

1 **Analysis of blood and nasal epithelial transcriptomes to identify**
2 **mechanisms associated with control of SARS-CoV-2 viral load in the**
3 **upper respiratory tract**

4
5 Mahdi Moradi Marjaneh^{1,2,3*}, Joseph D Challenger⁴, Antonio Salas^{5,6,7}, Alberto Gómez-
6 Carballa^{5,6,7}, Abilash Sivananthan¹, Irene Rivero-Calle^{6,7,8}, Gema Barbeito-Castiñeiras⁹,
7 Cher Y Foo¹⁰, Yue Wu¹¹, Felicity Liew¹², Heather R Jackson^{1,2}, Dominic Habgood-Coote^{1,2},
8 Giselle D’Souza^{1,2}, Samuel Nichols^{1,2}, Victoria J Wright^{1,2}, Michael Levin^{1,2}, Myrsini
9 Kaforou^{1,2}, Ryan S Thwaites¹², Lucy C Okell⁴, Federico Martín-Torres^{6,7,8}, Aubrey J
10 Cunningham^{1,2*}, on behalf of the PERFORM Consortium[†] and GEN-COVID Study Group
11 (<http://gencovid.eu>)

12

13 ¹Section of Paediatric Infectious Disease, Department of Infectious Disease, Imperial College
14 London, London, UK

15 ²Centre for Paediatrics and Child Health, Imperial College London, London, UK

16 ³Section of Virology, Department of Infectious Diseases, Faculty of Medicine, Imperial
17 College London, London, UK

18 ⁴Medical Research Council Centre for Global Infections Disease Analysis, Department of
19 Infectious Disease Epidemiology, Imperial College London, London, UK

20 ⁵Unidade de Xenética, Instituto de Ciencias Forenses, Facultade de Medicina, Universidade
21 de Santiago de Compostela, and GenPoB Research Group, Instituto de Investigación
22 Sanitaria (IDIS), Hospital Clínico Universitario de Santiago (SERGAS), Galicia, Spain

23 ⁶Genetics, Vaccines and Infections Research Group (GENVIP), Instituto de Investigación
24 Sanitaria de Santiago, Universidade de Santiago de Compostela, Santiago de Compostela,
25 Galicia, Spain

26 ⁷Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBER-ES),
27 Madrid, Spain

28 ⁸Translational Pediatrics and Infectious Diseases, Department of Pediatrics, Hospital Clínico
29 Universitario de Santiago de Compostela, Santiago de Compostela, Galicia, Spain

30 ⁹Servicio de Microbiología y Parasitología, Complejo Hospitalario Universitario de Santiago
31 de Compostela, Santiago de Compostela, Galicia, Spain

32 ¹⁰School of Medicine, Imperial College London, London, UK

33 ¹¹Department of Surgery and Cancer, Imperial College London, St. Mary's Hospital, London,
34 UK

35 ¹²National Heart and Lung Institute, Imperial College London, London, UK

36 †Personalised Risk Assessment in Febrile Illness to Optimise Real-Life Management
37 (PERFORM), London, UK.

38

39 *Co-corresponding Authors

40

41

42

43

44

45

46

47

48

49

50 **Abstract**

51 **Background:** The amount of SARS-CoV-2 detected in the upper respiratory tract (URT
52 viral load) is a key driver of transmission of infection. Current evidence suggests that
53 mechanisms constraining URT viral load are different from those controlling lower
54 respiratory tract viral load and disease severity. Understanding such mechanisms may help
55 to develop treatments and vaccine strategies to reduce transmission. Combining
56 mathematical modelling of URT viral load dynamics with transcriptome analyses we aimed
57 to identify mechanisms controlling URT viral load.

58 **Methods:** COVID-19 patients were recruited in Spain during the first wave of the
59 pandemic. RNA sequencing of peripheral blood and targeted NanoString *nCounter*
60 transcriptome analysis of nasal epithelium were performed and gene expression analysed in
61 relation to paired URT viral load samples collected within 15 days of symptom onset.
62 Proportions of major immune cells in blood were estimated from transcriptional data using
63 computational differential estimation. Weighted correlation network analysis (adjusted for
64 cell proportions) and fixed transcriptional repertoire analysis were used to identify
65 associations with URT viral load, quantified as standard deviations (z-scores) from an
66 expected trajectory over time.

67 **Results:** Eighty-two subjects (50% female, median age 54 years (range 3-73)) with COVID-
68 19 were recruited. Paired URT viral load samples were available for 16 blood transcriptome
69 samples, and 17 respiratory epithelial transcriptome samples. Natural Killer (NK) cells were
70 the only blood cell type significantly correlated with URT viral load z-scores ($r = -0.62$, $P =$
71 0.010). Twenty-four blood gene expression modules were significantly correlated with URT
72 viral load z-score, the most significant being a module of genes connected around *IFNA14*
73 (Interferon Alpha-14) expression ($r = -0.60$, $P = 1e-10$). In fixed repertoire analysis,
74 prostanoid-related gene expression was significantly associated with higher viral load. In

75 nasal epithelium, only *GPLY* (granulysin) gene expression showed significant negative
76 correlation with viral load.

77 **Conclusions:** Correlations between the transcriptional host response and inter-individual
78 variations in SARS-CoV-2 URT viral load, revealed many molecular mechanisms plausibly
79 favouring or constraining viral load. Existing evidence corroborates many of these
80 mechanisms, including likely roles for NK cells, granulysin, prostanoids and interferon
81 alpha-14. Inhibition of prostanoid production, and administration of interferon alpha-14 may
82 be attractive transmission-blocking interventions.

83

84 **Keywords**

85 COVID-19, SARS-CoV-2, upper respiratory tract, viral load, mathematical modelling,
86 transcriptome, gene network analysis

87

88 **Background**

89 The advent of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) leading to
90 the coronavirus disease 2019 (COVID-19) has placed an enormous burden on affected
91 individuals, healthcare systems, and economies worldwide. SARS-CoV-2 is highly
92 transmissible and causes a wide range of severity from asymptomatic infection to severe
93 disease and death. The amount of SARS-CoV-2 detected in the upper respiratory tract of
94 infected individuals (URT viral load) is a key driver of transmission of infection (1). High
95 URT viral loads can increase household and non-household transmissions by up to nearly
96 60% and 40%, respectively (2). Interestingly, URT viral load does not necessarily correlate
97 with severity of illness, nor is it determined by established risk factors for poor outcome
98 such as age and sex (3, 4). This suggests that the host immune mechanisms involved in
99 constraining the virus in the URT are different from those determining the severity of

100 illness, although such mechanisms have not been fully elucidated. In contrast, high and
101 persistent SARS-CoV-2 shedding in the lower respiratory tract (LRT) is associated with
102 severe disease (5), indicating differences in the mechanisms underlying control and
103 pathogenesis of SARS-CoV-2 in the URT and LRT. Understanding the mechanisms
104 controlling the viral load in the URT could illuminate new strategies to prevent transmission
105 from infected individuals and might also enable control of the localised infection before it
106 progresses to the LRT, triggering more serious illness.

107 URT viral load is highly dynamic. It changes over the course of illness due to dynamic
108 interactions with the host immune response; it peaks around the time of symptom onset and
109 then gradually decreases to a low level over the following 10 days (6, 7). Moreover, the
110 kinetics of viral load vary between individuals, presumably determined by variation in
111 immune responses (3). The host response constraining viral load includes both an immediate
112 innate component and a later adaptive response (3, 8, 9). With limited *in vivo* data,
113 researchers have attempted to mathematically model and explain the viral-host interaction
114 and host immune responses to better understand the dynamics of SARS-CoV-2 viral load.

115 We have recently developed a within-host model that has been successful in interpreting
116 URT viral load kinetics in a wide range of data including 2172 serial measurements from
117 605 subjects, collected from 17 different studies (3).

118 Despite the dynamic interaction between the virus and host immune system during SARS-
119 CoV-2 infection and the diversity in such interaction observed between individuals, the
120 immune response involves conserved elements which can be reflected in host transcriptomes
121 (10). While gene expression is a dynamic process, and a single transcriptomic experiment
122 usually captures only a “snapshot” in time, using robust transcriptional analyses we can
123 pinpoint key biological mechanisms underlying the immune response. The host
124 transcriptomic response in human infection is often studied in peripheral blood leukocytes.

125 This is because peripheral blood leukocytes mount cell-intrinsic responses to pathogens but
126 also mount transcriptional responses to signals arising from the organs through which they
127 circulate. Evaluating the host transcriptome in the context of the dynamics of host-pathogen
128 interaction can be a powerful approach to elucidate mechanisms responsible for control of
129 pathogen load (11).

130 Here we sought to combine mathematical modelling of URT SARS-CoV-2 viral load
131 dynamics in individual subjects with the analysis of peripheral blood and nasal epithelium
132 transcriptomes to identify mechanisms associated with the control of viral load. The
133 mechanistic correlates of URT viral load identified herein may be important to develop new
134 therapeutic and vaccine strategies to block transmission of SARS-CoV-2.

135

136 **Results**

137 **Participants**

138 We performed transcriptome analyses for 82 COVID-19 patients (50% female, median age
139 54 years (range 3-73 years)) recruited during the “first wave” of COVID-19 in Spain, before
140 vaccination and natural infection became determinants of the immune response to SARS-
141 CoV-2. Whole blood RNA sequencing (RNA-Seq) and nasal epithelium *n*Counter
142 NanoString gene expression assay data were generated (see Methods) for 68 and 24
143 subjects, respectively, with 10 subjects being included in both analyses. Clinical
144 characteristics of all subjects are provided in **Supplementary Table 1**.

145 The whole blood transcriptome profiles were used to construct gene co-expression networks
146 and detect clusters of interconnected genes (see below). For gene module discovery, to
147 optimise the generalisability of modules, we included all 68 COVID-19 subjects with RNA-
148 Seq data (regardless of whether they had co-infections), and an additional 18 uninfected
149 healthy control subjects and 9 subjects with non-COVID-19 infections (4 bacterial and 5

150 viral) (**Supplementary Table 1**), all sequenced in the same batch. However, of the COVID-
 151 19 cases who were free from suspected or proven bacterial co-infections, only 16 had URT
 152 viral load measurement and RNA samples collected on the same day and within 15 days of
 153 symptom onset (a time window during which the replicating virus can be isolated (12)).
 154 Only these 16 subjects were included in analyses correlating URT viral load with the whole
 155 blood transcriptome (**Table 1, Figure 1A**). Subjects were mostly female (57.1%), with ages
 156 ranging from 3 to 78 years (median = 55 years) (**Figures 1B and 1C**). The disease severity
 157 was mild (n = 3; 19%), moderate (n = 7; 44%), and severe (n = 6; 37%).

158

159 **Table 1.** Samples used to correlate URT viral load and whole blood transcriptome
 160

Days of illness ¹	Age range (year)	Sex	Severity	Average Ct value ²	Calculated viral load (log ₁₀ (viral load/ml))	Viral load z-score
3	11-15	Female	Mild	27.72	5.49	-0.32
4	51-55	Male	Moderate	37.08	2.52	-2.89
5	66-70	Male	Severe	31.53	4.34	-0.99
8	71-75	Female	Severe	35.87	5.60	0.80
8	41-45	Male	Moderate	27.41	2.91	-1.71
9	51-55	Male	Severe	35.43	3.04	-1.39
9*	51-55	Male	Moderate	32.36	3.83	-0.65
9	36-40	Female	Mild	33.1	3.98	-0.51
10	1-5	Female	Mild	21.03	3.23	-1.00
10	51-55	Female	Severe	34.39	7.45	2.94
11	71-75	Female	Moderate	29.97	4.77	0.64
11	56-60	Male	Severe	33.24	3.73	-0.33
12	41-45	Female	Moderate	36.64	2.90	-0.91
13*	51-55	Male	Severe	28.62	7.93	4
13	71-75	Male	Moderate	19.57	5.54	1.76
15	71-75	Female	Moderate	37.17	2.48	-0.69

161 ¹ How many days after symptom onset the viral load was measured.

162 ² Cycle threshold value

163 * These two samples are from the same subject
 164

165 We also performed a *nCounter* NanoString gene expression analysis on nasal epithelium
 166 samples from 24 COVID-19 patients, including 17 with URT viral load measurement within
 167 15 days from symptom onset (**Table 2**). The subjects' ages ranged from 16 to 80 years
 168 (median = 47 years), and most had mild disease (n = 9; 53%) or severe disease (n = 6;

169 35.3%).

170

171 **Table 2.** Subjects used to correlate URT viral load with nasal epithelium NanoString profiles

172

Days of illness	Age range (year)	Sex	Severity	Average Ct value	Calculated viral load (log ₁₀ (viral load/ml))	Viral load z-score
0	26-30	Female	Mild	27.53	5.94	-0.74
3	16-20	Female	Mild	27.72	5.74	-0.32
4	76-80	Female	Severe	25.36	6.46	0.56
5	66-70	Male	Severe	31.53	4.57	-0.99
6	41-45	Male	Mild	38.16	2.39	-2.8
7	46-50	Male	Mild	29.9	5.01	-0.17
8	71-75	Female	Severe	27.41	5.85	0.8
8	31-35	Female	Mild	28.93	5.33	0.33
8	31-35	Male	Mild	35.06	3.41	-1.45
9	51-55	Male	Severe	35.43	3.26	-1.39
9	36-40	Male	Mild	36.67	2.45	-2.14
10	71-75	Male	Severe	28.86	5.32	0.72
10	46-50	Female	Mild	28.78	5.48	0.87
11	26-30	Female	Mild	36.92	2.54	-1.65
12	26-30	Male	Moderate	27.97	5.61	1.4
12	61-65	Male	Severe	32.49	4.29	0.18
13	66-70	Female	Moderate	25.3	6.36	2.3

173

174

175 **Conversion of viral load measurements to z-scores using viral load regression model**

176 We recently developed a regression model fitted to viral load measurements within the first

177 15□days of symptoms across 16 datasets, capturing the viral load variation during the

178 course of infection between different individuals (3). Here, to determine whether individual

179 subjects in the current study had higher or lower than average viral load measurements

180 relative to their duration of illness, we used the previously published regression model to

181 calculate a z-score for each data point representing the deviation of the data point from the

182 mean viral load trajectory i.e. the regression line (**Tables 1 and 2, Figures 1D and 1E**; see

183 **Methods**). Viral load z-scores calculated from the data were not associated with the severity

184 of illness (**Figure 1F**). In our previous large-scale analysis of COVID-19 subjects (3), we

185 showed that age and sex did not significantly influence URT viral load dynamics and that

186 URT viral load dynamics did not affect the severity of illness. Therefore, in the present
187 study, we did not adjust the viral load z-scores for these variables.

188

189 **Exploring molecular correlates of SARS-CoV-2 viral load using whole blood** 190 **transcriptomics**

191 We aimed to identify groups of genes for which expression correlated with viral load z-
192 score, providing insights into the mechanisms controlling viral load. We first performed a
193 gene signature-based deconvolution (13), as in our previous studies (11, 14). Interestingly,
194 the computed proportion estimate of natural killer (NK) cell population was negatively
195 correlated with viral load z-score ($r = -0.62$ and $P = 0.010$, **Supplementary Figure 1**).
196 There was insufficient evidence to conclude a significant linear relationship between other
197 leukocyte populations (B-cells, monocytes, neutrophils, $CD4^+$ T-cells, and $CD8^+$ T-cells)
198 and viral load z-score. Gene expression counts were then adjusted for leukocyte mixture to
199 remove the confounding effect of differences in blood leukocyte proportions between
200 individuals. To make best use of the relatively small sample size of the selected 16 samples,
201 we performed dimensionality reduction using weighted correlation network analysis
202 (WGCNA) (15). First, we clustered the RNA-Seq profiles ($n = 96$) and removed an outlier
203 (**Supplementary Figure 2**). Then, a gene co-expression network was constructed, and
204 modules (clusters of highly co-expressed genes) were detected using the remaining 95
205 whole blood RNA-Seq profiles. Next, we correlated the first principal component of each
206 module (module eigengene) to viral load z-scores in the group of 16 samples with paired
207 data, reasoning that inducible mechanisms which restrict viral load would be enriched
208 amongst the most strongly correlated modules. Twenty-four modules were significantly
209 correlated with viral load ($P < 0.01$; **Figure 2, Table 3, and Supplementary Table 2**). To
210 aid interpretation, we represented each module by its hub gene (gene with the highest

211 connectivity within the module). Fourteen modules were positively correlated with viral
 212 load z-score and 10 were negatively correlated (**Figure 2A**). *IFNA14* (Interferon Alpha-14)
 213 and *AIPL1* (Aryl Hydrocarbon Receptor Interacting Protein Like 1) modules showed the
 214 strongest negative correlation with viral load ($r = -0.60$ with $P = 1e-10$ and $r = -0.60$ with P
 215 $= 2e-10$, respectively). The largest positive correlation was observed for the *AC011455.2*
 216 module ($r = 0.60$, $P = 2e-10$).

217

218 **Table 3:** Modules significantly correlated with viral load z-score. Each module is represented by its hub gene.
 219 Complete lists of module genes and their information are provided in **Supplementary Table 2**.
 220

Module	Gene count ¹	Genes with highest contribution to the module ²
IFNA14	45	<i>IFNA14, ADAD1, IL22, LIN28A, TMEM270</i>
AIPL1	43	<i>AIPL1, CFAP100, KCNK3, MOGAT2, OR5D3P, SLC13A2, SPEM1</i>
AC011455.2	41	<i>AC011455.2, FRMD1, TTC6, KCNK18, KIF12, SNAI2</i>
C7orf33	36	<i>C7orf33, DGKB, GK2, GHRH, PDZD9, SEPTIN14, SMIM40, TTL2</i>
IFNL3	50	<i>IFNL3, STRA8</i>
GALNT17	37	<i>GALNT17, GCSAML-AS1, IP6K3, NPY, OR6B2, SIM2, SSU72P2, SSX4</i>
CRYAA	57	<i>CRYAA*, AC104581.2*, ACSM4*, FABP6*, GBX2*, PASD1*, PCDHA12*, PDYN*, SSMEM1*</i>
ANGPTL7	43	<i>ANGPTL7, CALML3, GFY, HNRNPCL2, SLC6A1, TPTE, UGT1A3</i>
TUBB1	349	<i>TUBB1, NRG1, SELP, MPIG6B, SPARC, GP9, PTGS1, GP6, CTTN, ABLIM3, ARHGAP6, CMTM5, GP1BB, TSPAN9, ITGB5, GUCY1B1, TREML1, PGRMC1, MYLK, ITGB3, ITGA2B, PTCRA</i>
NAGA	638	<i>NAGA, RASSF4, ZNF385A</i>
NQO1	89	<i>ABCG5*, AC003688.1*, AL121899.2*, CCL26*, DRD1*, SERPINB13*, SP5*</i>
CHRNA4	51	<i>CHRNA4, CPN2, EPYC, GGTL2, HDGFL1, MARCOL, MUC21, OR1D2, OR4F15, RNASE11, SCEL, SLC35F4, XAGE2</i>
TRIM51	32	<i>TRIM51, AL583836.1, EPHA5, KRTAP10-7, LUZP2, TCF23</i>
CHRD2	31	<i>CHRD2, EGFLAM, OR51L1, PPFIA2, SKOR2, SLC01B1, TM4SF4</i>
TMPRSS7	48	<i>AC097636.1, APOF, BPIFB2, FOXR1, MMP10, MMP12, FLG</i>
C1QL4	42	<i>C1QL4, DNMT3L, KRTAP10-8, OR14J1, OVOL3, PMIS2, PPP1R14D, TFAP2D, UBTFL1, H2AC18*</i>
CYP2A7	87	<i>CYP2A7, DCX, NGB, NR0B1, SLC17A6, SPHKAP, SPRR2D, CEACAM18, OOSP4A</i>
AKR1C4	37	<i>AKR1C4, RGS21, CCDC63, DSG1, GLRA1, IL20, TMEM174</i>
AL049839.2	38	<i>AL049839.2, CCDC190, FAM236C, FAM236D, VSNL1, MAGEB1, NRAP, RHCG</i>
ADRA1D	54	<i>ADRA1D, DIO3, EPHA6, H2BC1, IL31, PDE6C, SPATA31D4, TCEAL5, UNCX</i>
ACADL	172	<i>ACADL, AC008770.4, ADCY8, APOA4, ASZ1, BTBD16, CASQ2, CCK, CNTN1, COMP, CYP3A7, CYP4A11, DAZ1, DCAF12L1, DSG4, DUX4, DYDC2, FOXG1, FOXR2, FSHR, GRIA1, GRIA2, GRM6, IFNA13, KIF2B, KLK3, KRTAP2-2, MISP, NPY2R, NRK, NTF3, NTSR2, NXPH1, NXPH2, OR13J1, OR14A2, OR14C36, OR1C1, OR51G2, OR5L1, OR5W2, OR7A10, OR7E24, OTX2, PAX1, PAX7, PHOX2B, PRAMEF7, PRDM9, PRM3, PSG7,</i>

		<i>RPTN, SCYGR2, SERPINA5, SLC19A3, SLC2A7, ST8SIA3, SULT2A1, TBPL2, TRIML1, TSPY10, TSPY2, TSPY3, TSPY8, UGT1A4, ZIC1, ZSCAN5C</i>
ZMIZ2	711	N/A ³
NR3C2	208	N/A ³
AL132671.2	97	<i>ALI32671.2, C2orf72, CCDC166, FGF23, MBL2, MUC3A, MYOG, SPINK6, STRA6</i>

221 ¹The number of genes involved in each module.

222 ² Genes with the absolute value of module membership (the correlation between the module eigengene and
223 gene expression values) > 0.9. Genes with negative module membership are marked with an asterisk.

224 ³ No genes had module membership higher than 0.9 or lower than -0.9

225

226 We selected the top 6 significant modules for further analysis: *IFNA14*, *AIPL1*,
227 *AC011455.2*, *C7orf33* (Chromosome 7 Open Reading Frame 33), *IFNL3* (Interferon
228 Lambda 3), and *GALNT17* (Polypeptide N-Acetylgalactosaminyltransferase 17).
229 *AC011455.2*, *C7orf33*, and *GALNT17* modules were positively correlated with viral load z-
230 score and positioned very close to each other in the hierarchical clustering (**Figure 2B**).
231 Therefore, we merged their gene sets to form a metamodule
232 (*AC011455.2/C7orf33/GALNT17*; total gene count = 114) for further data analysis,
233 assuming that the higher gene count would increase power to detect biologically relevant
234 changes. We used Qiagen's Ingenuity Pathway Analysis (IPA) for biological understanding
235 of the modules (16). Correlation coefficients between module genes and viral load z-score
236 were used to infer the activity pattern (activation or inhibition) of the biological processes
237 involved such as enriched pathways and upstream regulators.

238 **Figure 3A** illustrates top enriched canonical pathways for the *IFNA14*, *AIPL1*,
239 *AC011455.2/C7orf33/GALNT17*, and *IFNL3* modules (details provided in **Supplementary**
240 **Table 3**). Of these, 3 pathways showed an enrichment *P*-value of below 0.001 including
241 'pathogen induced cytokine storm signalling' (*P* = 5e-4) and 'IL-22 (Interleukin-22)
242 signalling' (*P* = 8e-4) enriched in the *IFNA14* module, and 'melatonin degradation' (*P* = 2e-
243 4) enriched in the *IFNL3* module. The pathogen induced cytokine storm signalling pathway
244 encompasses the highest number of genes from the tested module (*IFNA14*, *IL22*, *CCL4* (C-

245 C Motif Chemokine Ligand 4), *CD70*, and *COL4A4* (Collagen Type IV Alpha 4 Chain)
246 from the *IFNA14* module) compared to the other enriched pathways identified. *IFNA14* and
247 *IL22*, two main members of the *IFNA14* module (module membership = 0.98 and $P = 5e-64$
248 for both genes), are key components of the pathway. Interestingly in our dataset both genes
249 were negatively correlated with viral load z-score ($r = -0.59$ and $P = 3e-10$ for both genes),
250 whereas *CCL4*, *CD70*, and *COL4A4* were positively correlated with viral load z-score. The
251 IL-22 signalling pathway involves two genes, *IL22* and *IL22RA2* (Interleukin 22 Receptor
252 Subunit Alpha 2), with a high contribution to the *IFNA14* module (module membership =
253 0.98 and 0.75 with $P = 5e-64$ and $1e-18$, respectively) both negatively correlated with viral
254 load z-score ($r = -0.59$ and -0.63 with $P = 3e-10$ and $7e-12$, respectively). The superpathway
255 of melatonin degradation includes three genes (*CYP2F1* (Cytochrome P450 Family 2
256 Subfamily F Member 1), *IL4I1* (Interleukin 4-induced gene-1), and *UGT1A1* (UDP
257 Glucuronosyltransferase Family 1 Member A1) from the *IFNL3* module. However, this
258 pathway is likely to be of less importance here as the three genes each show relatively weak
259 individual correlations with viral load z-score (**Supplementary Table 2**).

260 *STING1* (Stimulator of Interferon Response CGAMP Interactor 1), *RBP3* (Retinol Binding
261 Protein 3), and *RORC* (RAR Related Orphan Receptor C) were predicted to be the most
262 significant upstream regulators of *IFNA14* module genes ($P = 6e-5$, $3e-4$, and $4e-4$,
263 respectively) (**Figures 3B and 3C**). They interact with *IL22*, *IL22RA2*, *CCL4*, *CD70*, and
264 *FEZ1* (Fasciculation and Elongation Protein Zeta 1), members of the *IFNA14* module which
265 are mapped to the pathogen induced cytokine storm signalling and IL-22 signalling
266 pathways (**Supplementary Tables 4 and 5**). None of the identified pathways and regulators
267 showed reliable evidence of activation (absolute value of IPA activation z-score > 2) and
268 therefore we were not able to infer an overall directionality (activation or inhibition) with
269 respect to viral load.

270 We also applied the *BloodGen3Module* tool (17) to identify gene modules associated with
271 viral load. Unlike WGCNA which detects modules from the analysed gene expression
272 dataset, *BloodGen3Module* uses fixed functionally pre-annotated modules characterising
273 different biological responses of distinct blood cell types. We used RNA-Seq data without
274 adjustment for leukocyte-mixture and evaluated differential expression of these modules
275 between samples with positive and negative viral load z-scores ($n = 5$ and $n = 11$,
276 respectively). We identified an aggregate of five modules showing high ‘module response’
277 and higher module expression in subjects with positive viral load z-score (aggregate module
278 A34; **Figure 4A and Supplementary Tables 6**). A module response is defined as the
279 percentage of genes for a given module showing significant differential expression between
280 the groups. From the module aggregate, the Prostanoids module showed the highest module
281 response (97%). Interestingly, we observed a significant overlap between the A34 aggregate
282 module and *TUBBI* (Tubulin Beta 1 Class VI) module which was found to be significantly
283 positively correlated with viral load z-score by WGCNA ($r = 0.39$ with $P = 9e-05$; **Figure**
284 **2**). Seventy genes including *TUBBI* were common between the A34 and *TUBBI* modules
285 (**Figure 4B and Supplementary Tables 6**) while A34 did not overlap with any other
286 WGCNA module correlated with viral load z-score. Among the A34 modules, the
287 Prostanoids module showed the highest overlap with the *TUBBI* module (from 36 genes
288 involved in the Prostanoids module, 30 were also included in the *TUBBI* module).

289

290 **Exploring molecular correlates of SARS-CoV-2 viral load using NanoString assay of** 291 **nasal epithelium**

292 We analysed RNA isolated from nasal epithelium samples of 24 COVID-19 patients using a
293 NanoString panel of 579 genes involved in core pathways and processes of human immune
294 responses (**Supplementary Table 7**). Seventeen subjects also had paired URT viral load

295 measurement within 15 days from symptom onset and had no evidence of bacterial co-
296 infection. Using WGCNA we identified seven gene co-expression networks which we refer
297 to as “pseudo-modules” since they were detected using a relatively low number of genes
298 included in the NanoString panel (**Supplementary Table 8**). Only one pseudo-module was
299 correlated with viral load z-score at significance threshold of 0.05 (*PTK2* (Protein Tyrosine
300 Kinase 2) module; $P = 0.016$). Additionally, correlation analysis between individual gene
301 expression and viral load z-score detected significant correlation (absolute correlation
302 coefficient > 0.5 and $P < 0.05$; **Figure 5**) in 12 individual genes, 11 of which were
303 positively correlated and one (*GPLY* (Granulysin)) was negatively correlated.

304

305 **Discussion**

306 Understanding mechanisms controlling SARS-CoV-2 viral load in the URT can provide
307 valuable leads towards treatment and vaccine strategies aimed at reducing viral
308 transmission. Such strategies have recently been highlighted as potential “game changers”
309 as societies adapt to living with COVID-19 (18). Current evidence suggests that the
310 mechanisms controlling URT viral load may be different from those controlling LRT viral
311 load and disease severity (3-5). However, our current knowledge concerning control of URT
312 viral load is far from complete. To unravel the biological complexity underlying the control
313 of SARS-CoV-2 viral load, we sought to identify correlates of the variation in viral load
314 which occurs in naturally infected individuals. Of note, we used samples from the first wave
315 of infection in Europe, prior to vaccination, infection-induced immunity, and circulation of
316 important variants of SARS-CoV-2. To account for the dynamic nature of URT viral load,
317 which rapidly increases to a peak just before symptom onset and then declines more slowly,
318 we quantified viral loads by their standardised deviation (z-scores) from a previously-
319 derived average trajectory (3). We correlated viral load z-scores with paired peripheral

320 blood and nasal epithelium transcriptomes.

321 After excluding individuals with proven or suspected bacterial co-infection and URT SARS-
322 CoV-2 viral load samples taken more than 15 days after symptom onset, relatively small
323 numbers of subjects for blood and nasal transcriptome analysis (n = 16 and n = 17,
324 respectively) remained. This prompted the use of gene modules rather than individual genes
325 for our primary analysis. This would reduce the complexity of large gene networks into
326 relevant modules and increase the statistical power to detect those correlating with viral
327 load. An individual module comprises genes that are more densely connected than
328 expected by chance and often involved in the same biological functions (19). We applied
329 two different methods to detect gene clusters, *WGCNA* and *BloodGen3Module*. The first
330 identifies modules directly from the gene expression data and the later uses pre-annotated
331 modules.

332 The peripheral blood module most significantly associated with URT viral load, had
333 *IFNA14* as its hub gene. *IFNA14* encodes the type I interferon, interferon α 14. Interferons
334 are glycoprotein cytokines made and released by host lymphocytes and considered to be key
335 effectors in antiviral responses. However, their pattern of expression and function during
336 SARS-CoV-2 infection is controversial. While some studies suggest protective effects of
337 interferons in severe COVID-19 (20-22), others indicate poor clinical outcomes in those
338 with increased production of interferons (23-26). There is limited data available on the
339 correlation between SARS-CoV-2 viral load and interferon expression. Sposito et al.
340 evaluated nasopharyngeal swabs of COVID-19 patients and showed that the expression of
341 type I and III interferons was significantly associated with viral load in patients under 70
342 years old (26). However, those aged over 70 years showed no association and/or showed a
343 significantly lower correlation coefficient. This evaluation did not include *IFNA14*. Also, it
344 appears that their viral load measurements were not adjusted for the time between sample

345 collection and symptom onset. *IFNA14* has been shown to activate a potent antiviral
346 response *via* binding to IFNAR1 and IFNAR2 (Interferon Alpha and Beta Receptor
347 Subunits 1 and 2) receptors (27, 28). This triggers the activation of JAK/STAT (Janus
348 Kinase/Signal Transducer and Activator of Transcription) signalling complexes which
349 subsequently induces the expression of ISGs (interferon-stimulated genes) that inhibit virus
350 infection (29). The strong negative correlation between the *IFNA14* module as well as
351 *IFNA14* as an individual gene and viral load in our data suggests that IFNA14 signalling
352 could play a key role in controlling SARS-CoV-2 viral load, i.e. increased expression of
353 *IFNA14* restricts viral replication. Schuhenn et al. recently showed that IFNA14 is one of
354 the most potent interferon alpha subtypes inhibiting SARS-CoV-2 replication and can cause
355 a significant reduction of SARS-CoV-2 viral titre by up to 10⁵-fold (30). Furthermore,
356 unpublished data suggest that, compared to IFNA2 (Interferon Alpha-2), which was used to
357 treat COVID-19 patients in an uncontrolled exploratory study in China (31), IFNA14 is
358 more efficient at preventing the infection while less detrimental to the immune system (32).
359 Not only is *IFNA14* important as an individual gene, but it also represents a network of
360 highly connected genes in our data, the *IFNA14* module, which showed a high enrichment
361 of two canonical pathways ‘pathogen induced cytokine storm signalling’ and ‘IL-22
362 signalling’. Although SARS-CoV-2 can trigger a ‘cytokine storm’ (33, 34), the changes in
363 expression of genes in this pathway were not consistently associated with activation or
364 inhibition of the pathway, agreeing with previous findings that ability to control of URT
365 viral load is dissociated from severity of illness (3, 4). IL-22 is a cytokine released by
366 several immune cells such as Th22 (T helper cells type 22) and plays an important role at
367 mucosal barriers, orchestrating the interaction between the epithelial cell layer and local
368 immune system in response to infections (35). IL-22 stimulates the IL-22 receptor complex
369 on epithelial cells resulting in downstream activation of JAK-STAT signalling pathway

370 which induces multiple antiviral responses and therefore can be protective during SARS-
371 CoV-2 infection (35-37). Elevated levels of IL-22 in the plasma have been implicated as a
372 hallmark of severe COVID-19 (24). Taken together there is compelling evidence that the
373 genes in the *IFNA14* module act to reduce URT viral load, and add to the evidence that
374 interferon α 14 should be considered as a candidate treatment to reduce viral load in the URT
375 and decrease transmissibility of SARS-CoV-2.

376 The *AIPL1* module was the second top module negatively correlated with viral load z-score.
377 Unlike *IFNA14*, *AIPL1* is not known to be involved with the pathogenesis of COVID-19,
378 and the enriched pathways for this module contained relatively few genes. Nevertheless,
379 enrichment of the ‘ α -tocopherol degradation’ pathway suggests a potential role of α -
380 tocopherol (also known as Vitamin E) in the control of viral load. α -tocopherol is an
381 antioxidant which may enhance the function of innate and adaptive immune cells, for
382 example increasing NK cell activity and the phagocytic capacity of leukocytes, which could
383 bolster the immune response to reduce pathogen load as observed in influenza (38, 39).
384 Emerging evidence suggests that water soluble derivatives of α -tocopherol have a potent
385 antiviral response especially when they are used synergistically with remdesivir to inhibit
386 SARS-CoV-2 RNA-dependent RNA polymerase (40).

387 We also identified modules positively correlated with viral load, possibly indicating that
388 these modules are induced in response to increasing amounts of virus or that expression of
389 these genes favour an increase in viral replication. The most significant of these modules
390 was the *IFNL3* module. The hub gene, *IFNL3*, encodes a type III interferon which is a
391 cytokine activated in response to mucosal viral infections and signals through the
392 heterodimeric IFNLR (Interferon Lambda Receptor) that is expressed distinctly in the URT
393 epithelial cells. This stimulates the activation of several transcription factors which
394 upregulate ISGs. Type III interferon signalling pathway is considered slower and induces a

395 weaker ISG response than type 1 interferons (41, 42). The most significant upstream
396 regulator of the *IFNL3* module is TLR9 (Toll Like Receptor 9), which may be stimulated by
397 unmethylated CpG (Cytosine-phosphate-Guanine) sequences during SARS-CoV-2 infection
398 (43) and result in the observed upregulation of *IFNL3* module genes.

399 Using *BloodGen3Module*, we identified a cluster of blood pre-annotated transcriptional
400 modules (A34) positively correlated with viral load. This was particularly interesting as
401 these modules showed a significant overlap with the *TUBBI* module found to be
402 significantly positively correlated with viral load by WGCNA. From the A34 modules, the
403 Prostanoids module showed the highest module response and overlap with the *TUBBI*
404 module. Prostanoids are a subclass of eicosanoids and regulate the inflammatory response
405 (44). The observed association between the prostanoids module expression and viral load z-
406 score suggests that high levels of prostanoids may suppress processes which constrain viral
407 load and therefore promote high viral load levels. This is supported by a recent study
408 showing that abrogation of eicosanoid signalling reduces viral load and rescues mice from
409 fatal SARS-CoV-2 infection (45).

410 In addition to peripheral blood samples, we studied samples taken from the primary
411 infection site, the nasal epithelium. Both blood and nasal transcriptomes can reflect the host
412 immune response to the infection. In a respiratory infection, epithelial cells are directly
413 infected, and peripheral blood leukocytes also respond to signals arising from the site of
414 infection (46). However, the difference in the transcriptomic analysis approach we used for
415 each dataset (RNAseq for peripheral blood and NanoString assay for nasal samples) made it
416 difficult to compare the results directly. The NanoString assay analysed a relatively small
417 number of genes (579 genes involved in immune response) and therefore the data did not
418 yield reliable module level results. However, individual genes correlated with URT viral
419 load z-score were identified that may be of interest. *GPLY* was the only negatively

420 correlated gene representing a likely role in the control of viral load. It is produced by a
421 variety of killer cells such as cytotoxic T lymphocytes and NK cells, and it has both
422 cytolytic and proinflammatory activity (47). Indeed, the expression of *GNLY* in lymphocytes
423 has been reported to be associated with recovery in COVID-19 suggesting it may play a
424 major role in clearance of infected cells and termination of infection (48). In agreement with
425 the correlation between *GNLY* and viral load, we also showed that cell proportion estimates
426 of NK cells in peripheral blood were negatively correlated with viral load z-score,
427 highlighting the importance of this cell population in constraining the virus. For example,
428 Witkowski et al. showed COVID-19 patients with normal NK cell numbers demonstrated a
429 more rapid decline of viral load compared to those with low NK cell numbers (49).
430 Our study was limited by the relatively small sample size which may have reduced the
431 statistical power and resulted in missed opportunities to capture some biological signals.
432 Additionally, we cannot establish from this data whether the molecular mechanisms
433 identified are cause or consequence of the viral load, although there are plausible
434 mechanisms which suggest causal roles in some cases.

435

436 **Conclusions**

437 To our knowledge, this is the most comprehensive study focusing on identifying molecular
438 correlates of the SARS-CoV-2 viral load control in the URT. We identified numerous
439 molecular processes which may contribute to the control of URT viral load. These candidate
440 mechanisms can be the focus of further functional studies and may lead to new strategies to
441 prevent COVID-19 and reduce SARS-CoV-2 transmission.

442

443 **Methods**

444 **Study design and participants**

445 We studied 82 COVID-19 patients recruited through the GEN-COVID study
446 (www.gencovid.eu), a multi-center and prospective cohort designed to evaluate the effect of
447 genetic factors on SARS-CoV-2 infection. Subjects were recruited at Hospital Clínico
448 Universitario de Santiago de Compostela (Galicia, Spain) between March 2020 and May
449 2020, during the first wave of infections in Spain, before significant levels of infection- and
450 vaccine-induced immunity in the community. COVID-19 was defined according to the
451 Spanish national guidelines
452 (<https://www.mscbs.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/documentos.htm>). The severity of the disease was defined as mild, moderate, and severe based on
453 WHO scoring for COVID-19 patients and as described previously (50, 51). We also
454 included 18 uninfected controls, and 9 subjects with non-COVID-19 infections recruited
455 through the PERFORM Consortium.

457

458 **Sample collection**

459 Blood samples and nasal epithelium specimens were collected at the same time at hospital
460 for moderate and severe COVID-19 subjects and at home for subjects with mild disease.
461 Whole blood was collected into PAXgene blood RNA tubes (PreAnalytiX) and nasal
462 epithelium samples were collected in Oragene CP-190 kit (DNA Genotek). Samples were
463 processed as described previously (51, 52).

464 One COVID-19 subject contributed two paired sets of samples (viral load and blood RNA-
465 Seq; **Table 1**) collected 3 days apart. We included both as they showed a noticeable
466 difference in viral load z-score and hence were informative.

467

468 **RNA isolation**

469 Total RNA was isolated from blood and nasal epithelium samples using PAXgene blood

470 miRNA extraction and RNeasy microkit, respectively, according to the manufacturer's
471 protocols (Qiagen). RNA amount and integrity were assessed using TapeStation 4200
472 (Agilent). RNA quality was checked based on DV200 metric to ensure that sufficient
473 percentage (over 50%) of RNA fragments were greater than 200 nucleotides in length and
474 also to estimate the optimal sample input for the *nCounter* NanoString analysis.

475

476 **Viral load measurements**

477 **Viral load quantification:** Nasopharyngeal samples were collected in Universal Transport
478 Medium (UTM) tubes and assessed for the presence and viral load of SARS-CoV-2. We
479 detected viral particles using a multiplex real-time PCR with the Allplex™SARS-CoV-2
480 Assay (Seegene). Viral load values (viral copies per ml) were computed from the Ct values
481 as described previously (3).

482 **Calculation of viral load z-scores:** A regression model of the average trajectory of viral
483 load over time and quantification of variation between individuals, using data from 16
484 datasets, was reported previously (3). Viral load values from the present study were
485 compared to the regression line to assess whether a particular viral load measurement,
486 sampled a certain number of days after symptom onset, was higher or lower than average.
487 A 'z-score' was calculated for each data point by calculating its deviation from the mean
488 trajectory and dividing by the standard deviation of the variation in viral load around the
489 mean trajectory (**Fig. 1D**).

490

491 **RNA sequencing**

492 Paired-end sequencing was performed at The Wellcome Centre for Human Genetics in
493 Oxford, UK as described previously (51). Sequencing was carried out using Novaseq6000
494 platform providing 150 bp paired end reads.

495

496 **RNA-Seq upstream analyses:** Adapter trimming and quality control of sequencing reads
497 were performed with Trimmomatic version 0.36 and FastQC version 0.11.7, respectively
498 (53, 54). The reads were then mapped against hg38 reference genome using STAR version
499 2.7.1a (55). RSEM version 1.3.1 was used for transcript quantification (56). Next, we
500 performed a gene signature-based deconvolution using CellCODE as in our previous work
501 and adjusted gene expression for leukocyte (B cells, monocytes, neutrophils, NK cells,
502 CD4⁺ T cells, and CD8⁺ T cells) mixture (11, 13, 14).

503

504 **NanoString experiment**

505 **NanoString *n*Counter assay**

506 We analysed immunological gene expression profiles of nasal epithelium using
507 the *SPRINT nCounter* system (NanoString Technologies) with the Human Immunology V2
508 Panel (579 genes covering the core pathways and processes of the immune response, and 15
509 internal reference genes for data normalization). The detail of the assay is described
510 previously (52).

511

512 **Differential gene expression analysis**

513 The gene expression counts adjusted for leukocyte mixture were correlated with viral load z-
514 scores using edgeR (57).

515

516 **Weighted correlation network analysis**

517 Gene counts were normalised using variance stabilizing transformation (*VST*) function of
518 DESeq2 R package (58) and adjusted for leukocyte mixture using *removeBatchEffect*
519 function of limma R package (59). We used WGCNA version 1.71 R package for weighted

520 correlation network analysis (15).

521

522 **Module repertoire analysis**

523 We applied *BloodGen3Module* version 1.4.0 R package (17) to the normalised gene
524 expression counts unadjusted for cell-mixture from 16 samples with paired viral load,
525 collected in the absence of bacterial co-infection. The package encompasses 382
526 functionally annotated blood transcriptional modules which have been grouped into 38
527 “aggregates” (A1-A38). The differential expression of the modules was compared between
528 two groups with positive and negative viral load z-scores ($n = 5$ and $n = 11$, respectively)
529 using t-test with fold change and p-value cut-off of 0.5 and 0.05, respectively. For each
530 module, we computed ‘module response’ as the percentage of genes for the module showing
531 significant differential expression between the two groups.

532

533 **Further statistical analysis**

534 The normality of distributions was assessed using the Shapiro-Wilk normality test. Pearson
535 correlation was used to analyse the degree of association between two continuous
536 variables. An independent-samples t-test and one-way ANOVA with Tukey’s post hoc test
537 were used to compare continuous variables between two and multiple groups, respectively.

538

539 **Figure Legends**

540 **Figure 1.** Overview of peripheral blood gene expression and viral load in subjects with
541 COVID-19. A) PCA (principal component analysis) plot of peripheral blood gene
542 expression determined by RNA-Seq. Samples with paired URT viral load measurement are
543 coloured as blue. B) and C) PCA plots represent samples with paired RNA-Seq and viral
544 load data coloured by age and sex, respectively. D) Calculation of viral load z-scores. In the

545 upper panel, the viral load data of the present study (black circles) are plotted against the
546 time since symptom onset. The green line indicates a linear regression model fitted to the
547 viral load data from 16 different datasets previously studied. The shaded green area
548 represents the 95% confidence interval for the regression model. As shown in the lower
549 panel, for each data point, a z-score is calculated as the distance of the data point from the
550 mean trajectory (green line). E) PCA plot of samples with paired data coloured based on
551 viral load z-score. F) Viral load z-score is compared between groups of different COVID-19
552 severity. Red dots and whiskers represent mean and 1 standard deviation.

553 **Figure 2.** Peripheral blood gene expression modules correlated with viral load z-score. A)
554 For each module, the Pearson correlation with viral load z-score and corresponding p-value
555 are displayed. The Pearson correlation scale is depicted on the right. B and C) Module
556 network and relationship with viral load z-score. The hierarchical clustering dendrogram of
557 the module eigengenes (B) was generated using all genes in the modules and shows the
558 dissimilarity of eigengenes with the distance measure being one minus correlation. Modules
559 coloured in red and blue are, respectively, positively and negatively correlated with viral
560 load. The heatmap (C) represents module eigengene adjacency calculated as $(1 +$
561 $\text{correlation})/2$.

562 **Figure 3.** Ingenuity pathway analyses of peripheral blood gene expression modules most
563 strongly correlated with viral load. A) For each module, the top 5 significant pathways are
564 illustrated in descending order of statistical significance as indicated by colour. For each
565 pathway, the size of the corresponding circle represents the number of module genes that
566 map to the pathway. The x-axis shows the ratio of the number of genes common between
567 the corresponding module and pathway divided by the total number of genes that map to the
568 same pathway. B) and C) For each module, the 5 most significant upstream (B) and master
569 regulators (C) are illustrated in descending order of statistical significance as indicated by

570 colour. For each regulator, the size of the corresponding circle represents the number of
571 module genes downstream to the regulator. The x-axis shows the ratio of the number of
572 module genes downstream to the corresponding regulator divided by the total number of
573 module genes.

574 **Figure 4.** Pre-annotated blood gene expression modules associated with viral load. A)
575 Module fingerprint grid plot. The differential expression of the modules is compared
576 between two groups with positive and negative viral load z-score using t-test with fold
577 change and p-value cut-off of 0.5 and 0.05, respectively. Each block corresponds to a
578 module position. Each row represents a ‘module aggregate’ including modules with the
579 same pattern of differential expression across reference datasets. Red and blue spots
580 represent modules with increased and decreased abundance in the positive vs negative viral
581 load z-score group, respectively. The gradient represents ‘module response’ which is the
582 percentage of genes for a given module showing significant change in abundance between
583 the two groups. Only modules with at least 15% response have been shown. B) Overlap of
584 genes between A34 and *TUBB1* module.

585 **Figure 5.** Correlation between nasal epithelium transcriptome and viral load z-score. The
586 volcano plot illustrates correlation coefficients and corresponding p-values. Each dot
587 represents a gene included in the NanoString panel. Genes strongly correlated with viral
588 load z-score (absolute correlation coefficient > 0.5 and $P < 0.05$) are coloured as red
589 (positive correlation) and blue (negative correlation).

590

591 **List of abbreviations**

592 AIPL1: Aryl Hydrocarbon Receptor Interacting Protein Like 1

593 C7orf33: Chromosome 7 Open Reading Frame 33

594 CCL4: C-C Motif Chemokine Ligand 4

595 COL4A4: Collagen Type IV Alpha 4 Chain
596 COVID-19: coronavirus disease
597 CpG: Cytosine-phosphate-Guanine
598 Ct value: cycle threshold value
599 CYP2F1: Cytochrome P450 Family 2 Subfamily F Member 1
600 FEZ1: Fasciculation and Elongation Protein Zeta 1
601 GALNT17: Polypeptide N-Acetylgalactosaminyltransferase 17
602 GNLY: Granulysin
603 IFNA14: Interferon Alpha 14
604 IFNA2: Interferon Alpha-2
605 IFNAR1 and IFNAR2: Interferon Alpha and Beta Receptor Subunits 1 and 2
606 IFNL3: Interferon Lambda 3
607 IFNLR: Interferon Lambda Receptor
608 IL-22: Interleukin-22
609 IL22RA2: Interleukin 22 Receptor Subunit Alpha 2
610 IL4I1: Interleukin 4-induced gene-1
611 IPA: Ingenuity Pathway Analysis
612 ISGs: interferon-stimulated genes
613 JAK/STAT: Janus Kinase/Signal Transducer and Activator of Transcription
614 LRT: lower respiratory tract
615 NK cells: natural killer cells
616 PCA: principal component analysis
617 PTK2: Protein Tyrosine Kinase 2
618 RBP3: Retinol Binding Protein 3
619 RNA-Seq: RNA sequencing

620 RORC: RAR Related Orphan Receptor C
621 SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus 2
622 STING1: Stimulator of Interferon Response CGAMP Interactor 1
623 Th22: T helper cells type 22
624 TLR9: Toll Like Receptor 9
625 TUBB1: Tubulin Beta 1 Class VI
626 UGT1A1: UDP Glucuronosyltransferase Family 1 Member A1
627 URT: upper respiratory tract
628 VL: viral load
629 WGCNA: weighted correlation network analysis

630

631 **Declarations**

632 **Ethics approval and consent to participate**

633 The GEN-COVID and PERFORM (Personalised Risk assessment in Febrile illness to
634 Optimise Real-life Management across the European Union; perform2020.org/) studies were
635 conducted according to the guidelines of the Declaration of Helsinki and approved by the
636 Ethics Committee of Galicia (CEIC, ref 2020/178, 18/03/2020) and St Mary's Research
637 Ethics Committee (16/LO/1684, 25/02/2013), respectively. A written informed consent was
638 obtained for each participant. If this was not done at the time of sampling, a retrospective
639 consent was sought at the earliest appropriate opportunity.

640 Here, we have replaced sample and subject IDs with identifiers that cannot reveal the
641 identity of the study subjects.

642

643 **Consent for publication**

644 Not applicable

645

646 **Availability of data and materials**

647 Raw RNA-Seq data and corresponding metadata are available on ArrayExpress under the
648 accession E-MTAB-12791. NanoString nCounter data and corresponding metadata are
649 available at https://github.com/MahdiMoradiMarjaneh/COVID19_viral_load. Codes used in
650 the analyses can be accessed on the same GitHub repository.

651

652 **Competing interests**

653 The authors declare that they have no competing interests.

654

655 **Funding**

656 This work was supported by UKRI (MRC) and the DHSC (NIHR) (Grant Ref:
657 MR/V027409/1). This study also received support from Instituto de Salud Carlos III
658 ([ISCIII] TRINEO: PI22/00162; DIAVIR: DTS19/00049; Resvi-Omics: PI19/01039 [AS];
659 ReSVinext: PI16/01569 [F.M.-T.]; Enterogen: PI19/01090 [F.M.-T.]; OMI-COVI-VAC
660 (PI22/00406 [F.M.-T.] cofinanciados FEDER), GAIN: Grupos con Potencial de Crecimiento
661 (IN607B 2020/08, [A.S.]); ACIS: BI-BACVIR (PRIS-3, [A.S.]), and CovidPhy (SA 304 C,
662 [A.S.]); and consorcio Centro de Investigación Biomédica en Red de Enfermedades
663 Respiratorias (CB21/06/00103; F.M.-T.); GEN-COVID (IN845D 2020/23, F.M.-T.) and
664 Grupos de Referencia Competitiva (IIN607A2021/05, F.M.-T.). The funders were not
665 involved in the study design, collection, analysis, interpretation of data, the writing of this
666 article or the decision to submit it for publication.

667 M.M.M. is supported in part by the NIHR Biomedical Research Centre of Imperial College
668 NHS Trust.

669 JD.C. and L.C.O. acknowledge funding from the MRC Centre for Global Infectious Disease

670 Analysis (reference MR/R015600/1), jointly funded by the UK Medical Research Council
671 (MRC) and the UK Foreign, Commonwealth & Development Office (FCDO), under the
672 MRC/FCDO Concordat agreement and is also part of the EDCTP2 programme supported by
673 the European Union.

674 F.L. is supported by an MRC clinical training fellowship [award MR/W000970/1]. F.L. and
675 R.T. are supported by the UK Coronavirus Immunology Consortium (UKCIC).

676 H.R.J. received support from the Wellcome Trust (4-year PhD programme, grant number
677 215214/Z/19/Z).

678 M.K. acknowledges support from the Wellcome Trust and the Medical Research Foundation
679 Grants (206508/Z/17/Z and MRF-160-0008-ELP-KAFO-C0801).

680 L.C.O. declares grant funding from Merck Group on an unrelated project.

681

682 **Authors' contributions**

683 Conceptualization: A.J.C, J.D.C., L.C.O., R.S.T.; Data curation: M.M.M., I.R.-C., A.G.-C.,
684 D.H.-C., H.R.J., C.Y.F., Y.W., F.L.; Formal analysis: M.M.M., J.D.C, A.S., A.G.-C., M.K.;
685 funding acquisition: A.J.C, J.D.C, L.C.O., R.S.T., M.L., M.K., F.M.-T., I.R.-C., A.G.-C.,
686 A.S.; Investigation: I.R.-C., A.G.-C., V.J.W., S.N., G.D'S., D.H.-C.; Methodology:
687 M.M.M., J.D.C., A.G.-C., A.S., H.R.J., V.J.W., D.H.-C.; Project administration: A.J.C, A.S.,
688 A.G.-C., F.M.-T.; Resources: A.J.C., F.M.-T., I.R.-C., A.G.-C., M.K.; Software: M.M.M,
689 H.R.J., D.H.-C.; Supervision: A.J.C, L.C.O., R.S.T., M.K., V.J.W., M.L., F.M.-T.;
690 Validation: M.M.M.; Visualisation: M.M.M.; Writing—original draft: M.M.M., A.J.C.;
691 Writing—review and editing: ALL.

692

693 **Acknowledgements**

694 The members and affiliations of the PERFORM consortium and the GEN-COVID

695 (www.gencovid.eu) study group are listed in the Supplementary text file.

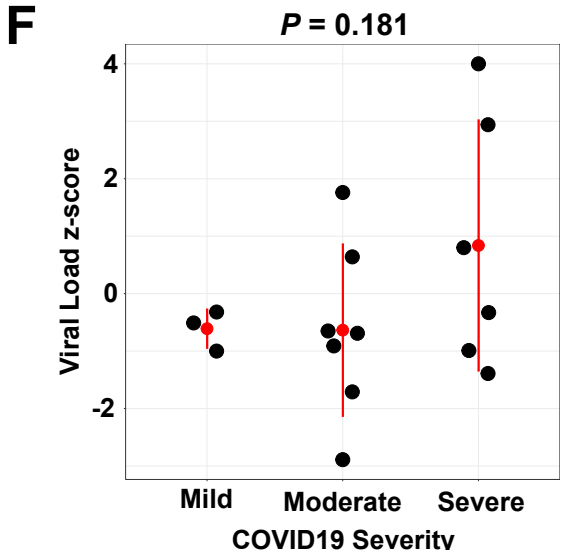
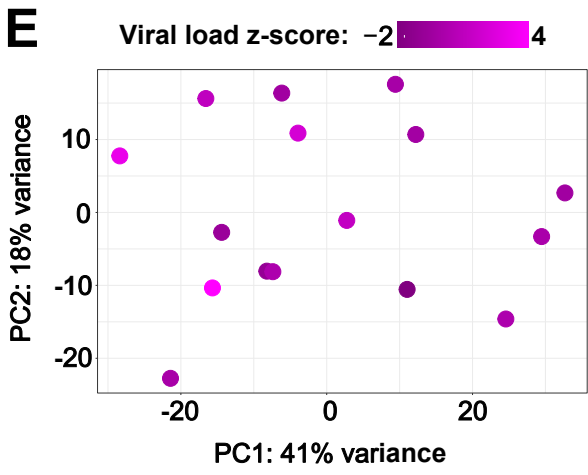
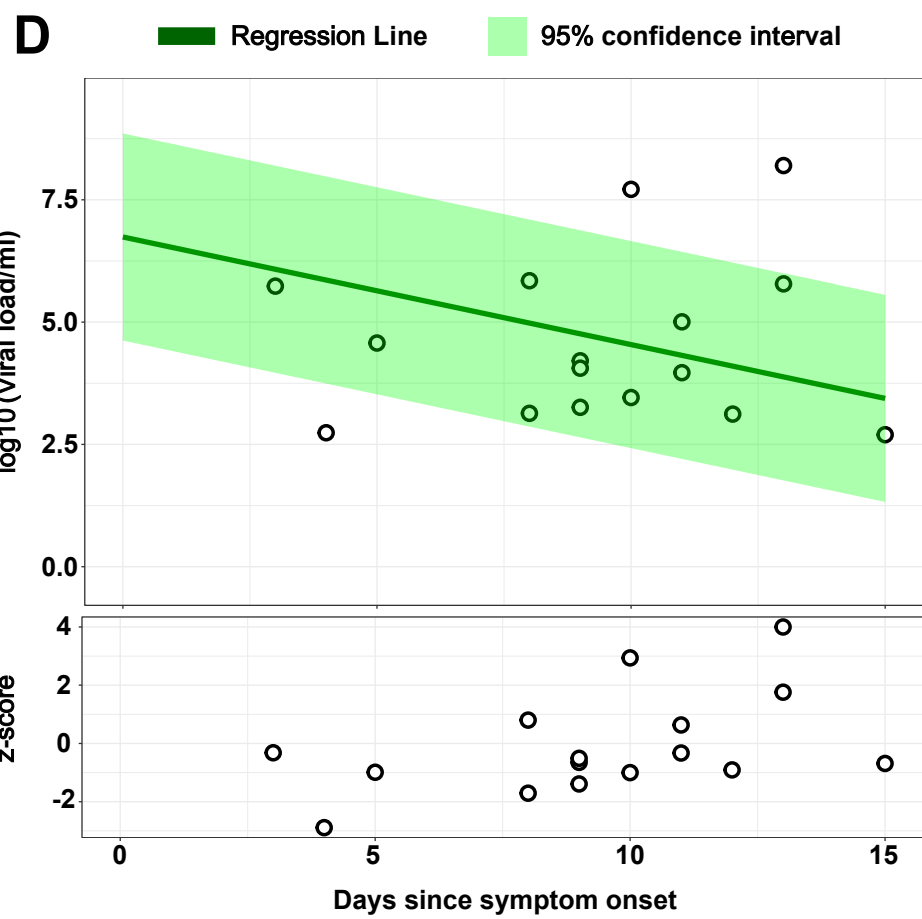
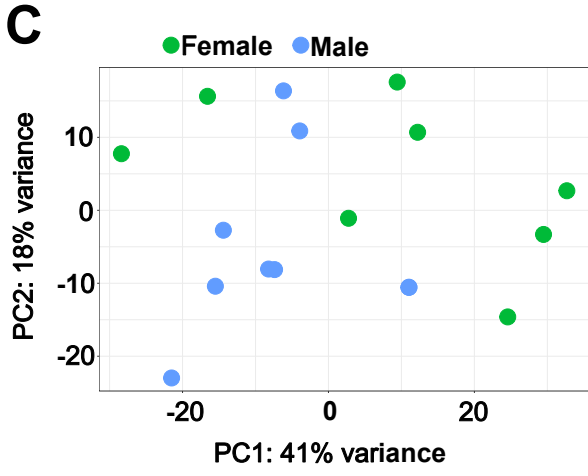
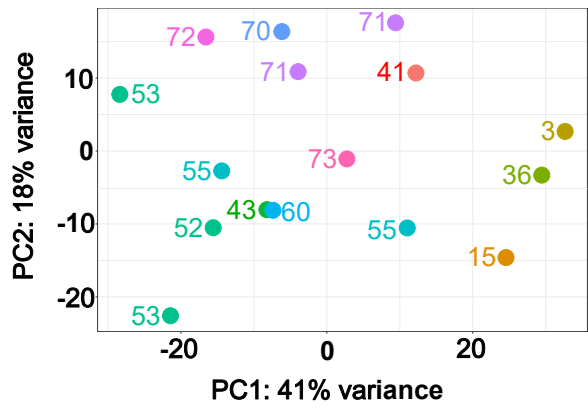
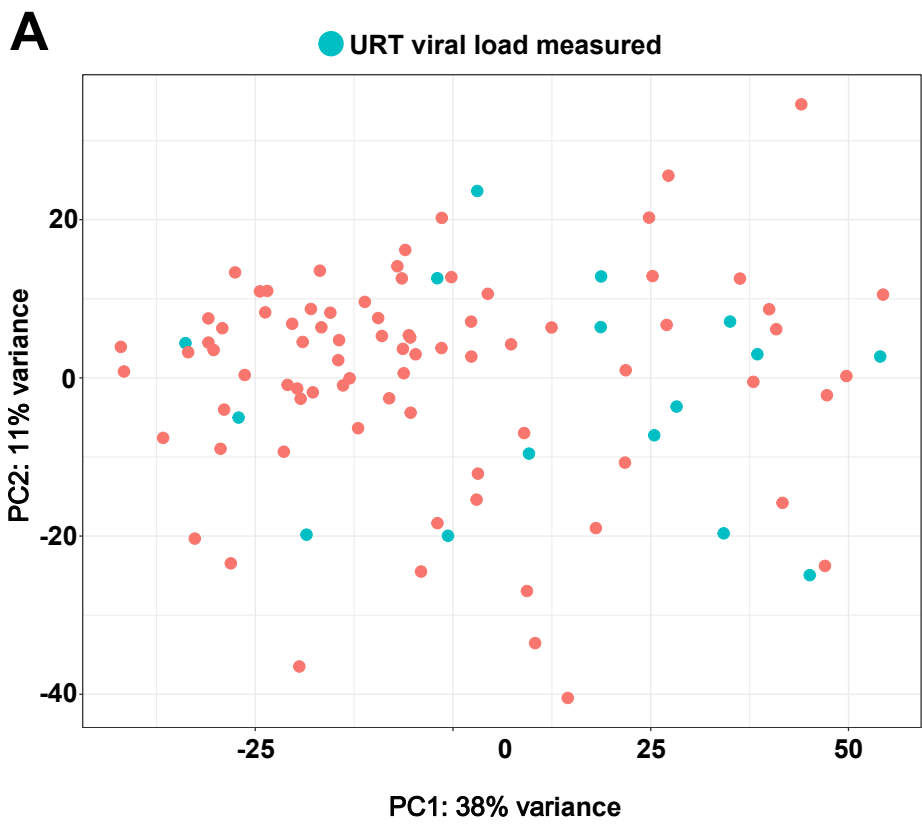
696

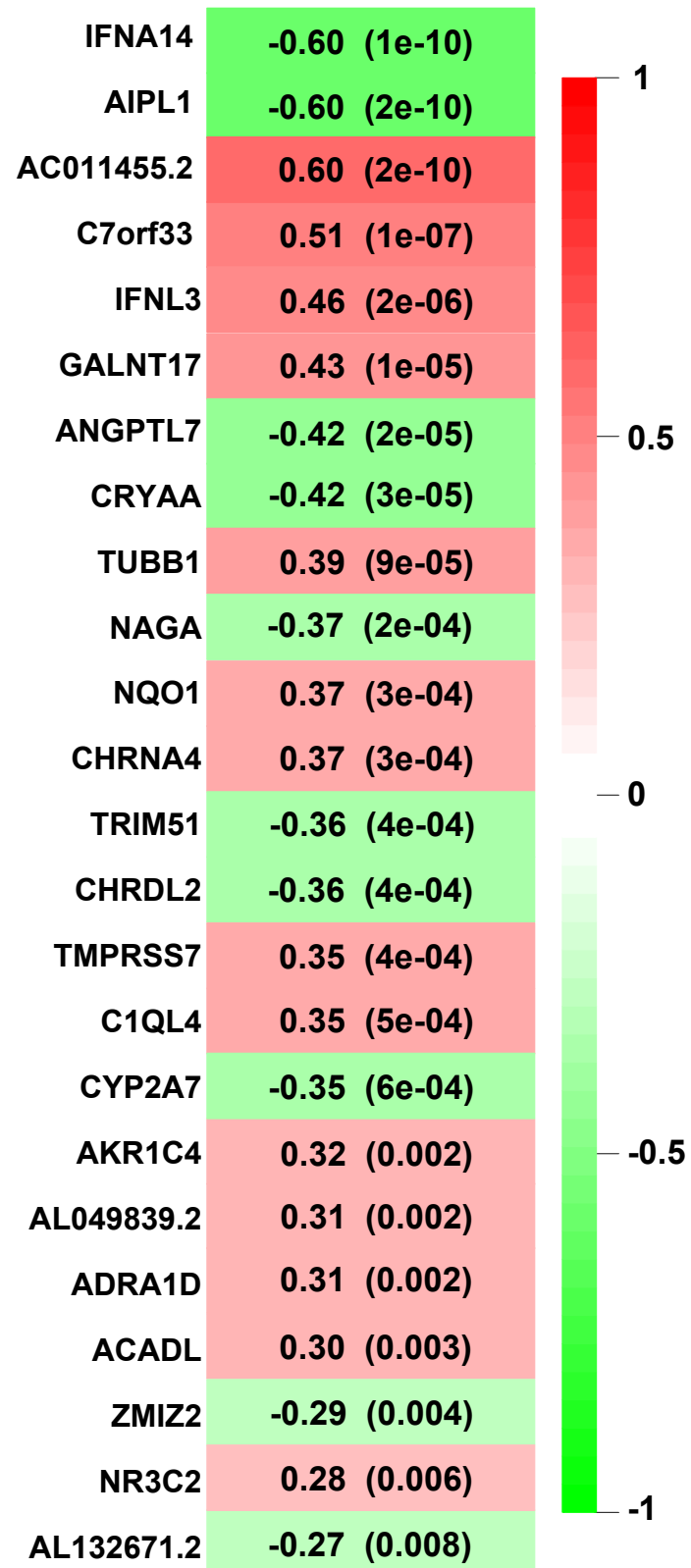
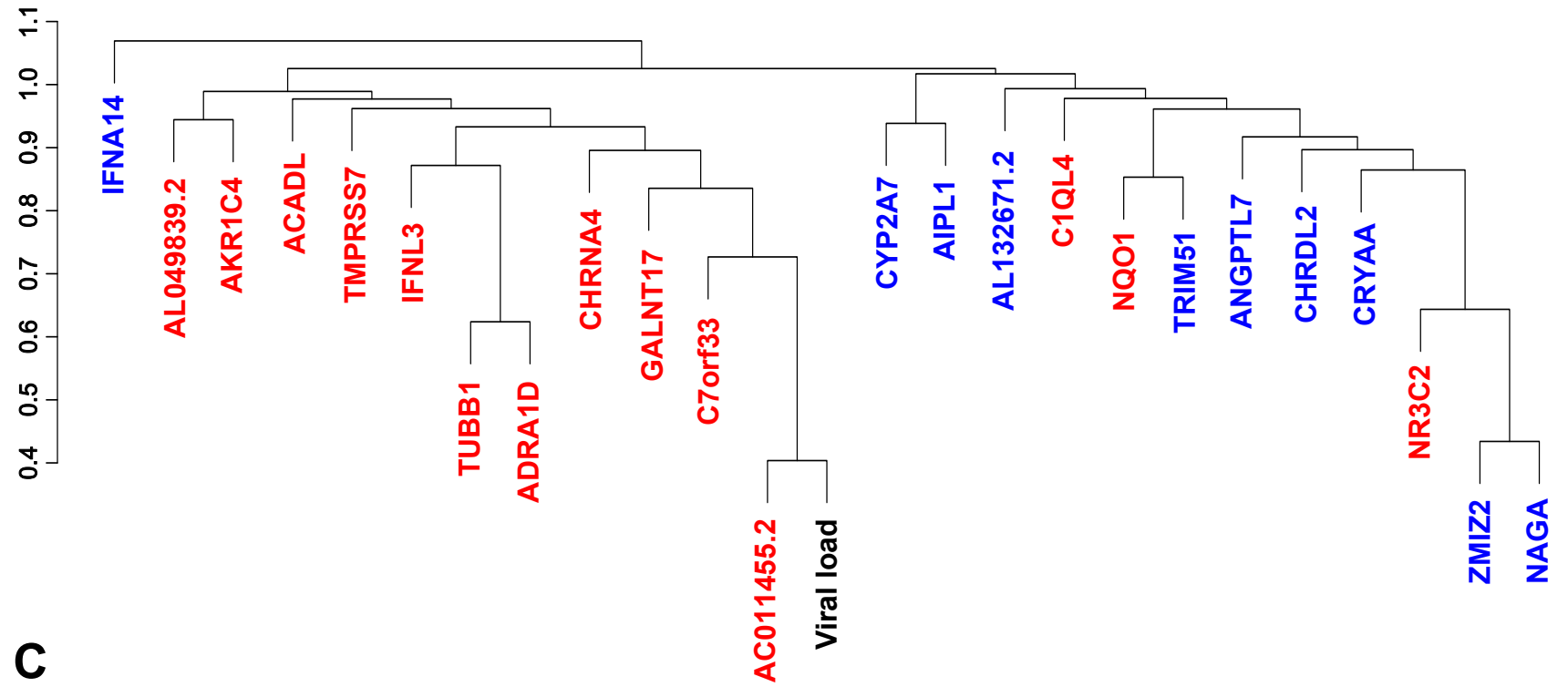
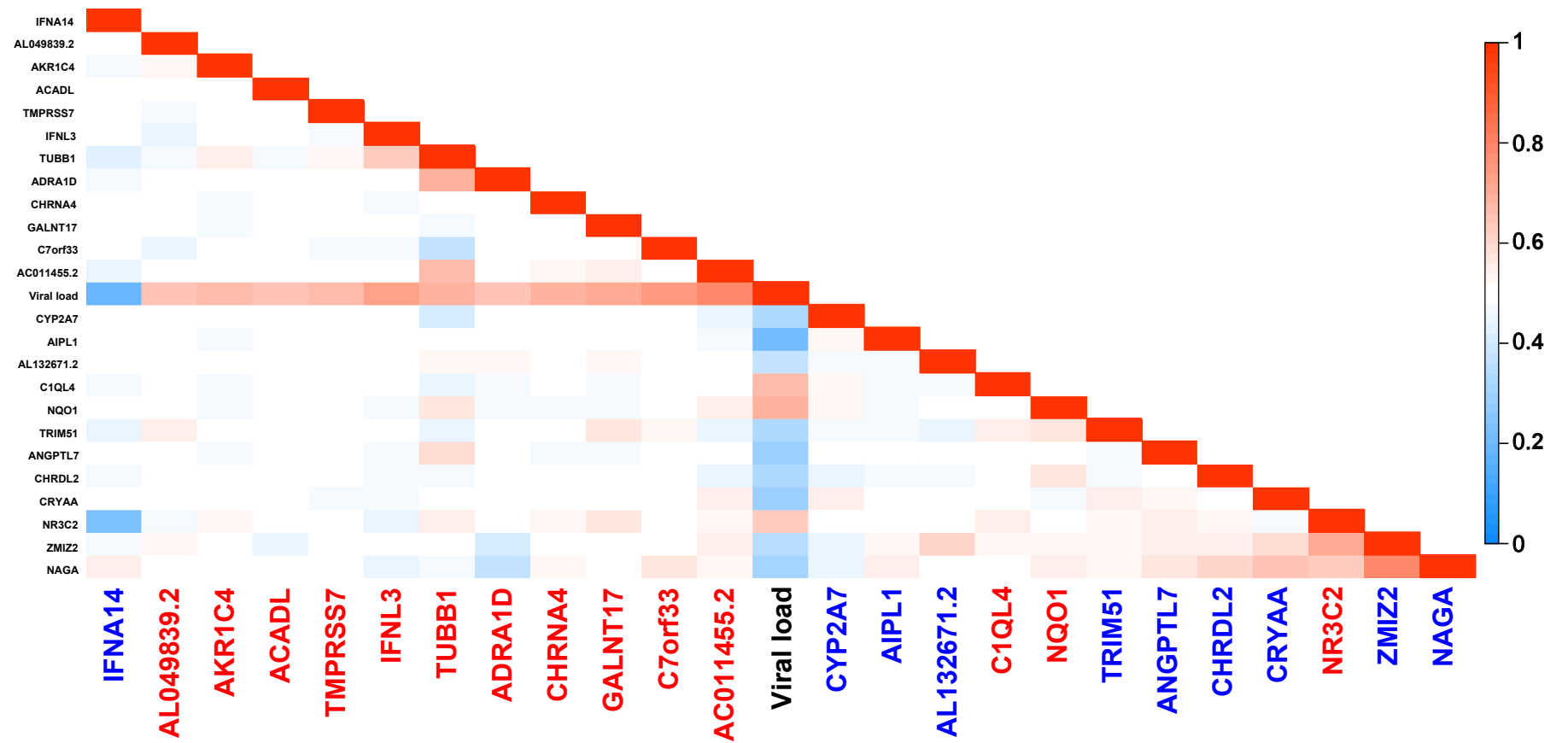
697 **References**

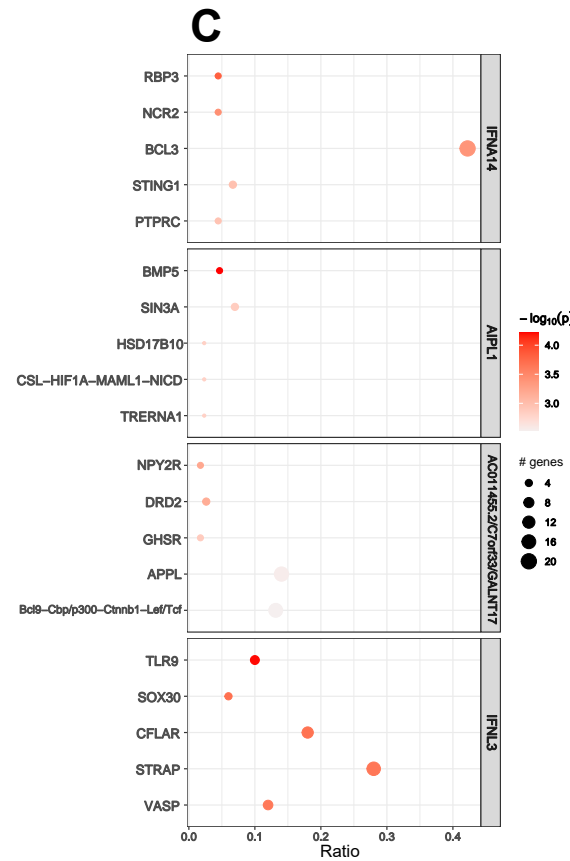
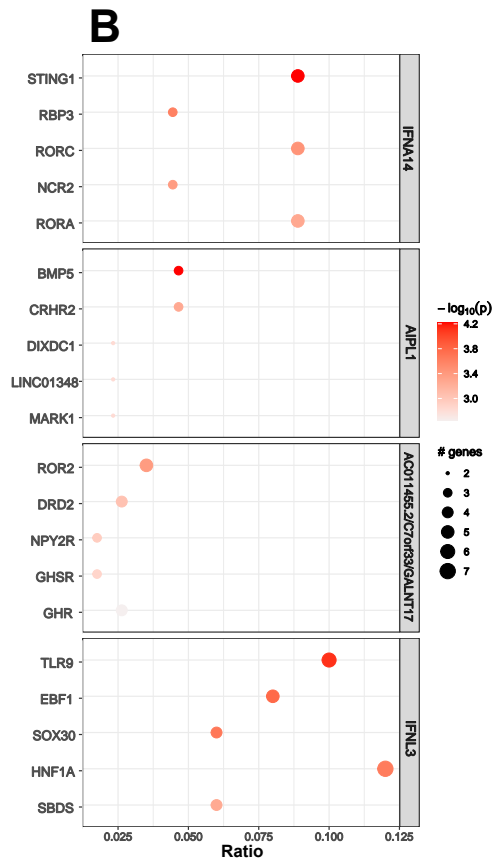
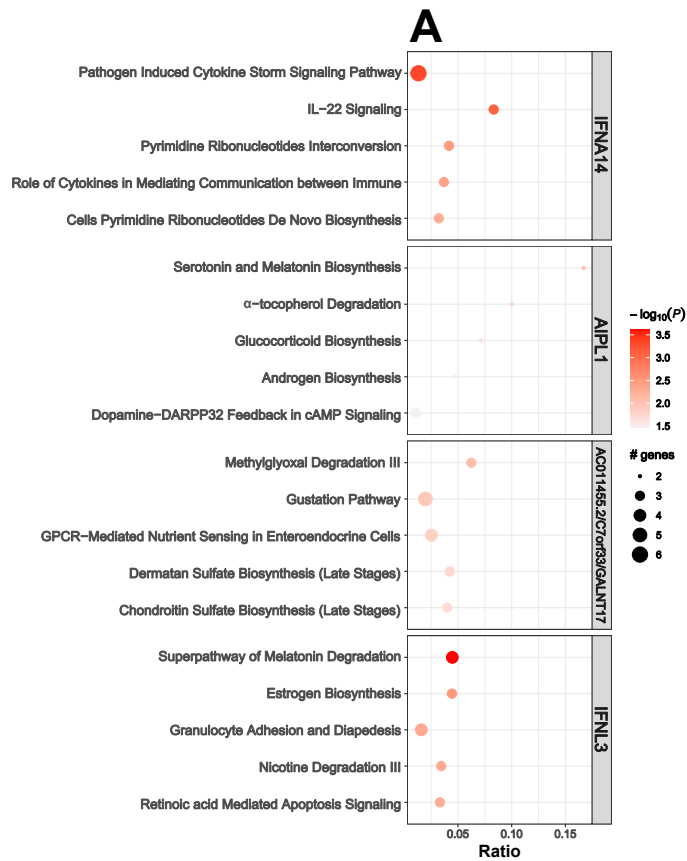
- 698 1. Cornelissen L, Andre E. Understanding the drivers of transmission of SARS-CoV-2. *Lancet Infect Dis.* 2021;21(5):580-1.
- 699 2. Marc A, Kerioui M, Blanquart F, Bertrand J, Mitja O, Corbacho-Monne M, et al. Quantifying
700 the relationship between SARS-CoV-2 viral load and infectiousness. *Elife.* 2021;10.
- 701 3. Challenger JD, Foo CY, Wu Y, Yan AWC, Marjaneh MM, Liew F, et al. Modelling upper
702 respiratory viral load dynamics of SARS-CoV-2. *BMC Med.* 2022;20(1):25.
- 703 4. Knight SR, Ho A, Pius R, Buchan I, Carson G, Drake TM, et al. Risk stratification of patients
704 admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol:
705 development and validation of the 4C Mortality Score. *BMJ.* 2020;370:m3339.
- 706 5. Chen PZ, Bobrovitz N, Premji ZA, Koopmans M, Fisman DN, Gu FX. SARS-CoV-2 shedding
707 dynamics across the respiratory tract, sex, and disease severity for adult and pediatric COVID-19.
708 *Elife.* 2021;10.
- 709 6. Cevik M, Tate M, Lloyd O, Maraolo AE, Schafers J, Ho A. SARS-CoV-2, SARS-CoV, and MERS-
710 CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and
711 meta-analysis. *Lancet Microbe.* 2021;2(1):e13-e22.
- 712 7. Arons MM, Hatfield KM, Reddy SC, Kimball A, James A, Jacobs JR, et al. Presymptomatic
713 SARS-CoV-2 Infections and Transmission in a Skilled Nursing Facility. *N Engl J Med.*
714 2020;382(22):2081-90.
- 715 8. Diamond MS, Kanneganti TD. Innate immunity: the first line of defense against SARS-CoV-2.
716 *Nat Immunol.* 2022;23(2):165-76.
- 717 9. Blanco-Melo D, Nilsson-Payant BE, Liu WC, Uhl S, Hoagland D, Moller R, et al. Imbalanced
718 Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell.* 2020;181(5):1036-45 e9.
- 719 10. McClain MT, Constantine FJ, Henao R, Liu Y, Tsalik EL, Burke TW, et al. Dysregulated
720 transcriptional responses to SARS-CoV-2 in the periphery. *Nat Commun.* 2021;12(1):1079.
- 721 11. Georgiadou A, Lee HJ, Walther M, van Beek AE, Fitriani F, Wouters D, et al. Modelling
722 pathogen load dynamics to elucidate mechanistic determinants of host-Plasmodium falciparum
723 interactions. *Nat Microbiol.* 2019;4(9):1592-602.
- 724 12. Singanayagam A, Patel M, Charlett A, Lopez Bernal J, Saliba V, Ellis J, et al. Duration of
725 infectiousness and correlation with RT-PCR cycle threshold values in cases of COVID-19, England,
726 January to May 2020. *Euro Surveill.* 2020;25(32).
- 727 13. Chikina M, Zaslavsky E, Sealfon SC. CellCODE: a robust latent variable approach to
728 differential expression analysis for heterogeneous cell populations. *Bioinformatics.*
729 2015;31(10):1584-91.
- 730 14. Lee HJ, Georgiadou A, Walther M, Nwakanma D, Stewart LB, Levin M, et al. Integrated
731 pathogen load and dual transcriptome analysis of systemic host-pathogen interactions in severe
732 malaria. *Sci Transl Med.* 2018;10(447).
- 733 15. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis.
734 *BMC Bioinformatics.* 2008;9:559.
- 735 16. Kramer A, Green J, Pollard J, Jr., Tugendreich S. Causal analysis approaches in Ingenuity
736 Pathway Analysis. *Bioinformatics.* 2014;30(4):523-30.
- 737 17. Altman MC, Rinchai D, Baldwin N, Toufiq M, Whalen E, Garand M, et al. Development of a
738 fixed module repertoire for the analysis and interpretation of blood transcriptome data. *Nat*
739 *Commun.* 2021;12(1):4385.
- 740 18. Kozlov M. Could a nose spray a day keep COVID away? *Nature.* 2022.
- 741 19. Hartwell LH, Hopfield JJ, Leibler S, Murray AW. From molecular to modular cell biology.
- 742

- 743 Nature. 1999;402(6761 Suppl):C47-52.
- 744 20. Combes AJ, Courau T, Kuhn NF, Hu KH, Ray A, Chen WS, et al. Global absence and targeting
745 of protective immune states in severe COVID-19. *Nature*. 2021;591(7848):124-30.
- 746 21. Pairo-Castineira E, Clohisey S, Klaric L, Bretherick AD, Rawlik K, Pasko D, et al. Genetic
747 mechanisms of critical illness in COVID-19. *Nature*. 2021;591(7848):92-8.
- 748 22. Wang EY, Mao T, Klein J, Dai Y, Huck JD, Jaycox JR, et al. Diverse functional autoantibodies in
749 patients with COVID-19. *Nature*. 2021;595(7866):283-8.
- 750 23. Broggi A, Ghosh S, Sposito B, Spreafico R, Balzarini F, Lo Cascio A, et al. Type III interferons
751 disrupt the lung epithelial barrier upon viral recognition. *Science*. 2020;369(6504):706-12.
- 752 24. Lucas C, Wong P, Klein J, Castro TBR, Silva J, Sundaram M, et al. Longitudinal analyses reveal
753 immunological misfiring in severe COVID-19. *Nature*. 2020;584(7821):463-9.
- 754 25. Major J, Crotta S, Llorian M, McCabe TM, Gad HH, Priestnall SL, et al. Type I and III
755 interferons disrupt lung epithelial repair during recovery from viral infection. *Science*.
756 2020;369(6504):712-7.
- 757 26. Sposito B, Broggi A, Pandolfi L, Crotta S, Clementi N, Ferrarese R, et al. The interferon
758 landscape along the respiratory tract impacts the severity of COVID-19. *Cell*. 2021;184(19):4953-68
759 e16.
- 760 27. Shim JM, Kim J, Tenson T, Min JY, Kainov DE. Influenza Virus Infection, Interferon Response,
761 Viral Counter-Response, and Apoptosis. *Viruses*. 2017;9(8).
- 762 28. Sun J, Wang J, Yuan X, Wu X, Sui T, Wu A, et al. Regulation of Early Host Immune Responses
763 Shapes the Pathogenicity of Avian Influenza A Virus. *Front Microbiol*. 2019;10:2007.
- 764 29. Hoffmann HH, Schneider WM, Rice CM. Interferons and viruses: an evolutionary arms race
765 of molecular interactions. *Trends Immunol*. 2015;36(3):124-38.
- 766 30. Schuhenn J, Meister TL, Todt D, Bracht T, Schork K, Billaud JN, et al. Differential interferon-
767 alpha subtype induced immune signatures are associated with suppression of SARS-CoV-2 infection.
768 *Proc Natl Acad Sci U S A*. 2022;119(8).
- 769 31. Zhou Q, Chen V, Shannon CP, Wei XS, Xiang X, Wang X, et al. Interferon-alpha2b Treatment
770 for COVID-19. *Front Immunol*. 2020;11:1061.
- 771 32. Ball K. Assessing the efficacy of non-canonical IFNA-subtypes as inhibitors of SARS-CoV-2
772 replication: Medical Research Scotland; 2020 [Available from:
773 [https://www.medicalresearchscotland.org.uk/awards/assessing-the-efficacy-of-non-canonical-ifna-
774 subtypes-as-inhibitors-of-sars-cov-2-replication/](https://www.medicalresearchscotland.org.uk/awards/assessing-the-efficacy-of-non-canonical-ifna-subtypes-as-inhibitors-of-sars-cov-2-replication/)].
- 775 33. Ragab D, Salah Eldin H, Taeimah M, Khattab R, Salem R. The COVID-19 Cytokine Storm; What
776 We Know So Far. *Front Immunol*. 2020;11:1446.
- 777 34. Ye Q, Wang B, Mao J. The pathogenesis and treatment of the 'Cytokine Storm' in COVID-19. *J*
778 *Infect*. 2020;80(6):607-13.
- 779 35. Ahn D, Prince A. Participation of the IL-10RB Related Cytokines, IL-22 and IFN-lambda in
780 Defense of the Airway Mucosal Barrier. *Front Cell Infect Microbiol*. 2020;10:300.
- 781 36. Albayrak N, Orte Cano C, Karimi S, Dogahe D, Van Praet A, Godefroid A, et al. Distinct
782 Expression Patterns of Interleukin-22 Receptor 1 on Blood Hematopoietic Cells in SARS-CoV-2
783 Infection. *Front Immunol*. 2022;13:769839.
- 784 37. Klooster JPT, Bol-Schoenmakers M, van Summeren K, van Vliet ALW, de Haan CAM, van
785 Kuppeveld FJM, et al. Enterocytes, fibroblasts and myeloid cells synergize in anti-bacterial and anti-
786 viral pathways with IL22 as the central cytokine. *Commun Biol*. 2021;4(1):631.
- 787 38. Han SN, Meydani SN. Impact of vitamin E on immune function and its clinical implications.
788 *Expert Rev Clin Immunol*. 2006;2(4):561-7.
- 789 39. Lewis ED, Meydani SN, Wu D. Regulatory role of vitamin E in the immune system and
790 inflammation. *IUBMB Life*. 2019;71(4):487-94.
- 791 40. Pacl HT, Tipper JL, Sevalkar RR, Crouse A, Crowder C, UAB Precision Medicine Institute, et al.
792 Water-soluble tocopherol derivatives inhibit SARS-CoV-2 RNA-dependent RNA polymerase. *bioRxiv*.
793 2021.

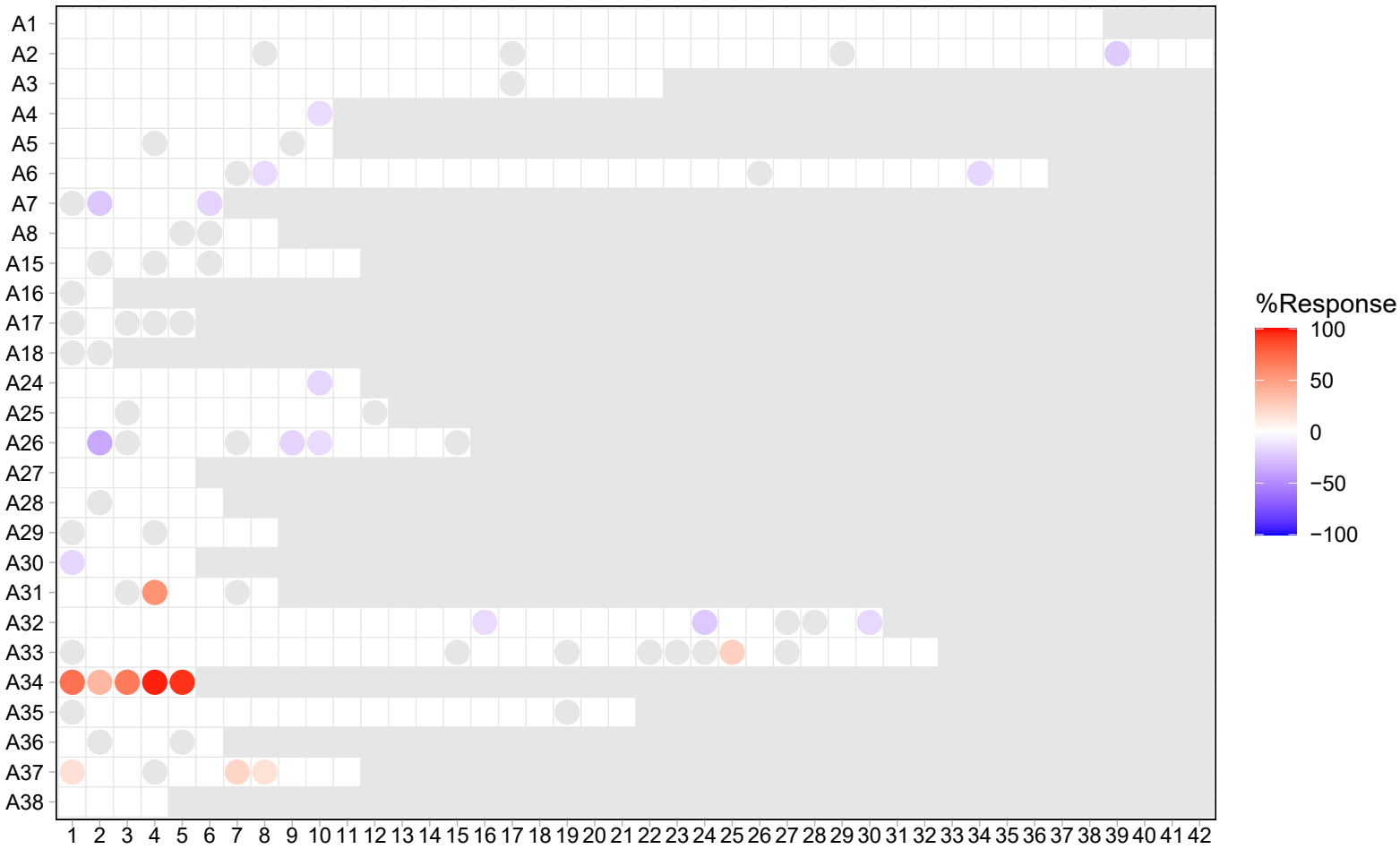
- 794 41. Pervolaraki K, Rastgou Talemi S, Albrecht D, Bormann F, Bamford C, Mendoza JL, et al.
795 Differential induction of interferon stimulated genes between type I and type III interferons is
796 independent of interferon receptor abundance. *PLoS Pathog.* 2018;14(11):e1007420.
- 797 42. Peterson ST, Kennedy EA, Brigleb PH, Taylor GM, Urbanek K, Bricker TL, et al. Disruption of
798 Type III Interferon (IFN) Genes *Ifnl2* and *Ifnl3* Recapitulates Loss of the Type III IFN Receptor in the
799 Mucosal Antiviral Response. *J Virol.* 2019;93(22).
- 800 43. Bezemer GFG, Garssen J. TLR9 and COVID-19: A Multidisciplinary Theory of a Multifaceted
801 Therapeutic Target. *Front Pharmacol.* 2020;11:601685.
- 802 44. Schmid T, Brune B. Prostanoids and Resolution of Inflammation - Beyond the Lipid-Mediator
803 Class Switch. *Front Immunol.* 2021;12:714042.
- 804 45. Wong LR, Zheng J, Wilhelmsen K, Li K, Ortiz ME, Schnicker NJ, et al. Eicosanoid signalling
805 blockade protects middle-aged mice from severe COVID-19. *Nature.* 2022;605(7908):146-51.
- 806 46. Yu J, Peterson DR, Baran AM, Bhattacharya S, Wylie TN, Falsey AR, et al. Host Gene
807 Expression in Nose and Blood for the Diagnosis of Viral Respiratory Infection. *J Infect Dis.*
808 2019;219(7):1151-61.
- 809 47. Dotiwala F, Lieberman J. Granulysin: killer lymphocyte safeguard against microbes. *Curr Opin*
810 *Immunol.* 2019;60:19-29.
- 811 48. Zhang JY, Wang XM, Xing X, Xu Z, Zhang C, Song JW, et al. Single-cell landscape of
812 immunological responses in patients with COVID-19. *Nat Immunol.* 2020;21(9):1107-18.
- 813 49. Witkowski M, Tizian C, Ferreira-Gomes M, Niemeyer D, Jones TC, Heinrich F, et al. Untimely
814 TGFbeta responses in COVID-19 limit antiviral functions of NK cells. *Nature.* 2021;600(7888):295-301.
- 815 50. WHO. Clinical management of COVID-19: interim guidance, 27 May 2020. 2020.
- 816 51. Jackson H, Rivero Calle I, Broderick C, Habgood-Coote D, D'Souza G, Nichols S, et al.
817 Characterisation of the blood RNA host response underpinning severity in COVID-19 patients. *Sci*
818 *Rep.* 2022;12(1):12216.
- 819 52. Gomez-Carballa A, Rivero-Calle I, Pardo-Seco J, Gomez-Rial J, Rivero-Velasco C, Rodriguez-
820 Nunez N, et al. A multi-tissue study of immune gene expression profiling highlights the key role of
821 the nasal epithelium in COVID-19 severity. *Environ Res.* 2022;210:112890.
- 822 53. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data.
823 *Bioinformatics.* 2014;30(15):2114-20.
- 824 54. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data 2010
825 [Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>].
- 826 55. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal
827 RNA-seq aligner. *Bioinformatics.* 2013;29(1):15-21.
- 828 56. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without
829 a reference genome. *BMC Bioinformatics.* 2011;12:323.
- 830 57. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential
831 expression analysis of digital gene expression data. *Bioinformatics.* 2010;26(1):139-40.
- 832 58. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-
833 seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
- 834 59. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential
835 expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47.



A**B****C**



A



B

