1 **High-content screening of coronavirus genes for innate immune suppression reveals**

2 **enhanced potency of SARS-CoV-2 proteins**

3

4 **Authors:** Erika J. Olson[1,2]*, David M. Brown[4]*, Timothy Z. Chang[1-2], Lin Ding[5], Tai L. Ng[1], H.

5 Sloane Weiss[1], Yukiye Koide[1,2], Peter Koch[1], Nathan Rollins[1], Pia Mach[1,2], Tobias Meisinger[1],

6 Trenton Bricken[1], Joshua Rollins[1], Yun Zhang[5], Colin Molloy[1], Yun Zhang[5], Bridget N.

7 Queenan[1], Timothy Mitchison[1], Debora Marks[1], Jeffrey C. Way[1-3#], John I. Glass[5#], Pamela A.

8 Silver[1-3#**]

9
10 [1]Department of Systems Biology, Harvard Medical School, 220 Longwood Avenue, Boston, MA
11 02115 USA.
12
13 [2]Wyss Institute, Harvard University. 3 Blackfan St., Boston, MA 02115 USA.
14
15 [3]Laboratory of Systems Pharmacology, Harvard Medical School, 200 Longwood Avenue,
16 Boston, MA 02115 USA.
17
18 [4] J. Craig Venter Institute, 9605 Medical Center Drive, Suite 150, Rockville, MD 20850.
19
20 [5] J. Craig Venter Institute, 4120 Capricorn Lane, La Jolla, CA 92037.
21
22 *Co-first authors
23
24 #Co-corresponding authors
25
26 **Lead contact
27
28 Correspondence: pamela_silver@hms.harvard.edu (P.A.S.), jglass@jcvi.org (J.G.),
29 jeff.way@wyss.harvard.edu (J.C.W.)
30
31
32
33
34

Summary

Suppression of the host intracellular innate immune system is an essential aspect of viral replication. Here, we developed a suite of medium-throughput high-content cell-based assays to reveal the effect of individual coronavirus proteins on antiviral innate immune pathways. Using these assays, we screened the 196 protein products of seven coronaviruses (SARS-CoV-2, SARS-CoV-1, 229E, NL63, OC43, HKU1 and MERS). This includes a previously unidentified gene in SARS-CoV-2 encoded within the Spike gene. We observe immune-suppressing activity in both known host-suppressing genes (e.g., NSP1, Orf6, NSP3, and NSP5) as well as other coronavirus genes, including the newly identified SARS-CoV-2 protein. Moreover, the genes encoded by SARS-CoV-2 are generally more potent immune suppressors than their homologues from the other coronaviruses. This suite of pathway-based and mechanism-agnostic assays could serve as the basis for rapid *in vitro* prediction of the pathogenicity of novel viruses based on provision of sequence information alone.

Key words: coronavirus, SARS-CoV-2, IRF-3, NFkB, STAT1, innate immunity, virus, cell-based screening

51 **Introduction**

52   Innate immunity is the human body's first line of defense against pathogens. We and

53 others hypothesize that viral pathogenicity may correlate with the degree to which a virus

54 inhibits the antiviral innate immune response. For example, Ebola virus encodes several genes

55 that suppress innate immune signaling, with the result that this virus can reach extremely high

56 titers before the immune system mounts an effective response (Kuhl et al., 2012). Hoffmann et

57 al. (2015) described the interaction between the host innate immune response and its

58 suppression by viral genes as a 'molecular arms race' and summarized how most or all viruses

59 encode such suppressing factors. Mechanisms of suppression include: inhibition of nuclear

60 protein transport, proteolysis of signaling factors by linkage to ubiquitin ligases, and broad

61 inhibition of host gene expression through inhibition of mRNA export, splicing and/or translation.

62   Increasing human population density and travel enhance the probability of viral

63 pandemics. In the case of SARS-CoV-2, there was a ~30 day lag time between determination

64 of the viral sequence and the full appreciation of its pandemic potential based on epidemiology.

65 It would be ideal to rapidly estimate the pandemic potential of a virus based on its sequence.

66 We envision a reductionist approach, in which gene activities are measured, mechanisms are

67 inferred, and a calculation of pathogenicity is made. While this goal has not yet been realized,

68 essentially all new viruses fall into known families, in which the pathogenicity of some members

69 is known. By assaying a new virus in parallel to similar viruses of known pathogenicity, it should

70 be possible to rapidly and accurately assess the danger of a new strain *in vitro*.

71   With this issue in mind, we generated a suite of 'high-content' cell-based assays which

72 reveal the ability of viral proteins to inhibit the innate immune system. These assays can be

73 employed within a month or less, enabling rapid identification of pandemic potential for newly

74 detected viruses. Medium and high-throughput 'high content' cell-based screens have been

75 successfully used to identify drugs that affect cellular pathways or phenotypes agnostic to

76 specific protein targets. For example, Kau et al. (2003) executed a cell-based screen for

77 modulators of the nuclear/cytoplasmic localization of a Forkhead transcription factor based on

78 immunostaining; this effort ultimately led to the development of selinexor, an approved anti-

79 cancer drug (Syed et al., 2019). Similarly, Mayer et al. (1999) screened compounds for the

80 ability to arrest cells in mitosis using a cell-based assay for enhancement of mitotic

81 phosphorylation sites; that effort led to development of Eg5 inhibitors, which entered clinical

82 trials for cancer treatment (Rath et al., 2012).

83   To identify and understand the mechanisms that distinguish SARS-CoV-2 from other

84 coronaviruses, we constructed mammalian and yeast expression vectors for genes encoded by

3

85   SARS-CoV-2, SARS-CoV-1, MERS, 229E, NL63, HKU1, and OC43, and tested these genes for

86   their ability to disrupt the innate immune system. Two assays are target-based screens that

87   provide mechanistic insight into how individual viral genes disable the innate immune response

88   measuring each gene's effect on specific components of the Type I interferon response (IRF-3,

89   NFkB, and STAT1 signaling pathways), and inflammatory gene expression (an NFkB-regulated

90   reporter gene) (Koch et al., 2018). A third assay is a mechanism-agnostic screen that reveals

91   the overall ability of viral proteins to successfully inhibit the immune response to a degree that

92   supports viral replication (Brown et al., 2016). In doing so, we confirm the immune-suppressing

93   role of certain coronavirus genes, implicate others, and report the particularly potent effect of

94   SARS-CoV-2 on the innate immune system. In addition, we identify a novel protein encoded by

95   SARS-CoV-2 which demonstrates potent immune-suppressing ability via the IRF-3 pathway.

96

97

98

99

100

Results

101

102   We designed mammalian expression vectors for all proteins encoded by seven human-

103   infecting coronaviruses: SARS-CoV-2, SARS-CoV-1, MERS-CoV, HCoV 229E, HCoV NL63,

104   HCoV HKU1, and HCoV OC43 (Supplemental Information). The RNA of coronaviruses can be

105   conceptually divided into two parts: (i) the 5' two-thirds of the genome which encodes a ~7,000

106   amino acid polyprotein termed ORF1ab that is cleaved into proteins NSP1-16, and (ii) the 3'

107   one-third of the genome which has multiple separate ORFs. This latter region encodes the

108   universal proteins Spike, Envelope, Membrane, and Nucleocapsid, plus a number of small so-

109   called "accessory" proteins that vary from virus to virus, some of which are encoded by

110   overlapping ORFs. Of the three highly pathogenic coronaviruses, SARS-CoV-2 and SARS-CoV-

111   1 are closely related, while MERS-CoV is somewhat related (Figure 1A). Four human-infecting

112   coronaviruses (HCoV OC43, HCoV HKU1, HCoV NL63, and HCoV 229E) are quite distantly

113   related; these viruses generally do not cause severe disease in healthy adults, although it is

114   unclear whether their lack of lethality is because they do not suppress the innate immune

115   system as severely as the SARS/MERS group or an alternative explanation such as

116   suppression by pre-existing immunity to related viruses. The vectors were constructed by Twist

117   Biosciences (Methods, Supplemental Information) and used directly for mammalian cell

118   transfection, or individual viral genes were inserted into yeast for the yeast-mammalian cell

119   fusion assay.

120   To ensure we screened all SARS-CoV-2 genes, we scanned the genome (Wuhan-hu-1,

121   MN908947) for potential ORFs by searching for similarity to known proteins. Enumerating all

122   translations in the 6 reading frames and aligning to sequence profiles of protein families from

123   PFAM (Mistry et al. 2020 doi: 10.1093/nar/gkaa913) (Methods), we identified a candidate ORF

124   15 (CaORF15) encoded in an alternative reading frame within the Spike sequence (Figure 1).

125   CaORF15 is 87 amino acids long (by comparison, Orf9b is encoded in an alternate frame of

126   Nucleoprotein and is 98 amino acids long). We predict CaORF15 to contain three

127   transmembrane alpha helices 1-21, 36-55, and 64-87 (Hofmann & Stoffel 1993) and share

128   similarity to the GtrA-like protein family (Guan et al., 1999; Gandini et al., 2017). The

129   corresponding region in SARS-CoV-1 is very divergent (56% identity) and does not encode

130   CaORF15. However, CaORF15 is encoded in the SARS-CoV-2-related bat virus RaTG13 and

131   shares higher identity than the remainder of Spike (98% versus 93% respectively) (Figure 1C).
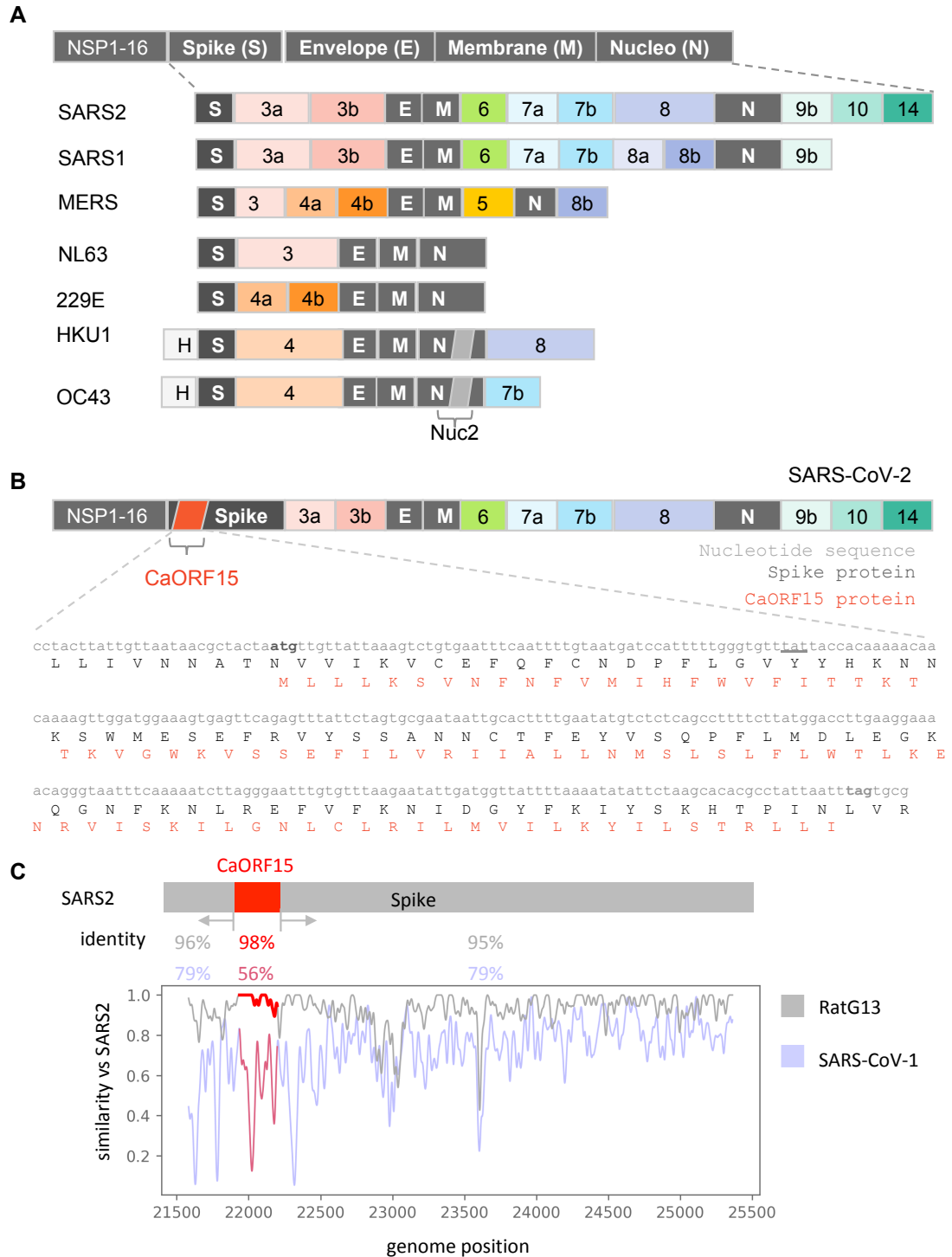
5

A.

NSP1-16	Spike (S)	Envelope (E)	Membrane (M)	Nucleo (N)	

SARS2: S 3a 3b E M 6 7a 7b 8 N 9b 10 14

SARS1: S 3a 3b E M 6 7a 7b 8a 8b N 9b

MERS: S 3 4a 4b E M 5 N 8b

NL63: S 3 E M N

229E: S 4a 4b E M N

HKU1: H S 4 E M N 8

OC43: H S 4 E M N 7b

Nuc2

B. SARS-CoV-2

NSP1-16 Spike 3a 3b E M 6 7a 7b 8 N 9b 10 14

CaORF15

Nucleotide sequence
Spike protein
CaORF15 protein

```
cctacttattgttaataacgctacta atgttgttattaaagtctgtgaatttcaatttgtaatgatccattttttgggtgtttattaccacaaaaacaa
          L  L  I  V  N  N  A  T  N  V  V  I  K  V  C  E  F  Q  F  C  N  D  P  F  L  G  V  Y  Y  H  K  N  N
                              M  L  L  L  K  S  V  N  F  N  F  V  M  I  H  F  W  V  F  I  T  T  K  T

caaaagttggatgggaaagtgagttcagagttttattctagtgcgaataattgcacttttgaatatgtctctcagcctttttcttatggaccttgaaggaaa
  K  S  W  M  E  S  E  F  R  V  Y  S  S  A  N  N  C  T  F  E  Y  V  S  Q  P  F  L  M  D  L  E  G  K
     T  K  V  G  W  K  V  S  S  E  F  I  L  V  R  I  I  A  L  L  N  M  S  L  S  L  F  L  W  T  L  K  E

acagggtaatttcaaaaatcttagggaatttgtgtttaagaatattgatggttattttaaaatatattctaagcacacgcctattaatt tagtgcg
  Q  G  N  F  K  N  L  R  E  F  V  F  K  N  I  D  G  Y  F  K  I  Y  S  K  H  T  P  I  N  L  V  R
     N  R  V  I  S  K  I  L  G  N  L  C  L  R  I  L  M  V  I  L  K  Y  I  L  S  T  R  L  L  I
```

C.

CaORF15

SARS2: Spike

identity: 96% 98% 95%
         79% 56% 79%

132

Figure 1. Identification of an additional ORF within the Spike coding sequence of SARS-CoV-

2. A. Schematic of protein coding regions of the seven coronaviruses tested. The 5' two-thirds of the

genome encodes a polyprotein found in all coronaviruses which is cleaved into proteins NSP1-16.

The 3' one-third of the genome encodes the universal proteins Spike, Envelope, Membrane, and

Nucleocapsid, plus several small proteins that vary from virus to virus. **B.** To identify candidate

138   ORFs in SARS-CoV-2, we translated all six reading reading frames and identified a possible 87-
139   amino acid ORF within Spike in an alternative reading frame. Scanning the sequences against
140   known PFAM protein domains (ref PFAM database 2021 doi: 10.1093/nar/gkaa913) revealed a
141   weak similarity to part of the "GTRA-like" PFAM domain (PF04138) (Guan et al., 1999; Gandini et
142   al., 2017). **C.** Similarity plot for the SARS-CoV-2 genomic sequence within the Spike region as
143   aligned with the bat coronavirus RaTG13 and SARS-CoV-1. The region encoding CaORF15 is more
144   conserved than the other segments of the Spike coding sequence for RaTG13 and less conserved
145   with respect to SARS-CoV-1.
146
147   **Construction of cell-based assay for suppression of Type I interferon pathways**
148   A striking feature of SARS-CoV-2 is the ability to infect humans and produce virus that
149   can replicate and spread to other humans without inducing symptoms. Typically, 'flu-like
150   symptoms' resulting from viral infection are a consequence of Type I interferon expression
151   (Maher et al., 2007). Patients who receive interferon α for hepatitis or interferon β for multiple
152   sclerosis experience flu-like symptoms from these proteins alone.
153   We therefore designed a cell-based assay to reveal the effect of individual coronavirus
154   proteins on Type 1 interferon pathways. The innate immune response to viruses relies on
155   cytokine signaling induced through the following cascade (Figures 2A and 2B): (i) out-of-place
156   nucleic acids (e.g. dsDNA or dsRNA in the cytoplasm) are detected by pattern-recognition
157   receptors (PRRs) (Hoffmann et al., 2015); (ii) PRRs initiate signaling pathways that cause
158   translocation of the transcription factors Interferon Response Factor 3 (IRF-3) and Nuclear
159   Factor κB (NFκB) to the nucleus; (iii) activated IRF-3 and NFκB in the nucleus induce the
160   expression of Type I interferons and other cytokines to warn the surrounding cells of incoming
161   viral attack, initiate the adaptive immune response, and attract specialized immune cells; and
162   (iv) signaling by Type I interferon receptors on the surface of surrounding cells causes
163   translocation of Signal Transducer And Activator Of Transcription (STAT) transcription factors to
164   trigger a state of antiviral hypervigilance (Hoffmann et al., 2015).
165   Combined activation of the transcription factors IRF-3 and NFkB is central to
166   development of an antiviral response. Koch et al. (2018) recently developed conditions in which
167   the nuclear/cytoplasmic localization of both proteins could be simultaneously visualized using
168   non-cross-reacting primary and secondary antibodies. This system has been adapted for high-
169   throughput screening to identify compounds that activated one or both factors (e.g., candidate
170   anti-cancer or antiviral compounds), or inhibited the response to added cGAMP (candidate anti-
171   inflammatory compounds). We adapted this screen for the present work.

7

172   Assay 1 was designed to test coronavirus genes for suppression of innate immunity via

173   disruption of IRF-3, NFkB, and STAT1 (Figures 2 and 3). We used a non-transformed, tert-

174   immortalized fibroblast cell line, BJ-5ta cells because they are thought to have an intact innate

175   immune system. However, in normal cells, DNA transfection alone stimulates the cGAS/STING

176   pathway (Figure 2E): cGAS binds to cytoplasmic DNA and produces the cyclic dinucleotide

177   cGAMP as a second messenger that activates STING and downstream signaling steps (Xiao

178   and Fitzgerald et al., 2013). We therefore generated a CRISPR knockout of cGAS (Methods).

179   Because the coronaviruses are RNA viruses, removal of cGAS is not expected to be relevant in

180   the present experiments; cGAS and cGAMP are inhibited or inactivated by certain genes of

181   DNA viruses, such as poxin (Eaglesham et al., 2019), which cleaves cGAMP, but such an

182   activity is not expected to benefit an RNA virus.

183
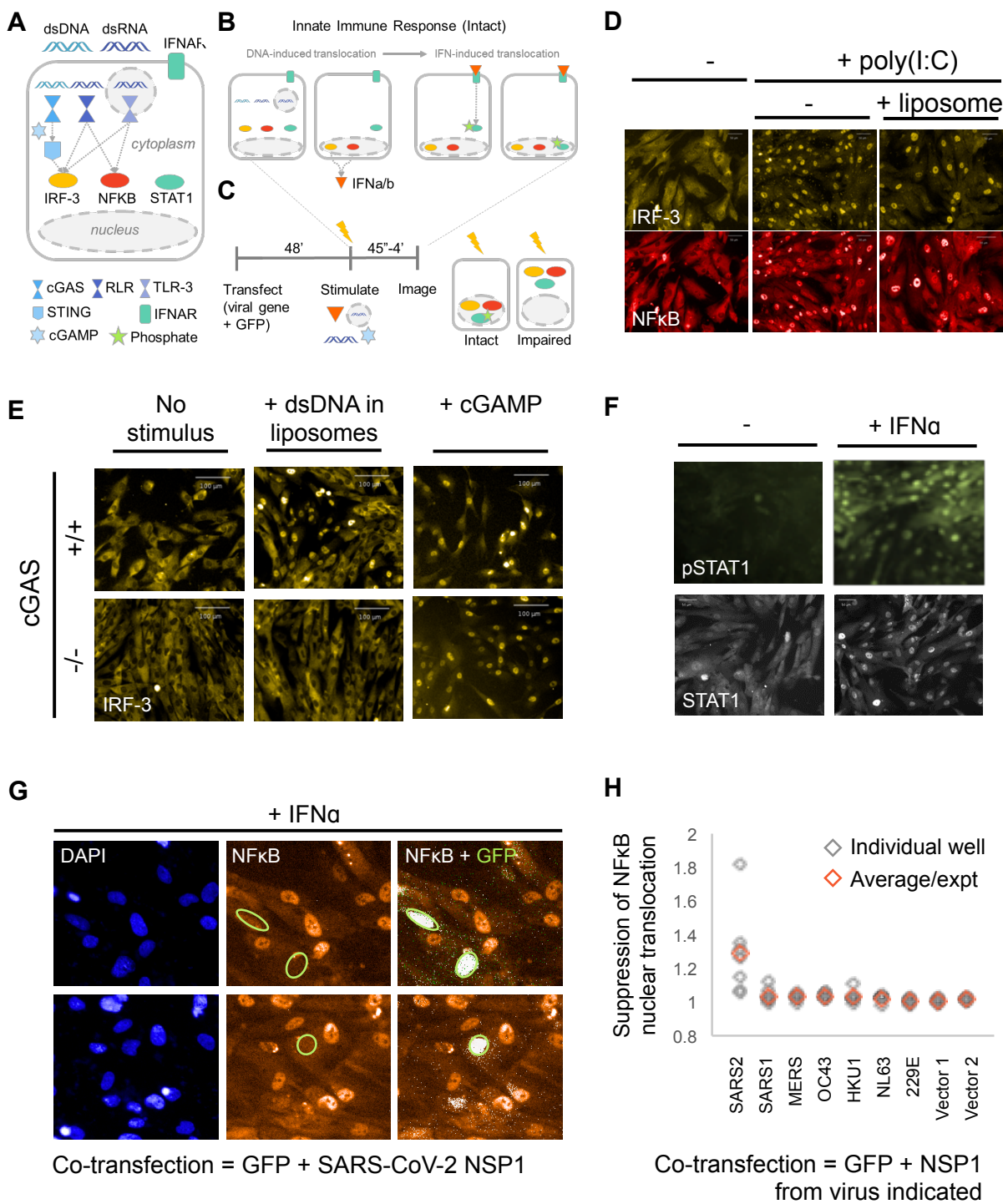
Figure 2. Construction of high-throughput antiviral innate immune response assay. A. Innate
immune system signaling pathways that respond to viral infection. Viruses encode proteins that
disrupt these pathways, by directly or indirectly inhibiting nuclear translocation or causing
degradation of three key transcription factors: IRF-3, NFκB, and/or STAT1. B. Viral infection (i.e.,

184
185
186
187
188

9

189 presence of nucleic acids in an inappropriate cellular context) activates the transcription factors IRF-
190 3 and NFκB. Together, IRF-3 and NFκB activate expression of type I interferons, which are secreted
191 and act locally to stimulate the STAT1 pathway. IRF-3, NFκB and STAT1 all translocate into the
192 nucleus during signaling. **C.** Experimental paradigm for testing virus genes for modulation of the IRF-
193 3, NFκB, and STAT1 signaling pathways. BJ-5ta cGAS-/- cells are transiently transfected with a
194 viral gene expression vector. The innate immune response pathways can be initiated *in vitro* by
195 addition of cGAMP (a second messenger), extracellular double-stranded RNA (dsRNA), cytoplasmic
196 dsRNA (by transfection), or interferon alpha (IFNα). Cells are then fixed and stained with antibodies
197 against IRF-3, NFκB, STAT1, and/or pTyr701-STAT1, incubated with secondary antibodies, and
198 then imaged. The effect of the viral protein on the innate immune response can be determined by
199 quantifying the degree of nuclear accumulation. **D.** BJ-5ta cells treated with poly(I:C) or with
200 liposomal poly(I:C) show translocation of both IRF-3 and NFκB into the nucleus. **E.** Knockout of the
201 cGAS gene prevents transfection-mediated stimulation of IRF-3 translocation, while allowing
202 downstream activation of IRF-3 via cGAMP. Double-stranded DNA in the cytoplasm activates the
203 enzyme cGAS to create the cyclic dinucleotide cGAMP, which then acts on STING to activate IRF-3.
204 Parental BJ-5ta cGAS+/+ cells show nuclear translocation of IRF-3 in response to transfected DNA
205 and also to exogenous cGAMP (top), while BJ-5ta cells with a CRISPR knockout of cGAS show an
206 IRF-3 response only after treatment with cGAMP (bottom). Thus, knockout of the cGAS gene allows
207 assessment of elements of the STING/IRF-3 pathway without interference by transfected DNA. **F.**
208 Treatment with IFNα causes STAT1 phosphorylation and translocation into the nucleus. **G.** Primary
209 image data showing inhibition of NFκB translocation into the nucleus after polyI:C stimulation of BJ-
210 5ta *cGAS$^{-/-}$* cells co-transfected with SARS-CoV-2 NSP1 and GFP expression constructs. **H.**
211 Results of image processing of sets of seven wells per condition, for cells transfected with NSP1
212 genes. See Supplemental Information for full details.
213
214   In the cGAS mutant cell line, DNA transfection-mediated stimulation of IRF-3 nuclear
215 transport and Type I interferon were essentially undetectable (Figure 2E), and the functional
216 transfection efficiency was increased (Supplemental Information). In one set of experiments,
217 cells were simultaneously stained with antibodies to NFkB and IRF-3; in other experiments
218 either anti-STAT1 or anti-phosphoSTAT1(Tyr701) were used (Methods). Treatment with either
219 naked polyIC, lipofectamine/polyIC, or cGAMP was used to stimulate signaling via the TLR3,
220 RIG-I/MAVS and STING pathways respectively; these stimuli cause a dramatic relocalization of
221 IRF-3 and/or NFkB from the cytoplasm to the nucleus (Figures 2D and 2E). In other
222 experiments, BJ-5ta *cGAS-/-* cells were treated with IFNalpha to initiate STAT1 phosphorylation
223 and nuclear translocation (Figure 2F).

224   We therefore used the mutant cell line to conduct the following assay (Figure 2C): cells

225   in 384-well plates were co-transfected with a viral gene expression vector and a GFP

226   expression plasmid. Two days later, cells were treated with cGAMP, IFNα, or a form of polyI:C

227   to stimulate an innate immune response. Based on the timing of the induced response (distinct

228   for the various stimuli), cells were fixed, stained with antibodies directed against relevant

229   transcription factors, and imaged. Fields of cells were scored by automated image processing

230   for expression of the GFP marker, indicating functional transfection, and for the total amount

231   and nuclear/cytoplasmic distribution of NFkB, IRF-3, STAT1, or pTyr701-STAT1. Evaluation of

232   effects of virus protein expression is based on comparison of cell phenotypes of transfected

233   versus untransfected cells in the same well and subsequent comparison to mock-transfected

234   wells as well as virus gene-transfected wells that were mock-stimulated. Apparent 'hits' are

235   identified and inhibition strengths are calculated (Figures 2G and 2H; Methods and

236   Supplemental Information). Other virally encoded suppressors of these pathways generally

237   work by either suppressing nuclear translocation of a given factor, or by inducing its

238   degradation; both mechanisms would be detected by our image-processing calculations.

239   In the present analysis, the data were aggregated to reveal the most clear-cut

240   observations (see Supplementary Information for details of the data processing and rationale for

241   the methods used to aggregate the data). Enhancement of transcription factor degradation and

242   blockage of nuclear import are both identified through these analyses. We summed data for

243   different versions of a given gene from different viruses to draw increased confidence about

244   inferred activities of those genes (vertical bar graphs in Figures 3A-3C).

245

246   **SARS CoV-2 genes most strongly modulate Type I interferon-related signaling pathways**

247   Many individual coronavirus proteins inhibited the nuclear accumulation of IRF-3 (Figure

248   3A) and/or NFkB (Figure 3B). Proteins with the strongest effects were NSP1, NSP3, NSP5, and

249   Orf7a; proteins NSP4, NSP10, NSP12, Env, and CaORF15 had more modest effects.

250   We determined which genes had consistent effects across coronaviruses. To make

251   these inferences, data for each gene within each virus were first averaged across multiple

252   experiments and summed across multiple input signal types to generate heat maps of all the

253   results; these data were further aggregated by averaging the scores for individual genes across

254   viruses (Figures 3A-3C; See Supplemental Information for breakdowns). This analysis revealed,

255   for example, that NSP5 from all of the tested viruses had an inhibitory effect on IRF-3 signaling.

256   The NSP10 and NSP12 genes from several different viruses also inhibited both IRF-3 and NFkB

257   nuclear accumulation. While the effects of these genes are modest, the fact that we observe

258   effects of these genes from multiple different viruses suggests that the effects are not due to

259   noise in the assay.

260   In contrast to the broad suppression of IRF-3 and NFkB translocation, coronaviral

261   proteins largely did not act upon the STAT1 pathway in our assay, with the noticeable exception

262   of two proteins: NSP1 and ORF6 (Figure 3C). Both of these are known inhibitors of host

263   functions: NSP1 binds to the ribosome and inhibits translation of a subset of host genes (Tidu

264   et al., 2020), while Orf6 protein binds to and functionally inhibits Nup98, a nuclear transport

265   protein that carries certain proteins into the nucleus (Miorin et al., 2020) and whose expression

266   is induced by inflammatory stimuli (Enninga et al., 2002).

267   A number of viral proteins appeared to stimulate, rather than inhibit, the STAT1 pathway

268   (Figure 3C). Such stimulation may represent situations where the innate immune system

269   successfully recognizes expression of a foreign or deleterious protein, such as by poor protein

270   folding that leads to creation of inflammasomes (Masters and O'Neill, 2011). Apparent

271   increases in transcription factor levels, as detected by immunofluorescence, may also occur

272   when a virus protein stabilizes the target host protein in an inactive state; for example, US3 of

273   HSV-1 hyperphosphorylates IRF-3 and appears to stabilize this protein in the cytoplasm (Wang

274   et al., 2013).

Figure 3. Inhibition of IRF-3, NFkB and STAT1 active nuclear levels by coronavirus genes. A-C. Statistically significant average inhibition scores from two to five experiments for each gene from each virus (e.g., Figure 2H) were combined. *Top:* Vertical bar graphs were generated by averaging the inhibition scores for a given gene across all viruses with that gene. *Left:* Horizontal bar graphs

280   show the sum of all of the inhibition scores of genes from a given virus. *Bottom:* Heat map of

281   average scores across all relevant stimuli in all experiments (each gene for each virus). Red

282   indicates inhibition of innate immune signaling, green shows enhancement; white indicates a given

283   gene is not present in a given virus; gray indicates a gene not tested. **A.** Inhibition of IRF-3

284   translocation and enhancement of degradation in cells stimulated by cGAMP plus low molecular

285   weight polyI:C in lipofectamine plus free high molecular weight polyI:C. **B.** Inhibition of NFkB

286   translocation and enhancement of degradation in cells stimulated by LMW-polyI:C in lipofectamine

287   plus free HMW polyI:C. **C.** Inhibition of STAT1 translocation, enhancement of degradation, and

288   formation of phospho-STAT1 in cells stimulated by interferon-alpha. See Supplementary Information

289   for full details.

290

291   We next estimated the cumulative capacity of each virus to inhibit the innate immune

292   system by summing the inhibitory effects of each individual gene within each virus (Figures 3A-

293   3C; horizontal bar graphs). This analysis indicates that the genes from SARS-CoV-2, in

294   aggregate, suppress IRF-3 and NFkB signaling more than the aggregate of genes from other

295   viruses. Part of this effect is accounted for by the larger number of genes within SARS-CoV-2

296   (i.e., the less pathogenic strains have fewer C-terminal ORFs, see Figure 1A). However, the

297   immune-suppressing proteins encoded by SARS-CoV-2 within this region also generally

298   showed stronger effects than their SARS-CoV-1 homologues. In particular, Orf3a, Orf6, and

299   Orf7a from SARS-CoV-2 showed stronger immune-suppressing effects (Figures 3A-3C). These

300   proteins have no homologues in MERS-CoV, HCoV 229E, HCoV NL63, HCoV HKU1 and HCoV

301   OC43, which encode a different set of accessory proteins that generally seem to be less active

302   in our assays. Collectively, we conclude that the individual and cumulative ability of coronaviral

303   proteins to suppress the Type 1 interferon-mediate innate immune response is strongest in

304   SARS-CoV-2.

305

306   **NSP1 from SARS-CoV-2 strongly inhibits NFkB-driven gene expression**

307   We determined which coronaviral proteins act by suppressing the expression of

308   inflammatory genes within the host cell by a second cell-based assay to test whether

309   coronavirus genes interfere with expression of a TNFα-inducible reporter element (Figure 4A).

310   Specifically, we constructed a stable pool of HEK293 cells (which have a high transfection

311   efficiency) with a DNA construct in which a degradation-tagged fluorescent mScarlet protein

312   was expressed from an artificial promoter with five NFkB binding sites upstream of a 'minimal

313   CMV' promoter element (Figures 4B and 4C; Supplemental Information). Cells were transiently

14

314   co-transfected with virus gene and GFP expression plasmids, stimulated with TNF-α, and then

315   assayed microscopically and by flow cytometry for expression of the mScarlet reporter in GFP-

316   expressing and non-expressing cells (Figures 4B-4D).

317   The NSP1 protein from several of the coronaviruses inhibited reporter gene expression,

318   with the effect being strongest in SARS-CoV-2 (Figures 4D-4E). Figure 4D shows typical

319   results for NSP1 of SARS-CoV-2: cells that express the GFP transfection reporter do not

320   express the TNF-inducible mScarlet reporter, and vice versa. NSP1 is known to bind to the

321   ribosome and disrupt translation of a subset of host mRNAs (Narayanan et al. 2015). Other

322   genes showing inhibition of gene expression from more than one virus include NSP4, NSP6,

323   NSP8, and NSP13, but the magnitude was much smaller than for the NSP1 genes

324   (Supplemental Information). Notably, NSP1 was the only gene implicated in both the nuclear

325   localization and gene expression assays. The difference in assay results may be because this

326   transcription-based assay used HEK293T cells, which have a defective STING pathway and

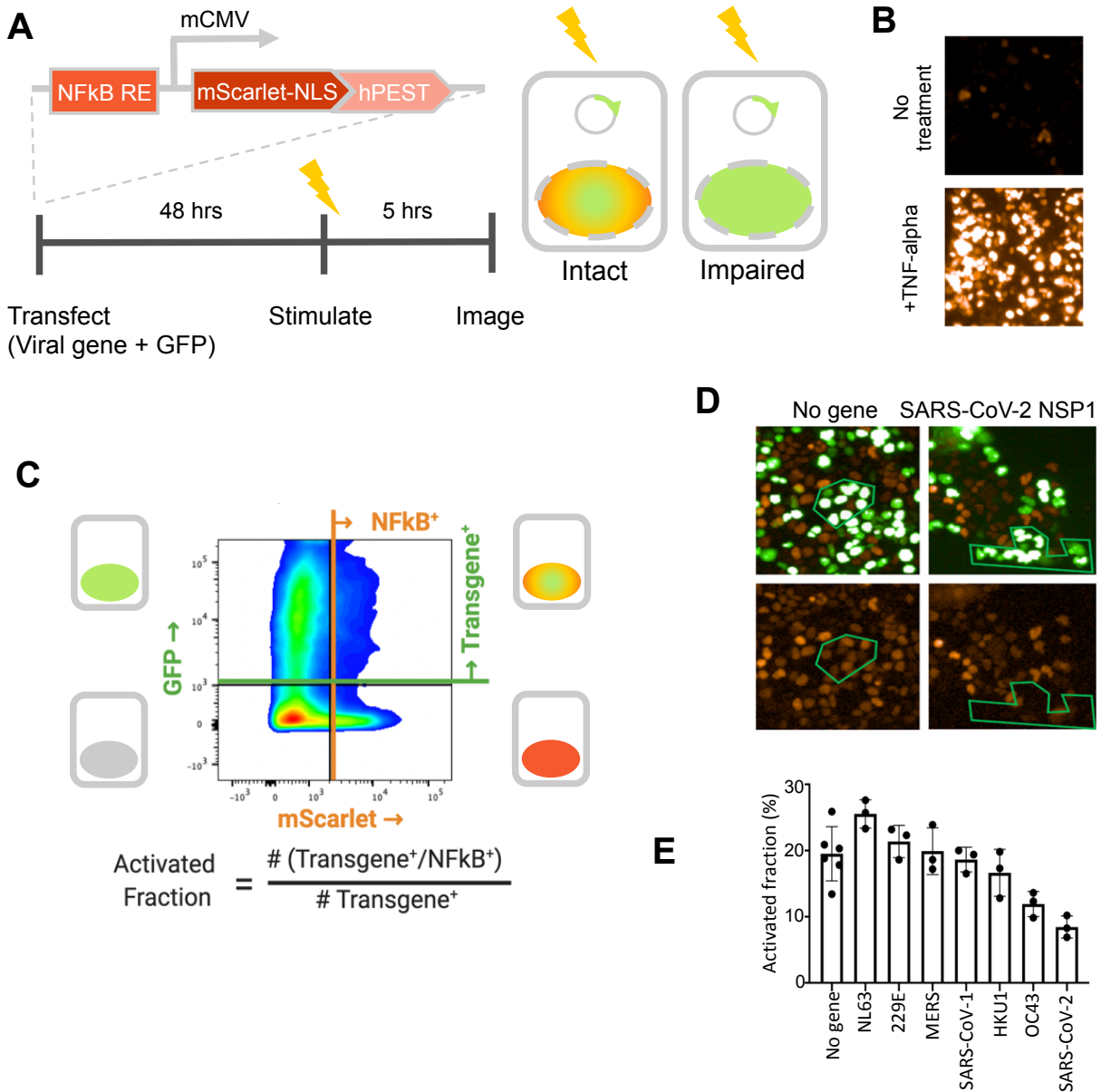327   possibly other innate immune signaling defects (Sun et al., 2013).

Figure 4. SARS-CoV-2 NSP1 strongly inhibits TNF-alpha activation of an NFkB reporter gene.

A. A fluorescent NFkB reporter consisting of a 5x repeat of the NFkB consensus sequence (NFkB RE) upstream of a minimal CMV promoter (mCMV), and a human codon-optimized, nuclear-localized mScarlet (mScarlet-NLS) fused to an hPEST degradation tag (hPEST) was constructed and stably integrated into HEK293 cells. Reporter HEK293T cells were co-transfected with expression plasmids encoding GFP and a virus gene; 48 hours later, 5 ng/ml TNF-α is added; and 5 hours later, red and green fluorescence was quantified. **B.** HEK293T cells stably transduced with the reporter responded to human TNF-α by expression of mScarlet. **C.** Flow cytometry analysis of transfected, TNFα-treated cells. Upper quadrants are transfected cells. The upper right quadrant

338   represents double-positive cells that are transfected and also express the inducible transgene. The

339   activated fraction was calculated as the ratio of double positive cells to all GFP-positive cells. **D.**

340   Reporter HEK293T cells were co-transfected with GFP and no-gene or SARS-CoV-2 NSP1

341   expression vectors, followed by TNF-α stimulation and imaging of red and green fluorescence.

342   Green polygons highlight identical populations of cells in both images for representative

343   comparisons. **E.** Activated fractions compared among NSP1s from several coronaviruses. Error bars

344   represent the standard deviation of at least three technical replicates.

345

346   **Cell-based assay for immune suppression-enabled viral replication**

347   Because of the complex nature of human antiviral innate immunity, we developed

348   multiple assays that could evaluate whether a viral protein was able to interfere with antiviral

349   defenses. The assays above test the capacity of viral proteins to affect specific pathways

350   involved in innate immunity. In another approach, we created a mechanism-agnostic test of

351   innate immune suppression by viral genes, where the indicator of a viral protein that interferes

352   with antiviral innate immunity is replication of a reporter virus. To do this, we built on our yeast-

353   mammalian cell fusion system reported in Brown et al. (2016). There, we found that yeast

354   spheroplasts could be fused to mammalian cells to deliver large DNAs to the mammalian

355   nucleus, and also yeast-expressed proteins to the mammalian cytoplasm. In one variation of

356   this experiment, a strain was constructed that contained the entire 155-kb herpes simplex type-1

357   (HSV-1) viral genome on a yeast centromeric (CEN) plasmid. When spheroplasted yeast were

358   fused with HEK293 cells, viral particles were produced that could then productively infect other

359   HEK293 cells. Typically, viral genomes in capsids and nucleocapsids have associated 'non-

360   structural' proteins that are carried with the genome into the target cell; these proteins often

361   suppress the innate immune system in the infected cell, facilitating replication without requiring

362   viral gene expression in the cell. In the yeast-mammalian cell fusion experiment, such proteins

363   are lacking. Fusion of such yeast with HeLa cells, which have a more intact innate immune

364   system than HEK293 cells, did not allow viral replication. However, when yeast cells were

365   further engineered to express the Ebola virus VP35 protein, fusion of such yeast with HeLa cells

366   yielded infectious virus. As Ebola VP35 is a suppressor of intracellular innate immunity, this

367   experimental paradigm establishes an assay for innate immune suppressors (Zampieri et al.,

368   2007). We have used this assay system in the present work.

369   In the methodology used here, we first inserted most of the coronavirus genes described

370   above into a galactose-inducible yeast expression vector in yeast by TAR cloning (Kouprina and

371   Larionov, 2016) to construct one set of yeast for delivery of viral proteins. We constructed a

372   second yeast strain in which the entire HSV-1 genome was inserted into a yeast *CEN* plasmid;

373   this HSV-1 also contained a reporter gene in which green fluorescent protein (GFP) was

374   expressed from a viral late promoter, so that viral replication could be easily assayed in a plate

375   reader. The extent of viral replication as inferred from the GFP signal was highly correlated with

376   direct titration of virus by a $TCID_{50}$. Fusion of the HSV-1-containing yeast with HeLa cells does

377   not generate live virus, but when the other yeast strain expresses Ebola virus VP35 protein,

378   fusion of both yeast with HeLa cells yields replicating virus; this configuration serves as a

379   positive control (Figure 5A; see Methods and Supplemental Information for details).

380   The virus genes with the strongest effects were, in rough order, ORF6, ORF7b,

381   CaORF15, NSP8, NSP9, NSP6, NSP3, ORF5, ORF3a, and Membrane (Figure 5B). A number

382   of these showed effects in the transcription factor nuclear accumulation assays, such as ORF6,

383   CaORF15, NSP3, and ORF3a. The fact that proteins such as ORF7b, NSP6, NSP8, NSP9,

384   ORF5 and Membrane were clear positives in this mechanism agnostic assay but not in the

385   mechanism-specific assays may point to the complex nature of both the innate immune system

386   and the process of HSV-1 infection, which involves 77 viral proteins. Conversely there were

387   proteins with no or minimal activity in the yeast fusion assays that were clearly positive in the

388   other assays. NSP1 from SARS-CoV-2 and SARS-CoV-1 were negative in the yeast fusion

389   assay, while the seemingly weaker NSP1 proteins (based on the transcription factor nuclear

390   accumulation assays; Figure 4) from HKU1 and 229E were positive; this might be rationalized if

391   the translation-inhibition function of the more potent versions of NSP1 interferes with expression

392   of HSV-1 genes. The same reasoning could apply to other genes, for example proteins

393   involved in coronavirus RNA processing and could interfere with HSV-1 gene expression. Thus,

394   positive results obtained in the yeast-based assay may be more meaningful than negative
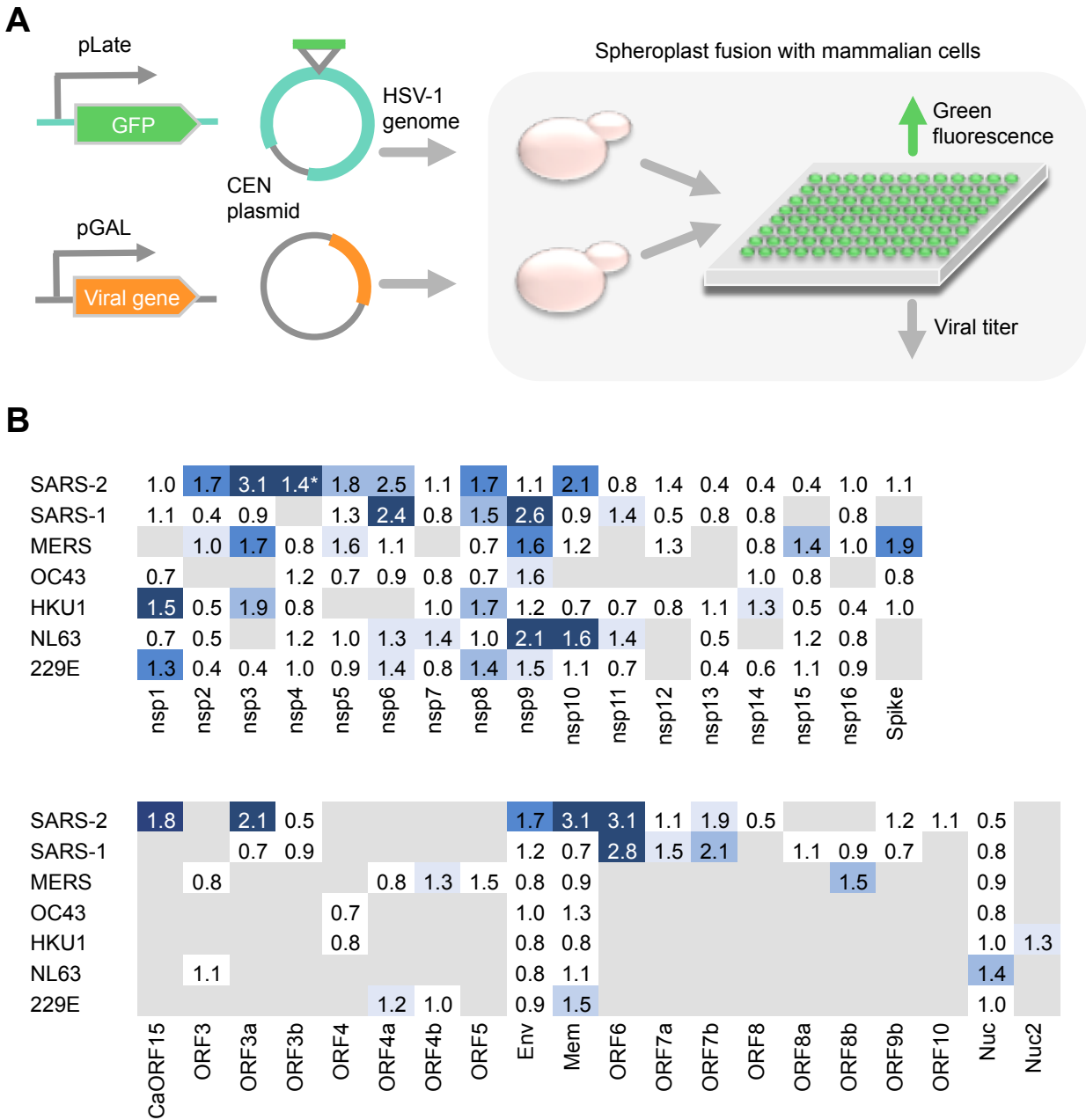
395   results.

Figure 5. Assay for innate immune suppression based on protein delivery via yeast-mammalian cell fusion. **A.** Assay concept. A yeast cell with a plasmid expressing a virus gene from the *GAL1* promoter is induced with galactose to produce the viral protein in the yeast cytoplasm. A second yeast cell has a *CEN* plasmid containing the entire Herpes Simplex Virus-1 genome that also includes a GFP gene transcribed from a late promoter. The strains are mixed, spheroplasted, and fused with HeLa cells, which allows replication of the HSV-1 only if the viral protein delivered by the first yeast can suppress the innate immune system. The Ebola virus VP35 gene is a positive control (Brown et al., 2016). Immune suppression is measured via viral expression of GFP. **B.** Results from

	nsp1	nsp2	nsp3	nsp4	nsp5	nsp6	nsp7	nsp8	nsp9	nsp10	nsp11	nsp12	nsp13	nsp14	nsp15	nsp16	Spike
SARS-2	1.0	1.7	3.1	1.4*	1.8	2.5	1.1	1.7	1.1	2.1	0.8	1.4	0.4	0.4	0.4	1.0	1.1
SARS-1	1.1	0.4	0.9		1.3	2.4	0.8	1.5	2.6	0.9	1.4	0.5	0.8	0.8		0.8	
MERS		1.0	1.7	0.8	1.6	1.1		0.7	1.6	1.2		1.3		0.8	1.4	1.0	1.9
OC43	0.7			1.2	0.7	0.9	0.8	0.7	1.6					1.0	0.8		0.8
HKU1	1.5	0.5	1.9	0.8			1.0	1.7	1.2	0.7	0.7	0.8	1.1	1.3	0.5	0.4	1.0
NL63	0.7	0.5		1.2	1.0	1.3	1.4	1.0	2.1	1.6	1.4		0.5		1.2	0.8	
229E	1.3	0.4	0.4	1.0	0.9	1.4	0.8	1.4	1.5	1.1	0.7		0.4	0.6	1.1	0.9	

	CaORF15	ORF3	ORF3a	ORF3b	ORF4	ORF4a	ORF4b	ORF5	Env	Mem	ORF6	ORF7a	ORF7b	ORF8	ORF8a	ORF8b	ORF9b	ORF10	Nuc	Nuc2
SARS-2	1.8		2.1	0.5					1.7	3.1	3.1	1.1	1.9	0.5			1.2	1.1	0.5	
SARS-1			0.7	0.9					1.2	0.7	2.8	1.5	2.1		1.1	0.9	0.7		0.8	
MERS		0.8				0.8	1.3	1.5	0.8	0.9						1.5			0.9	
OC43					0.7				1.0	1.3									0.8	
HKU1					0.8				0.8	0.8									1.0	1.3
NL63			1.1						0.8	1.1									1.4	
229E						1.2	1.0		0.9	1.5									1.0	

396

405   testing coronavirus genes. Numbers indicate the average fold increase in the GFP assay. Blue-

406   highlights indicate a enhancement of viral replication (dark: $p<0.05$; medium: $0.05<p<0.1$; light:

407    $0.1<p<0.15$; very light: $0.15<p<0.25$); white indicates no stimulation detected; gray indicates that the

408   genes were not tested or are not present. A number of genes score positively in this assay and also

409   in the assays of Figure 2-3, including NSP3, NSP5, ORF6, and CaORF15.

410

411   **Discussion**

412   In this work, we characterized the ability of the proteins encoded by seven

413   coronaviruses to suppress host intracellular innate immune signaling pathways using high-

414   content, medium-throughput cell-based assays. The first assay examined the nuclear

415   localization of the transcription factors IRF-3, NFkB, and STAT1, which play key roles in the

416   elaboration of the Type I interferon response to viral infection, and which are major targets for

417   inhibition by other viruses (Figure 2; Hoffmann et al., 2015). Viruses act on these pathways by

418   a variety of mechanisms, such as specific inhibition of upstream signal transduction proteins,

419   general inhibition of nuclear import, and enhancement of degradation. Our assays involved

420   transfection of a non-transformed cell line with expression vectors encoding each gene followed

421   by addition of a stimulator of innate immune signaling, and then immunofluorescence staining

422   and automated image processing. We found that:

423   1. Localization of IRF-3 and NFkB to the nucleus was inhibited by a number of coronavirus

424   genes, including known innate immune suppressors such as NSP5, NSP3, Orf7a and

425   NSP1 (Freitas et al., 2020; Lei et al., 2018), and proteins not previously known to have

426   such an activity, including CaORF15 and NSP9.

427   2. Two genes were found to have an impact on STAT1 activation: NSP1 and Orf6. Other

428   groups previously found that that NSP1 binds to the ribosome and inhibits translation of

429   a subset of host proteins (Tidu et al., 2021), and that Orf6 protein inhibits the nuclear

430   protein import factor Nup98 (Frieman et al., 2007). Our results are consistent with these

431   observations.

432   3. In the 3' region of the coronavirus genome, which encodes accessory proteins that vary

433   among the coronaviruses, SARS-CoV-2 proteins are more potent in their inhibition of

434   innate immune signaling than the corresponding SARS-CoV-1 protein (Figure 3).

435   4. SARS-CoV-2 appears to encode a novel protein ("CaORF15"), which is predicted to

436   have three transmembrane domains. CaORF15 is encoded in an alternate reading

437   frame within the Spike coding sequence, and is not found in SARS-CoV-1 but is present

438   in the more closely related bat virus RaTG13. CaORF15 inhibits IRF-3 nuclear

439   accumulation particularly through the STING pathway (Figure 3, Supplemental

440   Information) and also suppresses the innate immune system in the yeast-mammalian

441   cell fusion assay (Figure 5).

442   5. When the inhibitory effects of coronavirus genes are examined in aggregate, SARS-

443   CoV-2 appears to have more potential to suppress the innate immune system. This is

444   based on summation of effects on IRF-3 and NFkB nuclear localization (Figure 3A and

445   3B), the strength of NSP1 inhibition of NFkB-mediated gene expression (Figure 4), and

446   the genes showing effects in the yeast cell fusion assay (Figure 5).

447

448   NSP1 strongly blocks the activation of a TNF-inducible expression of a reporter

449 construct with multiple NFkB binding sites (Figure 4). We found that only the NSP1 proteins

450 from the various viruses strongly inhibited transcription of this reporter. NSP1 from SARS-CoV-

451 2 had the strongest inhibitory effect in the TNF-inducible reporter assay (Figure 4F); this effect

452 may contribute to SARS-CoV-2's ability to blunt the host response. Similar transcription-based

453 screens showed SARS-CoV-2 NSP1 inhibition of the IFN-beta promoter, which also depends on

454 NFkB for activation (Lei et al 2020).

455   Our pathway-agnostic assay also identified NSP1, NSP3, CaORF15, and ORF6 as likely

456 suppressors of innate immune signaling. Yeast containing both an attenuated HSV-1 genome

457 and expressed coronavirus protein are fused with a mammalian cell; fusion of the yeasts with

458 HeLa cells yields essentially no virus unless an active innate immune suppressor is also

459 delivered, which then allows replication of the HSV-1 genome and production of live virus. The

460 assay thus mimics an early step of virus infection, wherein 'non-structural' proteins are typically

461 bound to the viral nucleic acid and are co-delivered. Such non-structural proteins often

462 suppress the innate immune system; Ebola virus VP35 is an example (Kuhl and Pohlmann,

463 2012). This assay may detect proteins that act by distinct mechanisms from the pathway-

464 focused tests. For example, NSP8 and NSP9, which were not detected in the other assays,

465 score as positive in this test. Conversely, proteins that strongly inhibit host functions, such as

466 SARS-CoV-2 NSP1, may interfere with HSV-1 replication in this assay and thus score as

467 negative.

468   Viral proteases often act on virally encoded proteins and host proteins involved in innate

469 immunity (Lei and Hilgenfeld, 2017). The NSP5 gene from each of the seven viruses tested had

470 an inhibitory effect on IRF-3, and NSP5 from SARS-CoV-1, MERS, HKU1 and NL63 had an

471 inhibitory effect on NFkB. In agreement with our findings, Freitas et al. (2020) predicted that the

472   NSP5 papain-like protease would suppress innate immune signaling based on its cleavage and

473   reversal of ubiquitin and ISG15 modifications. NSP3 is a large protein with many domains,

474   including a papain-like protease. We also found that NSP3 genes from most of the viruses

475   tested had inhibitory effects on IRF-3 and NFkB. NSP3 has a number of additional domains

476   that may suppress innate immunity based on modulation of the ubiquitin ligase system (Lei et

477   al., 2018).

478   We also found that NSP9, NSP10 and NSP12 had innate immune-suppressing activity.

479   NSP9 from SARS-CoV-1, MERS, and NL63, promoted replication of HSV-1 in the yeast fusion

480   assay, as did NSP10 from NL63 and SARS-CoV-2. NSP10 and NSP12 inhibited IRF-3 and

481   NFkB. Lei et al. (2020) found that NSP12 inhibited the activation of a reporter gene containing

482   the IFN-beta promoter. All of these proteins are involved with mRNA metabolism. NSP12 is the

483   RNA-dependent RNA polymerase, so at a mechanistic level it is not apparent how this protein

484   might modulate IRF-3 and NFkB. It is possible that, by analogy to influenza NS1 with its

485   multiple surfaces that bind to unrelated host proteins, that NSP12 has surfaces that are not

486   involved with RNA polymerization and can therefore carry out an unrelated function of titrating a

487   host factor. Alternatively, these proteins may bind to host RNAs that express proteins involved

488   in innate immunity.

489   As part of our preparation of expression vectors for coronavirus genes, we identified a

490   new ORF in SARS-CoV-2, which we provisionally term "CaORF15" and which falls into the

491   GTRA-like set of protein folds (PFAM family PF04138). The protein is encoded in an alternative

492   open reading frame within the Spike coding sequence (Figure 1), specifically within the most N-

493   terminal domain of Spike. Three lines of evidence suggest that CaORF15 encodes a functional

494   protein and is not a fortuitous ORF. First, this region of the Spike coding region has an unusual

495   sequence conservation pattern: both SARS-CoV-2 and the bat coronavirus RaTG13 encode

496   this protein, and the coding region is more conserved at the nucleotide level than the rest of the

497   aligned Spike sequence from these two viruses. In contrast, in the more distantly related

498   viruses SARS-CoV-1 and M789 from pangolins that lack CaORF15, the nucleotide sequence is

499   more divergent than for the rest of the Spike gene (Figure 1; Supplemental Information). These

500   observations suggest a pressure to evolve CaORF15 and then to maintain it. Second,

501   expression of this protein in BJ-5ta cells inhibited the nuclear translocation of IRF-3 in response

502   to cGAMP (Figure 3, Supplemental Information). Third, delivery of this protein in the yeast

503   fusion assay stimulates replication of yeast-delivered HSV-1. These results suggest that

504   CaORF15 encodes a protein that functions to suppress the intracellular innate immune system.

505   The presence of CaORF15 within the Spike coding sequences means that nucleic acid-based

506   vaccines based on the natural Spike coding sequence – depending on their design - could also

507   encode the CaORF15 protein, which may be immunosuppressive in a vaccine context.

508   However, Pfizer's vaccine is based on a codon-optimized version of Spike in which the AUG of

509   CaORF15 is changed to ACG (WHO-INN, 2020).

510   In sum, we characterized the behavior of proteins from seven different coronaviruses in

511   three different assays for suppression of intracellular innate immune signaling. We found that a

512   number of proteins showed inhibitory activity. At least two of these, NSP9 and SARS-CoV-2

513   "CaORF15," do not appear to have been previously identified as such.

514   Innate immune suppression may correlate with asymptomatic spread rather than

515   pathogenicity per se. Infection with MERS, for example, is fatal much more often than infection

516   with SARS-CoV-2 or SARS-CoV-1; here the severe symptoms, presumably a result of a robust

517   Type I interferon response, prevent spread of the virus within the human population. In carrying

518   out these experiments, we sought to address whether the pandemic potential of a virus could be

519   estimated based on medium-throughput analysis of an entire viral genome for suppression of

520   the human intracellular innate immune system.

521   Emerging viruses generally fall into known families. To set the stage for inferring

522   pandemic potential in future emerging viruses, we compared the genes of SARS-CoV-2 with

523   several other coronaviruses of known pathogenicity. The results of this analysis indicate that

524   SARS-CoV-2 genes cumulatively appear to have a greater potential for immune pathway

525   suppression than other coronaviruses, including SARS-CoV-1. This is admittedly an

526   approximate statement, since the relative importance of each gene in these viruses is not

527   generally known, but the data collectively are likely to suffice for a rapid assessment of whether

528   to initiate a vaccine program. Taken together, these results suggest that rapid testing of viral

529   genes in assays for innate immune suppression, performed with genes from related viruses,

530   could be used for early-stage evaluation of the pandemic potential for emerging viruses.

531
532

533

534
535

STAR METHODS

RESOURCE AVAILABILITY

Lead Contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact Pamela Silver.

Materials Availability
Plasmids and strains are available upon reasonable request.

Data and Code Availability
All image datasets generated during this study are available upon request via OMERO and Dropbox.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Searching for candidate ORFs

In early 2020, when annotation of the SARS-CoV-2 genome was poor, we chose to search the raw sequence (Wuhan-hu-1, MN908947) for any overlooked open reading frames. First, we translated in all six frames and collected all 399 translations longer than or equal to 10 amino acids (Supplementary Table 1). We reasoned that genuine proteins may show similarity to established protein families and therefore aligned all 399 translations against PFAM sequence profiles using HMMscan (Mistry et al. 2020 doi: 10.1093/nar/gkaa913; Eddy 1998 doi: 10.1093/bioinformatics/14.9.755).

Of the translations, 15 correspond to known SARS-CoV-2 proteins and each resulted in significant alignments to one or more PFAM profiles (all E-values < 1 E-14). At the time, ORF14 was missing from available SARS-CoV-2 annotation, nevertheless our approach identified the translation as a likely protein because of significant alignment to PF17635, now also called bCoV_Orf14 (E-value 6.2 E-35). In addition, while ORF3b is split in SARS-CoV-2 relative to SARS-CoV-1 by an early stop codon, we found alignments to the resulting fractional translations (E-values = 0.0024 and 0.17). There remained 11 candidate ORFs not known to be SARS-CoV-2 proteins and that aligned to some PFAM profile (E-values=6.4 E-5 to 0.94). Of these candidates, one stood out as the longest (87aa) and most significant (E-value = 6.4 E-5) by a large margin. We dubbed this translation "Candidate ORF 15" (CaORF15) and decided to test the sequence in our assays. CaORF15 encodes 87 amino acids at genome coordinates 21936-22199, within Spike protein coding region.

Conservation analysis of CaORF15

576   To find out whether CaORF15 is conserved in other coronaviruses, we performed a

577   MAFFT alignment of the SARS-CoV-2, SARS, pangolin PCoV GX-P3B, and RatG13 genomes

578   (MN908947, AY274119, MT072865, MN996532 respectively). We then computed the identity of

579   the aligned CaORF15, and surrounding regions of Spike, versus nucleotides in SARS-CoV-2

580   (Fig 1C, top). We found that RatG13 encodes an amino acid sequence 93% identical to

581   CaORF15 in SARS-CoV-2, whereas SARS-CoV and PCoV GX-P3B do not. Instead, SARS-

582   CoV and PCoV GX-P3B encode shorter sequences (29aa and 51aa) with low identity to

583   CaORF15 (24% and 25% respectively). For visualization, we scanned a 40-nt window over the

584   aligned Spike regions, computing the non-gapped identity of each genome versus that of SARS-

585   CoV-2 (Fig 1C, bottom). To emphasize local identity, we smoothed the windows by weighting

586   matches with a centered normal distribution (sigma=8).

587

588   **Preparation of mammalian expression plasmids containing virus genes**

589   Genes encoding proteins from the seven coronaviruses SARS-CoV-2, SARS-CoV,

590   MERS-CoV, HCoV 229E, HCoV NL63, HCoV HKU1, and HCoV OC43 were ordered from Twist

591   Biosciences (Twist) or Integrated DNA Technologies (IDT). Sequences are listed in the Excel

592   file [Table SX]. Coronavirus gene fragments corresponding to proteins NSP2-16 post-cleavage

593   proteins were designed with added start codons for individual expression; normally these

594   proteins are generated from cleavage of the NSP1-16 polyprotein (Fehr and Pearlman 2015).

595   DNA sequences between 300 and 5000 bp were obtained from Twist Biosciences in a custom-

596   onboarded vector based on pSecTag2 [Addgene #V90020], but lacking the signal sequence of

597   that vector. Specifically, the methionine start codon deriving from pSecTag2 was also the start

598   coding for the cloned viral genes. The same methods were used to construct expression

599   vectors for controls: NS1 protein from influenza A virus [accession # NP_040984.1], the VP35

600   protein from Ebolavirus [accession # NP_690581.1], the V protein from parainfluenza virus 5

601   [accession # YP_138513.1], the V protein from measles virus [accession # YP_003873249],

602   and the M protein from vesicular stomatitis virus [accession # NP_041714]. DNA sequences

603   under 300bp were ordered from IDT with adapters for Gibson cloning using the same vector.

604   The only genes over 5000 bp were the nsp3 gene fragments from SARS-CoV-2, SARS-CoV,

605   MERS-CoV, HCoV HKU1, and HCoV OC43. These were ordered as clonal genes in two

606   fragments, referred to as nsp3-1 and nsp3-2, and assembled with Gibson assembly into the

607   wild-type full-length gene in pSecTag2. Most cloned genes were transformed into chemically

608   competent DH5α *E. coli* (New England Biolabs C2987I). The NSP3 genes were difficult to

609 construct as whole genes using DH5α, so we used *E. coli* PY1182 (an MM294-derived strain

610 with a *pcnB80* mutation to reduce copy-number of ColE1 vectors; a gift of R. Losick) as a host.

611   DNA for transfection was prepared according to the QIAGEN endotoxin-free maxi-prep

612 manufacturer protocol. After preparation, purified DNA was diluted in TE buffer to 200 ng/μL,

613 aliquoted and stored at -20 °C.

614

615 **High-Content Screening of IRF-3, NFkB and STAT1 signaling in BJ-5ta cells**

616 **Culturing BJ-5ta human fibroblast cell line**

617   BJ-5ta cells were purchased from ATCC (CRL-4001;

618 https://www.atcc.org/products/all/CRL-4001.aspx) and cultured in the manufacturer

619 recommended medium of 4:1 DMEM:M199 with 10% FBS (72% DMEM [ATCC 30-2002], 18%

620 M199 [Thermo Fisher Scientific 11150059], 10% U.S Origin FBS (GenClone #25-514) with 10

621 μg/mL hygromycin B ([Invivogen ant-hg-1). Cells were cultured in T-25, T-75, and T-150 cell

622 culture-treated flasks with vented caps (Corning) at 37 °C and 5% CO_2. Cells were passaged

623 every 3-4 days at 70-90% confluency to 30% confluency.

624

625 **Seeding BJ-5ta cells into 384-well plates**

626   BJ-5ta cells were lifted from T-150 culture flasks using 0.25% trypsin (VWR 45000-664)

627 for 5-10 minutes, then the trypsin was quenched with 2x volume of culture medium and

628 transferred to 50 mL Falcon tubes. The cell suspension was centrifuged in a swinging-bucket

629 rotor at 300g for 6 minutes at room temperature. The supernatant was discarded by aspiration

630 and cells were resuspended in a small amount of culture volume and counted with a Bio-Rad

631 TC20 Automated Cell Counter. Cells were diluted to 60,000/mL and 40 μL of diluted cell

632 suspension was added from a reservoir to all wells except the outer row (i.e. rows B-O, columns

633 2-23 were used) of tissue-culture treated black CellCarrier-384 Ultra Microplates (Perkin Elmer

634 6057302) using a 12-channel electronic multichannel 200 μL pipettor [Sartorius]. Plates were

635 then centrifuged at 200g for 4 minutes at room temperature and incubated overnight. Typically,

636 about 2-5,000 cells per well are seeded, and about 200-800 are transfected as defined by

637 expression of GFP.

638

639 **Co-transfection of virus gene plasmids and GFP into BJ-5ta cells**

640   30-60 minutes prior to transfection BJ-5ta culture medium (4:1 DMEM:M199 with 10%

641 FBS) was replaced with an equal volume of pre-warmed antibiotic-free transfection medium (4:1

642   DMEM:M199 with 20% FBS; 64% DMEM, 16% M199, 20% FBS). Transfection mixes were

643   prepared according to manufacturer protocol (GeneXPlus, ATCC ACS-4004) with final

644   concentrations of 1 µg DNA, 4 µL GeneXPlus in 100 µL of Opti-MEM I Reduced-Serum

645   Medium. Briefly, GeneXPlus, plasmid DNA (200 ng/µL), and Opti-MEM I Reduced-Serum

646   Medium (ThermoFisher #31985062) were warmed to room temperature and vortexed gently.

647   Plasmid DNA was aliquoted into sterile microcentrifuge tubes at a ratio of 3:1 virus gene

648   plasmid : GFP-containing plasmid. Opti-MEM was quickly mixed with GeneXPlus and the

649   appropriate volume was added to each DNA aliquot and mixed briefly by gentle pipetting.

650   GeneXPlus:DNA complexes were formed at room temperature for 15-20 minutes. Transfection

651   mixtures were then added to each well at 10% final volume (4.4 µL transfection mixture was

652   added to 40 µL transfection medium). Plates were centrifuged at 200 rcf, 4 minutes at room

653   temperature to collect all transfection mixture into the medium and briefly mixed by tilting plate

654   back and forth. Cells were incubated with transfection mixture at 37 °C, 5% CO_2 for 24 hours to

655   allow DNA to enter cells. Transfection medium was then exchanged for fresh culture medium

656   and cells were further incubated for another 24 hours prior to stimulation with innate immune

657   stimuli and fixation as described above.

658

659   **Homozygous knockout of cyclic GMP-AMP synthetase (cGAS) in BJ-5ta cells**

660   CRISPR was used to introduce a frameshift mutation at position 13 of exon 1 of cyclic

661   GMP-AMP synthetase (cGAS) in BJ-5ta cells. Three different Synthego-designed guide RNAs

662   were each co-transfected with Cas9-containing plasmid (Synthego) into BJ-5ta cells (ATCC

663   CRL-4001) using Lipofectamine 3000. After 48 hours, samples were removed from each

664   knockout pool for Inference of CRISPR Edits (ICE) analysis to assess gRNA efficiency, which

665   was between 1% and 6%. The knockout pool with 6% gRNA efficiency (gRNA 2) was diluted to

666   a density of 0.5 cells/100uL and plated into 96-well plates for clonal expansion. Colonies grown

667   from a single cell were visually identifiable after 3 weeks. After 8 weeks, the cGAS locus was

668   sequenced in each clonal colony to identify colonies with homozygous indels. One

669   homogygous knockout colony was identified from 20 screened. Homozygous knockout in the

670   successful colonies was confirmed via Western Blot for cGAS protein.

671

672   **Stimulation of innate immune signaling**

673   The BJ-5ta $cGAS^{-/-}$ cell line was used for these experiments. This cell line showed a

674   very reduced level of background innate immune signaling that otherwise resulted from

675   introduction of transfecting DNA. In addition, the transfection efficiency of these cells was

676   improved relative to the parental BJ-5ta cells. Cells intended for stimulation with poly(I:C) HMW

677   or liposome-encapsulated poly(I:C) LMW were primed 48 h in advance of stimulation by treating

678   with interferon α1 (Cell Signaling #8927) or interferon α2b (PBL Assay Science #11100-1) at 50

679   ng/mL (final concentration in the well 5 ng/mL). 24 h after treatment with interferon, the cell

680   medium was exchanged to remove external interferon from the cell environment. Different

681   innate immune stimuli were applied to cell medium at 10% culture volume as follows. Low

682   molecular weight (LMW) Poly(I:C) (Invivogen #tlrl-picw) at a concentration of 100 ng/μL (final

683   concentration in the well 10 ng/μL) complexed with Lipofectamine 3000 (ThermoFisher

684   #L3000015) in Opti-MEM reduced-serum medium (ThermoFisher #31985062) according to

685   manufacturer's instructions was used to stimulate RIG-I-Like Receptor (RLR) activity by

686   incubation at 37 °C, 5% CO_2 for 4 h. High molecular weight (HMW) Poly(I:C) (Invivogen #tlrl-

687   pic) at a concentration of 1mg/mL (final concentration in the well 100 ug/mL) was used to

688   stimulate TLR-3 activity by incubation at 37 °C, 5% CO_2 for 2 h. 2',3'-cyclic GMP-AMP

689   (cGAMP) (Invivogen #tlrl-nacga23-5) at a concentration of 1 mg/mL (final concentration in the

690   well 100 ug/mL) was used to stimulate STING pathway activity by incubation at 37 °C, 5% CO_2

691   for 2 h. Interferon α1 (Cell Signaling #8927) or α2b (PBL Assay Science #11100-1) at a

692   concentration of 50 ng/mL (final concentration in the well 5 ng/mL) was used to stimulate IFNAR

693   activity by incubation at 37 °C, 5% CO_2 for 45-50 min. Cell signaling was stopped by fixation as

694   described below.

695

696   **Cell fixation and immunofluorescent staining**

697   Cells were fixed with 15 μL of 16% methanol-free formaldehyde (ThermoFisher #28908)

698   added directly to the 45 μL of cell medium in the wells for a final fixation solution of 4%

699   formaldehyde. After a 20 minute incubation, the 4% formaldehyde solution was aspirated and

700   the cells were washed 3x with 60 μL PBS using an automated plate washer (BioTek EL406).

701   Cells to be stained for phospho-STAT1 were further permeabilized with ice-cold 100% methanol

702   (Sigma Aldrich #34860) and incubated at -20 °C for 10-15 minutes, then washed 3x with 60 μL

703   PBS using an automated plate washer. All primary and secondary antibodies were diluted

704   1:400 in PBS containing either 2.25% bovine serum albumin [Millipore Sigma #A2058] or 5%

705   normal goat serum (Abcam #ab7481) for blocking and 0.15% Triton X-100 (Sigma Aldritch

706   #T8787) for permeabilization. Fixed cells were stained with 40 μL diluted primary antibody

707   solution overnight at 4 °C. Cells were then washed 4x with 60 μL PBS using an automated

708   plate washer and stained with 40 μL diluted secondary antibody solution with DAPI

709   (ThermoFisher #D1306) added to a final concentration of 0.2 μg/mL. Cells were finally washed

710   4x with 60 μL PBS using an automated plate washer and sealed using impermeable black plate

711   seals. If not imaged immediately, fixed and stained cells were stored at 4 °C for a maximum of

712   4-7 days.

713   Cells treated with interferon α were stained either for phospho-STAT1 (Cell Signaling

714   Technology #9167) or for STAT1 (Cell Signaling Technology #14994). Cells treated with

715   cGAMP, poly(I:C) HMW, or poly(I:C) LMW were stained simultaneously for IRF-3 (Cell Signaling

716   Technology #11904) and NFκB (Santa Cruz Biotechnology #sc-8008). IRF-3, phospho-STAT1,

717   and STAT1 primary antibodies were detected using an Alexa-Fluor 647-conjugated goat anti-

718   rabbit IgG antibody (ThermoFisher #A21245). NFkB primary antibody was detected using an

719   Alexa-Fluor 568-conjugated donkey anti-mouse IgG secondary antibody (ThermoFisher

720   #A10037).

721

722   **High-content imaging and image segmentation**

723   Fluorescently stained plates were imaged on a PerkinElmer Operetta CLS High-Content

724   Imaging System with a 20x, numerical aperture 0.75 objective. 20-25 sites were imaged in each

725   well, covering 90-100% of the well. Each well was imaged for DAPI, GFP, and Alexa 647.

726   Wells treated with cGAMP or poly(I:C) were also imaged for Alexa 568. Image segmentation

727   was performed using Columbus software (PerkinElmer). Nucleus areas were identified with

728   Columbus Method C based on the DAPI channel and cytoplasmic areas were assigned with

729   Columbus Method D based on the Alexa 647 channel. Average intensity in the GFP, Alexa 647,

730   and Alexa 568 (if applicable) channels was calculated for the cytosol and nuclear areas of each

731   computationally identified cell. Single-cell results were exported from Columbus in CSV format

732   and can be viewed at [insert Harvard Dataverse data ID here upon acceptance].

733

734   **Data processing**

735   Single-cell results were analyzed using a custom Python script which can be found at

736   [insert GitHub URL here upon acceptance]. Briefly, nuclear objects identified by Columbus that

737   correspond to cell debris and artifacts were eliminated based on nuclear morphology. For each

738   transcription factor in each cell, Nuclear Localization (Nucleus intensity / Cytosol intensity) and

739   Total Cell intensity (Nucleus intensity + Cytosol intensity) were calculated. Within each well,

740   cells were sorted into GFP positive (GFP+) or GFP negative (GFP-) (as a proxy for expression

741   of virus protein) based on average Nucleus GFP intensity. The GFP positive or negative cutoff

742   was set at twice the median Nuclear GFP intensity (the median being within the distribution of

743   the more numerous GFP-negative cells). Within each well, the average Nuclear Localization or

744   Total Cell intensity was calculated for the GFP+ or GFP- subsets of cells. Subsequently, for

745   each well, the average Nuclear Localization or Total Cell intensity for the GFP+ cells was

746   normalized to the corresponding average for the GFP- cells to obtain a single normalized Mean

747   Nuclear Localization or Mean Total Cell intensity.

748   Quality control was performed on a plate-by-plate basis as follows. If the mean of either

749   of the two sets of controls containing no virus gene was outside the 20th-80th percentiles of the

750   plate data as a whole, the data for the aberrant control was discarded. If both no-gene controls

751   were non-aberrant, the two sets of no-gene control data were combined for the following

752   primary "fold change" calculation and normalization purposes. Additionally, for each individual

753   set of 7 technical replicates, if any data point was more than 3 times the interquartile range

754   higher than the 75th percentile or lower than the 25th percentile, it was removed from the

755   analysis.

756   Within each plate, the mean and standard deviation of the 7 technical replicates of each

757   virus gene/innate immune stimulus (one type of innate immune stimulus or mock stimulus per

758   plate) combination were calculated. The fold change and corresponding significance of each

759   set of 7 wells for a given virus gene that were treated with innate immune stimulus were

760   calculated according to one of two different methods. Attached spreadsheet(s) in the folders

761   "coronavirus_IRNF-raw" and "coronavirus_STAT-raw" contain these data of the 7 techical

762   replicates. To further process the data, by a first method the fold change of virus gene-

763   transfected wells treated with an innate immune stimulus was normalized relative to the fold

764   change calculated for wells transfected with the identical vector lacking a virus gene. For the

765   second method, virus gene-transfected wells were normalized by dividing by the mean of the

766   mock-transfected wells for both the stimulated and the mock stimulated conditions. The fold

767   change of normalized virus gene-transfected, innate immune-stimulated wells was then

768   calculated relative to normalized virus gene-transfected, mock-stimulated wells. The more

769   inhibitory of the metrics calculated by the two different methods was taken as the inhibitory

770   score for that virus gene for that stimulus in that experiment. In addition, for subsequent

771   calculations, fold change scores with p>0.1 were considered not significant and were

772   represented as 0 for averaging and summing purposes in subsequent calculations.

773   Data were further aggregated as follows to generate data representations in Figure 3

774   and corresponding Supplementary Figures. For each gene from each virus, data from two to

775   five experiments were averaged; in general, genes that showed no effects in the first two

776   experiments were not re-tested. Inhibition scores were calculated from average fold-change

777   scores by: (a) inverting the positive/negative sign, and (b) scaling the score to be in a range of

778   roughly 0 to 1. The purpose of these calculations is simply to make the inhibition scores more

779   intuitive. The rationale for changing the sign is that, for example, the response to an immune

780   stimulus will be reduced relative to negative controls if a virus gene inhibits that response, so

781   the sign of the fold change score will be negative. The rationale for re-scaling the scores is that

782   the primary fold change scores are very small; this is in part a result of the calculations

783   performed by the proprietary Columbus software, and likely in part results from the fact that we

784   do not perform a background subtraction when assessing signal levels in the initial image

785   processing. Thus, the primary fold change scores are thus highly artificial numbers; our

786   confidence in their meaning results from the fact that differences from controls are statistically

787   significant, that positive control genes such as parainfluenza virus V protein score correctly in

788   our assays, and that coronavirus genes identified by others as major inhibitors of innate immune

789   signaling also behave as expected in our assays.

790   Data represented by heat-maps in Supplementary Figures show the results for all of the

791   coronavirus genes tested for each immunostimulatory input (cGAMP, high-molecular weight

792   polyI:C, low molecular-weight polyI:C in lipofectamine, and interferon-alpha) and transcription

793   factor output (IRF-3, NFkB, STAT1, and phospho-STAT1).

794   Data in Figure 3A-C were generated by summing these data across inputs: specifically,

795   scores for IRF-3 were generated by summing the inhibition scores seen with cGAMP, hi-MW

796   polyI:C, and low-MW polyI:C; and scores for NFkB were summed from hi-MW polyI:C, and low-

797   MW polyI:C scores. These choices are based on the current understanding of innate immune

798   signaling pathways: that cGAMP activates only IRF-3 while the forms of polyI:C stimulate both

799   IRF-3 and NFkB. We considered the summation of these results to be legitimate because the

800   outputs and corresponding staining antibodies are the same across values being summed.

801   Data in Figure 3C result from summing results that derive from staining with an anti-STAT1

802   antibody and an anti-phospho-STAT1 antibody; these are different antibodies and could have

803   different signal/noise properties and thus might yield quantitatively non-comparable results.

804   However, we note that: (a) for each set of staining results, the inhibition scores fall into

805   comparable ranges and that the scores for these two metrics are generally correlated; and (b)

806   both ranges of inhibition scores can be bracketed by the empty vector negative control and the

807   PIV5 V protein positive control.

808

809   **Blinded testing and analysis**

810   For each protein, at least three different samples of the corresponding prepped DNA

811   were given to a third party, who randomized and blinded them. The samples were then tested,

812 analyzed, and the identity of each protein was assigned based on comparison to unblinded

813 samples run concomitantly.

814

High-Content Screening of TNFα signaling in HEK cells

Cell culture (Figure 4 and associated text in Results)

817   HEK293 [ATCC CRL-1573]and HEK293T/17 [ATCC CRL-11268] cells were maintained

818 in DMEM + 10% fetal bovine serum at 37°C in humidified 5% CO_2.

819

NFKB Reporter Lentivirus Construction

821   The fluorescent NFkB reporter construct was made by a Gibson assembly insertion of

822 the mScarlet-hPEST sequence (fpbase.org) into the pGreenfire1 NFkB lentivector (System

823 Biosciences). VSV-G pseudotyped lentiviral particles were made from the resulting lentivector.

824 Briefly, $2*10^6$ HEK293T/17 cells (ATCC) were cotransfected with the NFkB reporter lentivector

825 (2 μg) with pMD2.G (0.5 μg) and psPAX2 (1.5 μg) (Addgene), using Lipofectamine 3000

826 (ThermoFisher) according to the manufacturer's instructions. After an overnight transfection, the

827 media was replaced and cells incubated for 2 days. Lentiviruses were harvested and

828 concentrated from 10 mL cell culture media to 1 mL concentrated lentivirus using Lenti-X

829 concentrator according to the manufacturer's instructions (Takara). Lentivirus aliquots were

830 stored at -80C for further use.

831   The sequence of the resulting construct, "GF_NFKb-mSc-mPGK-Puro," is in

832 Supplemental Information.

833

NFKB Reporter Cell Line Generation

835       $1*10^6$ HEK293 cells, seeded in a 6-well plate, were transduced with 200 μL of reporter

836 lentivirus in the presence of 5 μg/mL Polybrene (Millipore) for two days. The NFkB reporter

837 cells were then cultured in fresh media with 5 μg/mL puromycin (Gibco) for two days to generate

838 a stable pool of reporter cells.

839

High-throughput NFkB Reporter Cell Screening

841   Stably-transduced HEK293 cells (passages 2-8 after transduction) were seeded in

842 100μL in 96 well plates at a density of $3*10^5$ cells/mL. The following day, co-transfection was

843 performed with 75 ng virus gene and 25 ng GFP-NLS per well using Lipofectamine 3000

844 (Thermo Fisher Scientific) according to the manufacturer's instructions. Each gene was

845 transfected into three replicate wells. After 24 hours of transfection, the media + lipofectamine

846 complexes was replaced with fresh media, and the cells were incubated overnight. The next

847 day, cells were stimulated with 5 ng/mL human TNF-α (Peprotech) for 5 hours. The negative

848 transfection control vector consisted of the empty backbone vector (No gene, or 'EV' for empty

849 vector), while positive control vector consisted of a dominant-negative (DN) IkBα (S32A, S36A),

850 a well-established inhibitor of NFkB.

851

852 **Flow Cytometry**

853 Following cytokine stimulation, the media was removed and the cells were stained with

854 Zombie Violet Fixable Viability dye (Biolegend) to quantify live cells according to the

855 manufacturer's instructions. After removing the dye, cells were briefly trypsinized, collected by

856 centrifugation, and resuspended in ice-cold FACS buffer (10% FBS in Dulbecco's phosphate-

857 buffered saline (DPBS) with 5 mM EDTA). Transfected cells were gated based on GFP

858 fluorescence, and statistics on 10,000 transfected cells were collected per well for analysis.

859

860 **Analysis**

861 Flow cytometry data was analyzed in FlowJo. Data was plotted in Graphpad Prism. Each

862 gene was assayed in 3 technical replicates and screened in at least two independent

863 experiments.

864

865 **HSV-1 Assay Protocol (Figure 5 and associated text in Results).**

866 **Preparation of yeast expression strains containing virus genes**

867 We constructed yeast strains that express viral genes from the galactose-inducible

868 *GAL1* promoter by using TAR cloning to combine the virus genes described above with a

869 plasmid, pLDJIF15, that supplies the yeast promoter, 5'end material, 3' end material, and a

870 selectable marker. In this process, the virus gene is amplified from the mammalian expression

871 vector, mixed with pLDJIF15, and the mixture transformed into an auxotrophic yeast.

872 Prototrophic yeast strains will be those in which both DNA fragments have entered the cell and

873 undergone homologous recombination to construct a circular, non-integrated plasmid. Thus,

874 there is no intermediate E. coli carrier of the constructed plasmids.

875 DNA encoding virus proteins was amplified from the mammalian plasmids with primers

876 D298 5'-aatatacctctatactttaacgtcaaggagCTATAGGGAGACCCAAGCTGGCTAGCCACC-3' and

877 D299 5'-aataaaaatcataaatcataagaaattcgcAGAAGGCACAGTCGAGGCTGATCAGCGGGT-3'

878 (lower case: sequences in the yeast expression vector pLDJIF15; upper case: sequences

879 flanking the virus genes in the mammalian expression plasmids) using PrimeSTAR® Max DNA

880   Polymerase (Takara R045A) in a 10 µl reaction. The common Kozak sequence used in the

881   mammalian expression vectors was designed to be consistent with the S. cerevisiae consensus

882   Kozak sequence, since this element is transferred into the yeast expression vectors.

883   Mammalian expression plasmids (0.5 µl) described above were used as template for each

884   reaction.

885   The backbone was amplified from pLDJIF15 with primers D300 5'-

886   GGTGGCTAGCCAGCTTGGGTCTCCCTATAGctccttgacgttaaagtatagaggtatatt-3' and

887   D301 5'-ACCCGCTGATCAGCCTCGACTGTGCCTTCTgcgaatttcttatgatttatgatttttatt-3' using

888   PrimeSTAR® Max DNA Polymerase (Takara R045A). The amplified backbone was treated with

889   DpnI (NEB R0176S) overnight at 37 °C and purified with NucleoSpin® Gel and PCR Clean-up

890   Kit (MN 740609.5). One hundred ng of backbone and 1 µl of virus gene insert linear fragments

891   from the 10 µl reaction were mixed and transformed into yeast (W303α: MATα ade2-1 ura3-1

892   his3-11 trp1-1 leu2-3 leu2-112 can1-00) via electroporation. Transformed yeast cells were

893   plated on -TRP (Teknova C71731) plates for selection. Trp+ colonies were verified by colony

894   PCR with primers: D66 5'-CAACCATAGGATGATAATGCGATTAG-3' and D67 5'-

895   TGAGAAAGCAACCTGACCTACAG-3' using QIAGEN Multiplex PCR Kit (Qiagen 206143). At

896   least 2 size-confirmed transformants were chosen and used in a spheroplast fusion experiment.

897   The sequence of pLDJIF15 is attached as a separate file.

898   For the HSV-1 assay-based experiments, we generally followed the protocol developed

899   by Brown et al (2016) with the following variations. First, we used two yeast strains to fuse with

900   the mammalian cells – one to deliver the viral protein

901   The process generally involves four steps, (1) preparing yeast spheroplasts, (2)

902   preparing the mammalian cells, (3) fusing cells, and (4) analyzing the results. First HSV-1

903   genome containing yeast spheroplasts and the yeast spheroplasts containing the expression

904   plasmid for each gene were prepared, 8 fusions were performed for each gene. To properly

905   synchronize preparing yeast spheroplasts and the HeLa cells, yeast spheroplasts were

906   prepared first and then frozen then thawed and fused with HeLa cells once ready. To prepare

907   yeast spheroplasts; first, inoculate 20 ml of –Trp media with selected yeast strain in a 50 ml

908   falcon tube for the viral test genes or -URA media for HSV-1 strain. Twenty-two different

909   samples were prepared at once at 30°C and grown overnight. Yeast cells were resuspended in

910   40 ml YPG media (or YPD media for HSV-1 samples) by adding directly to tube. Then grown

911   for 5-6 h at 30°C. Yeast cells were collected by centrifugation at 3,000 rpm for 3 min (50 ml

912   Corning Inc. centrifuge tube) and re-suspend in 20 ml 1M Sorbitol. Cells were kept at 4°C 18h.

913   Yeast cells were collected by centrifugation 3,000 rpm/3 min, and re-suspend in 20 ml SPEM. A

914 1:10 dilution of OD should be ~ 0.9. 20µl ß-Mercaptoethanol (Sigma) and 20µl Zymolyase-20T

915 [Zymolyase stock: 200 mg Zymolyase, 9 ml H2O, 1ml 1M Tris pH7.5, 10 ml 50% glycerol. Store

916 at –20ºC] were added. Yeast cells were incubated at 37°C for 23-25 min with gentle agitation.

917 Checking the OD600 of the yeast suspension: 0.1ml+0.9 ml 1M Sorbitol; 0.1 ml+0.9 ml 2%

918 SDS. Compare the OD readings of A/B ratio. Sheroplasting was stopped when the difference is

919 3-4 fold by adding 30 ml 1M sorbitol and gently mix (inverting). Cells were collected by

920 centrifugation at 1800 rpm, 8 min, 4ºC. Yeast spheroplasts were resuspended with 2 ml 1M

921 Sorbitol. Yeast spheroplasts counted using the OD calculation worksheet to determine the

922 amount of yeast cells needed for each fusion. Two samples worth of spheroplast solution

923 calculated from worksheet was added to a 1.5 ml tube (typically, 40 – 400 µl for each 2-sample

924 tube).   Yeast spheroplasts resuspended with 200µl 1M Sorbitol +15% DMSO and stored at -

925 80°C to be used when the mammalian cells are ready.

926   To prepare Mammalian cells, HeLa cells were grown to 70 to 80% confluence, in T150

927 flasks with 35 ml DMEM media + 10% FBS + penicillin streptomycin and amphotericin B. Next

928 200 µl Colcemid was added to each flask and incubated overnight at 37°C. Flask was striked 3

929 times to dislodge loose cells. Media removed from HeLa cells and placed in an empty flask and

930 cells trypsinized to remove cells from flask. Once cells have detached, cells were resuspended

931 in saved spent media. To recover the frozen yeast spheroplasts: they were thawed on ice 5

932 min, 22µl 10X TC added to each tube and mixed, centrifuged yeast spheroplasts 4000 rpm 1

933 min at 4°C, resuspended yeast spheroplasts in 400 µl STC to yield 2 samples worth per tube

934 and kept at 4°C until ready.

935   Next the cell fusion was performed. PEG solution (50% PEG 2000 + 10%DMSO in

936 75mM HEPES buffer) was prepared: 12.5g of PEG to a 50 ml tube, add 75mM HEPES to ~22.5

937 ml mark, invert on a shaker for ~30 min until dissolved, top with 75mM HEPES to 25ml, add

938 2.77 ml DMSO and mix. Mammalian cells counted using hemocytometer and add 3×10^5 HeLa

939 cells added to tube. HeLa cells were centrifuged down and resuspended in 500µl saved spent

940 media. Next, 200 µl of yeast spheroplasts from previous step was added to tube, and cell

941 mixture incubated 5 min at room temperature. Pelleted cells at 5000 rpm for 30 sec media

942 removed. Next, 500µl of PEG solution was added and incubated for 5 min at room temperature.

943 Fusion was ended by adding 1ml serum free DMEM and spinning down at 4000 rpm in a

944 tabletop centrifuge for 30 sec, media removed, and 1ml of DMEM media supplemented with

945 10% FBS, penicillin streptomycin and amphotericin B added to the sample. Next, 500 µl of

946 DMEM media supplemented with 10% FBS, penicillin streptomycin and amphotericin B was

947 added to a 24 well dish, and 500 µl of the fused, recovered sample was added to the 24 well

948   dish, after 4 h after cells had settled and reattached to plate so media was removed and

949   replaced with 500 µl fresh media to remove excess yeast. Evidence of fused cells can be seen

950   96h post fusion for HSV-1 replication and Tecan using fluorescence protocol analyzed sample

951   plate.

952

953

954

958

959   **AUTHOR CONTRIBUTIONS**

960   Erika J. Olson performed and supervised data collection for the screens in Figures 2 and 3,

961   consolidated the data from these screens, performed data analysis and curation, and

962   contributed to the writing; Tai L. Ng performed some the screens in Figures 2 and 3; H. Sloane

963   Weiss constructed plasmids used in the experiments and performed some the screens in

964   Figures 2 and 3; Timothy Z. Chang performed the screens in Figure 4; David M. Brown and Lin

965   Ding performed screens and constructed plasmids for data in Figure 5; Yukiye Koide

966   constructed the BJ-5ta cGAS knockout strain and expression vector design; Peter Koch, Yukiye

967   Koide, Pia Mach, Tobias Meisinger, and Timothy Mitchison contributed to the screening

968   methodology; Nathan Rollins, Joshua Rollins and Debora Marks performed to the sequence

969   analysis and identification of CaORF15, Trenton Bricken and Colin Molloy contributed to the

970   image processing; Yun Zhang contributed to the statistical analysis; Bridget N. Queenan

971   contributed to the writing and figure design; Jeffrey C. Way was primarily responsible for funding

972   acquisition and initial project conceptualization with assistance from Pamela A. Silver, John I.

973   Glass, Timothy Mitchison, and Debora Marks; Pamela Silver was responsible for the overall

974   project administration.

975

976   **DECLARATION OF INTERESTS**

977   The authors declare that they have no conflicts of interest.

978

979

980 **REFERENCES**
981
982 Boni, M.F., Lemey, P., Jiang, X., Lam, T.T.Y., Perry, B.W., Castoe, T.A., Rambaut, A.,
983 Robertson, D.L. (2020) Evolutionary Origins of the SARS-CoV-2 Sarbecovirus Lineage
984 Responsible for the COVID-19 Pandemic. Nat. Microbiol. *5*, 1408-1417.
985
986 Brown, D.M., Chan, Y.A., Desai, P.J., Grzesik, P., Oldfield, L.M., Vashee, S., Way, J.C., Silver,
987 P.A., Glass, J.I. (2016) Efficient Size-Independent Chromosome Delivery from Yeast to Cultured
988 Cell Lines. Nucleic Acids Research *45*: doi: 10.103/nar/gkw1252.
989
990 Eaglesham, J.B., Pan, Y., Kupper, T.S., Kranzusch, P.J. (2019) Viral and Metazoan Poxins are
991 cGAMP-Specific Nucleases that Restrict cGAS-STING signalling. Nature *566,* 259-263.
992
993 Enninga, J., Levy, D.E., Bloble, G., Fontoura, B.M.A. (2002) Role of Nucleoporin Induction in
994 Releasing an mRNA Nuclear Export Block. Science *295*: 1523-1525.
995
996 Freitas, B.T., Durie, I.A., Murray, J., Longo, J.E., Miller, H.C., Crich, D., Hogan, R.J., Tripp, R.A.,
997 Pegan, S.D. (2020) Characterization and Noncovalent Inhibition of the Deubiquitinase and
998 delSGylase Activity of SARS-CoV-2 Papain-Like Protease. ACS Infect. Dis. *6*: 2099-2109.
999
1000 Frieman, M., Yount, B., Heise, M., Kopecky-Bromberg, S.A., Palse, P., Baric, R.S. (2007)
1001 Severe Acute Respiratory Syndrome Coronavirus ORF6 Antagonizes STAT1 Function by
1002 Sequestering Nuclear Import Factors on the Rough Endoplasmic Reticulum/Golgi Membrane. J.
1003 Virol. *81*: 9812-9824.
1004
1005 Guan S, Bastin DA, Verma NK. Functional analysis of the O antigen glucosylation gene cluster
1006 of Shigella flexneri bacteriophage SfX. Microbiology (Reading). 1999;145 (Pt 5):1263-1273.
1007 PMID: 10376843
1008
1009 Gandini R, Reichenbach T, Tan TC, Divne C. Structural basis for dolichylphosphate mannose
1010 biosynthesis. Nat Commun. 2017 Jul 25;8(1):120. PMID: 28743912
1011
1012 Hoffmann, H-H., Schneider, W.M., Rice, C.M. (2015) Interferons and Viruses: An Evolutionary
1013 Arms Race of Molecular Interactions. Trends Immunol. *36(3)*: 124-138.
1014
1015 Kau, T.R., Schroeder, F., Ramaswamy, S., Wojciechowski, C.L., Zhao, J.J., Roberts, T.M.,
1016 Clardy, J., Sellers, W.R., Silver, P.A. (2003) A Chemical Genetic Screen Identifies Inhibitors of
1017 Regulated Nuclear Export of a Forkhead Transcription Factor in PTEN-Deficient Tumor Cells.
1018 Cancer Cell *4(6)*: 463-476.
1019
1020 Koch, P.D., Miller, H., Yu, G., Tallarico, J., Sorger, P.K., Wang, Y., Feng, Y., Thomas, J.R.,
1021 Ross, N.T., Mitchison, T. (2018) A High Content Screen in Macrophages Identifies Small
1022 Molecule Modulators of STING-IRF3 and NFkB Signaling. ACS Chem. Biol. *13(4)*: 1066-1081.
1023
1024 Kouprina, N., Larionov, V. (2016) Transformation-Associated Recombination (TAR) Cloning for
1025 Genomic Studies and Synthetic Biology. Chromosoma *125(4)*: 621-632.
1026
1027 Kuhl, A., Pohlmann, S. (2012) How Ebola Virus Counters the Interferon System. Zoonoses
1028 Public Health *59*: 116-131.
1029

1030 Jian Lei 1, Rolf Hilgenfeld. (2017) RNA-virus proteases counteracting host innate immunity.
1031 FEBS Lett 591(20):3190-3210.
1032
1033 Lei, J., Kusov, Y., Hilgenfeld, R. (2018) Nsp3 of Coronaviruses: Structures of Functions of a
1034 Large Multi-Domain Protein. Antiviral. Res. *149*: 58-74.
1035
1036 Lei, X., Dong, X., Ma, R., Wang, W., Xiao, X., Tian, Z., Wang, C., Wang, Y., Li, L., Ren, L., Guo,
1037 F., Zhao, Z., Zhou, Z., Xiang, Z., Wang, J. (2020) Activation and Evasion of Type I Interferon
1038 Responses by SARS-CoV-2. Nat. Commun. *11*: 3810.
1039
1040 Maher, S.G., Romero-Weaver, A.L., Scarzello ,A. J., Gamero, A. M., (2007) Interferon: cellular
1041 executioner or white knight? Curr Med Chem 14(12):1279-89.
1042
1043 Masters, S.L., O'Neill, L.A.J. (2011) Disease-Associated Amyloid and Misfolded Protein
1044 Aggregates Activate the Inflammasome. Trends Mol. Med. *17*: 276-282.
1045
1046 Mayer, T.U., Kapoor, T.M., Haggarty, S.T., King, R.W., Schreiber, S.L., Mitchison, T.J. (1999)
1047 Small Molecular Inhibitor of Mitotic Spindle Bipolarity Identified in a Phenotype-Based Screen.
1048 Science *286*: 971-974.
1049
1050 Miorin, L., Kehrer, T., Sanchez-Aparicio, M.T., Zhang, K., Cohen, P., Patel, R.S., Cupic, A.,
1051 Makio, T., Mei, M., Moreno, E., et al. (2020) SARS-CoV-2 Orf6 Hijacks Nup98 to Block STAT
1052 Nuclear Import and Antagonize Interferon Signaling. PNAS *117*: 28344-28354.
1053
1054 Narayanan, K., Ramirez, S.I., Lokugamage, K.G., Makino, S. (2015) Coronavirus nonstructural
1055 protein 1: Common and distinct functions in the regulation of host and viral gene expression.
1056 Virus Res *202*:89-100.
1057
1058 Rath, O., Kozielski, F. (2012) Kinesins and Cancer. Nat. Rev. Cancer *12*: 527-539.
1059
1060 Sun, L., Wu, J., Du, F., Chen, X., Chen, Z. J. (2013) Cyclic GMP-AMP synthase is a cytosolic
1061 DNA sensor that activates the type I interferon pathway. Science *339*(6121):786-91.
1062
1063 Syed, Y.Y. (2019) Selinexor: First Global Approval. Drugs *79*: 1485-1494.
1064
1065 Tidu, A., Janvier, A., Schaeffer, L., Sosnowski, P., Kuhn, L., Hammann, P., Westhof, E., Eriani,
1066 G., Martin., F. (2020) The Viral Protein NSP1 Acts as a Ribosome Gatekeeper for Shutting
1067 Down Host Translation and Fostering SARS-CoV-2 Translation. RNA doi:
1068 10.1261/rna.078121.120.
1069
1070 Wang, S., Wang, K., Lin, R., Zheng, C. (2013) Herpes Simplex Virus 1 Serine/Threonine Kinase
1071 US3 Hyperphosphorylates IRF3 and Inhibits Beta Interferon Production. J. Virol. *87*: 12814-
1072 12817.
1073
1074 WHO INN (International Nonproprietary Names Program). (2020) Entry 11889. Messenger RNA
1075 encoding the full-length SARS-CoV-2 spike glycoprotein.
1076
1077 Xiao, T.S., Fitzgerald, K.A. (2013) The cGAS-STING Pathway for DNA Sensing. Mol. Cell *51*:
1078 135-139.
1079

1080   Carisa A Zampieri, Nancy J Sullivan, Gary J Nabel. (2007) Immunopathology of highly virulent
1081   pathogens: insights from Ebola virus. Nat Immunol 8(11):1159-64. doi: 10.1038/ni1519.

1082 **SUPPLEMENTAL INFORMATION**

1083 Supplementary Information for Olson et al., "High-content screening of coronavirus genes for

1084 innate immune suppression reveals enhanced potency of SARS-CoV-2 proteins."

1085

1086 **BLAST analysis of segments of the Spike coding sequence.**

1087   To characterize the rate of evolution of the segment of Spike gene that also encodes

1088 CaORF15, we performed a BLAST alignment of segments of the SARS-CoV-2 Spike coding

1089 sequence with the Spike coding sequence from the coronaviruses RaTG13, pangolin M789, and

1090 SARS-CoV-1. Numerical values presented here correspond to the bar graph in Figure 1.

1091

1092 Supplementary Figure 1.

	21541-21935	21936-22199	22200-22799	22800-23399	23400-239990	24000-24599	24600-25199	25200-25384
Bat RatG13								
% identity	340/373(91%)	258/264(98%)	563/600(94%)	519/600(87%)	550/600(92%)	571/600(95%)	571/600(95%)	179/185(97%)
Gaps	0/373(0%)	0/264(0%)	0/600(0%)	4/600(0%)	12/600(2%)	0/600(0%)	0/600(0%)	0/185(0%)
Pangolin M789								
% identity	254/381(67%)	Not detected	390/469(83%)	532/598(89%)	536/600(89%)	550/600(92%)	542/600(90%)	169/185(91%)
Gaps	23/381(6%)		0/469(0%)	0/598(0%)	12/600(2%)	0/600(0%)	0/600(0%)	0/185(0%)
SARS-CoV-1								
% identity	183/272(67%)	Not detected	332/455(73%)	430/601(72%)	443/599(74%)	494/599(82%)	500/598(84%)	150/183(82%)
Gaps	23/272(8%)		2/455(0%)	9/601(1%)	14/599(2%)	2/599(0%)	0/598(0%)	0/183(0%)

1093

1094

1095   These results are comparable to Figure 1b of Boni et al. (2020), which indicate that the

1096 nucleotide sequence of the 5'-most region of the Spike gene in SARS-CoV-2 (which also

1097 encodes CaORF-15) is the most divergent region of the genome relative to the viruses SARS-

1098 CoV Guangzhou 2002 (HSZ-Cc), Zhejiang 2012 (Longquan_140),Hong Kong 2005 (HKU3-1),

1099 Zhejiang 2017 (CoVZC45), Zhejiang 2015 (CoVZXC21), Pangolin Guangxi 2017 (P1E), and

1100 Pangolin Guangdong 2019. However, this region is quite similar to Yunnan 2013 (RaTG13),

1101 which also encodes a CaORF15.

1102

1103 **Conservation of CaORF15 in SARS-CoV-2 strains**

1104   We then asked whether caORF15 was conserved among current variants of SARS-CoV-

1105 2 by examining the multiple sequence alignment of SARS-CoV-2 genomes provided by GISAID

1106 (Elbe & Buckland-Merret 2017 doi: 10.1002/gch2.1018). As of January 2021, of 276,082 human

1107 SARS-CoV-2 samples, 98% contain intact caORF15 and 92% have wholly identical nucleotide

1108 sequence to caORF15 to Wuhan-hu-1.

1109

1110

1111 **Breakdown of effects of coronavirus genes on innate immune signaling**

1112   The following Figures are constructed in the same manner as Figure 3A-C, but represent

1113   the data for individual inputs and outputs.


1114
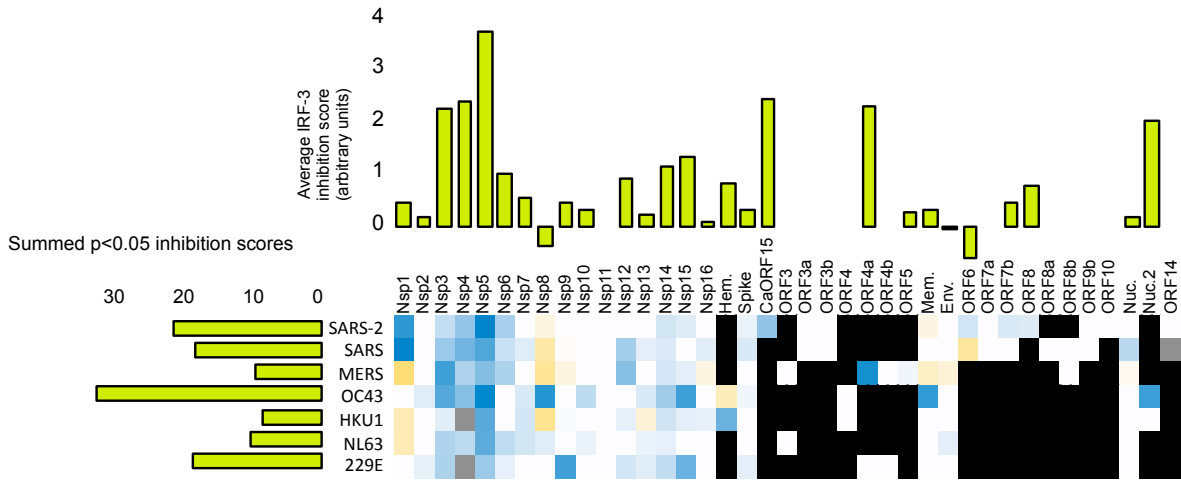
1115   Supplementary Figure 2A. Inhibition of cGAMP activation of IRF-3 by coronavirus genes. This

1116   represents signaling through the STING pathway. These data, along with Supplementary

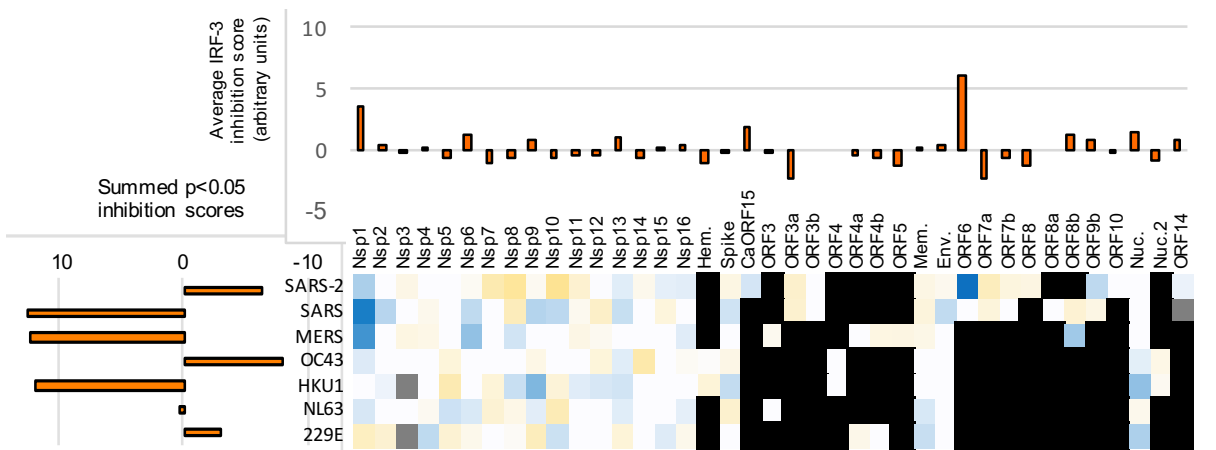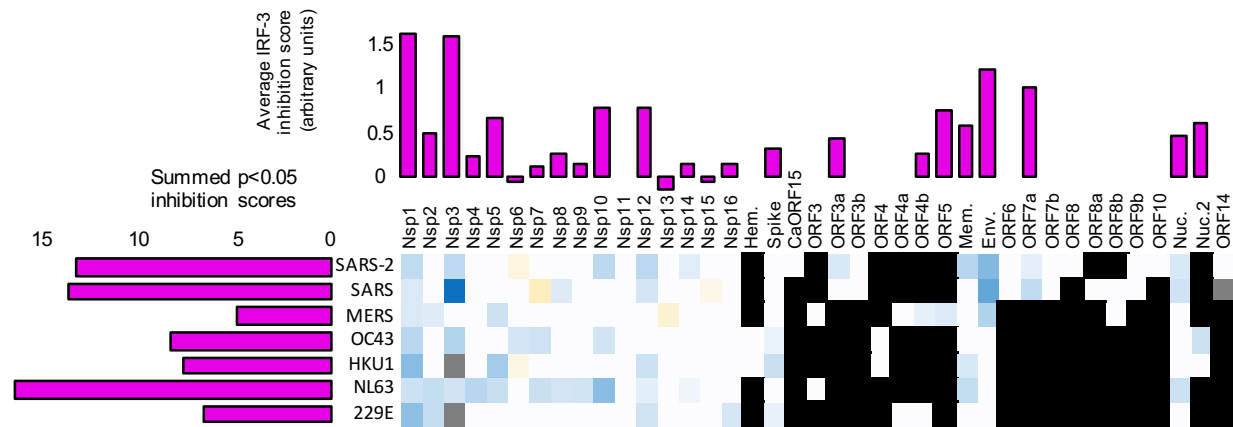1117   Figures 2B and 2C are summed to generate Figure 3A.

1118

1119
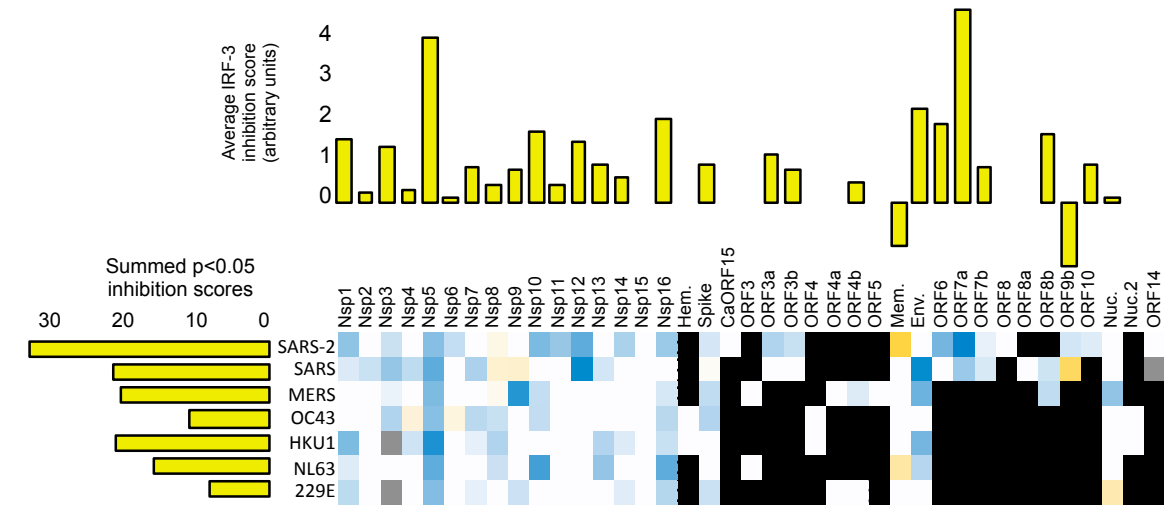
1120   Supplementary Figure 2B. Inhibition of high-molecular weight polyI:C activation of IRF-3 by

1121   coronavirus genes.   This represents signaling through the TLR-3 endosomal pathway.

1122
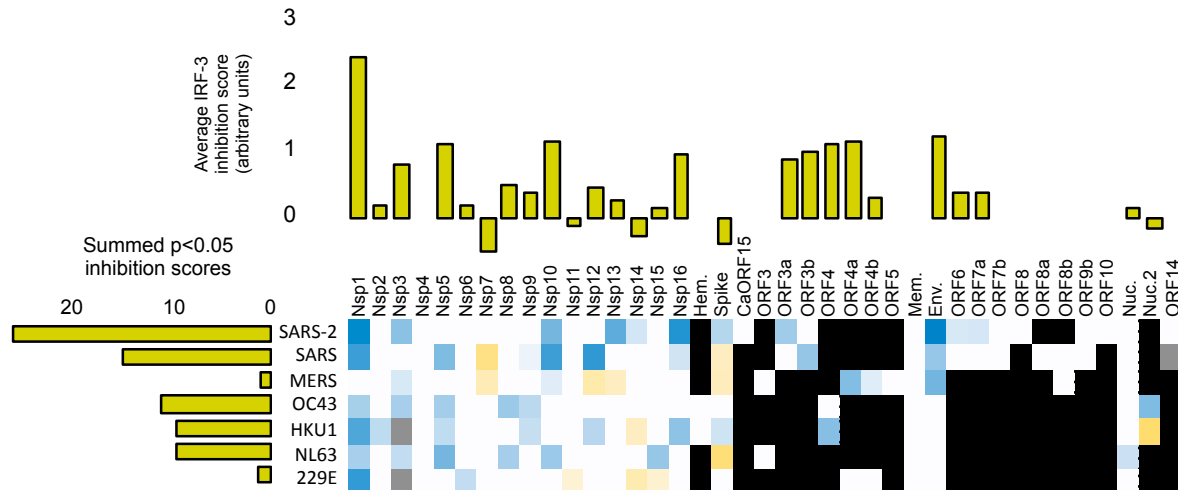
Supplementary Figure 2C. Inhibition of low molecular weight polyI:C/lipofectamine activation of IRF-3 by coronavirus genes. This represents signaling through the RIG-I-like receptor cytoplasmic pathway.
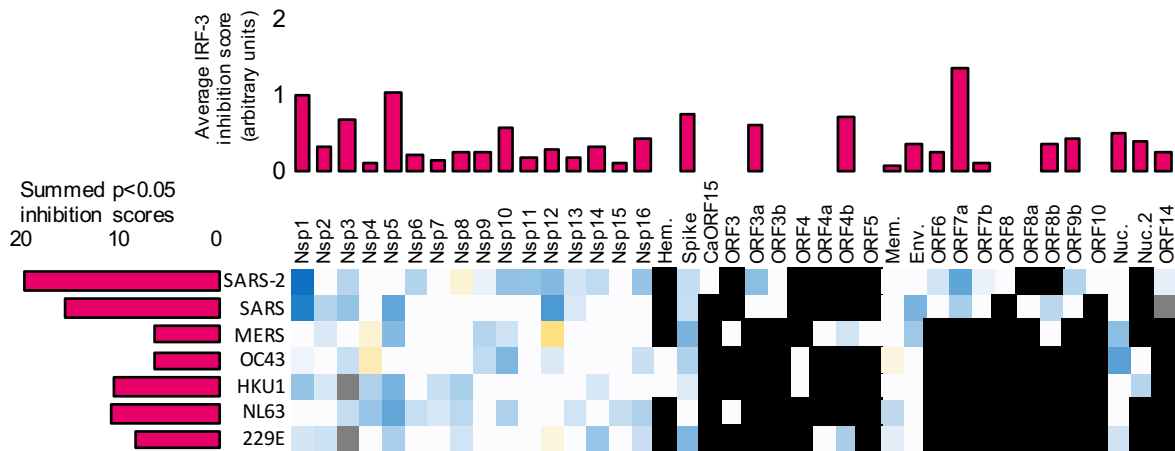

Supplementary Figure 2D. Inhibition of high-molecular weight polyI:C activation of NFkB by coronavirus genes. This presumably represents signaling through the TLR-3 endosomal pathway.

42

Supplementary Figure 2E. Inhibition of low molecular weight polyI:C/lipofectamine activation of IRF-3 by coronavirus genes. This presumably represents signaling through the RIG-I-like receptor cytoplasmic pathway.

Supplementary Figure 2F. Inhibition by coronavirus genes of an increase in the nuclear/cytoplasmic ratio of phospho-STAT1 induced by interferon alpha.
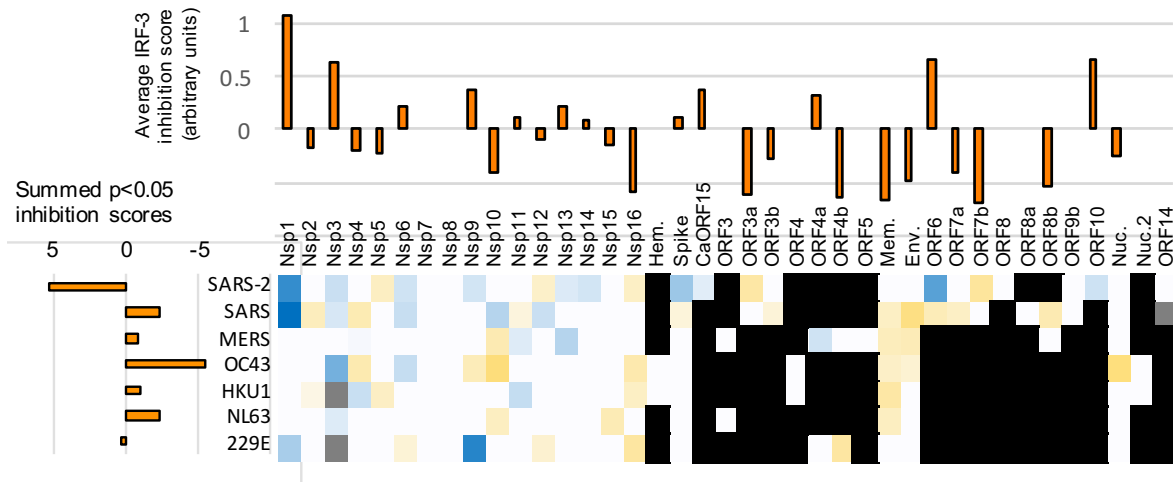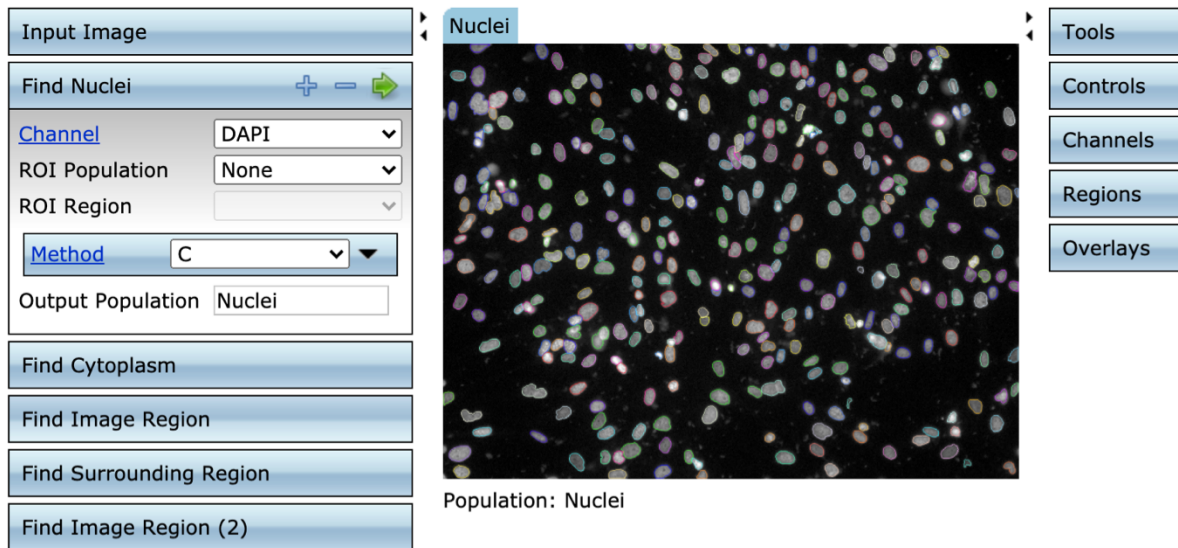
1146 Supplementary Figure 2G. Inhibition by coronavirus genes of an increase in the

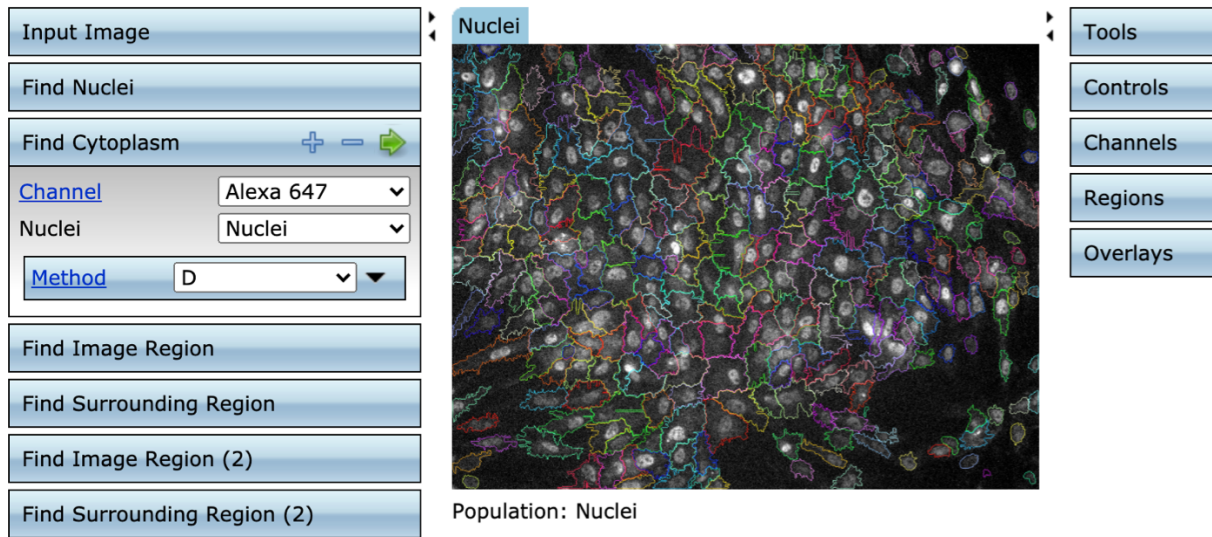1147 nuclear/cytoplasmic ratio of STAT1 induced by interferon alpha.
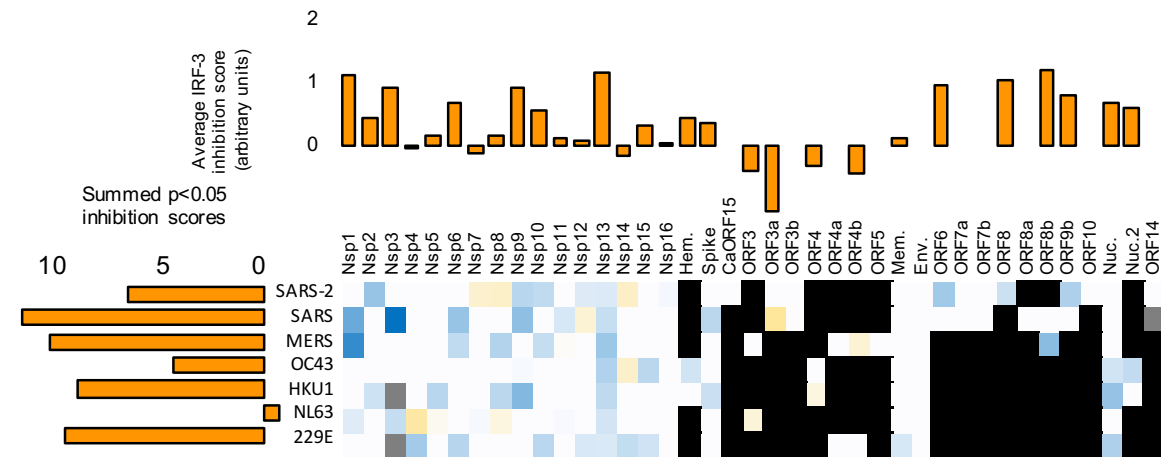
1150 Supplementary Figure 2H. Effect of coronavirus genes on the total cellular amount of phospho-

1151 STAT1 induced by interferon alpha.

1153
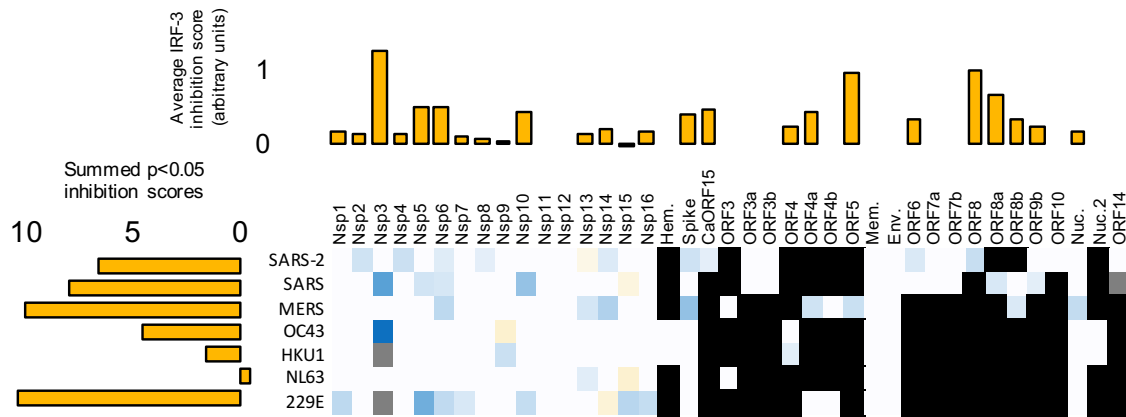
Supplementary Figure 2I. Effect of coronavirus genes on the total cellular amount of STAT1

after treatment by interferon alpha.

Supplementary Figure 3. Image processing. Typical screen shot using the Columbus software,

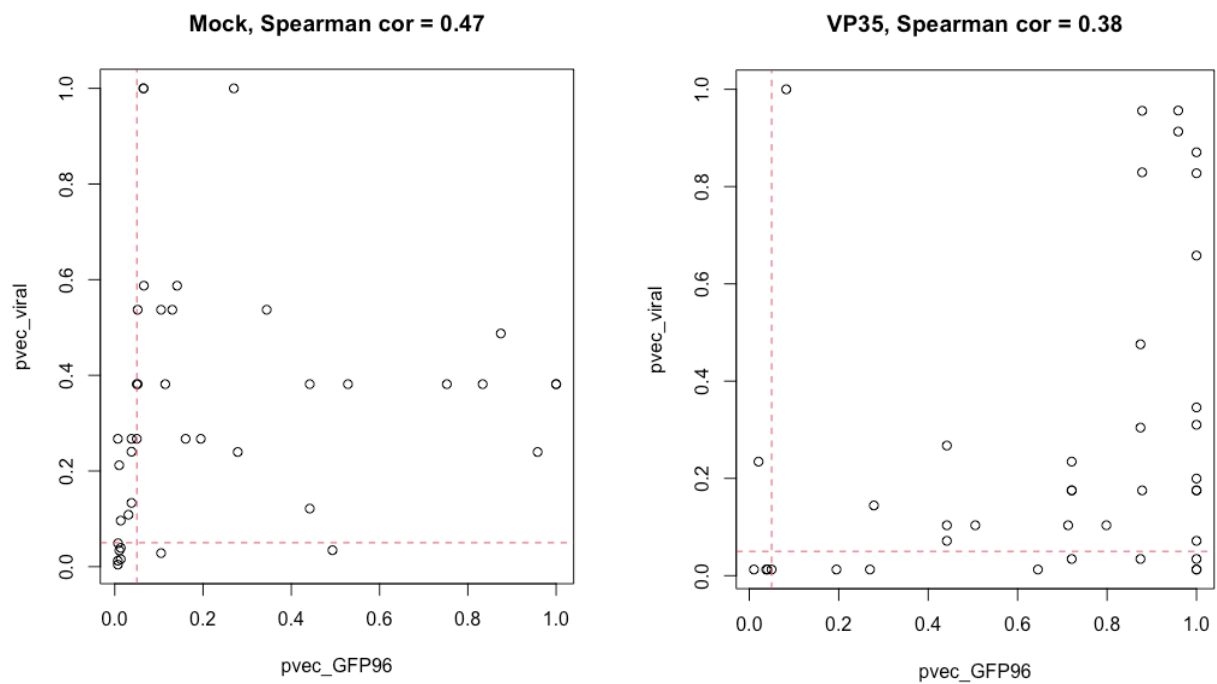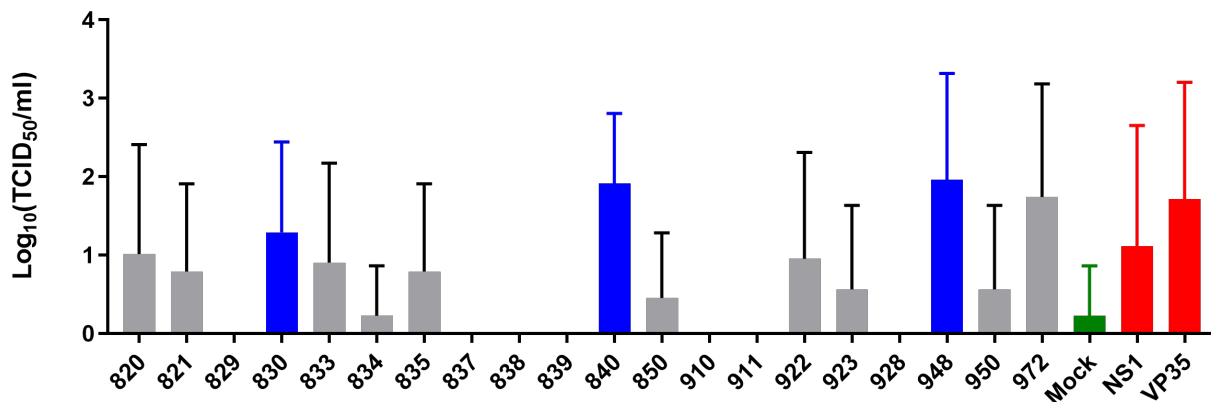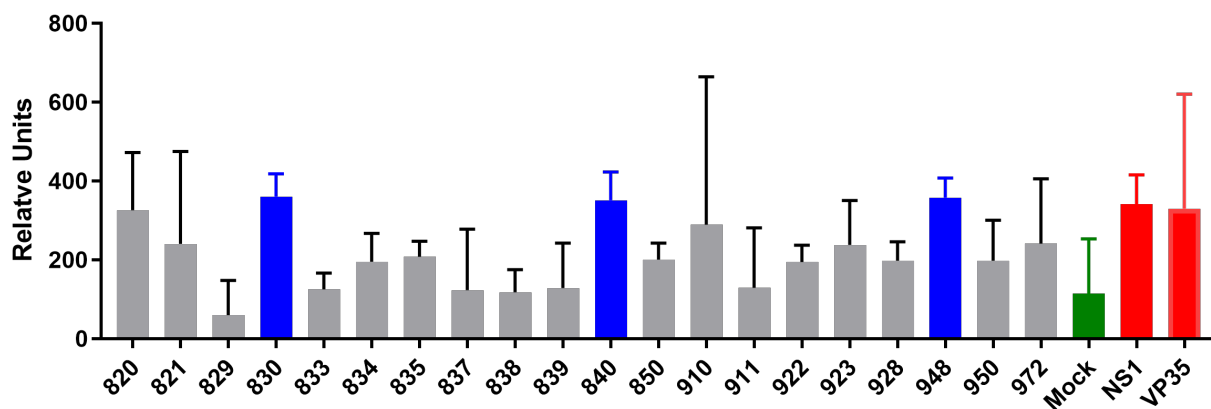showing nuclear segmentation (top) and cytoplasmic segmentation (bottom).

1159

Input Image		Nuclei		Tools
Find Nuclei				Controls
Find Cytoplasm	⊕ ➖ ➡			Channels
Channel	Alexa 647			Regions
Nuclei	Nuclei			Overlays
Method	D			
Find Image Region				
Find Surrounding Region				
Find Image Region (2)				
Find Surrounding Region (2)				

Population: Nuclei

1160

1161

1162   A.

Mock, Spearman cor = 0.47   **VP35, Spearman cor = 0.38**

pvec_viral vs pvec_GFP96

1163
1164   B.

46

1165

1166
1167 **Supplementary figure 4.** Comparison of HSV GFP relative fluorescence data to TCID$_{50}$ viral
1168 replication data. A Wilcoxon rank sum test was used to assess the detection power of testing
1169 the difference between each coronavirus gene and positive/negative control using viral titers
1170 and GFP fluorescence with in the HSV-1 based assay. **A.** The negative (Mock) and positive
1171 (Influenza virus NS1 and Ebola VP35) control p-values were plotted on panel A and B,
1172 respectively. In each graph, x-axis is p-value using GFP fluorescence; y-axis is p-value using
1173 viral titers; red dashed line indicates p-value = 0.05. The hypothesis that if the two assay types
1174 produce similar results, then their p-values would have high correlation, suggesting similar
1175 detection power. Results showed a spearman correlation of 0.47 for mock (negative control)
1176 (panel A) and 0.38 for VP35 (positive control) (panel B), both indicating moderate correlation.
1177 **B.** The TCID$_{50}$ viral replication titers from twenty coronavirus genes as well as the positive
1178 controls and negative control are shown in panel C and compared to the GFP data associated
1179 with each gene. Error bars indicate SD. Blue bars represent the coronavirus genes that were
1180 positive in the two assays and grey bars represent the coronavirus genes that were negative.
1181 Red bars are for the positive controls and the green bar is for the mock negative control.
1182