1   **SARS-CoV-2 RNA reverse-transcribed and integrated into the human genome**

2   Liguo Zhang[1], Alexsia Richards[1], Andrew Khalil[1, 2, 3], Emile Wogram[1], Haiting Ma[1], Richard A.

3   Young[1, 4], Rudolf Jaenisch[1, 4, #]

4

5   [1]Whitehead Institute for Biomedical Research, Cambridge, MA, USA.

6   [2]John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge,

7   MA, USA.

8   [3]Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, MA,

9   USA.

10   [4]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA.

11   [#]Correspondence: jaenisch@wi.mit.edu

12

13

14   **Summary**

15

16   Prolonged SARS-CoV-2 RNA shedding and recurrence of PCR-positive tests have been

17   widely reported in patients after recovery, yet these patients most commonly are non-infectious[1-

18   14]. Here we investigated the possibility that SARS-CoV-2 RNAs can be reverse-transcribed and

19   integrated into the human genome and that transcription of the integrated sequences might

20   account for PCR-positive tests. In support of this hypothesis, we found chimeric transcripts

21   consisting of viral fused to cellular sequences in published data sets of SARS-CoV-2 infected

22   cultured cells and primary cells of patients, consistent with the transcription of viral sequences

23   integrated into the genome. To experimentally corroborate the possibility of viral retro-

24   integration, we describe evidence that SARS-CoV-2 RNAs can be reverse transcribed in human

25   cells by reverse transcriptase (RT) from LINE-1 elements or by HIV-1 RT, and that these DNA

26   sequences can be integrated into the cell genome and subsequently be transcribed. Human

27   endogenous LINE-1 expression was induced upon SARS-CoV-2 infection or by cytokine

28   exposure in cultured cells, suggesting a molecular mechanism for SARS-CoV-2 retro-integration

29   in patients. This novel feature of SARS-CoV-2 infection may explain why patients can continue

30   to produce viral RNA after recovery and suggests a new aspect of RNA virus replication.

31

32

33   **Introduction**

34

35   Continuous or recurrent positive SARS-CoV-2 PCR tests have been reported in patients

36   weeks or months after recovery from an initial infection[1-14]. Although *bona fide* re-infection of

37   SARS-CoV-2 after recovery has been reported lately[15], cohort-based studies with strict

38   quarantine on subjects recovered from COVID-19 suggested "re-positive" cases were not caused

39   by re-infection[16,17]. Furthermore, no replication-competent virus was isolated or spread from

40   these PCR-positive patients[1-3,5,6,12]. The cause for such prolonged and recurrent viral RNA

41   production is unknown. As positive-stranded RNA viruses, SARS-CoV-2 and other beta-

42   coronaviruses such as SARS-CoV-1 and MERS employ an RNA-dependent RNA polymerase to

43   replicate their genomic RNA and transcribe their sub-genomic RNAs[18-20]. One possibility is that

44   SARS-CoV-2 RNAs could be reverse-transcribed and integrated into the human genome, and

45   transcription of the integrated DNA copies could be responsible for positive PCR tests.

46   Endogenous reverse transcriptase (RT) activity has been observed in human cells, and the

47   products of reverse transcription have been shown to become integrated into the genome[21,22]. For

48   example, *APP* transcripts have been shown to be reverse-transcribed by endogenous RT, with

49 resultant APP fragments integrated into the genome of neurons and transcribed[22]. Human LINE-

50 1 elements (~17% of the human genome), a type of autonomous retrotransposons, are a potential

51 source of endogenous RT, able to retro-transpose themselves and other non-autonomous

52 elements such as Alu[21,23].

53

54

55 **Results**

56

57 **Expression of viral-cellular chimeric transcripts in infected cultured and in patient-derived**

58 **cells is consistent with genomic integration of viral sequences.**

59   To investigate the possibility of viral integration into virus infected cells we analyzed

60 published RNA-Seq data from SARS-CoV-2 -infected cells for evidence of chimeric transcripts,

61 which would be indicative of viral integration into the genome and expression. Examination of

62 these data sets [24-30] (Fig. S1a-b) revealed a substantial number of host-viral chimeric reads (Fig.

63 1a-c, S1c). These occurred in multiple sample types, including cells and organoids from

64 lung/heart/brain/stomach tissues, as well as BALF cells directly isolated from COVID-19

65 patients (Fig. 1c). Chimeric read abundance was positively correlated with viral RNA level

66 across the sample types (Fig. 1c). Chimeric reads generally accounted for 0.004% - 0.14% of

67 total SARS-CoV-2 reads across the samples, with a 69.24% maximal number of reads in

68 bronchoalveolar lavage fluid cells derived from severe COVID19 patients and near no chimeric

69 reads from patient blood buffy coat cells (corresponding to almost no total SARS-CoV-2 reads).

70 A majority of chimeric junctions mapped to SARS-CoV-2 nucleocapsid (N) sequence (Fig. 1d-

71 e). This is consistent with the finding that nucleocapsid (N) RNA is the most abundant SARS-

72 CoV-2 sub-genomic RNA[31], and thus is most likely to be a target for reverse transcription and

73   integration. These analyses support the hypothesis that SARS-CoV-2 RNA may retro-integrate

74   into the genome of infected cells resulting in the production of chimeric viral-cellular transcripts.

75

76   **SARS-CoV-2 RNA can be reverse-transcribed and integrated into the human genome in**

77   **cells overexpressing a reverse transcriptase**

78   To provide experimental evidence for reverse-transcription and integration of SARS-

79   CoV-2 RNA, we overexpressed human LINE-1 or HIV-1 reverse transcriptase (RT) in

80   HEK293T cells and infected the transduced cells with SARS-CoV-2. The cells were tested 2

81   days after infection for viral sequences by PCR or fluorescence *in situ* hybridization (FISH) (Fig.

82   2a). Considering that the N RNA is the most abundant SARS-CoV-2 sub-genomic RNA[31] and is

83   most likely to be retro-integrated (Fig. 1d-e), we chose four N – targeting PCR primer sets that

84   are used in COVID-19 tests (primer source from WHO[32], Fig. 2a). PCR amplification of purified

85   cellular DNA showed positive gel-bands in cells with human LINE-1 or HIV-1 RT

86   overexpression (Fig. 2b) but not in non-transfected or non-infected cells. To test whether the

87   DNA copies of N sequences were integrated into the cellular genome, we gel-purified cell

88   genomic DNA (gDNA, >23 kb, Fig. S2a) and qPCR confirmed N sequences in gDNA of cells

89   with expression of all three types of RT (Fig. 2c). Cells with strong expression of LINE-1 driven

90   by a CMV promoter showed ~8-fold higher signals of N sequence detection suggesting a higher

91   copy-number of integrated N sequences than in cells expressing LINE-1 driven by its natural

92   promoter (5'UTR) or HIV-1 RT (Fig. 2c). We were able to clone full-length N DNA from gDNA

93   of cells overexpressing CMV-LINE-1 and confirmed its sequence by Sanger sequencing (Fig.

94   S2b). We did not detect the full-length N sequence from gDNA of cells transfected with 5'UTR-

95   LINE-1 or HIV-1 RT, which may be due to lower expression of RT in these cells (Fig. S2b). We

96   further confirmed that purified SARS-CoV-2 RNA from infected cells can be reverse-transcribed

97   *in vitro* by lysates of cells expressing either LINE-1 or HIV-1 RT (Fig. S2c-d).

98   We conducted single-molecule RNA-FISH (smRNA-FISH) using fluorophore-labeled

99   oligo-nucleotide probes targeting N (Fig. 2a) to confirm that viral N sequences were integrated

100   and detected their transcription in the nucleus. SARS-CoV-2 infected cells showed the expected

101   cytoplasmic FISH signals of N RNA (Fig. S3a). N RNA FISH signals were detected in cell

102   nuclei with cells overexpressing LINE-1 (Fig. 2d, S3b), indicating nascent transcription sites of

103   integrated N sequences. In the same cell population, a significantly higher fraction (~35%) of

104   infected cells overexpressing LINE-1, as indicated by LINE-1 ORF1p immunostaining, showed

105   nuclear N signals than cells not overexpressing LINE-1 (~12%) (Fig. 2e). A significantly higher

106   fraction of infected cells that were transfected with LINE-1 plasmid (~80% transfection

107   efficiency) showed positive nuclear N FISH signals (~30%) as compared to non-transfected cells

108   (13%; Fig. S3c). Infected but not transfected cells also exhibited nuclear N signals, albeit at a

109   lower frequency (~10%; Fig. 2e, S3c), implying integration of SARS-CoV-2 N RNA by cell

110   endogenous RT activity.

111

112   **Human endogenous LINE-1 expression induced by SARS-CoV-2 infection and cytokines**

113   **correlates with retro-integration**

114   Human LINE-1 elements are autonomous retro-transposons with their encoded reverse

115   transcriptase (ORF2p) and supporting protein (ORF1p) also aiding non-autonomous elements to

116   retro-transpose, such as Alu and other cellular RNAs[21]. We found that expression of LINE-1

117   elements was significantly up-regulated in published RNA-Seq data of cells upon infection with

118   SARS-CoV-2 and correlated with chimeric read abundance (Fig. 3a-b, S4a-d, compare Calu3

119   cells that are efficiently infected versus NHBE cells that are resistant to infection). Although the

120   upregulation in Calu3 was not higher than that in NHBE, multiple LINE-1 elements were

121   upregulated as compared to just one in NHBE (Fig. 3a, S4b, d). Expression analysis using LINE-

122   1 specific primers[33,34] showed a ~3-4-fold up-regulation of LINE-1 in Calu3 cells when infected

123   by SARS-CoV-2 (Fig. 3c). Moreover, PCR analysis on Calu3 cellular DNA showed retro-

124   integration of SARS-CoV-2 N sequences after infection (Fig. 3d-e), possibly by the activated

125   LINE-1 reverse transcriptase.

126   Patients infected with SARS-CoV-2 and other corona viruses show evidence of cytokine

127   induction associated with the immune response, and in severe cases experience a cytokine

128   storm[35-37], prompting us to investigate whether cytokines alone can induce LINE-1 activation.

129   We treated cells with cytokine-containing conditioned media from Myeloid, Microglia, or CAR-

130   T cell cultures and found a ~2-3-fold upregulation of endogenous LINE-1 expression by PCR

131   analysis (Fig. 3f, S5b). Expressed LINE-1 protein (ORF1p) was also confirmed by

132   immunofluorescence staining (Fig. 3g-h, S5a). In summary, our results show induced LINE-1

133   expression in cells stressed by viral infection or exposed to cytokines, suggesting a molecular

134   mechanism for SARS-CoV-2 retro-integration in human cells.

135

136
137 Discussion
138
139   In this study, we showed evidence that SARS-CoV-2 RNAs can be reverse-transcribed

140 and integrated into the human genome by several sources of reverse transcriptase such as

141 activated human LINE-1 or co-infected retrovirus (HIV). We found LINE-1 expression can be

142 induced upon SARS-CoV-2 infection or cytokine exposure, suggesting a molecular mechanism

143 responsible for SARS-CoV-2 retro-integration in patients. Moreover, our results suggest that the

144 integrated SARS-CoV-2 sequences can be transcribed, as shown by RNA-Seq and smRNA-FISH

145 data, providing a possible explanation for the presence of viral sequences at later times after

146 initial virus exposure and in the absence of detectable infectious virus[1-14]. The retro-inserted

147 SRAS-CoV-2 sequences are most likely sub-genomic fragments, as the integration junctions are

148 mostly enriched at the N sequence (Fig. 1d-e), excluding the production of infectious virus. Our

149 data may also explain that patients, after recovery from disease symptoms, may become again

150 positive for viral sequences as detected by PCR[1,8-14].

151   An important follow-up question is whether these integrated SARS-CoV-2 sequences can

152 express viral antigens. If so, it will be of clinical interest to assess whether viral antigens

153 expressed from integrated virus fragments could trigger an immune response in patients that

154 could affect the course and treatment of the disease. It is possible that the clinical consequences

155 of the integrated viral fragments may depend on their insertion sites in the human genome, and

156 on epigenetic regulation which has been shown in HIV patients[38]. Careful analysis on SARS-

157 CoV-2 retro-integration sites in patient samples and correlation with disease severity will help to

158 elucidate potential clinical consequences. Furthermore, immune response may vary depending on

159 an individual's underlying conditions. More generally, our results suggest a novel aspect of

160 infection possibly also for other common disease-causing RNA viruses such as Dengue, Zika or

161   Influenza virus, which could be subject to retro-integration and perhaps affect disease

162   progression.

163   Human LINE-1 accounts for ~17% of the human genome, ~100 out of 500,000 copies of

164   which are active[21,23]. LINE-1 – encoded reverse-transcriptase (ORF2p) and supporting protein

165   (ORF1p) are known to retro-transpose not only LINE-1 transcripts (in *Cis*), but also other RNA

166   species such as Alu (SINE) and cellular mRNA (in *Trans*, creating processed pseudogenes), with

167   a "target-site – primed reverse transcription" mechanism[21]. LINE-1 proteins have been shown as

168   nucleic acid chaperones with high RNA binding affinity[39], therefore it is perhaps not surprising

169   that they can retro-integrate exogenous viral RNAs. From an evolutionarily perspective, retro-

170   integration of viral RNA by LINE-1 could be an adaptive response by the host to provide

171   sustaining antigen expression possibly enhancing protective immunity. Conversely, retro-

172   integration of viral RNAs could be detrimental and cause a more severe immune response in

173   patients such as a "cytokine storm" or auto-immune reactions.

174   Our results may also be relevant for current clinical trials of antiviral therapies[40]. The

175   reliance of PCR tests to assess the effect of treatments on viral replication and viral load may not

176   reflect the efficacy of the treatment to suppress viral replication as the PCR assay may detect

177   viral transcripts from viral sequences stably integrated into the genome rather than infectious

178   virus.

179

180   **Methods**

181

182   **Cell culture and plasmid transfection**

183   HEK293T cells were obtained from ATCC (CRL-3216) and cultured in DMEM

184   supplemented with 10% heat-inactivated FBS (Hyclone, SH30396.03) and 2mM L-glutamine

185   (MP Biomedicals, IC10180683) following ATCC's method. Calu3 cells were obtained from

186   ATCC (HTB-55) and cultured in EMEM (ATCC 30-2003) supplemented with 10% heat-

187   inactivated FBS (Hyclone, SH30396.03) following ATCC's method.

188   Plasmid for HIV-1 reverse transcriptase expression: pCMV-dR8.2 dvpr was a gift from

189   Bob Weinberg (Addgene plasmid # 8455 ; http://n2t.net/addgene:8455 ;

190   RRID:Addgene_8455)[41]. Plasmids for human LINE-1 expression: pBS-L1PA1-CH-mneo

191   (CMV-LINE-1) was a gift from Astrid Roy-Engel (Addgene plasmid # 51288 ;

192   http://n2t.net/addgene:51288 ; RRID:Addgene_51288)[42]; EF06R (5'UTR-LINE-1) was a gift

193   from Eline Luning Prak (Addgene plasmid # 42940 ; http://n2t.net/addgene:42940 ;

194   RRID:Addgene_42940)[43]. Transfection was done with Lipofectamine™ 3000 (Invitrogen

195   L3000001) following manufacturer's protocol.

196

197   **SARS-CoV-2 infection**

198   SARS-CoV-2 USA-WA1/2020 (Gen Bank: MN985325.1) was obtained from BEI

199   Resources and expanded and tittered on Vero cells. Cells were infected in DMEM +2% FBS for

200   48 hrs using multiplicity (MOI) of 0.5 for infection of HEK293T cells and an MOI of 2 for

201   Calu3 cells. All sample processing and harvest with infectious virus were done in the BSL3

202   facility at the Ragon Institute.

203

Nucleic acids extraction, *in vitro* reverse transcription and PCR/qPCR

DNA extraction was following a published protocol[22]. For purification of genomic DNA, extracted total cellular DNA was run on 0.4% (w/v) agarose/1x TAE gel for 1.5 hrs with a 3V/cm voltage, with λ DNA-HindIII Digest (NEB N3012S) as size markers. Large fragment bands (>23.13 kb) were cut off, frozen in -80 °C and then crushed by a pipette tip. 3 times of volume (v/w) of high T-E buffer (10 mM Tris – 10 mM EDTA, pH 8.0) was added and then NaCl was added to 200 mM. Gel solution was heated at 70 °C for 15 mins with constant mixing and then extracted with Phenol:Chloroform:Isoamyl Alcohol (25:24:1, v/v) (Life Technologies 15593031) and Chloroform:Isoamyl alcohol 24:1 (Sigma C0549-1PT). DNA was then precipitated by sodium acetate and isopropyl alcohol. For small amount of DNA, glycogen (Life Technologies 10814010) was added as a carrier to aid precipitation.

RNA extraction was done with either TRIzol™ LS Reagent (Invitrogen 10296010) or RNeasy Plus Micro Kit (Qiagen 74034) following manufacturers' protocols. RNA reverse transcription was done with either SuperScript™ III First-Strand Synthesis SuperMix (oligo dT + random hexamer, Invitrogen 18080400) or qScript cDNA SuperMix (QuantaBio 95048-500), following manufacturers' protocols. *In vitro* reverse transcription assay for viral RNA by cell lysates was done following a published protocl[22].

PCR was done using AccuPrime Taq DNA Polymerase, high fidelity (Life Technologies 12346094). qPCR was done using SYBR™ Green PCR Master Mix (Applied Biosystems 4309155) or PowerUp™ SYBR™ Green Master Mix (Applied Biosystems A25742) in a QuantStudio™ 6 system (Applied Biosystems). See **Supplementary Table 1** for primer sequences used in this study. qPCR plots were generated with Prism 8 (Prism).

Immuno-fluorescence staining and single-molecule RNA-FISH

227

228   Cells subject to SARS-CoV-2 infection were grown in μ-Slide 8 Well (#1.5 polymer,

229   Ibidi 80826) and fixed with 4% paraformaldehyde/CMF-PBS at room temperature (RT) for 30

230   mins. Otherwise, cells were grown on 12 mm round coverslips (#1.5, Warner Instruments 64-

231   0712) and fixed with 1.6% paraformaldehyde/CMF-PBS at room temperature (RT) for 15 mins.

232   Cells were permeabilized with 0.5% (v/v) Triton X-100/PBS, blocked with 4% (w/v) BSA/CMF-

233   PBS at RT for 1 hr, incubated with 1:200 diluted anti-LINE-1 ORF1p mouse monoclonal

234   antibody (clone 4H1, Sigma MABC1152, Lot 3493991), and then with 1:400 diluted Donkey-

235   anti-Mouse-Alexa Fluor 594 second antibody (Invitrogen 21203).

236   Single-molecule RNA-FISH probes (Stellaris®) were ordered from LGC Biosearch

237   Technologies with Quasar® 670 Dye labeling. See **Supplementary Table 2** for probe

238   sequences. FISH procedure combining with immuno-fluorescence staining was following

239   previous publications[44,45].

240   Cells in μ-Slide were mounted with Ibidi Mounting Medium With DAPI (Ibidi 50011).

241   Cells on coverslips were mounted with VECTASHIELD® HardSet™ Antifade Mounting

242   Medium with DAPI (Vector Laboratories H-1500-10).

243

Microscopy and imaging analysis

244

245   3D optical sections were acquired with 0.2-μm z-steps using a DeltaVision Elite Imaging

246   System microscope system with a 100 × oil objective (NA 1.4) and a pco.edge 5.5 camera and

247   DeltaVision SoftWoRx software (GE Healthcare). Image deconvolution was done using

248   SoftWoRx. All figure panel images were prepared using FIJI software (ImageJ, NIH) and Adobe

249   Illustrator 2020 (Adobe), showing deconvolved single z-slices.

250   To measure the LINE-1 ORF1p immuno-staining signal intensity, we projected cell

251   optical sections (sum, 42 slices) with the "z projection" function in FIJI. We measured the sum

252   of intensity of the entire cell area in the z-projected image as the signal intensity, subtracted the

253   background intensity outside of cells and then divided by the mean of the "Basal media

254   treatment" group to have the normalized signal intensity, as previously described[44,45]. All images

255   from the same experiment were using the same exposure time and transmitted exciting light. All

256   intensity measurements were done with non-deconvolved raw images. Box plot was done in R

257   (version 4.0.3)[46].

258

259   **RNA-Seq data analysis**

260   RNA-Seq data were downloaded from GEO with the accession numbers GSE147507[24],

261   GSE153277[25], GSE156754[26], GSE157852[27], GSE153684[28], GSE145926[29], GSE154998[30]

262   (summarized in **Supplementary Figure 1a**).

263   To identify human – SARS-CoV-2 chimeric reads, raw sequencing reads were aligned to

264   concatenated human and SARS-CoV-2 genomes plus transcriptomes by STAR (version

265   2.7.1a)[47]. Human genome version hg38 with no alternative chromosomes and gene annotation

266   version GRCh38.97 were used. SARS-CoV-2 genome version NC_045512.2 and gene

267   annotation (http://hgdownload.soe.ucsc.edu/goldenPath/wuhCor1/bigZips/genes/) were used.

268   The following STAR parameters[31] were used to call chimeric reads unless otherwise specified

269   (**Supplementary Figure 1a**): --chimOutType Junctions SeparateSAMold WithinBAM HardClip

270   \ --chimScoreJunctionNonGTAG 0 \ --alignSJstitchMismatchNmax -1 -1 -1 -1 \ --

271   chimSegmentMin 50 \ --chimJunctionOverhangMin 50.

272   To analyze human LINE-1 expression in RNA-Seq data, a published method,

273   RepEnrich2[48], was used to map RNA-Seq reads to human repeat annotations, using human

274   repeat masker (hg38). Differential expression was analyzed using EdgeR package (version

275   3.30.3)[49,50] in R (version 4.0.3)[46].

276

277   **Conditioned media production and treatment**

278   As previously described[51], myeloid precursors were derived from human pluripotent stem

279   cells. Briefly, human embryonic stem cells were cultured in StemFlex (ThermoFisher) feeder-

280   free medium on Matrigel™- (Corning) coated tissue culture polystyrene. 24 hrs before single-

281   cell harvesting via TrypLE Express (ThermoFisher), cells were treated with 10 µM ROCK

282   Inhibitor (Y-27632) (Stem Cell Technologies) in Essential 8 (E8) medium (ThermoFisher). After

283   harvesting, cells were centrifuged at 300 g for 3 mins in non-adherent U-bottom 96-well plates

284   (Corning) at 10,000 cells per 150 µL/well of embryoid body (EB) medium consisting of 10 µM

285   ROCK Inhibitor, 50 ng/mL BMP-4 (Peprotech), 20 ng/mL SCF (Peprotech), 50 ng/mL VEGF

286   (Peprotech), and 100 U/mL Penn/Strep (ThermoFisher) in E8 base medium. EBs were cultured

287   in the 96-well plates for 4 days with 150 µL/well of EB medium added at day 2. After 4 days, 16

288   EBs/well were plated in a 6-well tissue culture polystyrene plated coated with Matrigel™ in

289   hematopoietic myeloid medium (HIM) consisting of 2mM GlutaMax (ThermoFisher), 55 µM

290   beta-mercaptoethanol, 100 ng/mL M-CSF (Peprotech), and 25 ng/mL IL-3 (Peprotech) in X-

291   VIVO 15 base medium (Lonza). HIM media was changed every 3-4 days for 2-3 weeks until

292   floating CD14-positive myeloid precursors emerged. Myeloid conditioned media consisted of

293   floating myeloid cells cultured in HIM media for 7 days at a concentration of $0.5 \times 10^6 - 1 \times 10^6$

294   cells/mL. Cells in conditioned media were removed by centrifugation and filtration through 0.2

295   µM filters. Calu3 cells were cultured in the myeloid conditioned media or HIM media (basal) for

296   two days with daily media change before harvest or fixation.

297   Microglia were differentiated from human induced pluripotent stem cells (hiPSCs) via

298   embryoid bodies and primitive macrophage precursors (PMPs)[51]. In brief, hiPSCs (cultured

299   feeder-free on matrigel in StemFlexTM (Gibco)) were dissociated with TrypLE Express (Gibco),

300   and 10,000 cells were plated per well in 96-well ultra-low attachment plates (Corning) in 100 µL

301   embryoid body medium (10 µM ROCK inhibitor, 50 ng/mL BMP-4, 20 ng/mL SCF, and 50

302   ng/mL VEGF-121 in StemFlex), before centrifugation at $300 \times g$ for 3 mins at 4 °C. Embryoid

303   bodies were cultured for 4 days, with adding 100 µL embryoid body medium after 2 days. 12 to

304   16 embryoid bodies were plated per well of tissue culture-treated 6-well plates and cultured in 3

305   mL hematopoetic medium (2 mM GlutaMax, 100 U/mL penicillin, 100 µg/mL streptomycin, 55

306   µM β-mercaptoethanol, 100 ng/mL M-CSF, 25 ng/mL IL-3, 100 U/mL penicillin, 100 µg/mL

307   streptomycin in X-VIVO 15 (Lonza, BW04418Q). From this point on, 2 mL medium was

308   exchanged every 4–7 days. PMPs were harvested from suspension during medium exchange and

309   plated in microglia differentiation media over 7-14 days to produce microglia like cell

310   monocultures (Neurobasal (Life Technologies 21103049) supplemented with Gem21 NeuroPlex

311   without Vitamin A (GeminiBio, 400-161), 2mM GlutaMAX (Gibco), 100 ng/mL IL-34, and 10

312   ng/mL GM-CSF, 100 U/mL penicillin, 100 µg/mL streptomycin). For microglia stimulation,

313   microglia differentiation media was exchanged with HEK293T media (DMEM + 10% heat-

314   inactivated FBS + final 2mM L-Glutamine) and supplemented with 100 hg/ml

315   lipopolysaccharide (LPS, Sigma Aldrich L4391-1MG) or PBS. After 24 hrs, the microglia

316   conditioned media was collected, centrifugated (1000 rpm 10min) and the supernatant was

317 directly applied to HEK293T cells. HEK293T cells received microglia conditioned media or

318 basal HEK293T media on three constitutive days before fixation.

319   Human anti-CD19 CAR-T cells were generated by transduction of primary T cells

320 purified from human peripheral blood mononuclear cells (PBMC) with CD19-CAR expressing

321 retrovirus[52]. Anti-CD19 CAR-T cells were co-cultured with CD19-expressing beta-like cells[52] or

322 WIBR3 cells with a luciferase-2A-CD19 expressing cassette integrated at the AAVS1 locus in

323 RPMI1640 medium with 10% human serum AB. Cells in the conditioned medium were removed

324 by filtration through 0.45 μM filters. RPMI1640 medium with 10% human serum AB was used

325 as basal media control. Calu3 cells were cultured in the CAR-T conditioned media with indicated

326 dilutions or in the basal media for two days before harvest.

327

328

329 **Data Availability**

330   The datasets generated during and/or analysed during the current study are available from

331 the corresponding author on reasonable request.

332

333

334 **Acknowledgements**

343

344

345   **Author contributions**

346   Project design by R.J. and R.A.Y, execution of experiments and data analysis by L.Z.,

347   A.R, R.J and R.A.Y; E.W., A.K., and H.M. generated cells and reagents; Manuscript preparation

348   by L.Z. and R.J. with input from all authors.

349

350

351   **Competing interests**

352   R.J. is an advisor/co-founder of Fate Therapeutics, Fulcrum Therapeutics, Omega

353   Therapeutics, and Dewpoint Therapeutics. R.A.Y. is a founder and shareholder of Syros

354   Pharmaceuticals, Camp4 Therapeutics, Omega Therapeutics, and Dewpoint Therapeutics. All

355   other authors declare no competing interests.

References

1 https://www.cdc.go.kr/board/board.es?mid=a30402000000&bid=0030.

2 Bullard, J. *et al.* Predicting Infectious Severe Acute Respiratory Syndrome Coronavirus 2 From Diagnostic Samples. *Clinical Infectious Diseases*, doi:10.1093/cid/ciaa638 (2020).

3 He, X. *et al.* Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine* **26**, 672-675, doi:10.1038/s41591-020-0869-5 (2020).

4 Li, N., Wang, X. & Lv, T. Prolonged SARS-CoV-2 RNA shedding: Not a rare phenomenon. *J Med Virol*, doi:10.1002/jmv.25952 (2020).

5 Mina, M. J., Parker, R. & Larremore, D. B. Rethinking Covid-19 Test Sensitivity — A Strategy for Containment. *New England Journal of Medicine* **383**, e120, doi:10.1056/NEJMp2025631 (2020).

6 Sethuraman, N., Jeremiah, S. S. & Ryo, A. Interpreting Diagnostic Tests for SARS-CoV-2. *JAMA* **323**, 2249-2251, doi:10.1001/jama.2020.8259 (2020).

7 Yang, J.-R. *et al.* Persistent viral RNA positivity during the recovery period of a patient with SARS-CoV-2 infection. *Journal of Medical Virology* **92**, 1681-1683, doi:https://doi.org/10.1002/jmv.25940 (2020).

8 An, J. *et al.* Clinical characteristics of recovered COVID-19 patients with re-detectable positive RNA test. *Ann Transl Med* **8**, 1084, doi:10.21037/atm-20-5602 (2020).

9 Chen, D. *et al.* Recurrence of positive SARS-CoV-2 RNA in COVID-19: A case report. *Int J Infect Dis* **93**, 297-299, doi:10.1016/j.ijid.2020.03.003 (2020).

10 Lan, L. *et al.* Positive RT-PCR Test Results in Patients Recovered From COVID-19. *JAMA* **323**, 1502-1503, doi:10.1001/jama.2020.2783 (2020).

11 Loconsole, D. *et al.* Recurrence of COVID-19 after recovery: a case report from Italy. *Infection*, doi:10.1007/s15010-020-01444-1 (2020).

12 Lu, J. *et al.* Clinical, immunological and virological characterization of COVID-19 patients that test re-positive for SARS-CoV-2 by RT-PCR. *EBioMedicine* **59**, 102960, doi:10.1016/j.ebiom.2020.102960 (2020).

13 Luo, S., Guo, Y., Zhang, X. & Xu, H. A follow-up study of recovered patients with COVID-19 in Wuhan, China. *Int J Infect Dis* **99**, 408-409, doi:10.1016/j.ijid.2020.05.119 (2020).

14 Ye, G. *et al.* Clinical characteristics of severe acute respiratory syndrome coronavirus 2 reactivation. *J Infect* **80**, e14-e17, doi:10.1016/j.jinf.2020.03.001 (2020).

15 To, K. K. *et al.* COVID-19 re-infection by a phylogenetically distinct SARS-coronavirus-2 strain confirmed by whole genome sequencing. *Clin Infect Dis*, doi:10.1093/cid/ciaa1275 (2020).

16 Huang, J. *et al.* Recurrence of SARS-CoV-2 PCR positivity in COVID-19 patients: a single center experience and potential implications. *medRxiv*, 2020.2005.2006.20089573, doi:10.1101/2020.05.06.20089573 (2020).

17 Yuan, B. *et al.* Recurrence of positive SARS-CoV-2 viral RNA in recovered COVID-19 patients during medical isolation observation. *Sci Rep* **10**, 11887, doi:10.1038/s41598-020-68782-w (2020).

18 Alanagreh, L., Alzoughool, F. & Atoum, M. The Human Coronavirus Disease COVID-19: Its Origin, Characteristics, and Insights into Potential Drugs and Its Mechanisms. *Pathogens* **9**, doi:10.3390/pathogens9050331 (2020).

19   Fehr, A. R. & Perlman, S. Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol Biol* **1282**, 1-23, doi:10.1007/978-1-4939-2438-7_1 (2015).

20   de Wit, E., van Doremalen, N., Falzarano, D. & Munster, V. J. SARS and MERS: recent insights into emerging coronaviruses. *Nat Rev Microbiol* **14**, 523-534, doi:10.1038/nrmicro.2016.81 (2016).

21   Kazazian, H. H., Jr. & Moran, J. V. Mobile DNA in Health and Disease. *N Engl J Med* **377**, 361-370, doi:10.1056/NEJMra1510092 (2017).

22   Lee, M. H. *et al.* Somatic APP gene recombination in Alzheimer's disease and normal neurons. *Nature* **563**, 639-645, doi:10.1038/s41586-018-0718-6 (2018).

23   Brouha, B. *et al.* Hot L1s account for the bulk of retrotransposition in the human population. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 5280-5285, doi:10.1073/pnas.0831042100 (2003).

24   Blanco-Melo, D. *et al.* Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell* **181**, 1036-1045 e1039, doi:10.1016/j.cell.2020.04.026 (2020).

25   Huang, J. *et al.* SARS-CoV-2 Infection of Pluripotent Stem Cell-Derived Human Lung Alveolar Type 2 Cells Elicits a Rapid Epithelial-Intrinsic Inflammatory Response. *Cell stem cell*, doi:10.1016/j.stem.2020.09.013 (2020).

26   Perez-Bermejo, J. A. *et al.* SARS-CoV-2 infection of human iPSC-derived cardiac cells predicts novel cytopathic features in hearts of COVID-19 patients. *bioRxiv*, doi:10.1101/2020.08.25.265561 (2020).

27   Jacob, F. *et al.* Human Pluripotent Stem Cell-Derived Neural Cells and Brain Organoids Reveal SARS-CoV-2 Neurotropism Predominates in Choroid Plexus Epithelium. *Cell stem cell*, doi:10.1016/j.stem.2020.09.016 (2020).

28   Giobbe, G. G. *et al.* SARS-CoV-2 infection and replication in human fetal and pediatric gastric organoids. *bioRxiv*, 2020.2006.2024.167049, doi:10.1101/2020.06.24.167049 (2020).

29   Liao, M. *et al.* Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat Med* **26**, 842-844, doi:10.1038/s41591-020-0901-9 (2020).

30   Gill, S. E. *et al.* Transcriptional Profiling of Leukocytes in Critically Ill COVID19 Patients: Implications for Interferon Response and Coagulation. *PREPRINT (Version 2) available at Research Square*, doi:10.21203/rs.3.rs-63632/v2 (2020).

31   Kim, D. *et al.* The Architecture of SARS-CoV-2 Transcriptome. *Cell* **181**, 914-921 e910, doi:10.1016/j.cell.2020.04.011 (2020).

32   https://www.who.int/docs/default-source/coronaviruse/whoinhouseassays.pdf?sfvrsn=de3a76aa_2.

33   Jones, R. B. *et al.* LINE-1 retrotransposable element DNA accumulates in HIV-1-infected cells. *J Virol* **87**, 13307-13320, doi:10.1128/JVI.02257-13 (2013).

34   Tiwari, B. *et al.* p53 directly represses human LINE1 transposons. *Genes & development*, doi:10.1101/gad.343186.120 (2020).

35   Channappanavar, R. & Perlman, S. Pathogenic human coronavirus infections: causes and consequences of cytokine storm and immunopathology. *Semin Immunopathol* **39**, 529-539, doi:10.1007/s00281-017-0629-x (2017).

36   Mehta, P. *et al.* COVID-19: consider cytokine storm syndromes and immunosuppression. *The Lancet* **395**, 1033-1034, doi:10.1016/s0140-6736(20)30628-0 (2020).

444 37 Costela-Ruiz, V. J., Illescas-Montes, R., Puerta-Puerta, J. M., Ruiz, C. & Melguizo-
445   Rodriguez, L. SARS-CoV-2 infection: The role of cytokines in COVID-19 disease. *Cytokine*
446   *Growth Factor Rev* **54**, 62-75, doi:10.1016/j.cytogfr.2020.06.001 (2020).
447 38 Jiang, C. *et al.* Distinct viral reservoirs in individuals with spontaneous control of HIV-1.
448   *Nature* **585**, 261-267, doi:10.1038/s41586-020-2651-8 (2020).
449 39 Naufer, M. N., Furano, A. V. & Williams, M. C. Protein-nucleic acid interactions of LINE-1
450   ORF1p. *Semin Cell Dev Biol* **86**, 140-149, doi:10.1016/j.semcdb.2018.03.019 (2019).
451 40 Baum, A. *et al.* REGN-COV2 antibodies prevent and treat SARS-CoV-2 infection in rhesus
452   macaques and hamsters. *Science* **370**, 1110, doi:10.1126/science.abe2402 (2020).
453 41 Stewart, S. A. *et al.* Lentivirus-delivered stable gene silencing by RNAi in primary cells.
454   *RNA* **9**, 493-501, doi:10.1261/rna.2192803 (2003).
455 42 Wagstaff, B. J., Barnerssoi, M. & Roy-Engel, A. M. Evolutionary conservation of the
456   functional modularity of primate and murine LINE-1 elements. *PloS one* **6**, e19672,
457   doi:10.1371/journal.pone.0019672 (2011).
458 43 Farkash, E. A., Kao, G. D., Horman, S. R. & Prak, E. T. Gamma radiation increases
459   endonuclease-dependent L1 retrotransposition in a cultured cell assay. *Nucleic acids*
460   *research* **34**, 1196-1204, doi:10.1093/nar/gkj522 (2006).
461 44 Zhang, L. *et al.* TSA-Seq reveals a largely "hardwired" genome organization relative to
462   nuclear speckles with small position changes tightly correlated with gene expression
463   changes. *bioRxiv*, 824433, doi:10.1101/824433 (2020).
464 45 Kim, J., Venkata, N. C., Hernandez Gonzalez, G. A., Khanna, N. & Belmont, A. S. Gene
465   expression amplification by nuclear speckle association. *The Journal of cell biology* **219**,
466   doi:10.1083/jcb.201904046 (2020).
467 46 R Core Team (2020). R: A language and environment for statistical computing. *R*
468   *Foundation for Statistical Computing, Vienna, Austria. URL* https://www.R-project.org/.
469 47 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21,
470   doi:10.1093/bioinformatics/bts635 (2013).
471 48 Criscione, S. W., Zhang, Y., Thompson, W., Sedivy, J. M. & Neretti, N. Transcriptional
472   landscape of repetitive elements in normal and cancer human cells. *BMC Genomics* **15**,
473   583, doi:10.1186/1471-2164-15-583 (2014).
474 49 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for
475   differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-
476   140, doi:10.1093/bioinformatics/btp616 (2010).
477 50 McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor
478   RNA-Seq experiments with respect to biological variation. *Nucleic acids research* **40**,
479   4288-4297, doi:10.1093/nar/gks042 (2012).
480 51 Brownjohn, P. W. *et al.* Functional Studies of Missense TREM2 Mutations in Human
481   Stem Cell-Derived Microglia. *Stem cell reports* **10**, 1294-1307,
482   doi:10.1016/j.stemcr.2018.03.003 (2018).
483 52 Ma, H., Jeppesen, J. F. & Jaenisch, R. Human T Cells Expressing a CD19 CAR-T Receptor
484   Provide Insights into Mechanisms of Human CD19-Positive beta Cell Destruction. *Cell*
485   *Rep Med* **1**, 100097, doi:10.1016/j.xcrm.2020.100097 (2020).
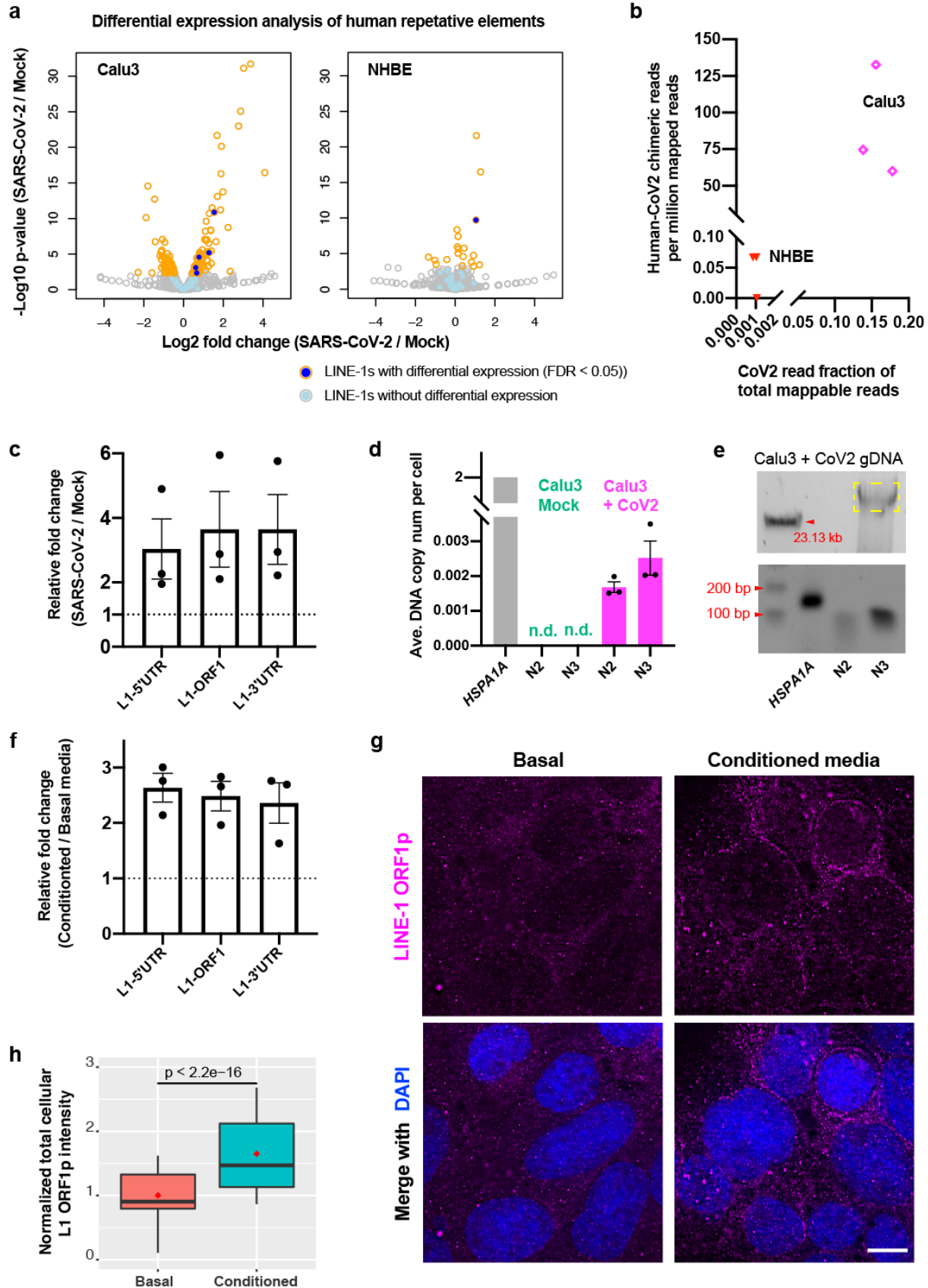486
487

488   **Figures**

489
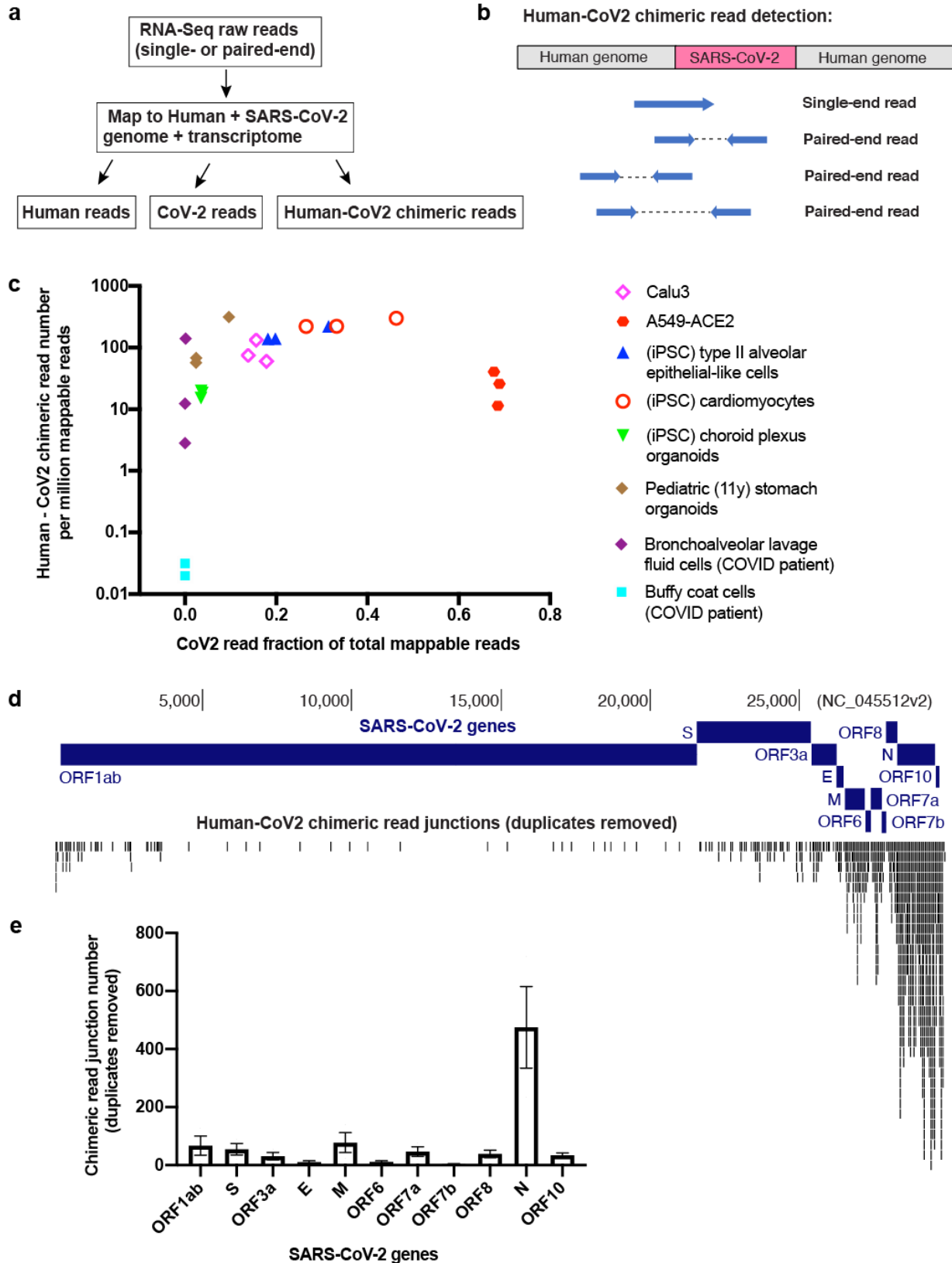
490   **Figure 1. Human – SARS-CoV-2 chimeric transcripts identified in published data sets of**

491   **infected cultured and patient-derived cells. a)** Pipeline to identify human-CoV2 chimeric

492   RNA-Seq reads. **b)** Diagram showing human-CoV2 chimeric reads mapped to potential SARS-

493   CoV-2 integration sites in the human genome from published RNA-Seq data. **c)** Scatter plot

494   showing human-CoV2 chimeric read number (per million total mappable reads, y-axis) versus

495   SARS-CoV-2 read fraction of total mappable reads (x-axis) in published RNA-Seq datasets

496   (summarized in **Supplementary Figure 1a**) from different bio-samples with SARS-CoV-2

497   infection. **d-e)** Human-CoV2 chimeric read junctions (duplicates removed) mapped to the SARS-

498   CoV-2 genome (**d**) and distribution among SARS-CoV-2 genes (**e**, three biological replicates;

499   mean ± s.e.m.). RNA-Seq data is from SARS-CoV-2 infected Calu3 cells (see **Supplementary**

500   **Figure 1a**). Chimeric read junction is defined as the "breaking point" of sequences mapped to

501   human or SARS-CoV-2 genome/transcriptome in a given RNA-Seq read.

502
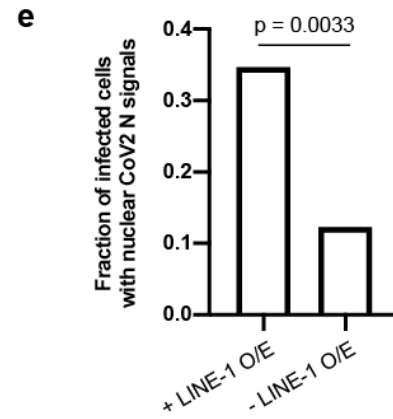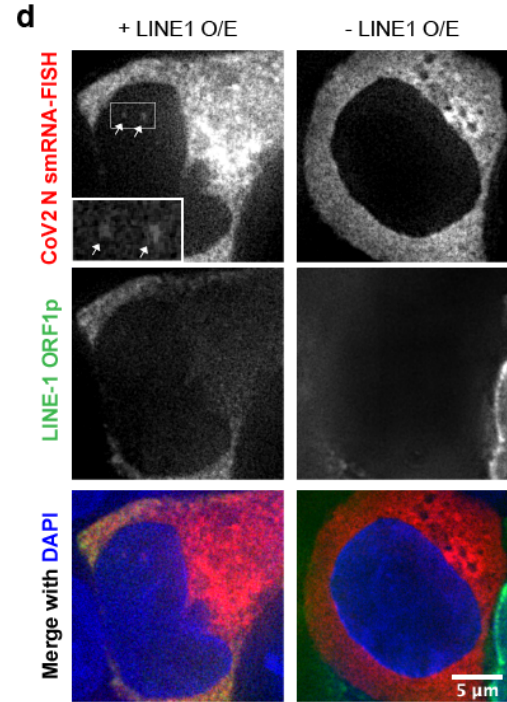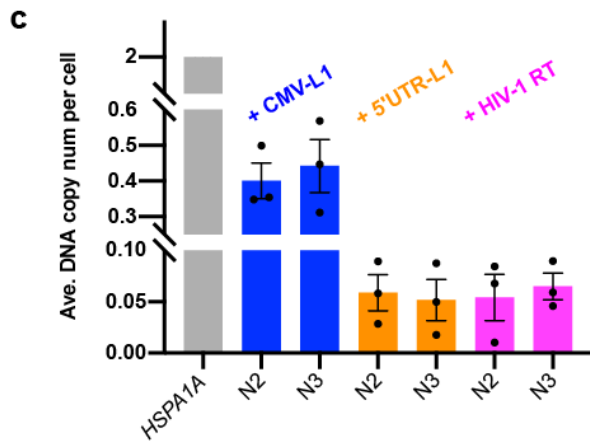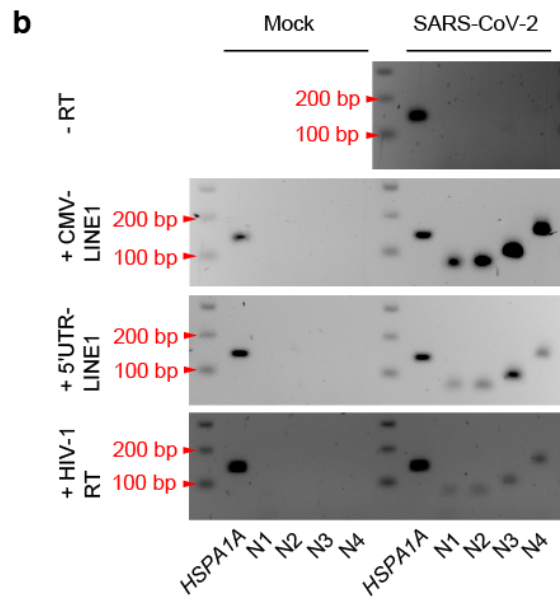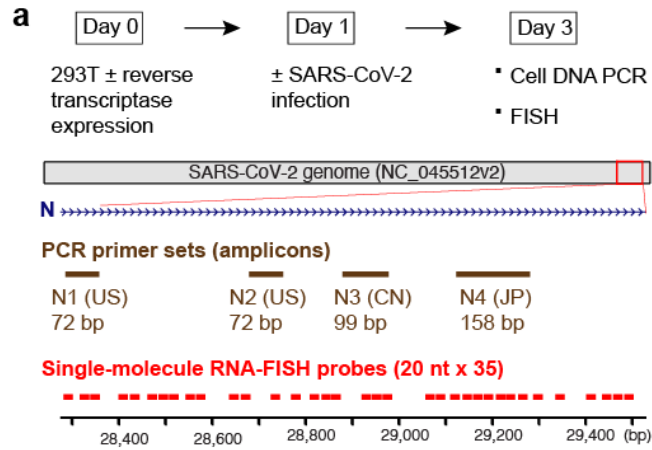
503 **Figure 2. SARS-CoV-2 RNA can be reverse-transcribed and integrated into the host**

504 **genome in cells with reverse transcriptase expression. a)** Experimental workflow (top), PCR

505 primer sets (shown as amplicons, brown) and single-molecule RNA-FISH probes (red) to detect

506 reverse-transcription and integration of SARS-CoV-2 nucleocapsid (N) sequence (middle, blue).

507 **b)** PCR detection of SARS-CoV-2 N sequences in cellular DNA purified from mock (left) or

508 SARS-CoV-2 (right) infected HEK293T cells without or with transfection of human LINE-1

509 (CMV-LINE1 or 5'UTR-LINE1) or HIV-1 RT. *HSPA1A*: human *HSPA1A* gene as control; N1 –

510 N4: SARS-CoV-2 N sequences as shown in **a)**. **c)** qPCR detection and copy-number estimation

511 of SARS-CoV-2 N sequences on gel-purified HEK293T genomic DNA. *HSPA1A*: human

512 *HSPA1A* gene as a reference; N2, N3: SARS-CoV-2 N sequences as shown in **a)**. Three

513 biological replicates; mean ± standard error of the mean (s.e.m.). **d)** Single-molecule RNA-FISH

514 (red) targeting SARS-CoV-2 N sequence using probes shown in **a)** plus LINE-1 ORF1 protein

515 immuno-staining (green) and merged channels with DAPI (blue) in SARS-CoV-2 infected

516 HEK293T cells with (left) or without (right) transfected LINE-1. Insets: 2.5x enlargement of

517 region in white-box to show nuclear signals of SARS-CoV-2 N sequence (white arrows). Images

518 were single z-slices from 3D optical sections (0.2-μm z-steps). **e)** Fraction of HEK293T cells

519 infected by SARS-CoV-2 (indicated by cytoplasmic FISH signals) showing nuclear FISH signals

520 of N sequence with (+ LINE-1 O/E, n = 75) or without (- LINE-1 O/E, n = 57) LINE-1

521 overexpression (indicated by LINE-1 ORF1 protein immuno-staining). Combination of two

522 independent cell samples; Chi-Square Test of Homogeneity.

a Differential expression analysis of human repetative elements

LINE-1s with differential expression (FDR < 0.05))
LINE-1s without differential expression

523

524 **Figure 3. LINE-1 expression as an endogenous reverse-transcriptase source in human cells**

525 **is induced by SARS-CoV-2 infection and cytokine-containing conditioned media treatment.**

526 **a)** RNA-Seq (GSE147507, see **Supplementary Figure 1a**) differential expression analysis for

527 all human repetitive elements in SARS-CoV-2 versus mock-infected Calu3 (left) or NHBE

528 (right) cells. Volcano plots showing -Log10 p-values (y-axis) versus Log2 fold-changes (x-axis)

529 for all human repetitive elements with (orange circle) or without (grey circle) significant

530 expression changes (SARS-CoV-2 versus mock-infected); dots: LINE-1 families with (dark

531 blue) or without (light blue) significant expression changes. **b)** Scatter plot showing human-

532 CoV2 chimeric read number (per million total mappable reads, y-axis) versus SARS-CoV-2 read

533 fraction of total mappable reads (x-axis) in published RNA-Seq (GSE147507, see

534 **Supplementary Figure 1a**) from infected Calu3 (magenta) or NHBE (red) cells. **c)** Endogenous

535 LINE-1 expression fold-changes between SARS-CoV-2 versus mock-infected Calu3 cells

536 measured by RT-qPCR with primers probing 5'UTR, ORF1, or 3'UTR regions of LINE-1.

537 Reference genes: *GAPDH* and *TUBB*. Three biological replicates; mean ± s.e.m. **d)** qPCR

538 detection and copy-number estimation of SARS-CoV-2 N sequences in mock (green) or SARS-

539 CoV-2 infected (magenta) Calu3 cellular DNA. *HSPA1A*: human *HSPA1A* gene as a reference;

540 N2, N3: SARS-CoV-2 N sequences as shown in **Figure 1a**. Three biological replicates; mean ±

541 s.e.m; n.d.: not detected. **e)** Gel purification of large-fragment genomic DNA (yellow box, top)

542 from SARS-CoV-2 infected Calu3 cells and PCR detection of SARS-CoV-2 N sequences in the

543 purified genomic DNA (bottom) with same primer sets as in **d)**. **f)** Endogenous LINE-1

544 expression fold-changes in Calu3 cells comparing Myeloid conditioned versus basal media

545 treatment measured by RT-qPCR with primers probing 5'UTR, ORF1, or 3'UTR regions of

546 LINE-1. Reference genes: *GAPDH* and *TUBB*. Three biological replicates; mean ± s.e.m. **g)**

25

547   LINE-1 ORF1 protein immuno-staining (magenta, same exposure and intensity scaling) plus

548   merged channels with DAPI (blue) in Calu3 cells cultured in basal or myeloid conditioned

549   media. Scale bar: 10 μm. **h)** Normalized cellular total LINE-1 ORF1p immuno-staining signals

550   of Calu3 cells cultured in basal (n = 84, mean = 1.0, median = 0.9) or myeloid conditioned media

551   (n = 126, mean = 1.7, median = 1.5). Combination of two independent cell samples. Box plots

552   show median (inside line), means (red dot), interquartile range (IQR, box), and upper/lower

553   quartile ± 1.5-times IQRs (whiskers). Welch's t-test.

26

554   **Supplementary Figures**

a

| Data ID (GEO)	Sample	SARS-CoV-2 infection	RNA-Seq read length (nt)	Min overhang (threshold)
GSE147507	Calu3, A549-ACE2, NHBE	MOI 2, 24 hrs	150	50
GSE153277	iPSC-derived type II alveolar epithelial-like cells	MOI 5, 4 days	76+76	50
GSE156754	iPSC-derived cardiomyocytes	MOI 0.1, 48 hrs	76+76	50
GSE157852	iPSC-derived choroid plexus organoids	MOI 0.1-0.05, 72 hrs	72	20
GSE153684	Cultured organoid from pediatric stomach tissue (age 11 year)	MOI 1	51+51	20
GSE145926	Bronchoalveolar lavage fluid (BALF) cells	COVID19+ patient (severe)	26+100	20
GSE154998	Blood buffy coat cells	COVID19+ patient (ICU)	72	20

b
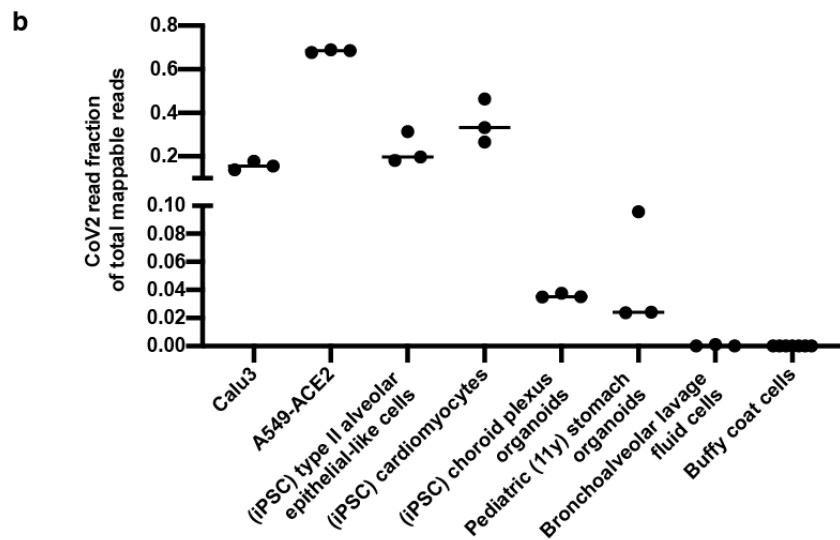
c   **Human - CoV2 chimeric read example from Calu3 (infected) RNA-Seq:**

57 nt mapped to human Chromosome X
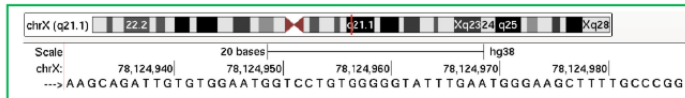
AAGCAGATTGTGTGGAATGGTCCTGTGGGGGTATTTGAATGGGAAGCTTTT
GCCCGGCGCGTAGTACGATCGAGTGTACAGTGAACAATGCTAGGGAGAGC
TGCCTATATGGAAGAGCCCTAATGTGTAAAATTAATTTTAGTAGTGCT

92 nt mapped to SARS-CoV-2 genome

555

27
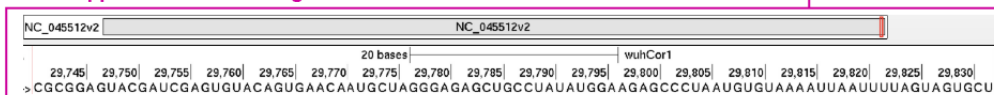
556 **Supplementary Figure 1. Human – SARS-CoV-2 chimeric reads identified from published**

557 **RNA-Seq data. a)** Published data used to identify human – CoV2 chimeric reads summarizing

558 GEO accession number (data ID), sample type, infection method/type (MOI: Multiplicity Of

559 Infection), RNA-Seq format (single or paired-end with read length), and threshold to call

560 chimeric reads (Min overhang: minimum number of bases mapped to either human or SARS-

561 CoV-2 genome/transcriptome to call a chimeric reads). **b)** Comparison of SARS-CoV-2 read

562 fraction of total mappable reads in the published RNA-Seq datasets as shown in **a)**. **c)** One

563 chimeric read example (149 nt) from Calu3 (infected) RNA-Seq with 57 nt mapped to human

564 Chromosome X (green) and 92 nt (magenta) mapped to the SARS-CoV-2 genome.

565
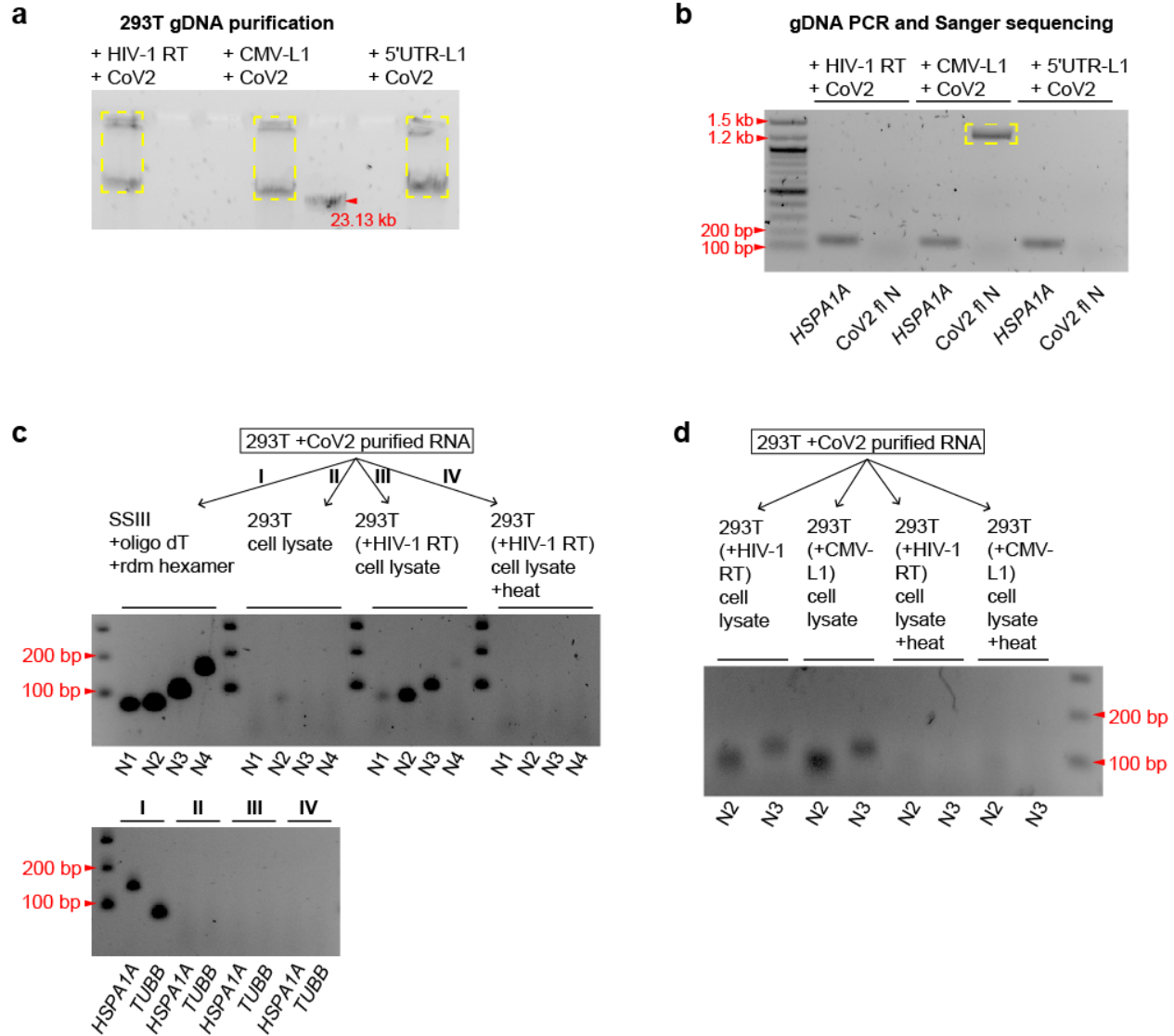
Supplementary Figure 2. SARS-CoV-2 RNA can be reverse-transcribed *in vivo* and *in vitro* by other sources of reverse transcriptase. a) Gel purification of large-fragment genomic DNA (yellow boxes) from SARS-CoV-2 infected HEK293T cells carrying transfected HIV-1 RT, CMV-LINE1 or 5'UTR-LINE1. **b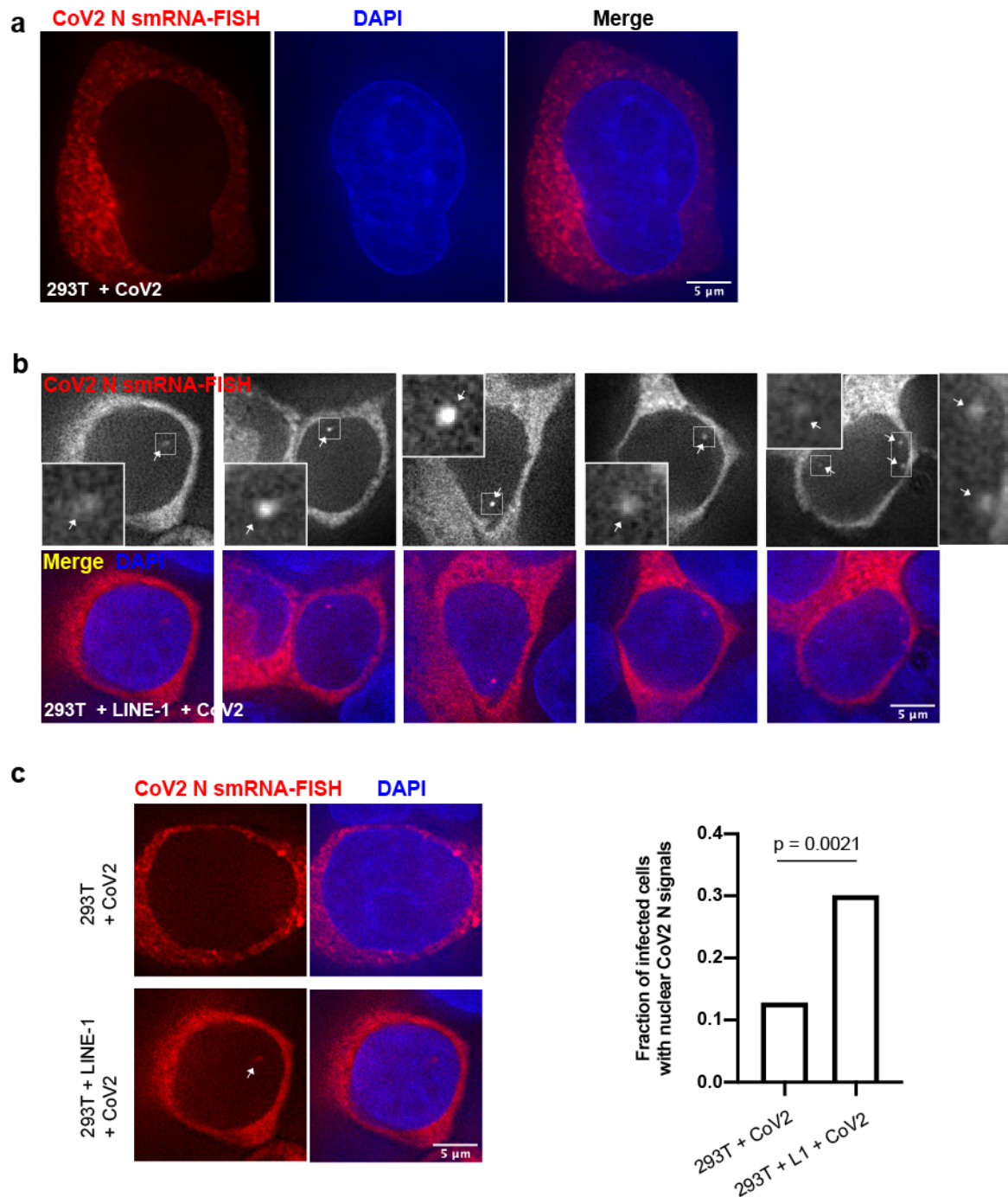)** Cloning and Sanger sequencing of DNA copy of full-length SARS-CoV-2 N sequence (CoV2 fl N, yellow box) from gel-purified HEK293T genomic DNA as shown in **a)**. CoV2 fl N: amplification of full-length N sequence (1.26 kb) by primers targeting the two ends of N. *HSPA1A*: human *HSPA1A* gene as a control. Note that we can only detect full-length N sequence in gDNA from cells with CMV-LINE-1 expression, corresponding

574   to the high copy-number of integrated N sequences as shown in **Figure 2c**. **c)** *In vitro* reverse

575   transcription of SARS-CoV-2 RNA by adding RNA purified from SARS-Cov-2 infected

576   HEK293T cells to a commercial reverse transcriptase (I, SSIII, with oligo dT and random

577   hexamer primers, positive control), or HEK293T cell lysate (II), or lysates of HEK293T cells

578   expressing HIV-1 reverse transcriptase without (III) or with (IV) heat inactivation. Gel images

579   showing PCR detection of SARS-CoV-2 N sequences from the *in vitro* reverse transcription

580   products using primer sets (N1 – N4) as shown in **Figure 2a**. *HSPA1A* and *TUBB*: PCR primer

581   sets against human *HSPA1A* and *TUBB* genes as controls. **d)** Same *in vitro* reverse transcription

582   and PCR detection setup as in **c)** using lysates of HEK293T cells expressing HIV-1 reverse

583   transcriptase or human LINE-1.

584

585   Supplementary Figure 3. SARS-CoV-2 N RNA signals detected in cell nuclei by single-

586   molecule RNA-FISH. a-b) Example images of single-molecule RNA-FISH (red/grey) targeting

587   SARS-CoV-2 N sequence using probes shown in **Figure 2a** and merged channels with DAPI

588   (blue) in SARS-CoV-2 infected HEK293T cells without (**a**) or with (**b**) human LINE-1

589   transfection. Insets in **b)**: 4x enlargement of regions in white-boxes to show nuclear signals of

590   SARS-CoV-2 N sequence (white arrows). **c)** Comparison of nuclear N RNA-FISH signals in

591   SARS-CoV-2 infected HEK293T cells without or with human LINE-1 transfection. Left:

592   example images as in **a)** and **b);** Right: fraction of HEK293T cells infected by SARS-CoV-2

593   (indicated by cytoplasmic FISH signals) showing nuclear N RNA-FISH signals in cell

594   populations without (left bar, n = 109) or with (right bar, n = 103) CMV-LINE-1 plasmid

595   transfection (~80% transfection efficiency). Combination of two independent cell samples; Chi-

596   Square Test of Homogeneity. All images shown were single z-slices from 3D optical sections

597   (0.2-μm z-steps).

598
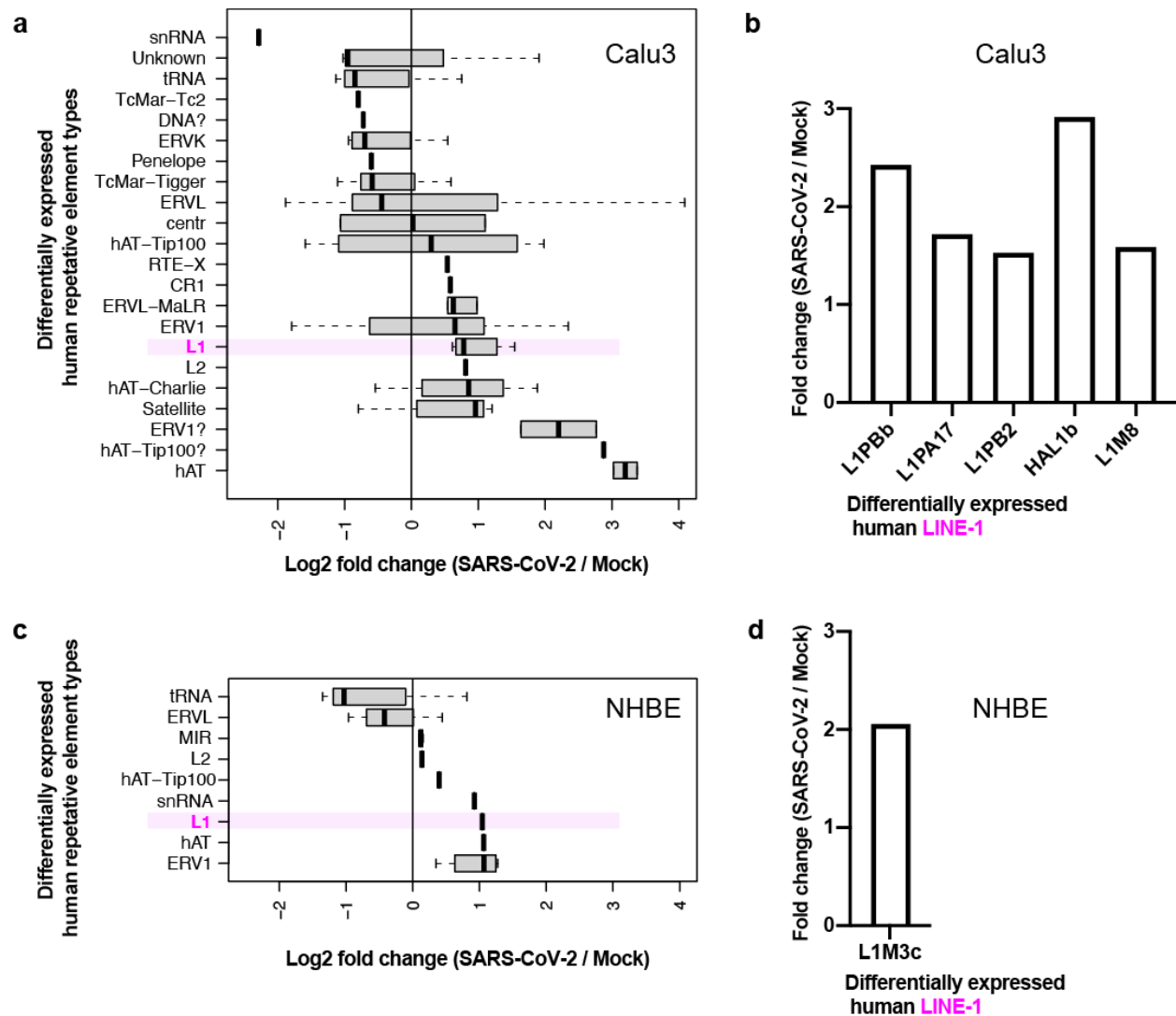
Supplementary Figure 4. LINE-1 induction in human cells correlates with SARS-CoV-2 infection. a, c) Log2 fold-changes (x-axis) of different types of human repetitive elements (y-axis) with significant (FDR < 0.05) expression changes in SARS-CoV-2 versus mock infected Calu3 (**a**) or NHBE (**c**) cells from published RNA-Seq data (GSE147507). **b, d)** Fold changes (y-axis) of different human LINE-1 families (x-axis) with significant (FDR < 0.05) expression changes in SARS-CoV-2 versus mock infected Calu3 (**b**) or NHBE (**d**) cells from published RNA-Seq data (GSE147507, see **Supplementary Figure 1a**).

606
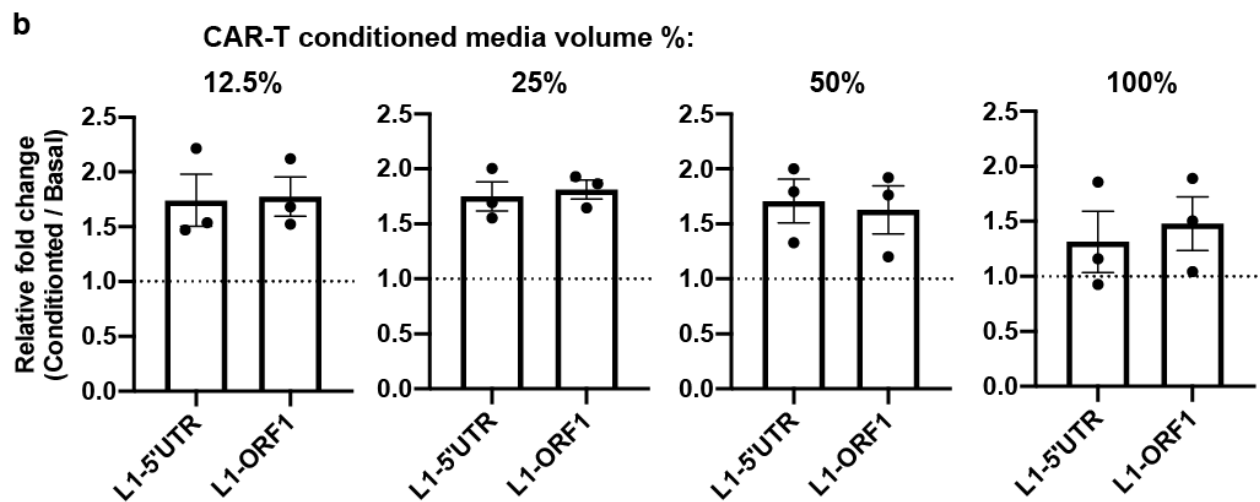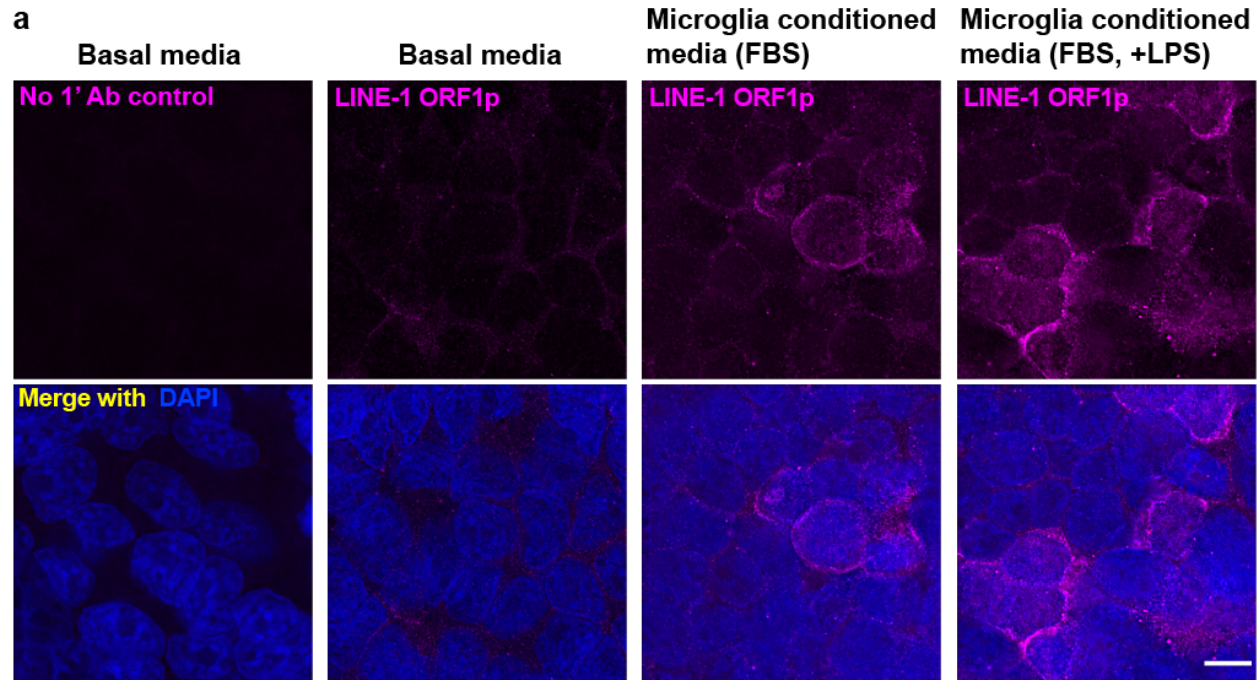
Supplementary Figure 5. Cytokine containing media treatment triggers LINE-1 expression in human cells. a) LINE-1 ORF1 protein immuno-staining (magenta, same exposure and intensity scaling, 1st column: no primary antibody control) plus merged channels with DAPI (blue) in HEK293T cells cultured in basal (1st and 2nd columns) or microglia conditioned media (3rd column) or LPS-treated microglia conditioned media (4th column). Scale bar: 10 μm. **b)** Endogenous LINE-1 expression fold-changes in Calu3 cells between CAR-T conditioned

613   (diluted with basal media at indicated percentage in volume) versus basal media treatment

614   measured by RT-qPCR with primers probing 5'UTR, ORF1, or 3'UTR regions of LINE-1.

615   Reference genes: *GAPDH* and *TUBB*. Three independent cell samples treated with two batches

616   of media; mean ± s.e.m.

617 **Supplementary Tables**

618

619 **Supplementary Table 1.** Primer sequences used in this study

Species	Name	Sequence
SARS-CoV-2	N1	Forward: GACCCCAAAATCAGCGAAAT Reverse: TCTGGTTACTGCCAGTTGAATCTG
SARS-CoV-2	N2	Forward: GGGAGCCTTGAATACACCAAAA Reverse: TGTAGCACGATTGCAGCATTG
SARS-CoV-2	N3	Forward: GGGGAACTTCTCCTGCTAGAAT Reverse: CAGACATTTTGCTCTCAAGCTG
SARS-CoV-2	N4	Forward: AAATTTTGGGGACCAGGAAC Reverse: TGGCACCTGTGTAGGTCAAC
SARS-CoV-2	N (full length)	Forward: ATGTCTGATAATGGACCCCAAAAT Reverse: TTAGGCCTGAGTTGAGTCAGC
Human	*HSPA1A*	Forward: ATCTCCACCTTGCCGTGTT Reverse: ATCCAGTGTTCCGTTTCCAG
Human	*TUBB*	Forward: TCCCTAAGCCTCCAGAAACG Reverse: CCAGAGTCAGGGGTGTTCAT
Human	*GAPDH*	Forward: GTCTCCTCTGACTTCAACAGCG Reverse: ACCACCCTGTTGCTGTAGCCAA
Human	LINE-1-5'UTR	Forward: GACGCAGAAGACGGTGATTT Reverse: TCACCCCTTTCTTTGACTCG
Human	LINE-1-ORF1	Forward: CTCGGCAGAAACCCTACAAG Reverse: CCATGTTTAGCGCTTCCTTC
Human	LINE-1-3'UTR	Forward: CATGGAATACTATGCAGCCATAAA Reverse: TCCCACCTATGAGTGAGAA

620

36

621 **Supplementary Table 2.** Single-molecule RNA FISH probe sequences used in this study

Name	Sequence
Cov2NC_1	tgattttgggggtccattatc
Cov2NC_2	agggtccaccaaacgtaatg
Cov2NC_3	tggttactgccagttgaatc
Cov2NC_4	ttattgggtaaaccttgggg
Cov2NC_5	tgagagcggtgaaccaagac
Cov2NC_6	cctcgagggaatttaaggtc
Cov2NC_7	ttggtgttaattggaacgcc
Cov2NC_8	aatttggtcatctggactgc
Cov2NC_9	accacgaattcgtctggtag
Cov2NC_10	gatctttcattttaccgtca
Cov2NC_11	tttgttagcaccatagggaa
Cov2NC_12	cagttgcaacccatatgatg
Cov2NC_13	gattgcagcattgttagcag
Cov2NC_14	tagaagcctttggcaatgt
Cov2NC_15	acgagaagaggcttgactgc
Cov2NC_16	actgttgcgactacgtgatg
Cov2NC_17	tgcctggagttgaatttctt
Cov2NC_18	cagcaaagcaagagcagcat
Cov2NC_19	agctggttcaatctgtcaag
Cov2NC_20	tttaccagacattttgctct
Cov2NC_21	tatgctttagtggcagtacg
Cov2NC_22	gccgaaagcttgtgttacat
Cov2NC_23	aatttccttgggtttgttct
Cov2NC_24	cttgtctgattagttcctgg
Cov2NC_25	ggccaatgtttgtaatcagt
Cov2NC_26	tgggggcaaattgtgcaatt
Cov2NC_27	cgacattccgaagaacgctg
Cov2NC_28	gtgtgacttccatgccaatg
Cov2NC_29	tgtgtaggtcaaccacgttc
Cov2NC_30	atttggatctttgtcatcca
Cov2NC_31	ttgtatgcgtcaatatgctt
Cov2NC_32	ggtaaggcttgagtttcatc
Cov2NC_33	gaagagtcacagtttgctgt
Cov2NC_34	atcatccaaatctgcagcag
Cov2NC_35	ggattgttgcaattgtttgg

622