100 ng of pIFΔ(-116)lucter and pJatLacZ were co-transfected with the polymerase expressing plasmids. Twenty-four hours post transfection, cells were harvested in PBS and resuspended in Reporter Lysis buffer (Promega). Luciferase activity was measured using a Luciferase Assay System (Promega) and a GloMax (Promega), and normalised using the β-galactosidase signal measured using ortho-Nitrophenyl-β-galactoside (ONPG) and a GloMax. The background was subtracted using signals obtained from cells transfected

12

390   with an empty pcDNA3. Luciferase levels were corrected for viral RNA levels obtained

391   with primer extensions and a ^{32}P-labelled NP-2 primer (Table S3). For total RNA

392   transfections, 100 ng of total RNA was transfected with 100 ng of pIFΔ(-116)lucter and

393   pJatLacZ using Lipofectamine2000. Analysis of luciferase expression was performed

394   as described above. Statistical analysis was carried out using ANOVA.

395

396   **Immunoprecipitations**

397   For myc-RIG-I immunoprecipitations, 10 cm dishes with HEK 293T cell were

398   transfected with 3 µg pcDNA3-NP, pcDNA3-PB2, pcDNA3-PB1, pcDNA3-PA,

399   pcDNA-myc-RIG-I or pcDNA-myc-EGF, and a pPOLI plasmid encoding either a full-

400   length or truncated vRNA template using Lipofectamine 2000. Twenty-four hours post

401   transfection, the cells were harvested in cold PBS and lysed in 600 µl Tris lysis buffer

402   (50 mM Tris-HCl, pH 8.0; 5% glycerol; 0.5% Igepal; 200 mM NaCl; 1 mM EDTA; 1

403   mM DTT; and 1x EDTA-free protease inhibitor (Roche)) on ice for 1 h. The lysates

404   were cleared at 10,000 g for 5 min. Six µg of anti-myc antibody (Sigma-Aldrich) was

405   added to 0.5 ml of cleared lysate and mixed at 4 ºC for 1.5 h. The lystate-antibody mix

406   was bound to Dynabeads (Novex) at 4 ºC for 1.5 h, washed 3 times with IgG wash

407   buffer (10 mM Tris-HCl pH 8.0; 150 mM NaCl; 0.1% Igepal; 1 mM PMSF; 1 mM

408   EDTA), and finally analysed for bound RNA and protein. Statistical analysis of data

409   from three independent experiments was carried out using ANOVA.

410

411   **ATPase assay**

412   For wild-type and mutant myc-RIG-I purification, HEK 293T cell were transfected with

413   5 µg pcDNA-myc-RIG-I or pcRNA-myc-RIG-I mut using Lipofectamine 2000.

414   Twenty-four hours post transfection, the cells were harvested in cold PBS and lysed in

415   lysis buffer (50 mM Hepes, pH 8.0; 5% glycerol; 0.5% Igepal; 200 mM NaCl; 2 mM

416   MgCl$_2$; 10 mM CaCl$_2$; 1 mM DTT; 1 U/ml Micrococcal Nuclease (Thermo Scientific);

417   and 1x EDTA-free protease inhibitor) on ice for 1 h. Three µg of anti-myc antibody

418   was next added per 0.5 ml of cleared lysate and mixed at 4 ºC for 1.5 h. The lystate-

419   antibody mix was bound to Protein G Mag Sepharose Xtra beads (GE Healthcare) at 4

420   ºC for 1.5 h, washed 6 times with 20 column volumes of RIG-I wash buffer (50 mM

421   Hepes, pH 8.0; 200 mM NaCl; 0.1% Igepal; 5% glycerol; 1 mM PMSF; 2 mM MgCl$_2$)

422   at 4 ºC for 10 min, and finally myc-RIG-I was eluted from beads in 1 column volume

wash buffer containing 0.5 mg/ml c-myc peptide (Pierce) for 15 min at 4 ºC. Activity assays were performed in 50 mM Hepes pH 8.0, 150 mM NaCl, 2 mM $MgCl_2$, 5 mM DTT, and 0.1 μM [γ-^{32}P]ATP. [γ-^{32}P]ATP and $^{32}P_i$ were resolved using PEI-cellulose TLC plates (Sigma-Aldrich) in 0.4 M KH_2PO_4 pH 3.4.

Cell and animal infections

HEK 293T cells were infected with influenza A/WSN/33 (H1N1) virus, free of DI RNAs, at a multiplicity of infection (MOI) of 5. RNA was extracted 5 h post infection and analysed using deep sequencing or qualitative RT-PCR. 6-wells containing A549 cells were infected with A/HK/68 (H3N2), A/OK/1992/05 (H3N2), A/Brisbane/59/2007 (H1N1), A/Vietnam/1203/04 (H5N1), or A/Vietnam/1203/04 (H5N1) with the V43I mutation with an MOI of 5. RNA was extracted 8 hours post infection and analysed using deep sequencing or qualitative RT-PCR. Ferret lung tissue was obtained from ferrets infected with A/Indonesia/5/2005 (H5N1) or A/Netherlands/602/2009 (H1N1) [29], or A/Brevig Mission/1/1918 (H1N1) [30]. A single ferret (lung titre = $7.6×10^1$ $log_{10}TCID_{50}$/g) was excluded from analysis on the basis of its apparent lack of infection. Ferret RNA was isolated from lung tissue samples using Trizol (Invitrogen) and analysed using qualitative RT-PCRs and next generation mvRNA sequencing with universal primers and quantitative mRNA sequencing.

Sequence alignment and structural modelling

PB2 amino acid sequences from influenza A viruses A/WSN/33 (H1N1), A/Brevig Mission/1/18 (H1N1), A/Northern Territory/60/68 (H3N2), and A/duck/Fujian/01/02 (H5N1) were aligned using Muscle 3.0 and visualised using ESPript[40]. The bat influenza A virus polymerase structure (PDB 4WSB) was visualised in Pymol 1.6.

Next generation sequencing of mvRNAs using adapters

Total cell RNA from transfected or infected cells was isolated using Tri Reagent (Sigma) or Trizol (Invitrogen) according to the manufacturer's instructions and fractionated into small (17-200 nt) and large (>200 nt) RNA fractions using an RNA Clean and Concentrator kit (Zymo Research). Next, the small RNA fraction was denatured at 70 ºC for 2 min and subsequently treated with 2 U of XRN-1 in NEB buffer 2 at 37 ºC for 15 min to deplete miRNAs. Next, XRN-1 was inactivated by adding 10

mM EDTA and incubating the reaction at 70 ºC for 10 min. Viral triphosphorylated RNAs were converted to monophosphorylated RNAs by adding 5 U of RNA 5' Pyrophosphohydrolase (RppH) and 10 mM $MgCl_2$ and incubating the reactions at 37 ºC for 15 min. RNA was purified using an RNA Clean and Concentrator kit and libraries for deep sequencing were prepared using the NEBNext Small RNA Library Prep Kit according to the manufacturer's instructions. To ensure accurate quantitation after PCR amplification, the concentration of each library was measured by qPCR on a StepOnePlus instrument (ABI) and the number of PCR cycles used to subsequently amplify the remaining library material was calibrated so as to ensure the PCR was in the early stage of exponential amplification and to not over-cycle the PCR reactions. Amplified sequencing libraries were purified on a 6% Novex TBE PAGE according to the manufacturer's instructions to remove primer-dimers. Paired-end sequencing (2x75bp) on an Illumina HiSeq 4000 was carried out by the Oxford Genomics Centre, Wellcome Trust Centre for Human Genetics (Oxford, UK).

Next generation sequencing of mvRNAs using universal primers

To spike viral RNA for quantitative sequencing, 0.2 µl of 100 pM spike RNA (Table S2) was added to 40 ng of the small RNA fraction (see above). The RNA mixture was next converted into cDNA using primers Lv3aa, Lv3ga and Lc3 and Superscript III (Invitrogen) at 37 ºC for 30 min. Second strand synthesis was performed using Q5 polymerase (NEB) and primers Lv5, Lc3a, and Lc3g at 47 ºC for 10 min, followed by a further extension at 72 ºC for 3 min. The excess of barcoded primers was removed by incubating the second strand reaction with 1 U of exonuclease VII (NEB) at 37 ºC for 1 h. The exonuclease was inactivated at 95 ºC for 10 min. Next, the DNA was amplified using Q5 polymerase, primer P5 and i7 index primers (Lexogen), and subsequently sequenced on a NextSeq 500 sequencer (Illumina).

Preparation of reference genome files for deep sequencing of mvRNAs

Prior to mapping, a reference genome file was prepared from relevant viral reference sequences in Genbank (see above). For the analysis of sequencing libraries prepared using universal influenza virus primers, the 5′ and 3′ viral promoter sequences of each segment were modified to match the degenerate universal primer sequences used in sample preparation (see Table S3), and the sequences of the spiked-in mvRNA quantitation standards (see Table S2) were appended to the reference genome. For

15

490   WSN, VN04, and VN04[hf] viruses, deep-sequencing data of mRNA generated in A549

491   cell infections (above) was exploited to generate updated reference genome files: the

492   *mpileup* and *consensus* commands in the *bcftools* software package[41] were used

493   following mapping of non-host mRNA reads to the relevant viral reference genome

494   using *STAR* aligner[42].

495

496   **Data processing pipeline for deep sequencing of mvRNAs using universal**

497   **influenza virus primers**

498   Raw sequencing reads were first trimmed to remove sequencing adaptor sequences and

499   reads with quality scores less than 20 using the *cutadapt* software package[43] and the 8-

500   nt unique molecular identifier (UMI) at the start of each read were removed from the

501   sequence and appended to the read ID line of the FASTQ file using the *extract*

502   command from the *umi_tools* software package[44]. Sequencing reads were then mapped

503   end-to-end to the appropriate viral reference genome using the *STAR* aligner[42], and

504   permitting sequencing reads to have long internal deletions (i.e. an mvRNA, interpreted

505   as splicing by STAR) with at least 16 nt anchored on either side of the deletion (--

506   *outSJfilterOverhangMin 16 16 16 16*). The default settings of *STAR* were modified so

507   that no alignment scoring penalty was given for an internal deletion and no preference

508   was given to internal deletions that overlapped particular sequence motifs (--

509   *scoreGapNoncan 0 --scoreGapGCAG 0 --scoreGapATAC 0*), and to ensure accurate

510   quantitation, only the top-scoring alignment was included in the outputted BAM file (-

511   -*outSAMmultNmax 1*), which was sorted and indexed using *samtools*[45]. An aligned read

512   was counted as an mvRNAs if it was anchored to the viral reference genome at the 5′

513   end in vRNA sense and contained an internal deletion (called as a splice junction by

514   *STAR*), and the total numbers of mvRNAs and spiked-in quantitation standards were

515   tallied using the *idxstats* command of *samtools*. mvRNA levels relative to the

516   quantitation were then reported as number of reads per million mapped (RPM)

517   quantitation standard. To validate the qPCR protocol used to prevent over-cycling of

518   sequencing libraries, the quantitation was then repeated following removal of PCR

519   duplicates, exploiting the UMI appended to each read, using the *umi_tools dedup*

520   command, and counting the various mvRNA species as unique using the --*spliced-is-*

521   *unique* option. The identities of the individual mvRNAs were then extracted from the

522   *SJ.out.tab* file generated by STAR.

523

16

Data processing pipeline for deep sequencing of mvRNAs using adapter ligation

Raw sequencing reads were first trimmed using the *cutadapt* software package[43] to remove RNA adapters, sequencing adapters, and reads with quality scores less than 20. Since adapter ligation captures both host-derived and virus-derived small RNA species, reads were first mapped end-to-end to the DASHR database of human small RNAs[46] using the *STAR* aligner[42], with spliced alignments disabled (*--alignIntronMax 1*). Non-human reads were outputted using the *--outReadsUnmapped Fastx* option, were then mapped to the appropriate viral reference genome to find mvRNAs as described above, and quantitated relative to the total number of viral reads (RPM viral) or host reads (RPM host).

Quantitative mRNA sequencing and differential gene expression analysis

Libraries for gene expression analysis were prepared using a QuantSeq 3' mRNA-Seq Library Prep Kit FWD for Illumina (Lexogen) according to the manufacturer's instructions and sequenced on a NextSeq 500 sequencer. mRNA reads were aligned to the reference genome (CRCh38, GRCm38, or MusPutFur1.0) using the STAR read aligner[42], exploiting the built-in trimming functions to remove the first 12 bases corresponding the Lexogen random primer (*--clip5pNbases 12*) and any contaminating poly(A) tails in the sequencing reads (*--clip3pAdapterSeq AAAAAAAAAAAAAAAAAA*), as well as requiring a minimum match to the reference genome of 40 bp (*--outFilterMatchNmin 40*). Gene counts were generated using reference genome annotations (Gencode v26 for CRCh38, and Ensembl 90 for MusPutFur1.0) using the STAR command *--quantMode GeneCounts*. Differential gene expression analysis was then carried out using the DEseq2 package in R[47] to identify genes that were up- or down-regulated as a function of mvRNA levels, independently of viral load or titre. Specifically, the likelihood ratio test (LRT) was used to compare a full model (in which gene expression varies as a function of both viral load or titre, and mvRNA levels) to a reduced model (in which changes in gene expression are fully explained by viral load or titre alone) using analysis of deviance (ANODEV) to generate a P-value for the log-fold-change of each gene, which were adjusted for multiple testing by controlling the false discovery rate (FDR) using Independent Hypothesis Testing[48] and reported as q-values. mvRNA levels were determined by deep sequencing using universal influenza virus primers, as detailed above. Viral load or titre was determined by segment 6 qRT-PCR or by using previously published values[29,30]. Subsequent enrichment analysis of

17

558   Gene Ontology terms specifically affected by mvRNA levels was carried out using

559   Parametric Analysis of Gene Set Enrichment[49] via the GAGE package in R, with data

560   from the above genome annotations, accessed via the biomaRt package[50]. Significance

561   of enrichment for GO terms was calculated using a one-sample z-test[49] in GAGE, and

562   P-values were adjusted for multiple testing using the Benjamini-Hochberg method and

563   were reported as q-values.

564

References

1.   Kash, J. C. & Taubenberger, J. K. The Role of Viral, Host, and Secondary Bacterial Factors in Influenza Pathogenesis. *The American Journal of Pathology* **185,** 1528–1536 (2015).

2.   Kash, J. C. *et al.* Genomic analysis of increased host immune and cell death responses induced by 1918 influenza virus. *Nature* **76,** 105–4 (2006).

3.   Kobasa, D. *et al.* Aberrant innate immune response in lethal infection of macaques with the 1918 influenza virus. *Nature* **445,** 319–323 (2007).

4.   de Jong, M. D. *et al.* Fatal outcome of human influenza A (H5N1) is associated with high viral load and hypercytokinemia. *Nat. Med.* **12,** 1203–1207 (2006).

5.   Pichlmair, A. *et al.* RIG-I-mediated antiviral responses to single-stranded RNA bearing 5'-phosphates. *Science* **314,** 997–1001 (2006).

6.   Velthuis, te, A. J. W. & Fodor, E. Influenza virus RNA polymerase: insights into the mechanisms of viral RNA synthesis. *Nat Rev Microbiol* **14,** 479–493 (2016).

7.   Kowalinski, E. *et al.* Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. *Cell* **147,** 423–435 (2011).

8.   Lee, M.-K. *et al.* Structural features of influenza A virus panhandle RNA enabling the activation of RIG-I independently of 5'-triphosphate. *Nucleic Acids Res* **44,** 8407–8416 (2016).

9.   Robb, N. C. *et al.* Single-molecule FRET reveals the pre-initiation and initiation conformations of influenza virus promoter RNA. *Nucleic Acids Res* **44,** 10304–10315 (2016).

10.   Russell, A. B., Trapnell, C. & Bloom, J. D. Extreme heterogeneity of influenza virus infection in single cells. *eLife* pii: e32303 (2017). doi:10.1101/193995

11.   Killip, M. J., Jackson, D., Pérez-Cidoncha, M., Fodor, E. & Randall, R. E. Single-cell studies of IFN-β promoter activation by wild-type and NS1-defective influenza A viruses. *J. Gen. Virol.* **98,** 357–363 (2017).

12.   Killip, M. J., Fodor, E. & Randall, R. E. Influenza virus activation of the interferon system. *Virus Research* **209,** 11–22 (2015).

13.   Jennings, P. A., Finch, J. T., Winter, G. & Robertson, J. S. Does the higher order structure of the influenza virus ribonucleoprotein guide sequence rearrangements in influenza viral RNA? *Cell* **34,** 619–627 (1983).

14.   Perez, J. T. *et al.* Influenza A virus-generated small RNAs regulate the switch from transcription to replication. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 11525–11530 (2010).

15.   Coloma, R. *et al.* The Structure of a Biologically Active Influenza Virus Ribonucleoprotein Complex. *PLoS Pathog* **5,** e1000491–10 (2009).

18

16. Li, H. *et al.* Internal genes of a highly pathogenic H5N1 influenza virus determine high viral replication in myeloid cells and severe outcome of infection in mice. *PLoS Pathog* **14,** e1006821 (2018).

17. Turrell, L., Lyall, J. W., Tiley, L. S., Fodor, E. & Vreede, F. T. The role and assembly mechanism of nucleoprotein in influenza A virus ribonucleoprotein complexes. *Nature Communications* **4,** 1591 (2013).

18. Kato, H. *et al.* Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature* **441,** 101–105 (2006).

19. Rehwinkel, J. *et al.* RIG-I Detects Viral Genomic RNA during Negative-Strand RNA Virus Infection. *Cell* **140,** 397–408 (2010).

20. Killip, M. J., Smith, M., Jackson, D. & Randall, R. E. Activation of the Interferon Induction Cascade by Influenza A Viruses Requires Viral RNA Synthesis and Nuclear Export. *J. Virol.* **88,** 3942–3952 (2014).

21. Dulin, D. *et al.* Backtracking behavior in viral RNA-dependent RNA polymerase provides the basis for a second initiation site. *Nucleic Acids Res* **43,** 10421–10429 (2015).

22. Cheung, P. P. H. *et al.* Generation and characterization of influenza A viruses with altered polymerase fidelity. *Nature Communications* **5,** 1–13 (2014).

23. Woodman, A., Arnold, J. J., Cameron, C. E. & Evans, D. J. Biochemical and genetic analysis of the role of the viral polymerase in enterovirus recombination. *Nucleic Acids Res* **44,** 6883–6895 (2016).

24. Forero, A. *et al.* The 1918 Influenza Virus PB2 Protein Enhances Virulence through the Disruption of Inflammatory and Wnt-Mediated Signaling in Mice. *J. Virol.* **90,** 2240–2253 (2016).

25. Miotto, O., Heiny, A. T., Tan, T., August, J. T. & Brusic, V. Identification of human-to-human transmissibility factors in PB2 proteins of influenza A by large-scale mutual information analysis. *BMC Bioinformatics* **9,** S18–18 (2008).

26. Graef, K. M. *et al.* The PB2 Subunit of the Influenza Virus RNA Polymerase Affects Virulence by Interacting with the Mitochondrial Antiviral Signaling Protein and Inhibiting Expression of Beta Interferon. *J. Virol.* **84,** 8433–8445 (2010).

27. Du, Y. *et al.* Genome-wide identification of interferon-sensitive mutations enables influenza vaccine design. *Science* **359,** 290–296 (2018).

28. Chan, F. K.-M., Luz, N. F. & Moriwaki, K. Programmed necrosis in the cross talk of cell death and inflammation. *Annu. Rev. Immunol.* **33,** 79–106 (2015).

29. van den Brand, J. M. A. *et al.* Comparison of Temporal and Spatial Dynamics of Seasonal H3N2, Pandemic H1N1 and Highly Pathogenic Avian Influenza H5N1 Virus Infections in Ferrets. *PLoS ONE* **7,** e42343–21 (2012).

30. de Wit, E. *et al.* 1918 H1N1 influenza virus replicates and induces pro-inflammatory cytokine responses in extra-respiratory tissues of ferrets. *J. Infect. Dis.* (2018). doi:10.1093/infdis/jiy003

31. Fodor, E. *et al.* A Single Amino Acid Mutation in the PA Subunit of the Influenza Virus RNA Polymerase Inhibits Endonucleolytic Cleavage of Capped RNAs. *J. Virol.* **76,** 8989–9001 (2002).

32. Kashiwagi, T., Leung, B. W., Deng, T., Chen, H. & Brownlee, G. G. The N-Terminal Region of the PA Subunit of the RNA Polymerase of Influenza A/HongKong/156/97 (H5N1) Influences Promoter Binding. *PLoS ONE* **4,** e5473 (2009).

19

33. Tumpey, T. M. *et al.* Characterization of the reconstructed 1918 Spanish influenza pandemic virus. *Science* **310,** 77–80 (2005).

34. Vreede, F. T., Jung, T. E. & Brownlee, G. G. Model Suggesting that Replication of Influenza Virus Is Regulated by Stabilization of Replicative Intermediates. *J. Virol.* **78,** 9568–9572 (2004).

35. Fodor, E. *et al.* Rescue of influenza A virus from recombinant DNA. *J. Virol.* **73,** 9679–9682 (1999).

36. Childs, K. S., Andrejeva, J., Randall, R. E. & Goodbourn, S. Mechanism of mda-5 Inhibition by Paramyxovirus V Proteins. *J. Virol.* **83,** 1465–1473 (2009).

37. Zettl, M., Adrain, C., Strisovsky, K., Lastun, V. & Freeman, M. Rhomboid Family Pseudoproteases Use the ER Quality Control Machinery to Regulate Intercellular Signaling. *Cell* **145,** 79–91 (2011).

38. Hertzog, J. *et al.* Infection with a Brazilian isolate of Zika virus generates RIG-I stimulatory RNA and the viral NS5 protein blocks type I IFN induction and signaling. *Eur. J. Immunol.* **46,** 509 (2018).

39. Oymans, J. & Velthuis, te, A. J. W. A mechanism for prime-realignment during influenza A virus replication. *J. Virol.* **92,** e01773–17 (2018).

40. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42,** W320–4 (2014).

41. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27,** 2987–2993 (2011).

42. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2012).

43. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10 (2011).

44. Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Research* **27,** 491–499 (2017).

45. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

46. Leung, Y. Y. *et al.* DASHR: database of small human noncoding RNAs. *Nucleic Acids Res* **44,** D216–D222 (2016).

47. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15,** 550 (2014).

48. Ignatiadis, N., Klaus, B., Zaugg, J. B. & Huber, W. Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nat Methods* **13,** 577–580 (2016).

49. Kim, S.-Y. & Volsky, D. J. PAGE: parametric analysis of gene set enrichment. *BMC Bioinformatics* **6,** 144 (2005).

50. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols* **4,** 1184–1191 (2009).

20

Author Contributions

J.C.L., E.F., J.S. and A.J.W.t.V showed that subgenomic influenza virus RNAs stimulate IFN-β production and are bound by RIG-I. A.J.W.t.V. and J.C.L. found that mvRNAs are produced by influenza virus polymerases. D.L.V.B. designed sequencing strategies. D.L.V.B. and A.J.W.t.V. performed deep sequencing experiments and analyses. M.J.K., M.J.O.-M., H.F., and R.E.R. contributed reagents and protocols. E.d.W., D.v.R., and J.Y.S. provided ferret lung tissues. R.L.Y.F., H.Y. and L.L.M.P. performed A549 infections. A.J.W.t.V., D.L.V.B., J.C.L., and E.F. analysed data. A.J.W.t.V., J.C.L., D.L.V.B., and E.F. wrote the manuscript with input from co-authors.

Competing interests: Authors declare no competing interests.

Materials and correspondence: All sequencing data have been deposited in the NCBI Sequence Read Archive (SRA) under accession number SUB3758924. Correspondence and requests for materials should be addressed to E.F. (ervin.fodor@path.ox.ac.uk) or A.J.W.t.V. (ajwt6@cam.ac.uk).
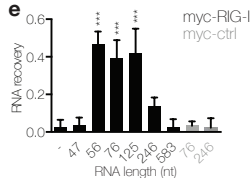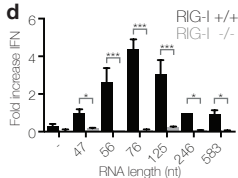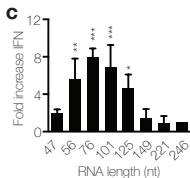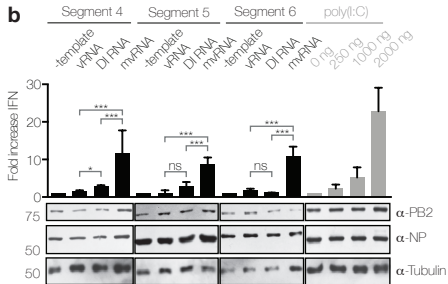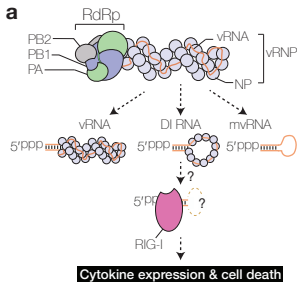
21

733

734

Figure 1. mvRNAs of influenza A virus are bound by RIG-I and induce IFN expression. (**a**) Models of the influenza virus ribonucleoprotein (vRNP) complex and potential activators of RIG-I. (**b**) Analysis of IFN-β promoter activity induced by the replication of segment 4, 5 or 6 vRNAs, DI RNAs or mvRNAs or by the transfection of poly(I:C). (**c**) IFN-β promoter activity induced by the replication of engineered, segment 5-based, short RNAs in HEK 293T cells or (**d**) in wild-type or HEK 293 RIG-I knockout cells. (**e**) Binding of segment 5-based RNAs to myc-tagged RIG-I or mouse EGF control protein (myc-ctrl). In all panels, error bars indicate standard deviation with n=3. P-values were determined using ANOVA, are shown compared to 246 nt RNA in Fig. 1C and myc-ctrl with 246 nt mvRNA in Fig.1E, and are indicated as * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$.
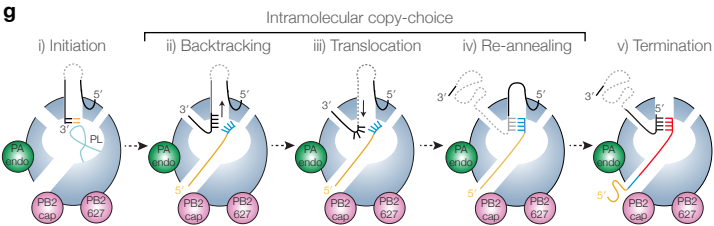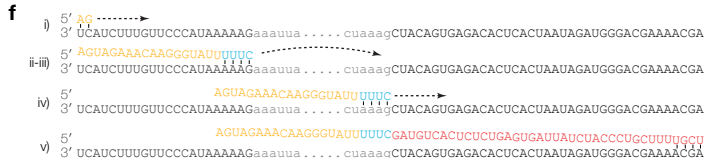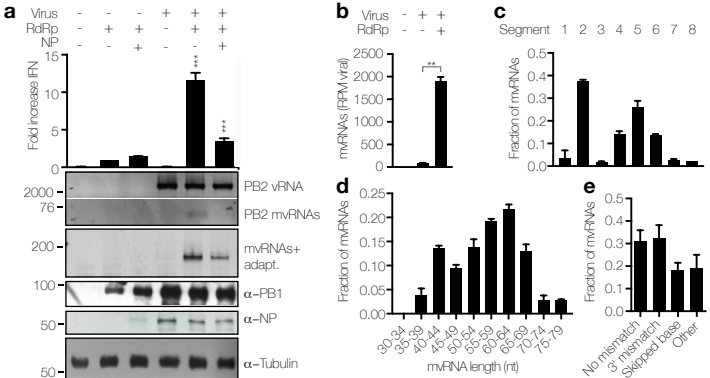
746

Figure 2. Dysregulation of RNA replication in cells infected with WSN results in the generation of mvRNAs. (**a**) Analysis of IFN-β promoter activity and steady state vRNA and mvRNA levels in WSN infections following overexpression of viral polymerase or viral polymerase and NP. mvRNAs were also amplified with universal primers containing adapters for sequencing (mvRNAs+adapt). (**b**) Quantitation of mvRNAs using deep sequencing, expressed as reads per million (RPM). (**c**) mvRNA distribution per genome segment. (**d**) Size distribution of mvRNAs. (**e**) mvRNA distribution per type of intramolecular copy-choice mechanism. (**f**) Example of mvRNA formation through an intramolecular copy-choice mechanism involving a 3′ mismatch. (**g**) Model of mvRNA formation by the polymerase (model adapted from[6]). In all panels, error bars indicate standard deviation with n=3. P-values were determined using ANOVA and shown compared to lane 3 in Fig. 2A and using an unpaired t-test in Fig. 2b, and are indicated as in Fig. 1.

760

Figure 3. The PB2 polymerase subunit of highly virulent influenza A viruses promotes mvRNA synthesis. (**a**) Analysis of mvRNA levels during the replication of a segment 5-based 246 nt RNA template by the WSN, BM18, NT60 and FJ02 polymerases, and IFN-β promoter activity induced by the transfection of total RNA isolated from these cells into reporter HEK 293T cells expressing luciferase. (**b**) Location of PB2 amino acid residues 9, 64 and 81 in the bat influenza A virus

767 polymerase structure (PDB 4WSB). (**c**) Analysis of the effect of PB2 mutations on

768 mvRNA formation and IFN-β promoter activity induced after transfection of total RNA

769 isolated from these cells into luciferase reporter HEK 293T cells. (**d**) IFN-β promoter

770 activity as function of PB2 mutation and mvRNA formation. In all panels, error bars

771 indicate standard deviation. P-values were determined using ANOVA in Fig. 1a and c,

772 or linear regression in Fig. 3d, and are indicated as * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq$

773 0.001, **** $P \leq 0.0001$.

774

775 **Figure 4. Levels of mvRNAs produced during infection correlate with innate**

776 **immune responses.** (**a**) Analysis of mvRNAs in A549 cells infected with WSN, VN04,

777 VN04[hf] using deep sequencing or PAGE. mvRNAs were amplified using universal

778 primers containing adapters for sequencing (mvRNAs+adapt). mvRNA counts were

779 normalised to mvRNA and mcRNA internal standards. (**b**) Analysis of mRNAseq of

780 infected A549 cells showing GO terms down-regulated (left) and GO terms up-

781 regulated (right) in VN04 infection as compared to VN04[hf] in response to mvRNA

782 levels. (**c**) Analysis of mvRNAs in lungs of ferrets one day after infection with IN05,

783 NL09 or BM18 using deep sequencing or PAGE. (**d**) Analysis of tissue mRNAseq

784 showing GO terms enriched as function of mvRNA levels in lungs of ferrets infected

785 with IN05, NL09 and BM18 influenza viruses. (**e**) Model for the expression of

786 cytokines in influenza virus infected cells. In all panels, error bars indicate standard

787 deviation. P-values were determined using ANOVA and are indicated in Fig. 4a and c

788 as in Fig. 3, and using a one-sample z-test (see Methods) for Gene Ontology analysis

789 in Fig. 4b and d, and are indicated by blue shading.

a RdRp

PB2, PB1, PA — vRNA / vRNP / NP

vRNA, DI RNA, mvRNA

5'ppp — vRNA
5'ppp — DI RNA
5'ppp — mvRNA

5'ppp — RIG-I

Cytokine expression & cell death

b

	Segment 4				Segment 5				Segment 6				poly(I:C)			
	-template	vRNA	DI RNA	mvRNA	-template	vRNA	DI RNA	mvRNA	-template	vRNA	DI RNA	mvRNA	0 ng	250 ng	1000 ng	2000 ng

Fold increase IFN

α-PB2
α-NP
α-Tubulin

c Fold increase IFN — RNA length (nt): 47, 56, 76, 101, 125, 149, 221, 246

d RIG-I +/+ / RIG-I -/- — Fold increase IFN — RNA length (nt): -, 47, 56, 76, 125, 246, 583

e myc-RIG-I / myc-ctrl — RNA recovery — RNA length (nt): -, 47, 56, 76, 125, 246, 583, 76, 246

a Virus / RdRp / NP panels with Fold increase IFN bar graph and western blots (PB2 vRNA, PB2 mvRNAs, mvRNAs+ adapt., α-PB1, α-NP, α-Tubulin)

b mvRNAs (RPM viral) bar graph — Virus / RdRp

c Fraction of mvRNAs by Segment (1–8)

d Fraction of mvRNAs by mvRNA length (nt): 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79

e Fraction of mvRNAs: No mismatch, 3' mismatch, Skipped base, Other

f

i) 5′ AG →
   3′ UCAUCUUUGUUCCCAUAAAAAGaaauua . . . . cuaaag CUACAGUGAGACACUCACUAUAGAUGGGACGAAAACGA

ii-iii) 5′ AGUAGAAACAAGGGUAUU UUUC
        3′ UCAUCUUUGUUCCCAUAAAAAGaaauua . . . . cuaaag CUACAGUGAGACACUCACUAUAGAUGGGACGAAAACGA

iv) 5′ AGUAGAAACAAGGGUAUU UUUC →
    3′ UCAUCUUUGUUCCCAUAAAAAGaaauua . . . . cuaaag CUACAGUGAGACACUCACUAUAGAUGGGACGAAAACGA

v) 5′ AGUAGAAACAAGGGUAUU UUUC GAUGUCACUCUCUGAGUGAUUAAUCUACCCUGCUUUUGGCU
   3′ UCAUCUUUGUUCCCAUAAAAAGaaauua

g Intramolecular copy-choice

i) Initiation | ii) Backtracking | iii) Translocation | iv) Re-annealing | v) Termination

PA endo, PB2 cap, PB2 627, PL

a

M, PB1a, WSN, BM18, N160, FJ02, BM18[iv]

vRNA
150 —
5S rRNA
100 —
90 —
70 —
mvRNAs
50 —
40 —

vRNA
100 — mvRNAs
PB1
75 —
NP
50 —

mvRNA level (%)
150 —
100 —
50 —
0 —
**** ** **** ** ****

Fold increase IFN
20 —
15 —
10 —
5 —
0 —
**** **** ****

b

N9D, 5′, 3′
PB1, PA-C
PA-N
M81T
Template exit channel
T64M
PB2

c

WSN PB2, N9D, T64M, M81T, N9D+T64M, N9D+M81T, BM18

300 —
100 — vRNA
mvRNAs
100 —
75 — PB2
50 — NP
50 — Tubulin

mvRNA level (%)
400 —
300 —
200 —
100 —
0 —

Fold increase IFN
25 —
20 —
15 —
10 —
5 —
0 —
* **** **** **** ****

d

$r^2 = 0.95$; ***

Fold increase IFN
20 — BM18
15 —
N9D+M81T
M81T
10 — N9D+T64M
T64M
5 — N9D
WSN PB2
0 —

mvRNA level (%)
0, 100, 200, 300, 400

a

Mock, WSN, VN04, VN04^PA

mvRNAs (RPM standard)

mvRNAs+adapt. primer dimers

NP vRNA

actin mRNA

b

-log₁₀(p-value) of enrichment: >10, 5.0, 0.0
genes in GO term: 400, 800, 1200, 1600

Reg. of transcription
Cell division
Phosphorylation
Reg. of translation

Reg. of immune response
SRP-dependent protein targeting to membrane
Type I interferon signaling

Log₂ fold change in response to mvRNAs

c

Mock, IND5, NL09, BM18

mRNAs (RPM standard)

mvRNAs+adapt. primer dimers

NP vRNA

actin mRNA

d

-log₁₀(p-value) of enrichment: 4.0, 2.0, 0.0
genes in GO term: 25, 50, 75, 100

Reg. of apoptotic process
Response to virus
Immune response
Innate immune response

Log₂ fold change in response to mvRNAs

e

host-specific residues
vRNP
5′ ppp mvRNA
5′ ppp RIG-I
RIG-I activation
Modulation by viral factors (e.g., NS1, PA-X, PB1-F2)
Cytokine expression & cell death